

TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Elektrische Antriebssysteme

Hochverstärkungsbasierte Regelung nichtlinearer Antriebssysteme

Hans Schuster

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing.(Univ. Tokio) Martin Buss

Prüfer der Dissertation: 1. Univ.-Prof. Dr.-Ing. Dr.-Ing. h.c. Dierk Schröder, i.R.
2. Univ.-Prof. Dr. rer. nat. Achim Ilchmann,
Technische Universität Ilmenau
3. Univ.-Prof. Dr.-Ing. Rolf Findeisen,
Otto-von-Guericke-Universität Magdeburg

Die Dissertation wurde am 22.10.2008 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 21.01.2009 angenommen.

Danksagung

Die vorliegende Dissertation entstand während meiner Forschungstätigkeit als Wissenschaftlicher Assistent am Lehrstuhl für Elektrische Antriebssysteme der Technischen Universität München.

Mein besonderer Dank gebührt Herrn Prof. Dr.–Ing. Dr.–Ing. h.c. Dierk Schröder, meinem Doktorvater, für die Ermöglichung und Förderung dieser Dissertation. An seinem Lehrstuhl eröffnete er mir in einem interessanten Forschungsgebiet den Freiraum zur Entfaltung und Erprobung eigener Ideen. Ebenso möchte ich für die großzügige Unterstützung danken, die mir Gelegenheit zur zahlreichen Teilnahme an internationalen wissenschaftlichen Fachtagungen gab.

Sehr herzlich bedanke ich mich bei Prof. Dr. rer. nat. habil. Achim Ilchmann vor allem für tatkräftige und wirkungsvolle Unterstützung. Mit der steten Bereitschaft, mich in zahllosen Diskussionen von seinem umfangreichen Fachwissen profitieren zu lassen, sowie mit seiner Begleitung und seinem Engagement hat er wesentlich zum Gelingen dieser Arbeit beigetragen. Ferner danke ich für die Übernahme des Zweitgutachtens.

Herrn Prof. Dr.–Ing. Rolf Findeisen gilt mein Dank für die Übernahme des Drittgutachtens und für das entgegengebrachte Interesse an dieser Arbeit. Für den Prüfungsvorsitz möchte ich Herrn Prof. Dr.–Ing.(Univ. Tokio) Martin Buss danken.

Für motivierende Impulse und die Hilfestellung bei verschiedensten Problemen bin ich allen derzeitigen und ehemaligen Mitgliedern des Lehrstuhls zu Dank verpflichtet. Besonders hervorheben möchte ich ausgiebige Diskussionen in freundschaftlicher Atmosphäre mit Frau Dr.–Ing. Anne Angermann, Herrn Dr.–Ing. Matthias Feiler und Herrn Dipl.–Ing. Christian Westermaier, dem ich darüberhinaus meinen Dank auch für seinen unermüdlichen Einsatz zur Optimierung der Messdatenerfassung am Prüfstand ausspreche.

Nicht zuletzt bedanke ich mich bei meiner Familie für das mir entgegengebrachte Verständnis, die uneingeschränkte Unterstützung und den dauerhaften Beistand, der mir stets ein großer Rückhalt war. Susanne und Jonas haben während der Ausarbeitungsphase mit unerschöpflicher Geduld hingenommen, bei mir oftmals nur auf geteilte Aufmerksamkeit zu stoßen.

Meinen Eltern danke ich für die Unterstützung und Förderung in meiner Ausbildungszeit, die dadurch den Grundstein für diese Dissertation gelegt haben.

München, 22.10.2008

Hans Schuster

Kurzzusammenfassung

Die vorliegende Arbeit befasst sich mit dem Entwurf und der theoretischen Analyse hochverstärkungsbasierter Regler für nichtlineare Antriebssysteme. Es wird eine realitätsnahe Prozessstruktur zugrunde gelegt, die eine elastische Kopplung von Antriebsmaschine und Last beschreibt (Zweimassensystem), unbekannte Parameterwerte zulässt, eine unbekannte nichtlineare Charakteristik aufweist (drehzahlabhängige Reibkennlinie), sowie unbekannte Störungen annimmt (Lastmomente, Messrauschen, Verfälschung der Stellgröße durch den Aktor). Die Komplexität der Strecke macht eine korrekte Parameteridentifikation unmöglich, so dass nicht auf Parameterwerte zurückgegriffen werden kann. Das entwickelte Regelungskonzept garantiert trotz bestehend einfacher Struktur die Einhaltung einer vorgegebenen Fehlertoleranzgrenze, zeigt stationäre Genauigkeit und dämpft aktiv mechanische Schwingungen. Neben dem Nachweis der Stabilität werden theoretische sowie simulative Untersuchungen durchgeführt und durch experimentelle Ergebnisse verifiziert.

Abstract

The objective of this thesis is the design and theoretical analysis of high-gain-based controllers for non-linear drive systems. Taken as a basis is a realistic process model which encompasses elastic coupling of drive engine and load (two-mass system), features unknown non-linear characteristic (speed-dependent friction), and allows for unknown parameter values as well as unknown disturbances (load, measurement noise, parasitics by the actuator). The complexity of the plant precludes the correct identification of parameters, thus suppressing availability of values. For all its simplicity the developed controller ensures tracking error evolution within a prespecified performance envelope, yields steady state accuracy and actively damps mechanical vibrations. In addition to a proof of stability, the approach is investigated by theoretical analysis as well as simulations which are verified by experimental results.

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Einführung | 1 |
| 1.1 | Allgemeines | 1 |
| 1.2 | Lineare Regler | 2 |
| 1.3 | Off-line-Identifikation | 5 |
| 1.4 | Adaption | 6 |
| 1.4.1 | Indirekte Adaption | 6 |
| 1.4.2 | Direkte Adaption | 8 |
| 1.5 | Zielsetzung | 9 |
| 1.6 | Gliederung der Arbeit | 9 |
| | | |
| 2 | Das Zweimassensystem | 11 |
| 2.1 | Lineares Modell | 12 |
| 2.2 | Analyse des Systemverhaltens | 15 |
| 2.2.1 | Motordrehzahl als Ausgang | 16 |
| 2.2.2 | Lastdrehzahl als Ausgang | 19 |
| 2.2.3 | Linearkombination als Ausgang | 20 |
| 2.2.3.1 | Beobachtbarkeit ohne Verwendung des Verdrehwinkels | 20 |
| 2.2.3.2 | Beobachtbarkeit bei Verwendung des Verdrehwinkels . | 21 |
| 2.3 | Stand der Technik bei der industriellen Regelung eines linearen ZMS . | 24 |
| 2.3.1 | Stromregelkreis | 25 |
| 2.3.2 | Drehzahlregelkreis | 26 |
| 2.3.3 | Regelung der Antriebsmaschinendrehzahl | 28 |
| 2.3.3.1 | Harte Ankopplung | 31 |
| 2.3.3.2 | Weiche Ankopplung | 32 |
| 2.3.4 | Regelung der Arbeitsmaschinendrehzahl | 34 |
| 2.3.4.1 | Harte Ankopplung | 34 |
| 2.3.4.2 | Weiche Ankopplung | 34 |
| 2.3.5 | Bewertung der Kaskadenstruktur | 35 |

| | | |
|----------|---|-----------|
| 2.3.6 | Zustandsregelung | 35 |
| 2.4 | Nichtlineares Modell | 42 |
| 2.4.1 | Reibmodell | 42 |
| 2.4.1.1 | Statische Reibkraftmodelle | 44 |
| 2.4.1.2 | Stick-Slip Effekt | 48 |
| 2.4.1.3 | Dynamische Reibkraftmodelle | 49 |
| 2.4.1.4 | Reibkennlinie des Prüfstandes | 50 |
| 2.4.1.5 | Reibmodell im Hinblick auf hochverstärkungsbasierte Regelung | 51 |
| 2.4.2 | Integration von Reibung in das Modell des ZMS | 51 |
| 3 | Hochverstärkungsbasierte adaptive Regler | 53 |
| 3.1 | Auswirkung hoher Kreisverstärkungen | 55 |
| 3.2 | Hohe Verstärkung im linearen Regelkreis | 56 |
| 3.3 | Hohe Verstärkung im nichtlinearen Regelkreis | 62 |
| 3.4 | Adaptive Stabilisierung | 64 |
| 3.4.1 | Ordnung $n = 1$ | 65 |
| 3.4.2 | Ordnung $n > 1$ | 71 |
| 3.5 | Adaptive Regelung | 75 |
| 3.5.1 | Festwertregelung für Integrale Regelstrecken | 76 |
| 3.5.2 | Festwertregelung für Proportionale Regelstrecken | 76 |
| 3.5.3 | Folgeregelung | 80 |
| 3.5.4 | Einfluss von Störungen | 81 |
| 3.6 | σ -Modifikation | 89 |
| 4 | Funnel-Control | 93 |
| 4.1 | Regelstrategie | 95 |
| 4.2 | Regelbare Systemklasse | 98 |
| 4.3 | Charakteristische Besonderheit der regelbaren Systemklasse | 101 |
| 4.4 | Hochverstärkungsbasierte Regelung mit nicht-monotoner Verstärkung . | 107 |
| 4.5 | Anwendung am Einmassensystem | 113 |
| 4.6 | Folgeregelung und Störgrößenunterdrückung | 115 |
| 4.7 | Festwertregelung | 126 |
| 4.7.1 | Führungsintegrator/Kaskadenstruktur | 127 |
| 4.7.2 | Integraler Anteil im Funnel-Regelkreis | 128 |
| 4.8 | Zeitvariante Regelung mit internem Modell bei linearen Prozessen . . . | 130 |

| | | |
|----------|--|------------|
| 5 | Einfluss hoher Relativgrade | 141 |
| 5.1 | Die Problematik hoher Relativgrade | 141 |
| 5.2 | Lösungsansätze im Falle hoher Relativgrade | 145 |
| 5.2.1 | Dynamischer Kompensator | 145 |
| 5.2.2 | Backstepping | 148 |
| 5.2.3 | Zustandsauskopplung | 153 |
| 6 | Hochverstärkungsbasierte Regelung des ZMS | 165 |
| 6.1 | Regelung des linearen Modells | 167 |
| 6.1.1 | Regelung der Motordrehzahl | 168 |
| 6.1.2 | Regelung der Lastdrehzahl | 171 |
| 6.1.3 | Zustandsregelung | 172 |
| 6.2 | Behebung des Zielkonfliktes durch dynamische Zustandsrückführung . . | 182 |
| 6.3 | Regelung des nichtlinearen Modells | 186 |
| 6.3.1 | Regelung der Motordrehzahl | 187 |
| 6.3.2 | Zustandsregelung | 187 |
| 6.3.3 | Zustandsregelung mit Hochpassfilter und I-Anteil | 192 |
| 6.4 | Messergebnisse | 205 |
| 7 | Zusammenfassung und Ausblick | 209 |
| A | Grundlagen und Definitionen | 211 |
| A.1 | Abkürzungen und Schreibweisen | 211 |
| A.2 | Existenz und Eindeutigkeit von Lösungen | 213 |
| A.3 | Instantane Verstärkung | 217 |
| A.3.1 | Instantane Verstärkung linearer Systeme | 217 |
| A.3.2 | Instantane Verstärkung nichtlinearer Systeme | 219 |
| A.4 | Stationäre/Statische Verstärkung | 220 |
| A.5 | Lie-Ableitung | 222 |
| A.6 | Relativgrad | 224 |
| A.7 | Basiswechsel mit diffeomorphen Abbildungen | 231 |
| A.8 | Byrnes-Isidori-Normalform | 233 |
| A.8.1 | Transformation nichtlinearer Systeme | 234 |
| A.8.2 | Transformation linearer Systeme | 237 |
| A.9 | Nulldynamik, Minimalphasigkeit | 239 |
| A.10 | Essentielles Supremum | 249 |
| A.11 | Norm | 251 |

| | |
|--|-----|
| A.11.1 Norm für Vektoren | 251 |
| A.11.2 Norm für Matrizen | 252 |
| A.11.3 Norm für Funktionen | 254 |
| A.12 Grundlegende Theoreme und Hilfssätze | 257 |
| A.12.1 Zusammenhang zwischen Minimalphasigkeit, Stabilisierbarkeit und Detektierbarkeit | 257 |
| A.12.2 Berechnung der Übertragungsfunktion aus der RNF | 258 |
| A.12.3 Gronwall-Bellman Ungleichung | 259 |
| A.12.4 Geršgorin-Theorem | 260 |

Kapitel 1

Einführung

1.1 Allgemeines

Der Ursprung der Regelungstechnik, bzw. der Regelung von Antrieben, liegt im Jahr 1788, als Boulton und Watt den Fliehkraftregler erstmals zur Drehzahlregelung bei Dampfmaschinen einsetzen. Diese mechanische Rückkopplung ist zum stabilen Betrieb der Anlage unabdingbar und wird als der erste Regler im heutigen Sinne betrachtet [10].

Seither vollzog sich zwar in diesem Gebiet ein tiefgreifender Wandel, es wurden fundamental wichtige mathematische Werkzeuge zur Beschreibung dynamischer Prozesse entwickelt. Neben diesen theoretischen Fortschritten hat sich auch die Realisierung der Regler stark verändert. Angefangen von mechanischen Reglern über zeitkontinuierliche, elektronische Regler aus Röhrenbausteinen und Operationsverstärkern ist die Entwicklung heute bei zeitdiskret arbeitenden Mikroprozessoren angelangt, welche durch steigende Leistungsfähigkeit immer aufwändigere Regelungsalgorithmen erlauben. Dennoch bleibt die Hauptaufgabe der Regelungstechnik unverändert in der gezielten Beeinflussung technischer Prozesse zur Verbesserung des dynamischen und statischen Verhaltens bestehen. Das Ziel einer Regelung ist der sichere, wirtschaftliche und ökologische Betrieb technischer Anlagen und physikalischer Prozesse. Stetig steigende Anforderungen an die Produktqualität stellen dabei zusammen mit zunehmendem Kostendruck den Hauptantrieb für die ständige Entwicklung verbesserter Regelungssysteme dar.

Im Bereich der Antriebstechnik schlagen sich die wachsenden Anforderungen in einer höheren Ordnung des Prozessmodells nieder: weil durch die Realisierung schneller Verfahrbewegungen einer Masse der Antriebsstrang mit hohen Drehmomenten bzw. Kräften belastet wird und gleichzeitig jede reale Kraftübertragungseinheit (Verbindungswelle, Riemen, Kette, Seilzug, etc.) in gewissem Maße elastisches Verhalten zeigt, können in unerwünschter Weise die Eigenschwingungen des mechanischen Systems angeregt werden. Wo sich früher die Aufgabenstellung in der Drehzahlregelung der Motorwelle erschöpfte, wird heute zum Zwecke einer dynamischen und gleichzeitig präzisen Betriebsführung die Regelung der Lastdrehzahl erforderlich sein. Daher muss in modernen Antriebssystemen das elastische Verhalten der mechanischen Kopplung zwischen Antriebs- und Arbeitsmaschine (Zweimassensystem) mit in den Reglerentwurf einbe-

zogen werden, um eine gewünschte Dämpfung von auftretenden Schwingungen zu erreichen. Aus diesem Grund sind für die Beschreibung und Beherrschung der komplexer gewordenen Systeme verbesserte Methoden notwendig.

Die rasante Entwicklung der Mikroprozessoren in den vergangenen Jahren gestattet prinzipiell die Implementierung rechenintensiver Regelalgorithmen bei hinreichend kurzen Abtastzeiten. Als Beispiel ist hier die modellbasierte prädiktive Regelung anzuführen, die in den siebziger Jahren chemischen Prozessen mit trägem Verhalten (Zeitkonstanten im Minutenbereich) vorbehalten war, inzwischen aber an elektrischen Antrieben mit Abtastzeiten im Millisekundenbereich eingesetzt werden kann [78]. Allerdings ist durch das Bestreben nach Wirtschaftlichkeit ein Preislimit gesetzt, das in vielen Fällen die Verwendung leistungsstarker und damit teurer Prozessoren ausschließt und stattdessen kostengünstige Hardware verlangt. Darüberhinaus steht für die eigentliche Stellgrößenberechnung selten die gesamte Rechenkapazität zur Verfügung, da ein Prozessrechner im Allgemeinen neben der Regelung weitere Aufgaben wie Messwertaufbereitung, Fehlerdiagnose, Fehlerbehandlung, Anlagensteuerung oder auch die simultane Regelung zusätzlicher Prozesse bewältigen muss. Aufgrund ökonomischer Vorgaben bleibt meist nur wenig Raum für aufwändige Stellgrößenberechnungen, wodurch der Entwurf des Regelgesetzes maßgeblich beeinflusst wird. So überschreitet beispielsweise der Einsatz lernender neuronaler Netze oftmals die Leistungsgrenze gängiger Industrie-PC's und führt zum Verlust der Echtzeitfähigkeit. Besonders wenn Netze mit einer hohen Zahl an Neuronen zur Approximation des Prozessverhaltens notwendig sind, ist die verfügbare Rechenleistung schnell ausgeschöpft. Zudem verlängert sich mit wachsender Größe des Netzes die benötigte Lernphase, die akzeptable Zeitbereiche oft übersteigt.

Ein zweites Argument, das die Ablehnung rechenintensiver Algorithmen erklärt, ist der meist komplexe Zusammenhang zwischen einer Vielzahl von Einstellparametern und resultierender Regelgüte. Wenn die Auswirkung einer veränderten Reglereinstellung für den Anlagenbetreiber nicht zu überblicken ist und daher bei jeder Modifikation des Prozesses eine Anpassung durch den regelungstechnischen Experten erforderlich wird, darf die Akzeptanz solcher Verfahren nicht erwartet werden.

1.2 Lineare Regler

Aus genannten Gründen ist verständlich, warum im industriellen Umfeld PID-ähnliche Konzepte etabliert sind und für den Einsatz an realen Anlagen überwiegend lineare Regler genutzt und basierend auf linearen Prozessmodellen ausgelegt werden. Nachdem sich ein solcher Regelkreis mit linearen Differentialgleichungen beschreiben lässt, stehen ausgereifte mathematische Methoden für die Analyse und Synthese der Regelung zur Verfügung. Aussagen zu relevanten Eigenschaften wie asymptotische Stabilität, Schwingfähigkeit, Eigenfrequenz, Einschwingverhalten, Dämpfungsgrad oder Überschwinger lassen sich sowohl im Frequenzbereich zum Beispiel anhand von Bodediagramm, Nyquist-Plot, Wurzelortskurve und Übertragungsfunktion, als auch im Zeitbereich mittels Zustandsbeschreibung treffen. Die aufgezählten Werkzeuge sind für die Regelungstechnik von fundamentaler Bedeutung und begründen in verständlicher

Weise die bevorzugte Verwendung linearer Regler.

Als gravierender Nachteil der konventionellen Regelungstheorie erweist sich jedoch die Voraussetzung, dass ein gegebenes Streckenmodell das reale Prozessverhalten exakt beschreibt, wobei Modellparameter als Größen mit bekannten Zahlenwerten betrachtet werden. Unter diesen Umständen lässt sich ein passender Regler entwerfen, der auf eine hinreichend gute Übereinstimmung zwischen Modell und realer Strecke angewiesen ist. Aus verschiedenen Gründen besteht jedoch eine Diskrepanz zwischen Modell und Realität, so dass die zugrunde liegende Annahme nicht in vollem Umfang zutrifft.

Einerseits entstehen Fehler in der Systemordnung durch eine beabsichtigte Untermollierung. Da jedes reale System bei beliebig genauer Betrachtung von beliebig hoher Ordnung ist, muss eine Ordnungsreduktion erfolgen. Diese Maßnahme berücksichtigt lediglich die dominante Systemdynamik, die durch ein praktikables Modell abgebildet wird. Je besser das zu erstellende Modell mit der Realität übereinstimmen muss, desto höher ist im Allgemeinen auch dessen Ordnung zu wählen.

Andererseits liegen Modellfehler auch in ungewollten Fehlern beim Ermitteln der Parameterwerte begründet. Weil aus Kostengründen meist auf präzise und daher aufwändige Messungen verzichtet werden muss, sind (erhebliche) Unsicherheiten bei der Parameterbestimmung unvermeidlich. Deshalb kann bei der Parametrierung eines Modells üblicherweise nur auf (grobe) Schätzwerte zurückgegriffen werden.

Eine weitere Fehlerquelle bilden die im realen System vorhandenen Nichtlinearitäten. Für die Beschreibung eines Regelkreises mittels linearer Differentialgleichungen müssen nichtlineare Effekte entweder vernachlässigt oder linearisiert werden. Häufig ist der nichtlineare Charakter einer Strecke nur schwach ausgeprägt, so dass eine solche Vorgehensweise zulässig ist und auf ein taugliches Modell führt. Jedoch ist für den Einzelfall eine Überprüfung der Modellqualität notwendig.

Beim Entwurf eines Reglers ist stets zu beachten, dass keine reale Regelstrecke mit vertretbarem Aufwand exakt modellierbar ist. Die unumgängliche Abweichung zwischen Modellverhalten und realem Prozess setzt einen Regler mit einem Mindestmaß an Robustheit gegenüber Parameterfehlern und Störungen voraus. Dieser muss in der Lage sein, eine ganze Klasse von Regelstrecken zu beherrschen, die sich durch unterschiedliche Parametrierungen auszeichnen. Wesentliche Erfolge auf dem Gebiet der robusten Regler sind der bekannte \mathcal{H}_∞ -Entwurf und die Beiträge von Ackermann [1] und Kharitonov [52], [1, Kap. 5.1], [79, S. 153f]. Wenn die Streckenparameter (z.B. wegen Messunsicherheiten) nicht durch exakte Zahlenwerte, sondern lediglich durch Zahlenintervalle angegeben werden können, entscheidet eine Auswertung der Kharitonov-Polynome, ob für alle möglichen Parameterkonstellationen die Stabilität eines linearen Regelkreises garantiert ist.

Wie die industrielle Praxis zeigt, besitzen lineare robuste Regler ein weitläufiges Einsatzspektrum, obwohl Grenzen für die maximal zulässige Variation der Parameter bestehen. Nicht in allen Anwendungsfällen existiert tatsächlich *ein* linearer Regler, der über die gesamte Schwankungsbreite aller Parameter (d.h. für sämtliche Strecken einer gegebenen Systemklasse) das gewünschte dynamische Verhalten erzielt. Ein bekann-

tes Beispiel für dieses Problem ist in der Flugmechanik zu finden [8, S. 12ff]. Für verschiedene Konstellationen von Flughöhe und Fluggeschwindigkeit muss zwischen geeigneten Reglereinstellungen gewechselt werden, wobei die Realisierung einer Auswahlstrategie die Messung einiger Prozessgrößen notwendig macht. Nachdem in Flugzeugen grundsätzlich die Fluggeschwindigkeit und Höhe sensorisch erfasst werden (und zwar unabhängig vom Regler, d.h. aus anderen Gründen als für eine gute Regelung erforderlich wäre), ist ein solches Vorgehen ohne Zusatzaufwand möglich.

Robustheit (im Sinne von universeller Eignung eines fest eingestellten Reglers für alle Elemente einer Systemklasse) bedeutet den Verzicht auf optimale Abstimmung des Reglers auf die tatsächliche Strecke. Ein typisches Anwendungsgebiet robuster Regler ist die automatisierte Containerverladung [1, S. 4ff]. Mit jedem Hubvorgang des Kranes variiert die Lastmasse, die durch einen Modellparameter abgebildet wird. Aus regelungstechnischer Sicht kommt daher jede Änderung der Lastmasse einem Austausch des Streckenmodells gleich. In diesem Beispiel ist ein Regler erforderlich, der alle Strecken aus einer gegebenen Klasse mit gewünschter Güte regelt und somit die Pendelbewegung des Containers innerhalb einer maximal zulässigen Amplitude hält und mit exponentieller Rate abdämpft. Dabei steht die Austauschbarkeit der Strecke im Vordergrund, eine Justierung der Reglerparameter auf ein einziges Modell (in diesem Falle für ein spezielles Containergewicht) entspricht nicht der Zielsetzung.

Hiervon abweichend ist die regelungstechnische Aufgabenstellung in einer Vielzahl von Anwendungsfällen gelagert. Im Kontext nichtlinearer Strecken mit Parameterunsicherheiten lässt sich das Streckenmodell mit konstanten Werten parametrieren, wobei deren Beträge unbekannt oder starken Unsicherheiten unterworfen sind. Die Gesamtheit aller möglichen Parameterkonstellationen spannt eine Systemklasse auf, die das geeignete Prozessmodell enthält. Ziel ist daher, nur ein einziges, aber nicht spezifiziertes Element aus einer vorgegebenen Systemklasse bestmöglich zu regeln. Diese Problemstellung verlangt nach einem Regler, der im Zusammenspiel mit jedem Element der Systemklasse stabiles Verhalten zeigt, möglicherweise anfänglich das Regelziel nicht erreicht, jedoch im Laufe der Zeit durch Adaption an das Verhalten der Strecke die Eigenschaften des Regelkreises verbessert. Sofern der Adaptionsvorgang nicht automatisiert abläuft, ist bei Inbetriebnahme der Anlage ein manuelles „Feintuning“ der Reglereinstellung notwendig. Die mit systematischen Methoden berechneten Reglerparameter dienen lediglich als Ausgangspunkt für die experimentelle Suche nach einem verbesserten Parametersatz. Die heuristische Vorgehensweise ist meist zeitintensiv und bringt sowohl Personalkosten als auch Verluste durch lange Stillstandszeiten bei Inbetriebnahme und Änderung der Produktionsanlage mit sich. Vor allem in Anwendungen mit zeitlich veränderlichen Parametern ist eine solche Vorgehensweise nicht zweckmäßig. Durch diese Tatsache ist die Entwicklung von Identifikationsmethoden, sowie selbst-einstellender und adaptiver Regler motiviert.

1.3 Off-line-Identifikation

Ansätze, die sich in eine Identifikationsphase, Reglersynthese und anschließende Regelungsphase gliedern, lassen sich unter dem Begriff „off-line-Methoden“ zusammenfassen. Kennzeichnend ist, dass zunächst eine Identifikation der unregulierten Strecke erfolgt. Dabei wird durch einen geeigneten Stellgrößenverlauf der Prozess ausreichend angeregt und dessen Reaktion erfasst. Der Zusammenhang zwischen Anregung und Prozessantwort (E/A-Verhalten) dient als Grundlage, um die Parameter eines Modells zu adaptieren. Sobald die Übereinstimmung zwischen dem Verhalten von Modell und Prozess als ausreichend erachtet wird, kann die Identifikationsphase beendet werden. Der anschließende Reglerentwurf erfolgt unter Nutzung des adaptierten Modells, welches voraussetzungsgemäß das Streckenverhalten hinreichend gut beschreibt. In der nachfolgenden Regelungsphase findet keine weitere Adaption des Modells statt, weshalb die Reglerparameter konstant bleiben. Off-line-Methoden sind daher im Sinne von [8, Kap. 1] keine adaptiven Regler, sondern nutzen Identifikationsverfahren lediglich für eine Modellparametrierung, die dem Reglerentwurf vorgelagert ist.

Um einerseits der Affinität zu linearen Regelkreisen nachzukommen und andererseits auch nichtlineare Prozesse regeln zu können, bietet sich die in [84, Kap. 5-6] beschriebene Vorgehensweise an. Oftmals sind in Antriebssystemen die Parameter des linearen Streckenteils bekannt, die nichtlinearen Effekte (wie z.B. Reibung oder Lose) jedoch nicht. Ziel ist, deren Charakteristik zu ermitteln und auf Basis der gewonnenen Systemkenntnis eine Kompensation zu entwerfen. Dadurch wird der Einfluss nichtlinearer Elemente reduziert bzw. unterdrückt, so dass die Strecke im Idealfall lineares Verhalten zeigt.

Als Funktionsapproximator eignen sich diverse Typen Neuronaler Netze, wie RBF (Radial Basis Function Network), GRNN (General Regression Neural Network) oder MLP (Multi Layer Perceptron Network). Mit dem Begriff „Lernen“ wird eine Parameteradaption bezeichnet, die das Verhalten des Funktionsapproximators an dasjenige der realen Nichtlinearität angleicht. Hierzu ist eine unter Umständen nicht-konvexe Optimierungsaufgabe zu lösen.

Sofern in der Strecke die Auswirkung der nichtlinearen Funktion hinreichend gut kompensiert ist, lässt sich ein Regler mit linearen Standardmethoden [85, Kap. 3-4] auslegen, der dem Prozess das gewünschte Verhalten einprägt. Allerdings zeigen die Untersuchungen in [86] und [92], dass eine teils ausgeprägte Empfindlichkeit gegenüber unvermeidbaren Ungenauigkeiten in der Kompensation besteht, so dass dieser Ansatz in einigen Problemstellungen versagt.

Allgemein besitzen off-line-Verfahren den entscheidenden Nachteil, dass für die Identifikation ein Betrieb der unregulierten Anlage notwendig ist. Hierfür muss stabiles Verhalten gegeben sein und die Regelstrecke durch eine Steuerung gefahrlos betrieben werden können. Bei instabilen Strecken ist ein Minimum an Streckenkenntnis erforderlich, um zumindest eine stabilisierende Regelung zu entwerfen. Diese muss und wird im Allgemeinen nicht die gewünschte Regelgüte erzielen, soll aber den notwendigen Identifikationsbetrieb sicherstellen. Eine solche Vorgehensweise ist jedoch sehr anwen-

dungsspezifisch und abhängig von der Problemstellung. Damit ist die Durchführbarkeit jeweils für den vorliegenden Einzelfall zu überprüfen.

Darüberhinaus stellt sich die Frage nach geeigneten Eingangssignalen. Die Bedingung der „persistence of excitation“ [8, S. 63ff] fordert eine beständige Anregung, die den Prozess in alle relevanten Betriebsbereiche zwingt, damit sich dessen charakteristische Eigenschaften zeigen. Diese Forderung steht möglicherweise im Widerspruch zum Schutz der Anlage, der die Aussparung einiger Betriebszustände verlangt. Zusätzlich ist an einer solchen Vorgehensweise nachteilig, dass Veränderungen der Strecke während der Regelungsphase nicht mehr durch eine Anpassung der Reglerparameter berücksichtigt werden.

1.4 Adaption

Das Gebiet der adaptiven Regelung ist in den vergangenen Jahren durch intensive Bearbeitung zu einem weitläufigen Themenfeld herangereift. Die Entwicklung einer Vielzahl unterschiedlicher Konzepte lässt die Einteilung in die Gruppen „indirekte Adaption“ und „direkte Adaption“ zweckmäßig erscheinen. Diese Klassifikation entscheidet anhand des Kriteriums, ob eine Methode zur Schätzung unbekannter Streckenparameter Verwendung findet. Indirekte Adaptionsverfahren ermitteln im Gegensatz zu off-line-Methoden das Streckenverhalten während des Regels (also on-line) und umgehen somit die beschriebenen Nachteile. Direkte Adaptionmethoden verzichten auf eine Identifikation der Strecke.

1.4.1 Indirekte Adaption

In diese Klasse sind adaptive Regler einzuordnen, die sich zur Bestimmung geeigneter Reglerparameter auf einen Identifikationsalgorithmus stützen. Abbildung 1.1 zeigt die zugrunde liegende Struktur: der mit „Parameterschätzung“ bezeichnete Block beinhaltet ein Modell der Strecke, dessen E/A-Verhalten durch Parameteradaption an dasjenige der realen Strecke angepasst wird. Die durch den Block „Adaptionsgesetz“ symbolisierte Parametrierung des Reglers bedient sich der ermittelten Kenngrößen der Strecke und berechnet – wie bei einem off-line-Entwurf (d.h. Reglersynthese für Prozesse mit bekannten Parametern) – abhängig von gegebenen Spezifikationen eine geeignete Einstellung des Reglers. Das Attribut „indirekt“ bezieht sich auf die Eigenschaft, dass die Berechnung der Reglerparameter über den Zwischenschritt der Streckenidentifikation erfolgt.

Das skizzierte Vorgehen basiert auf dem „certainty equivalence Prinzip“, das die Nutzung der geschätzten Werte (Modellparameter) anstelle der realen Streckenparameter erlaubt, selbst wenn zwischen Modell und Prozess eine große Diskrepanz besteht [8, S. 91]. Solange das aktuelle Identifikationsergebnis stark von den korrekten Parametern abweicht, wird die Reglereinstellung auf Grundlage ungeeigneter Schätzwerte vorgenommen. Es ist daher zu erwarten, dass der Regler in diesem Falle seine Aufgabe zwar

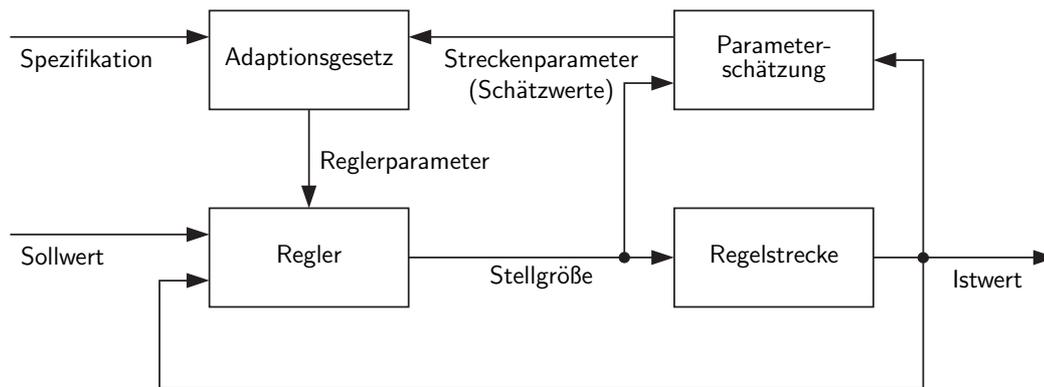


Abbildung 1.1: Prinzipskizze der indirekten Adaption.

nicht zufriedenstellend erfüllen kann, dabei jedoch eine beständige Anregung generiert. Jene beeinflusst den Fortgang der Identifikation positiv, weshalb sich die Qualität des Modells stetig verbessert. Dieser Mechanismus führt schlussendlich auf ein hinreichend gutes Identifikationsergebnis, so dass die daraus resultierende Einstellung des Reglers asymptotisch die gewünschte Regelgüte erreicht.

Die Ausführungen in [69, Kap. 5] rechtfertigen das beschriebene Vorgehen und weisen die Konvergenz der Parameterschätzung nach. Dabei bestimmt der gewählte Sollwertverlauf, ob asymptotisch das resultierende Modell in vollem Umfang mit der realen Strecke übereinstimmt. Unter der Annahme einer beständigen Anregung (persistent excitation) ist im gemessenen Prozessausgang ausreichend Information enthalten, um daraus die realen Parameter zu ermitteln. Ohne beständige Anregung ist keine eindeutige Bestimmung der tatsächlichen Werte möglich. Dennoch konvergiert die Schätzung und stellt geeignete Werte für die Reglerparametrierung bereit. Diese beschreiben das benötigte E/A-Verhalten der Strecke lediglich im vorliegenden Betriebsbereich, können jedoch den Prozess nicht in sämtlichen Situationen nachbilden. Weil die Modellparameter nicht zwingend die Werte der unbekanntenen Prozessparameter erreichen und daher keine Identifikation der Strecke im eigentlichen Sinne erfolgt, ist die Bezeichnung „Parameterschätzung“ treffender.

Nachdem anfänglich keine Übereinstimmung zwischen Schätzwerten und Streckenparametern erwartet werden kann und daher die Reglereinstellung auf einem nicht richtig adaptierten Modell basiert, muss eine transiente Phase mit schlechtem Regelverhalten in Kauf genommen werden. Zwar bleiben Stellgröße, Prozessausgang, Schätzwerte und Reglerparameter beschränkt; obere Schranken für diese Größen lassen sich jedoch vorab nicht ermitteln. Insofern sichert ein Stabilitätsbeweis weder das Überschreiten zulässiger Höchstwerte, noch das Einhalten des Regelzieles im Einschwingvorgang.

In [20] und [21] wird der genannte Ansatz zur Drehzahlregelung eines elektrischen Antriebes genutzt, der ein unbekanntes, aber stückweise konstantes Lastmoment zu überwinden hat. Dabei wird im Parameterschätzer ein ARMA-Modell verwendet, so dass dieses Konzept auf die Regelung linearer Strecken beschränkt ist.

Darüberhinaus ist die Parameterschätzung anfällig gegenüber unmodellierter Dynamik

und nicht messbaren, zeitvarianten Störungen [80]. Selbst geringe Störeinflüsse können zu einem unbrauchbaren Schätzergebnis führen und sich auf diese Weise in einem instabilen Regelkreis niederschlagen. Rohrs et al. kommen daher zu dem Schluss, dass solche Regler ohne Zusatzmaßnahmen zur Robustifizierung nicht für praktische Anwendungen einsetzbar sind.

1.4.2 Direkte Adaption

Konzepte, die auf eine explizite Parameterschätzung bzw. Identifikation verzichten und stattdessen den Regler unmittelbar in Abhängigkeit des Prozessverhaltens adaptieren, gliedern sich in die Klasse der „direkten Regler“ ein. Der in Abbildung 1.2 mit „Adaptionsgesetz“ beschriftete Block greift nicht auf einen Parameterschätzer zurück, worin sich der Unterschied zu indirekten Methoden zeigt.

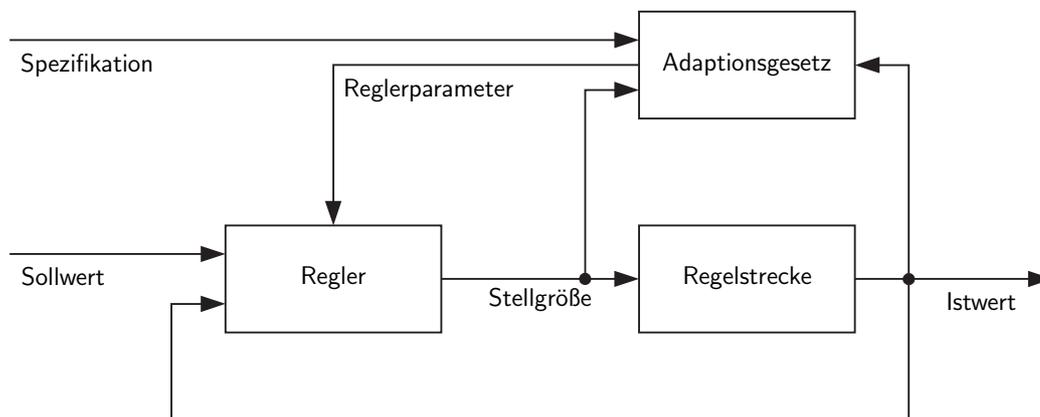


Abbildung 1.2: Prinzipskizze der direkten Adaption.

Bedingt durch den Verzicht auf Parameterbestimmung bzw. Identifikation entfällt auch deren Empfindlichkeit gegenüber Störungen. Ein weiterer Vorzug liegt darin, dass gewöhnlich Modelle technologischer Prozesse durch eine größere Anzahl an Parametern beschrieben werden als die eingesetzten Regler. Eine lineare Strecke der Ordnung $n \in \mathbb{N}$ benötigt mindestens $2n$ Parameter für eine vollständige Beschreibung. Dagegen besitzen PID-Regler mit drei Parametern im Allgemeinen erheblich weniger Einstellmöglichkeiten als das Streckenmodell. Deshalb ist die Adaption der Reglerparameter meist mit geringerem Aufwand verbunden, als die Identifikation der Strecke. Von Bedeutung ist diese Tatsache nicht nur bei der Implementierung, für die Rechenleistung und Speicherkapazitäten nur selten im Überfluss zur Verfügung stehen, sondern auch im Bezug auf den Entwicklungsaufwand [62]:

„Mathematical modelling remains the most difficult part of most control engineering projects, and it consumes the great majority of the control engineer’s time. [...] Modelling is an expensive activity, and in many potential application areas modelling is perceived to be uneconomic.“

Obwohl eine kostspielige Modellbildung zum Beispiel in den Bereichen Luft- und Raumfahrt (Modellierungsaufwand ist gering im Verhältnis zu den Gesamtkosten des Endprodukts), Automobilbranche (hohe Stückzahlen bei identischen Regelstrecken) oder auch petrochemische Industrie (die Regelgüte ist hohen Qualitätsanforderungen unterworfen) vertretbar sein kann, scheinen vor diesem Hintergrund nicht-modellbasierte Ansätze vorteilhaft. Ohne auf ein geeignet parametrisiertes Streckenmodell zurückzugreifen, ist beispielsweise die Adaption hochverstärkungsbasierter Regler möglich. Auf eine ausführliche Beschreibung wird an dieser Stelle verzichtet, weil solche Regler als Gegenstand der Kapitel 3 und 4 dort eingehend dargestellt sind. Es sei hier lediglich die Besonderheit erwähnt, dass eine Untergruppe der hochverstärkungsbasierten Regler die Einhaltung einer vorab festgelegten Fehlerschranke nicht nur asymptotisch, sondern bereits unmittelbar nach Inbetriebnahme der Regelung garantiert. Damit lässt sich das transiente Verhalten des Regelkreises an gegebene Anforderungen anpassen.

1.5 Zielsetzung

In der Antriebstechnik sind als Regelstrecken häufig schwingungsfähige mechanische Systeme gegeben, die üblicherweise durch nicht messbare Störungen (z.B. Lastmomente) beeinflusst sind, oftmals nichtlineare Elemente mit unbekannter Charakteristik (z.B. Reibkennlinien) beinhalten und nicht selten auch Parameter mit unbekanntem Betrag aufweisen. Darüberhinaus ist Robustheit gegenüber Messrauschen eine unverzichtbare Eigenschaft geeigneter Regler. Da Parameterschätzungen für solche Prozesse meist scheitern, wird in der vorliegenden Arbeit ein nicht-parameterbasierter Ansatz verfolgt. Ein solche Strategie, die nicht auf die Parameterwerte eines Modells zurückgreift, wird in der englischsprachigen Literatur (siehe z.B. [35]) mit dem Attribut „non-identifier-based“ gekennzeichnet.

Die Zielsetzung besteht im Entwurf eines Reglers für nichtlineare Antriebssysteme mit unbekanntem Parametern, der Störeinflüssen toleriert, nur geringen Implementierungsaufwand erfordert und daher das Potential für industrielle Anwendbarkeit besitzt. Darüberhinaus soll nur eine Modellierung struktureller Art erfolgen, auf die Nutzung geschätzter Modellparameter ist zu verzichten, so dass der Aufwand für die Modellparametrierung entfällt.

1.6 Gliederung der Arbeit

Zunächst wird in Kapitel 2 der Laboraufbau des Zweimassensystems beschrieben und ein lineares Modell erstellt. Daran schließt sich die Auslegung linearer Regler nach konventionellen Methoden an, die ein exaktes Modell voraussetzen, sowie eine ausführliche Analyse und Bewertung. Es folgt eine Darstellung diverser Reibmodelle und deren Integration in die Differentialgleichung des Zweimassensystems. Dabei ist zu betonen, dass sich die Nutzung des Modells allein auf die Untersuchung struktureller Eigenschaften beschränkt – die Parametrierung ist für den späteren Reglerentwurf unerheblich.

Kapitel 3 beginnt mit einer Betrachtung konstanter, hoher Verstärkungen im linearen und nichtlinearen Regelkreis. Für die Stabilisierung linearer Strecken mit unbekanntem Parametern werden adaptive hochverstärkungsbasierte Regler mit monotonem Anpassungsgesetz diskutiert. Das Verhalten solcher Regler im Falle von Festwert- und Folgeregelung, sowie unter dem Einfluss von Störgrößen wird im Anschluss beleuchtet.

Eine spezielle Form der hochverstärkungsbasierten Regelung ist Inhalt von Kapitel 4. Diese zeichnet sich durch eine nicht-monotone Reglerverstärkung aus und garantiert für den Regelfehler als Besonderheit die Einhaltung eines weitgehend frei wählbaren Grenzwertverlaufes. Die hiermit verbundene Aussage übertrifft gewöhnliche Stabilitätsbeweise, die lediglich die Existenz einer Schranke für den Regelfehler sicherstellen. Das Prinzip des internen Modells wird aufgegriffen, um nicht nur einen beliebig kleinen, sondern asymptotisch verschwindenden Regelfehler zu erreichen.

Die Anwendung der beschriebenen Regler setzt einen Prozess mit Relativgrad 1 voraus. Antriebssysteme entsprechen dieser Forderung meist nicht, weshalb der Relativgrad des betrachteten Zweimassensystems auf geeignete Weise zu reduzieren ist. Der Problematik hoher Relativgrade widmet sich Kapitel 5, beschreibt bekannte Ansätze und zieht eine Zustandsauskopplung als praktikable Lösungsmöglichkeit heran.

Kapitel 6 setzt auf den Grundlagen von Kapitel 4 auf und führt Zusatzmaßnahmen ein, welche sowohl die aktive Bedämpfung von Schwingungen in der elastischen Welle, als auch einen verschwindenden Regelfehler in der Ruhelage ermöglichen und dadurch die Regelgüte verbessern. Es werden ein Hochpassfilter, sowie ein PI-Regler in das theoretische Konzept eingebettet und der Nachweis geführt, dass diese Erweiterungen die strukturellen Eigenschaften des Zweimassensystems beibehalten. Abschließend erfolgt die experimentelle Validierung der entworfenen Regler an der Versuchsanlage.

Nach Kapitel 7 mit Zusammenfassung und Ausblick folgt im Anhang eine Zusammenstellung und Erläuterung mathematischer und systemtheoretischer Grundlagen. Ohne den Anspruch auf Vollständigkeit zu erheben, fokussiert sich die getroffene Auswahl auf Definitionen und Eigenschaften, die im Kontext der vorliegenden Arbeit Relevanz besitzen.

Kapitel 2

Das Zweimassensystem

Die Motivation für die regelungstechnische Betrachtung eines Zweimassensystems und den Aufbau eines entsprechenden Prüfstandes besteht vor allem darin, die Verhältnisse in Antriebssystemen realitätsnah nachzubilden. In mechatronischen Systemen ist der antreibende Motor üblicherweise nicht direkt an der Arbeitsmaschine angebracht, sondern ist über eine Kraftübertragungseinheit angekoppelt. Dabei können Wellen, Zahnriemen, Zugseile, Getriebe oder Kupplungen verwendet sein, die elastisches Verhalten in das System einbringen. Ebenso kann die Arbeitsmaschine selbst Elastizitäten besitzen, die für die Regelung des Antriebes berücksichtigt werden müssen. Hierfür sind Roboterarme gängige Beispiele, wenn bei deren Konstruktion Kompromisse zwischen Steifigkeit und Gewicht geschlossen werden müssen. Ebenso lässt sich die dominante Dynamik von Schreib-Leseköpfen in Computerfestplatten, Antrieben in kontinuierlichen Fertigungsanlagen oder von Vorschüben in Werkzeugmaschinen als Zweimassensystem modellieren. Damit steht ein Zweimassensystem beispielhaft für viele Antriebsanordnungen, da sich starre Verbindungen in der Praxis nicht bzw. nur selten näherungsweise realisieren lassen.

Besteht dagegen eine ideale, also starre Kopplung beider Massen, sind die Drehzahlen von Antriebsmasse und Lastmasse identisch und es liegt ein Einmassensystem vor. Das Massenträgheitsmoment der Gesamtanlage entspricht hier der Summe der beiden Trägheitsmomente von Antriebs- und Arbeitsmaschine. Um jedoch reale Kopplungen berücksichtigen zu können, muss eine elastische Verbindung betrachtet werden, wodurch ein schwingungsfähiges Gesamtsystem entsteht.

Für einen Laboraufbau ist als Last ein Motor zweckmäßig, der an Stelle einer komplexen und spezifischen mechanischen Anordnung die Aufgabe einer Arbeitsmaschine übernimmt und das entsprechende Widerstandsmoment einprägt. Dies stellt auf der Lastseite eine hohe Flexibilität bereit, so dass weitgehend beliebige Verläufe von Störmomenten auf der Arbeitsmaschinenseite aufgebracht werden können. Dadurch lassen sich verschiedene reale Situationen im Betrieb elektrischer Antriebssysteme sehr einfach nachbilden. Wie Abbildung 2.1 zeigt, besteht der Prüfstand „Zweimassensystem“ daher aus zwei Maschinen, die mit einer elastischen Welle verbunden sind und bietet damit eine wertvolle Möglichkeit, um entwickelte Regelungsstrategien experimentell zu validieren und deren Potential für eine industrielle Anwendung abschätzen

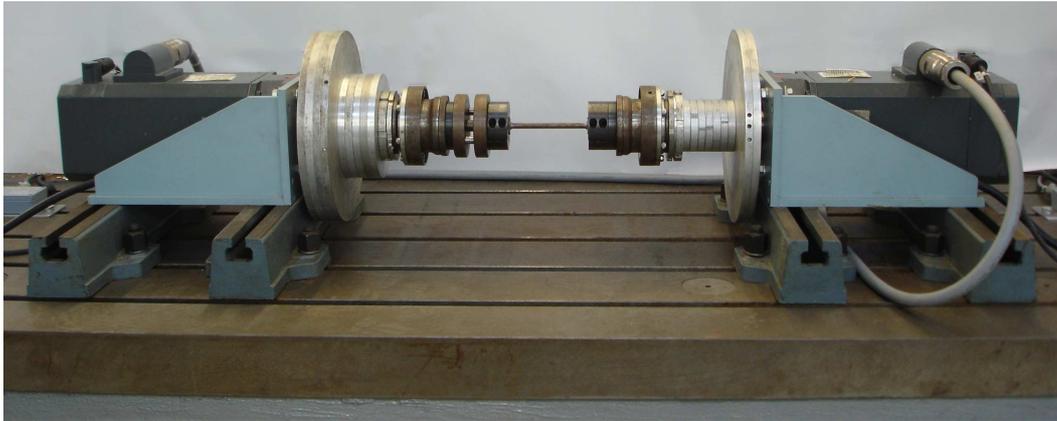


Abbildung 2.1: Laboraufbau des Zweimassensystems.

zu können. Die technischen Daten der beiden baugleichen Synchronmaschinen sind in Tabelle 2.1 angegeben.

| | |
|--------------|-------------------------------|
| Nennzahl | $N_N = 2000 \text{ min}^{-1}$ |
| Nennmoment | $M_N = 25 \text{ Nm}$ |
| Nennleistung | $P_N = 5.2 \text{ kW}$ |

Tabelle 2.1: Technische Daten der elektrischen Antriebe.

Das klassische regelungstechnische Problem bei schwingungsfähigen Antrieben stellt dabei die Bedämpfung der Eigenschwingung zusammen mit dem Erreichen eines guten Folgeverhaltens für die Drehzahl der Arbeitsmaschine dar.

2.1 Lineares Modell

Üblicherweise werden elastische Kopplungen als Feder-Dämpfer Systeme modelliert [85, Kap. 18]. Ein mechanisches Modell hierfür ist in Abbildung 2.2 angegeben. Eine An-

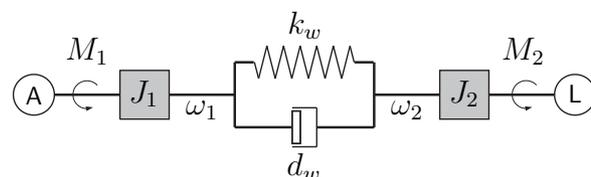


Abbildung 2.2: Mechanisches Ersatzmodell des Zweimassensystems.

triebsmaschine A bringt ein Drehmoment $M_1(t)$ auf, das durch ein elastisches Element an die Lastmaschine (=Arbeitsmaschine) L übertragen wird. Diese bildet das Verhalten der anzutreibenden Mechanik nach, indem deren Rückwirkung auf den Antriebsstrang durch ein entsprechendes Lastmoment $M_2(t)$ generiert wird. Dadurch können

ohne aufwändige Umbauarbeiten am Prüfstand unterschiedliche mechanische Systeme simuliert werden.

Die Geometrie und die Massen beider Maschinen werden durch die Massenträgheitsmomente $J_1 > 0$ und $J_2 > 0$ beschrieben. Durch die Einwirkung der Drehmomente $M_1(t)$ bzw. $M_2(t)$ auf diese Massen entstehen dort die beiden Drehzahlen $\omega_1(t)$ bzw. $\omega_2(t)$. Werden die Positionen der Massen zum Zeitpunkt t durch die Winkelstellungen $\varphi_1(t)$ und $\varphi_2(t)$ bezeichnet, gibt die Größe

$$\varphi_w(t) = \varphi_1(t) - \varphi_2(t) \quad (2.1)$$

den Differenzwinkel zwischen beiden Massen, also die Torsion der Welle an. Durch $\dot{\varphi}_j(t) = \omega_j(t)$ und $\ddot{\varphi}_j(t)$ wird die Drehgeschwindigkeit bzw. die Beschleunigung der entsprechenden Masse $j \in \{1, 2\}$ beschrieben.

Die Elastizität der Verbindungswelle wird durch eine mechanische Drehfeder dargestellt und durch ein Feder-Dämpfer System mit der Federhärte $k_w > 0$ und der Dämpfung $d_w > 0$ modelliert. Mit diesem Grundbaustein kann bereits das Verhalten einer Vielzahl von mechanischen Anordnungen hinreichend genau beschrieben werden, obwohl nicht-lineare Effekte (Lose im Getriebe bzw. Kupplung, Haft- und Gleitreibung) hier noch nicht berücksichtigt sind. Damit muss eine gewisse Diskrepanz zwischen dem Verhalten des Modells und der realen Anlage in Kauf genommen werden.

Durch die Betrachtung der Welle als Feder-Dämpfer System setzt sich das Wellenmoment aus zwei Einzelkomponenten zusammen und ist abhängig vom Torsionswinkel $\varphi_w(t)$. Nach dem Hooke'schen Gesetz wird das Moment $M_k(t) = k_w \cdot \varphi_w(t)$ von einer Drehfeder übertragen. Zum anderen tritt ein geschwindigkeitsabhängiges Moment $M_d(t) = d_w \cdot \dot{\varphi}_w(t)$ aufgrund des Dämpfers auf. Damit ergibt sich in Summe das Drehmoment

$$M_w(t) = M_k(t) + M_d(t) = k_w \cdot \varphi_w(t) + d_w \cdot \dot{\varphi}_w(t) \quad (2.2)$$

in der Welle, das vom Verdrehwinkel zwischen beiden Massen, sowie von dessen zeitlicher Änderung abhängt.

An der Masse der Antriebsmaschine greift das vom Motor erzeugte Antriebsmoment $M_1(t)$ an, die Masse der Arbeitsmaschine ist dem Widerstandsmoment $M_2(t)$ unterworfen. Wird eine Vorzeichenfestlegung derart getroffen, dass die Lastmasse durch ein positives Wellenmoment beschleunigt wird, so wirkt wegen des physikalischen Grundsatzes „actio gegengleich reactio“ ein positives Wellenmoment daher bremsend auf die Antriebsmasse. Damit lässt sich an beiden Massen eine Drehmomentenbilanz aufstellen, wobei nach d'Alembert der Term $J_j \ddot{\varphi}_j(t)$, $j \in \{1, 2\}$, dem Beschleunigungsmoment entspricht:

$$J_1 \ddot{\varphi}_1(t) = M_1(t) - M_w(t) = M_1(t) - [k_w \cdot \varphi_w(t) + d_w \cdot \dot{\varphi}_w(t)] \quad (2.3)$$

$$J_2 \ddot{\varphi}_2(t) = M_w(t) - M_2(t) = [k_w \cdot \varphi_w(t) + d_w \cdot \dot{\varphi}_w(t)] - M_2(t) \quad (2.4)$$

Diese Gleichungen können als Signalflussplan dargestellt werden, der in Abbildung 2.3 gegeben ist. Eine physikalisch motivierte Zustandsbeschreibung ergibt sich auf Basis

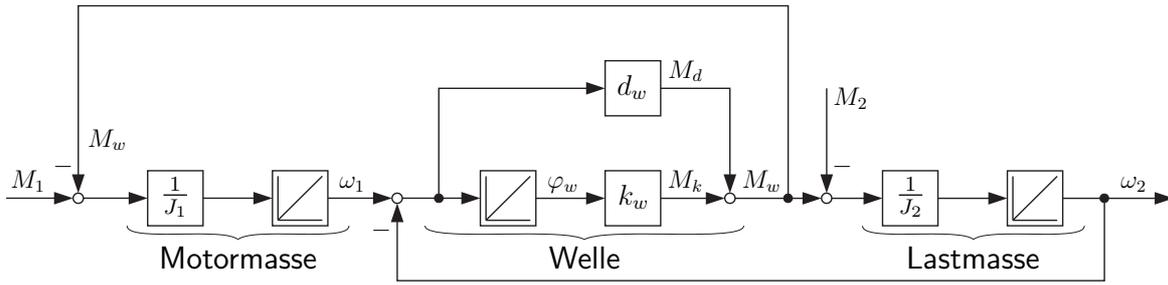


Abbildung 2.3: Signalflussplan für das Zweimassensystem in physikalischer Form.

der drei Größen $\omega_1(t)$, $\omega_2(t)$ und $\varphi_w(t)$. Mit dem Zustandsvektor

$$x(t) = \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ \omega_1(t) \end{pmatrix}, \quad \begin{array}{l} \omega_2(t) : \text{ Lastdrehzahl zum Zeitpunkt } t \\ \varphi_w(t) : \text{ Verdrehwinkel zum Zeitpunkt } t \\ \omega_1(t) : \text{ Motordrehzahl zum Zeitpunkt } t \end{array} \quad (2.5)$$

und zusammen mit der Momentenbilanz und der Beschreibung des Feder-Dämpfer Elementes folgt damit das lineare Zustandsraummodell [85, S. 821]

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t) + b_L M_2(t), & x(0) &\in \mathbb{R}^3 \\ y(t) &= cx(t) \end{aligned} \quad (2.6)$$

mit der Systemmatrix und den Einkoppelvektoren

$$A = \begin{bmatrix} -d_w/J_2 & k_w/J_2 & d_w/J_2 \\ -1 & 0 & 1 \\ d_w/J_1 & -k_w/J_1 & -d_w/J_1 \end{bmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ 1/J_1 \end{pmatrix}, \quad b_L = \begin{pmatrix} -1/J_2 \\ 0 \\ 0 \end{pmatrix}. \quad (2.7)$$

Der Auskoppelvektor $c \in \mathbb{R}^{1 \times 3}$ hingegen ist anwendungsabhängig durch

$$c = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \quad c = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}, \quad \text{oder durch} \quad c = \begin{pmatrix} c_1 & c_2 & c_3 \end{pmatrix} \quad (2.8)$$

gegeben. Die Eingangsgröße $u(t) = M_1(t)$ stimmt mit dem Antriebsmoment überein, die Störgröße $M_2(t)$ ist ein Lastmoment, das im Falle von Reibung von der Lastdrehzahl $\omega_2(t)$ abhängen kann. Der Modellierung in [85, S. 822] folgend, bleibt auch hier der Einfluss eines vorhandenen Umrichters unberücksichtigt, weil moderne Stellglieder mit hohen Taktfrequenzen arbeiten und dadurch die unvermeidliche Verzögerung vernachlässigbar gering halten.

Basierend auf bekannten Materialkonstanten und geometrischen Gegebenheiten der Anlage ermöglichen die Gesetze der Mechanik eine Berechnung der Massenträgheitsmomente. Sei $R \geq 0$ der Abstand eines Massenelementes von der Drehachse, ergibt das Integral über die rotierende Gesamtmasse $m > 0$ das Massenträgheitsmoment zu

$$J = \int_m R^2 dm \geq 0. \quad (2.9)$$

Verschiedene Messungen und Versuche erlauben die Bestimmung der Reibung beispielsweise durch Auslaufversuche oder die Bestimmung der Federhärte durch Belastungsversuche. Letztgenannte ergibt sich als Quotient aus aufgebrachtem Drehmoment und daraus resultierender Wellentorsion.

Neben solchen „Messungen“ der Parameter besteht als weitere Möglichkeit die Anwendung diverser Identifikations- und Schätzverfahren [5], [32], [78], [84].

Die für den Prüfstand ermittelten Parameterwerte sind in Tabelle 2.2 angegeben. Naturgemäß verhindern Messunsicherheiten eine exakte Bestimmung, dennoch zeigen die Parameter ausreichende Genauigkeit, um funktionsfähige Regler auslegen zu können und geben damit einen geeigneten Anhaltspunkt an, bzw. können als nominelle Werte betrachtet werden. Deshalb lassen sich am Laboraufbau diverse konventionelle Regler in Betrieb nehmen und ermöglichen somit einen Vergleich mit neuartigen Regelungskonzepten.

| | |
|--------------------------------|------------------------------|
| Trägheitsmoment des Antriebs | $J_1 = 0.166 \text{ kgm}^2$ |
| Trägheitsmoment des Abtriebs | $J_2 = 0.333 \text{ kgm}^2$ |
| Federhärte der Welle | $k_w = 410 \text{ Nm/rad}$ |
| Dämpfungskoeffizient der Welle | $d_w = 0.25 \text{ Nms/rad}$ |

Tabelle 2.2: Nominelle Parameterwerte des Zweimassensystems.

Weil allerdings die Parameteridentifikation im Allgemeinen hohen Aufwand erfordert, sind vor allem für die breitgefächerte Anwendung Regler erwünscht, die eine Kenntnis dieser Werte nicht verlangen. Insofern sollen die Systemparameter beim Entwurf einer Regelung als unbekannt angenommen werden. Jedoch ist offensichtlich, dass aufgrund physikalischer Gegebenheiten sämtliche Parameter auf positive Wertebereiche beschränkt sind und

$$J_1 > 0, \quad J_2 > 0, \quad k_w > 0, \quad d_w > 0 \quad (2.10)$$

gilt. Allein diese Information über die Parameter soll bei der Systemanalyse und Reglersynthese zur Verfügung stehen. Die Verwendung des Begriffes „unbekannter Parameter“ ist im Folgenden daher mit der korrekten, aber umständlichen Formulierung „unbekannter Betrag, aber positiver Wert des Parameters“ gleichzusetzen.

2.2 Analyse des Systemverhaltens

Das Verhalten des linearen Zweimassensystems lässt sich vorteilhaft an Übertragungsfunktionen analysieren. Hierzu müssen jedoch Eingangs- und Ausgangssignal festgelegt sein. Vor allem die Wahl des Ausgangs ist per se nicht eindeutig, sondern hängt vom Anwendungsfall ab, d.h. vom Zweck für den die Übertragungsfunktion aufgestellt werden soll. Üblicherweise werden Führungs- und Störübertragungsfunktionen bestimmt,

die sich durch die Festlegung des Eingangssignales unterscheiden. Im Falle des Zweimassensystems ist daher sowohl das Drehmoment der Antriebsmaschine als auch das Lastmoment als Systemeingang denkbar.

Als Ausgangssignal hingegen eignen sich die Drehzahlen $\omega_1(t)$ und $\omega_2(t)$, sowie eine Linearkombination sämtlicher Elemente des Zustandsvektors $x(t)$, die durch die unterschiedlichen Auskoppelvektoren in Gleichung (2.8) festgelegt werden.

Die Übertragungsfunktion beschreibt das E/A-Verhalten eines dynamischen Systems, ohne dabei auf die Systemzustände $x(t)$ zurückzugreifen. Deshalb ändert sich die Übertragungsfunktion bei einem Basiswechsel (z.B. Koordinatentransformation auf Rege­lungsnormform oder Byrnes-Isidori-Normform) nicht.

Im Folgenden werden die Übertragungsfunktionen angegeben und untersucht, wenn eines der drei genannten Signale als Ausgang festgesetzt wird.

2.2.1 Motordrehzahl als Ausgang

Wird die Differentialgleichung (2.6) mit dem Auskoppelvektor $c = (0, 0, 1)$ einer Laplace-Transformation unterzogen, ergibt sich gemäß

$$F(s) = \frac{y(s)}{u(s)} = \frac{\omega_1(s)}{u(s)} = c(sI_3 - A)^{-1}b \quad (2.11)$$

die Führungsübertragungsfunktion

$$F(s) = \frac{\frac{1}{J_1}s^2 + \frac{d_w}{J_1 J_2}s + \frac{k_w}{J_1 J_2}}{s^3 + d_w \frac{J_1+J_2}{J_1 J_2}s^2 + k_w \frac{J_1+J_2}{J_1 J_2}s} = \underbrace{\frac{1}{s(J_1 + J_2)}}_{\text{starre Kopplung}} \cdot \underbrace{\frac{1 + s \frac{d_w}{k_w} + s^2 \frac{J_2}{k_w}}{1 + s \frac{d_w}{k_w} + s^2 \frac{J_1 J_2}{J_1+J_2} \frac{1}{k_w}}}_{\text{elastische Welle}} \quad (2.12)$$

der Strecke. Diese besteht aus einem Anteil, der einer starren Kopplung entspricht, sowie einem Anteil, der durch die elastische Verbindung hervorgerufen wird. Bei starrer Kopplung ($k_w \rightarrow \infty$) entfällt der Einfluss der Welle und es verbleibt im Nenner ein Polynom erster Ordnung, welches das integrale Verhalten eines Einmassensystems beschreibt.

Es liegt ein Nennerpolynom der Ordnung $n = 3$ und ein Zählerpolynom der Ordnung $m = 2$ vor. Eine der drei Polstellen ist der starren Kopplung zugeordnet und befindet sich unabhängig aller Parameterwerte im Ursprung der komplexen s -Ebene, d.h. $p_1 = 0$. Die elastische Welle ruft zwei weitere Pole hervor, die durch

$$\begin{aligned} p_{2,3} &= \frac{-d_w \pm \sqrt{d_w^2 - 4 \frac{J_1 J_2}{J_1+J_2} k_w}}{2 \frac{J_1 J_2}{J_1+J_2}} = \\ &= -\frac{1}{2} \frac{J_1 + J_2}{J_1 J_2} d_w \pm \frac{1}{2} \frac{1}{J_1 J_2} \sqrt{(J_1 + J_2)[(J_1 + J_2)d_w^2 - 4J_1 J_2 k_w]} \quad (2.13) \end{aligned}$$

gegeben sind. Da sämtliche Parameter ausschließlich positive Werte annehmen, gilt für den Radikanten in Gleichung (2.13) die Ungleichung $d_w^2 - 4 \frac{J_1 J_2}{J_1 + J_2} k_w < d_w^2$. Daraus folgt, dass nur negative Realteile für die Polstellen auftreten können, das Zweimassensystem ist daher unabhängig von den exakten Parameterwerten stets stabil.

Technisch relevant sind vor allem Parameterkonstellationen, welche die Bedingung

$$d_w < 2\sqrt{k_w \frac{J_1 J_2}{J_1 + J_2}} \quad (2.14)$$

erfüllen. Kleine Dämpfungskoeffizienten führen auf einen negativen Radikanten in (2.13) und sind damit verantwortlich für ein konjugiert komplexes Polpaar. Erwartungsgemäß resultiert schwingungsfähiges Systemverhalten, das sich durch die Übertragungsfunktion eines schwingungsfähigen PT₂-Bausteins

$$F_{PT2}(s) = \frac{K\omega_o^2}{\omega_o^2 + 2D\omega_o s + s^2} = \frac{K}{1 + 2\frac{D}{\omega_o} s + \frac{1}{\omega_o^2} s^2}, \quad K, D \in \mathbb{R}, \quad \omega_o > 0 \quad (2.15)$$

beschreiben lässt. Deren Nenner wird als Normpolynom für einen Koeffizientenvergleich vorgegeben und mit dem Nenner des Faktors, der in der Übertragungsfunktion (2.12) die elastische Kopplung beschreibt, gleichgesetzt:

$$s^2 : \quad \frac{J_1 J_2}{J_1 + J_2} \frac{1}{k_w} = \frac{1}{\omega_o^2} \quad \Rightarrow \quad \omega_o = \sqrt{k_w \frac{J_1 + J_2}{J_1 J_2}} \quad (2.16)$$

$$s : \quad \frac{d_w}{k_w} = 2\frac{D}{\omega_o} \quad \Rightarrow \quad D = \frac{\omega_o d_w}{2 k_w} = \frac{d_w}{2} \sqrt{\frac{1}{k_w} \frac{J_1 + J_2}{J_1 J_2}} \quad (2.17)$$

Zusammen mit den nominellen Werten aus Tabelle 2.2 resultiert daraus die Eigenfrequenz $\omega_o = 60.8$ rad/s, bzw. $f_o = \omega_o/(2\pi) = 9.7$ Hz, sowie der Dämpfungsgrad $D = 0.0185$. Ein Vergleich des Zählers mit dem Normpolynom führt auf

$$\omega_{oZ} = \sqrt{\frac{k_w}{J_2}} = 35.1 \text{ rad/s} \quad \text{und} \quad D_Z = \frac{d_w}{2} \sqrt{\frac{1}{k_w J_2}} = 0.0107. \quad (2.18)$$

Damit bestehen die Verhältnisse

$$\frac{\omega_o}{\omega_{oZ}} = \sqrt{\frac{J_1 + J_2}{J_1}} > 1 \quad \text{und} \quad \frac{D}{D_Z} = \sqrt{\frac{J_1 + J_2}{J_1}} > 1 \quad (2.19)$$

zwischen Nenner- und Zählergrößen, die unabhängig von den Parameterwerten stets größer 1 sind. Die Resonanzfrequenz des Zählertermes liegt daher in jedem Falle unter derjenigen des Nenners.

Im Bereich sehr tiefer Frequenzen ($\omega \ll \omega_{oZ}$, bzw. $s \rightarrow 0$) verschwindet wegen

$$\lim_{s \rightarrow 0} \frac{1 + s \frac{d_w}{k_w} + s^2 \frac{J_2}{k_w}}{1 + s \frac{d_w}{k_w} + s^2 \frac{J_1 J_2}{J_1 + J_2} \frac{1}{k_w}} = 1 \quad (2.20)$$

der Einfluss der elastischen Welle, das System verhält sich wie starr gekoppelt. Im Bodediagramm in Abbildung 2.4 äußert sich dies im Verbleib des integralen Anteils,

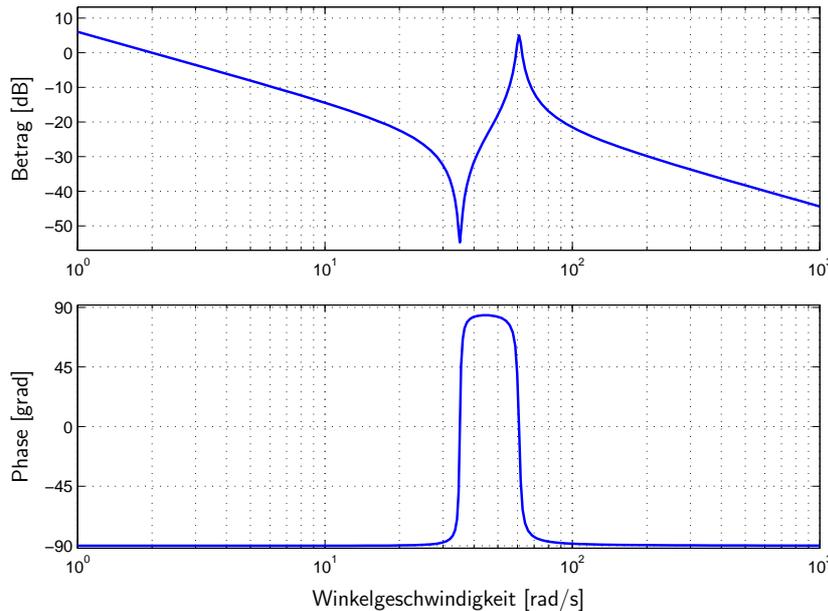


Abbildung 2.4: Bodediagramm für das Zweimassensystem mit Motordrehzahl als Ausgang.

$$F(j\omega) \approx \frac{1}{(J_1 + J_2)j\omega} \quad \text{für } \omega \ll 1 \quad (2.21)$$

der einen Phasenwinkel von -90° und einen Abfall im Betrag von 20dB/Dekade verursacht.

Der weitere Verlauf des Bodediagramms hängt entscheidend von der Lage der Resonanzfrequenzen von Zähler und Nenner ab. Wegen $\omega_{oZ} < \omega_o$ bedingt das Zählerpolynom zweiter Ordnung zunächst eine Verringerung des Betrages und ein Anheben der Phase um 180° , bevor das Nennerpolynom diesen beiden Effekten durch Betragsüberhöhung an der Resonanzstelle ω_o und Absenken der Phase um 180° entgegenwirkt.

Für schnelle Anregungen ($\omega \gg \omega_o$, bzw. $s \rightarrow \infty$) verhält sich wegen

$$\lim_{\omega \rightarrow \infty} \frac{1 + j\omega \frac{d_w}{k_w} - \omega^2 \frac{J_2}{k_w}}{1 + j\omega \frac{d_w}{k_w} - \omega^2 \frac{J_1 J_2}{J_1 + J_2} \frac{1}{k_w}} = \frac{J_1 + J_2}{J_1} \quad (2.22)$$

die Welle als proportionaler Faktor. Deshalb nähert sich der Verlauf des Bodediagrammes wiederum an integrales Verhalten an. Es ist allerdings wegen

$$F(j\omega) \approx \frac{1}{J_1 j\omega} \quad \text{für } \omega \gg 1 \quad (2.23)$$

ausschließlich die Antriebsmaschine maßgeblich, so dass die Betragskennlinie gegenüber den niedrigen Frequenzen um $20 \log_{10}([J_1 + J_2]/J_1) \approx 10$ dB parallel verschoben verläuft.

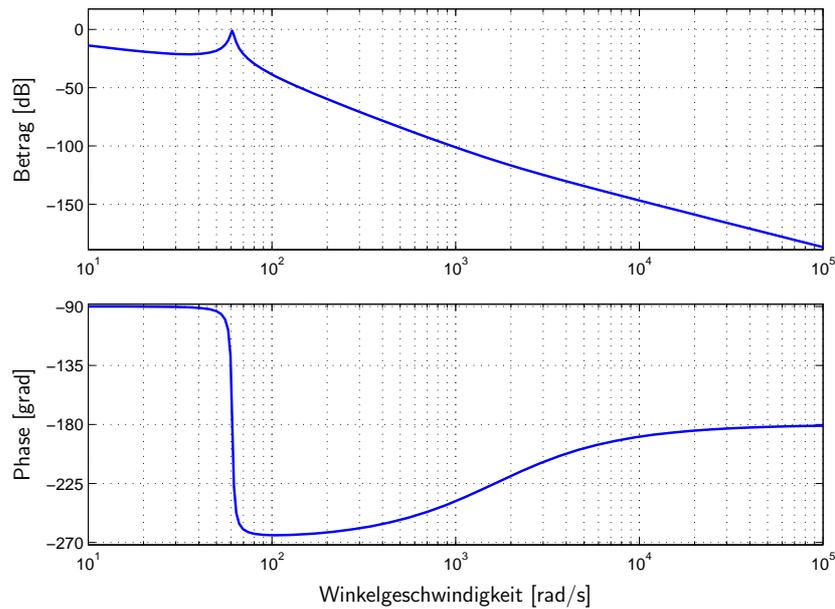


Abbildung 2.5: Bodediagramm für das Zweimassensystem mit Lastdrehzahl als Ausgang.

2.2.2 Lastdrehzahl als Ausgang

Wird die Differentialgleichung (2.6) mit dem Auskoppelvektor $c = (1, 0, 0)$ einer Laplace-Transformation unterzogen, ergibt sich gemäß

$$F(s) = \frac{y(s)}{u(s)} = \frac{\omega_2(s)}{u(s)} = c(sI_3 - A)^{-1}b \quad (2.24)$$

die Führungsübertragungsfunktion

$$F(s) = \frac{\frac{d_w}{J_1 J_2} s + \frac{k_w}{J_1 J_2}}{s^3 + d_w \frac{J_1 + J_2}{J_1 J_2} s^2 + k_w \frac{J_1 + J_2}{J_1 J_2} s} = \frac{1}{s(J_1 + J_2)} \cdot \frac{1 + s \frac{d_w}{k_w}}{1 + s \frac{d_w}{k_w} + s^2 \frac{J_1 J_2}{J_1 + J_2} \frac{1}{k_w}} \quad (2.25)$$

der Strecke. Nachdem sich im Vergleich zu Abschnitt 2.2.1 lediglich das gewählte Ausgangssignal unterscheidet, tritt eine Veränderung ausschließlich im Zählerpolynom auf, das sich auf die Ordnung $m = 1$ reduziert. Der Nenner und damit auch dessen Ordnung bleibt unverändert bestehen. Insofern sind Eigenfrequenz ω_o und Dämpfungsgrad D von der gewählten Auskopplung unabhängig.

Für langsame Frequenzen ist wegen

$$\lim_{s \rightarrow 0} \frac{1 + s \frac{d_w}{k_w}}{1 + s \frac{d_w}{k_w} + s^2 \frac{J_1 J_2}{J_1 + J_2} \frac{1}{k_w}} = 1 \quad (2.26)$$

die elastische Welle wiederum ohne Einfluss, es liegt das Verhalten einer starren Kopplung vor (Einmassensystem). Im Bodediagramm (Abbildung 2.5) fällt der Betrag mit 20dB/Dekade, der Phasenwinkel beginnt bei -90° . In Erscheinung tritt das elastische Verhalten vor allem an der Resonanzfrequenz ω_o . Das konjugiert komplexe Polpaar im

Nenner führt dort zu einer Resonanzüberhöhung im Betrag und zu einer weiteren Phasenabsenkung, die maximal 180° beträgt. Dieser wirkt jedoch der Zähler entgegen, der bei $\omega_{oZ} = k_w/d_w = 1640$ rad/s eine Anhebung um 90° bewirkt. Trotz der entgegenwirkenden Phasenhebung des Zählers geschieht eine Absenkung unter -180° . Im Bereich zwischen ω_o und ω_{oZ} fällt der Betrag um 60dB/Dekade, ab ω_{oZ} mit 40dB/Dekade. Die Phase dreht für höhere Frequenzen auf -180° zurück. Kritisch am Verhalten dieser Regelstrecke ist der schnelle Abfall des Phasenwinkels unter -180° .

2.2.3 Linearkombination als Ausgang

Wird die Differentialgleichung (2.6) mit dem Auskoppelvektor $c = (c_1, c_2, c_3) \in \mathbb{R}^{1 \times 3}$ einer Laplace-Transformation unterzogen, ergibt sich gemäß

$$F(s) = \frac{y(s)}{u(s)} = c(sI_3 - A)^{-1}b \quad (2.27)$$

die Führungsübertragungsfunktion

$$\begin{aligned} F(s) &= \frac{\frac{c_3}{J_1} s^2 + \frac{(c_2 J_2 + (c_1 + c_3) d_w)}{J_1 J_2} s + \frac{(c_1 + c_3) k_w}{J_1 J_2}}{s^3 + d_w \frac{J_1 + J_2}{J_1 J_2} s^2 + k_w \frac{J_1 + J_2}{J_1 J_2} s} = \\ &= \frac{1}{s(J_1 + J_2)} \cdot \frac{\frac{c_3 J_2}{k_w} s^2 + \frac{(c_2 J_2 + (c_1 + c_3) d_w)}{k_w} s + (c_1 + c_3)}{1 + s \frac{d_w}{k_w} + s^2 \frac{J_1 J_2}{J_1 + J_2} \frac{1}{k_w}} \quad (2.28) \end{aligned}$$

der Strecke. Die Auskopplung beeinflusst das Nennerpolynom nicht, die Koeffizienten des Zählerpolynoms können dagegen frei eingestellt werden. Im Folgenden wird die Fragestellung behandelt, welche Kombinationen der Rückführungen zu einem beobachtbaren System führen, bzw. unter welchen Bedingungen der Verlust der Beobachtbarkeit eintritt und daher eine Pol-Nullstellen-Kürzung in der Übertragungsfunktion (2.28) entsteht. Hierzu werden die beiden Fälle $c_2 = 0$ und $c_2 \neq 0$ unterschieden.

2.2.3.1 Beobachtbarkeit ohne Verwendung des Verdrehwinkels $c_2 = 0$

Die Eigenschaft der Beobachtbarkeit lässt sich am Rang der zugehörigen Beobachtbarkeitsmatrix ablesen. Unter der Restriktion $c_2 = 0$ lautet die Beobachtbarkeitsmatrix

$$\begin{aligned} Q &= \begin{bmatrix} c \cdot A^0 \\ c \cdot A^1 \\ c \cdot A^2 \end{bmatrix} = \\ &= \begin{bmatrix} c_1 & 0 & c_3 \\ -d_w \frac{c_1 J_1 - c_3 J_2}{J_1 J_2} & k_w \frac{c_1 J_1 - c_3 J_2}{J_1 J_2} & d_w \frac{c_1 J_1 - c_3 J_2}{J_1 J_2} \\ \frac{(k_w J_1 J_2 - d_w^2 J_1 - d_w^2 J_2)(c_3 J_2 - c_1 J_1)}{J_1^2 J_2^2} & \frac{k_w d_w (J_1 + J_2)(c_3 J_2 - c_1 J_1)}{J_1^2 J_2^2} & \frac{(d_w^2 J_1 - k_w J_1 J_2 + d_w^2 J_2)(c_3 J_2 - c_1 J_1)}{J_1^2 J_2^2} \end{bmatrix} \quad (2.29) \end{aligned}$$

und ist singular, wenn deren Determinante

$$\det(Q) = \frac{k_w^2}{J_1^2 J_2^2} (c_1 J_1 - c_3 J_2)^2 (c_1 + c_3) \quad (2.30)$$

den Wert 0 annimmt. Damit geht die Beobachtbarkeit in zwei Fällen verloren:

Fall 1:

Das Zweimassensystem ist nicht vollständig beobachtbar, wenn die Konstellation

$$c_1 = -c_3 \quad \Rightarrow \quad \det(Q) = 0 \quad (2.31)$$

angesetzt wird. Im Grenzfall (2.31) entsteht die Übertragungsfunktion

$$F(s) = \frac{c_3 J_2 s}{J_1 J_2 s^2 + (J_1 + J_2) d_w s + (J_1 + J_2) k_w} \quad (2.32)$$

deren Ordnung auf 2 reduziert ist, worin sich der Verlust der Beobachtbarkeit ausdrückt. Der integrale Anteil im Nenner ist entfallen, lediglich das konjugiert komplexe Polpaar verbleibt. Gegenphasige Schwingungen der Massen sind am Ausgang zwar sichtbar, nicht jedoch die Drehzahl im stationären Zustand. Weil in der Gleichgewichtslage $\omega_1(t) = \omega_2(t)$ gilt, ergibt sich für $c_1 = -c_3$ und $c_2 = 0$ das Ausgangssignal $y(t) = c_1 \omega_2(t) + c_3 \omega_1(t) = 0$, unabhängig davon, welche Drehzahl vorliegt.

Fall 2:

Bei Auskoppelkoeffizienten mit positivem Vorzeichen kann die Lösung

$$c_1 = c_3 \frac{J_2}{J_1} \quad \Rightarrow \quad \det(Q) = 0 \quad (2.33)$$

zutreffen. Die gekürzte Übertragungsfunktion

$$F(s) = \frac{c_3}{J_1} \cdot \frac{J_1 J_2 s^2 + (J_1 + J_2) d_w s + (J_1 + J_2) k_w}{J_1 J_2 s^2 + (J_1 + J_2) d_w s + (J_1 + J_2) k_w} \cdot \frac{1}{s} = \frac{c_3}{J_1 s} \quad (2.34)$$

weist die Ordnung 1 auf und verliert das konjugiert komplexe Polpaar. Hierin spiegelt sich die Tatsache wider, dass die Schwingungen in der Welle vom Ausgang aus nicht mehr sichtbar sind, wenn die Auskopplung der Gleichung (2.33) genügt.

Zusammenfassend ist festzuhalten, dass zwei Kombinationen von c_1 und c_3 existieren, für die unter der Voraussetzung $c_2 = 0$ die Beobachtbarkeit verloren geht.

2.2.3.2 Beobachtbarkeit bei Verwendung des Verdrehwinkels $c_2 \neq 0$

Für die Analyse der Beobachtbarkeit unter Berücksichtigung des Torsionswinkels mit $c_2 \neq 0$ kann wiederum der Rangabfall der Beobachtbarkeitsmatrix bestimmt werden. Allerdings entsteht bei der Berechnung der Determinante ein unübersichtliches Polynom dritter Ordnung in den Koeffizienten c_1 , c_2 und c_3 , so dass von dieser Vorgehensweise Abstand genommen werden soll. Die einfachere Methode ist hier, Zähler- und

Nennerpolynom der Übertragungsfunktion (2.28) zu faktorisieren und anschließend zu beantworten, für welche Konstellationen der Rückführkoeffizienten eine Kürzung der Linearfaktoren eintritt.

Aus der Berechnung der Polstellen p_1 , p_2 und p_3 in Abschnitt 2.2.1 können die drei Linearfaktoren des Nenners entnommen werden. Neben dem Faktor $s - p_1 = s$ treten die beiden Faktoren $s - p_2$ und $s - p_3$ auf, die dem konjugiert komplexen Polpaar entsprechen und durch Gleichung (2.13) gegeben sind.

Der Radikant

$$R_N := (J_1 + J_2)^2 d_w^2 - 4J_1 J_2 (J_1 + J_2) k_w < 0 \quad (2.35)$$

in

$$p_2 = \frac{-(J_1 + J_2)d_w + \sqrt{R_N}}{2J_1 J_2} \in \mathbb{C} \quad \text{und} \quad p_3 = \frac{-(J_1 + J_2)d_w - \sqrt{R_N}}{2J_1 J_2} \in \mathbb{C} \quad (2.36)$$

ist negativ, weil laut Annahme ein schwingungsfähiges System vorliegt. Das Zählerpolynom zweiter Ordnung besitzt die beiden Nullstellen

$$q_1 = \frac{-c_2 J_2 - (c_1 + c_3)d_w - \sqrt{R_Z}}{2c_3 J_2} \quad \text{und} \quad q_2 = \frac{-c_2 J_2 - (c_1 + c_3)d_w + \sqrt{R_Z}}{2c_3 J_2}, \quad (2.37)$$

wobei

$$R_Z = (c_2 J_2 + (c_1 + c_3)d_w)^2 - 4J_2 k_w c_3 (c_1 + c_3) \quad (2.38)$$

von den Auskoppelkoeffizienten und den Systemparametern abhängt. Daher kann über das Vorzeichen von R_Z zunächst keine Aussage getroffen werden, die Nullstellen können sowohl reelle als auch komplexe Werte annehmen.

Als Ergebnis lässt sich schließlich die Übertragungsfunktion in der gesuchten Form

$$F(s) = \frac{c_3 \cdot (s - q_1) \cdot (s - q_2)}{J_1 \cdot s \cdot (s - p_2) \cdot (s - p_3)} \quad (2.39)$$

darstellen.

Die Beobachtbarkeit geht verloren für alle Konstellationen der Auskopplungen, welche die Nullstellen derart platzieren, dass eine Kürzung mit den Polstellen eintritt. Daraus leiten sich prinzipiell sechs Möglichkeiten für eine Pol-Nullstellen-Kürzung ab, die jeweils eine Bedingung an den Realteil und eine an den Imaginärteil des zu kürzenden Linearfaktors stellen:

Fall 1:

Kürzung von $s - q_1$ mit s :

$$\Rightarrow c_2 J_2 + (c_1 + c_3)d_w + |\sqrt{R_Z}| = 0 \quad \text{und} \quad R_Z \geq 0$$

Fall 2:

Kürzung von $s - q_2$ mit s :

$$\Rightarrow c_2 J_2 + (c_1 + c_3)d_w - |\sqrt{R_Z}| = 0 \quad \text{und} \quad R_Z \geq 0$$

Fall 3:

Kürzung von $s - q_1$ mit $s - p_2$:

$$\Rightarrow \frac{c_2 J_2 + (c_1 + c_3) d_w}{2 c_3 J_2} = \frac{(J_1 + J_2) d_w}{2 J_1 J_2} \quad \text{und} \quad \underbrace{\frac{|\sqrt{-R_Z}|}{2 c_3 J_2}}_{>0} = \underbrace{\frac{-|\sqrt{-R_N}|}{2 J_1 J_2}}_{<0} \quad \text{mit} \quad R_Z < 0$$

Fall 4:

Kürzung von $s - q_2$ mit $s - p_2$:

$$\Rightarrow \frac{c_2 J_2 + (c_1 + c_3) d_w}{2 c_3 J_2} = \frac{(J_1 + J_2) d_w}{2 J_1 J_2} \quad \text{und} \quad \underbrace{\frac{-|\sqrt{-R_Z}|}{2 c_3 J_2}}_{<0} = \underbrace{\frac{-|\sqrt{-R_N}|}{2 J_1 J_2}}_{<0} \quad \text{mit} \quad R_Z < 0$$

Fall 5:

Kürzung von $s - q_1$ mit $s - p_3$:

$$\Rightarrow \frac{c_2 J_2 + (c_1 + c_3) d_w}{2 c_3 J_2} = \frac{(J_1 + J_2) d_w}{2 J_1 J_2} \quad \text{und} \quad \underbrace{\frac{|\sqrt{-R_Z}|}{2 c_3 J_2}}_{>0} = \underbrace{\frac{|\sqrt{-R_N}|}{2 J_1 J_2}}_{>0} \quad \text{mit} \quad R_Z < 0$$

Fall 6:

Kürzung von $s - q_2$ mit $s - p_3$:

$$\Rightarrow \frac{c_2 J_2 + (c_1 + c_3) d_w}{2 c_3 J_2} = \frac{(J_1 + J_2) d_w}{2 J_1 J_2} \quad \text{und} \quad \underbrace{\frac{-|\sqrt{-R_Z}|}{2 c_3 J_2}}_{<0} = \underbrace{\frac{|\sqrt{-R_N}|}{2 J_1 J_2}}_{>0} \quad \text{mit} \quad R_Z < 0$$

In den Fällen 1 und 2 kürzen die Linearfaktoren im Zähler die reelle Polstelle im Nenner. Daher dürfen die Nullstellen q_1 bzw. q_2 für eine Kürzung keine Imaginärteile besitzen, wofür die Bedingung $R_Z \geq 0$ notwendig und hinreichend ist.

Nachdem ein schwingungsfähiges Zweimassensystem zugrunde gelegt wurde, sind die Polstellen p_2 und p_3 komplexe Zahlen. Damit müssen für deren Kürzung auch die Nullstellen Imaginärteile enthalten. Deshalb wird in den Fällen 3 bis 6 nicht nur $R_N < 0$, sondern auch $R_Z < 0$ angenommen. Daraus ist ersichtlich, dass die beiden Fälle 3 und 6 keine Relevanz besitzen, da die Bedingung an den Imaginärteil keinesfalls zutreffen kann. Nur für den speziellen Fall $R_N = R_Z = 0$ können die Bedingungen für die Gleichheit des Imaginärteiles erfüllt werden. Das daraus folgende Zweimassensystem ist allerdings nicht schwingungsfähig und widerspricht somit der Annahme.

Weiterhin ist erkennbar, dass die Fälle 3 und 6 und die Fälle 4 und 5 jeweils identisch sind. Da sowohl q_1 und q_2 als auch p_2 und p_3 konjugiert komplex zueinander sind, ist die Gleichung für die Imaginärteile lediglich mit einem negativen Vorzeichen versehen. Eine Kürzung der Faktoren $(s - q_2)$ und $(s - p_2)$ bedingt gleichzeitig auch die Kürzung von $(s - q_1)$ mit $(s - p_3)$.

Aufgrund fehlender Relevanz in den Fällen 3 und 6 und Redundanz in den Fällen 4 und 5, müssen lediglich die Fälle 1, 2 und 5 untersucht werden:

zu Fall 1 bzw. 2:

Kürzung von $s - q_1$ mit s bzw. Kürzung von $s - q_2$ mit s :

Die Forderungen an den Realteil von q_1 bzw. q_2 in den Fällen 1 und 2 können durch

$$c_2 J_2 + (c_1 + c_3) d_w = \pm |\sqrt{R_Z}|, \quad \text{mit } R_Z \geq 0 \quad (2.40)$$

zusammengefasst werden. Durch Einsetzen der Größe R_Z aus Gleichung (2.38) und Quadrieren lässt sich die notwendige Bedingung

$$4J_2 k_w c_3 (c_1 + c_3) = 0 \quad \Leftrightarrow \quad c_3 (c_1 + c_3) = 0 \quad (2.41)$$

ableiten. Es ist zu entnehmen, dass die beiden Lösungen $c_3 = 0$ und $c_1 + c_3 = 0$ jeweils zu einem nicht vollständig beobachtbaren System führen. Im zulässigen Grenzfall $R_Z = 0$ errechnet sich die Konstellation

$$c_2 J_2 + (c_1 + c_3) d_w = 0 \quad \Leftrightarrow \quad c_2 = -\frac{d_w}{J_2} (c_1 + c_3), \quad (2.42)$$

die eine Kürzung der Polstelle im Ursprung verursacht.

zu Fall 5:

Kürzung von $s - q_1$ mit $s - p_3$:

Aus den Forderungen an Realteil und Imaginärteil in Fall 5 lassen sich durch Auflösen nach c_2 die beiden Gleichungen

$$c_2 = \frac{(c_3 J_2 - c_1 J_1) d_w}{J_1 J_2} \quad (2.43)$$

und

$$c_2 = -\frac{(c_1 + c_3) d_w}{J_2} \pm \frac{\sqrt{(J_1 + J_2)^2 d_w^2 c_3^2 + 4J_1 J_2 k_w c_3 (J_1 c_1 - J_2 c_3)}}{J_1 J_2} \quad (2.44)$$

formulieren. Gleichsetzen von (2.43) mit (2.44) führt auf den Zusammenhang

$$c_1 = c_3 \frac{J_2}{J_1}, \quad (2.45)$$

der sowohl die Realteil- als auch die Imaginärteilbedingung simultan erfüllt und somit für eine Kürzung in der Übertragungsfunktion sorgt. Durch Einsetzen von (2.45) in (2.43) errechnet sich daraus allerdings $c_2 = 0$. Im Umkehrschluss folgt, dass eine Kürzung der Linearfaktoren $(s - q_1)$ mit $(s - p_3)$ bzw. $(s - q_2)$ mit $(s - p_2)$ nicht auftreten kann, sofern $c_2 \neq 0$ vorausgesetzt wird.

2.3 Stand der Technik bei der industriellen Regelung eines linearen Zweimassensystems

Bei der industriellen Regelung von Zweimassensystemen kommen überwiegend konventionelle PI-Regler in Kaskadenstruktur zum Einsatz. Ein „innerer“ Regelkreis dient zur Stromregelung in der Antriebsmaschine und ist damit für das Drehmoment bestimmend. Diesem wird ein „äußerer“ Regelkreis überlagert, der die Regelung der Drehzahl übernimmt. Aufgrund ihrer Vorteile (v.a. einfache Struktur, einfache Reglerauslegung und schnelles Ausregeln von Störungen) wird die Kaskadenstruktur eingesetzt (siehe [85, S. 115-118]). Die Einstellung der Regler beruht bei dieser Vorgehensweise auf vollständig bekannten Systemparametern.

2.3.1 Stromregelkreis

Gewöhnlich lässt sich das Übertragungsverhalten zwischen Klemmenspannung der Maschine und momentbildendem Strom $i(t)$ durch einen Tiefpass erster Ordnung beschreiben. Ähnliches gilt für den Stromrichter, der die vom Stromregler vorgegebene Sollspannung $u_{soll}(t)$ in eine Klemmenspannung übersetzt, wobei auch hier in guter Näherung Tiefpassverhalten erster Ordnung angenommen werden kann. In Summe stellt sich daher die vom Stromregler zu beherrschende Strecke als PT₂-Glieder mit reellen Polstellen dar:

$$F_S(s) = \frac{i(s)}{u_{soll}(s)} = \frac{V_S}{(1 + sT_\sigma)(1 + sT_A)} \quad (2.46)$$

Die Konstante $V_S > 0$ bezeichnet hierbei die stationäre Verstärkung der Kombination von Stromrichter und Maschine, $T_\sigma > 0$ stellt die Zeitkonstante des Stromrichters dar, $T_A > 0$ beschreibt die Zeitkonstante, die aus dem Verhältnis von Induktivität und Ohm'schen Widerstand der Maschine resultiert.

Als Stromregler bietet sich ein PI-Baustein an, mit einer Auslegung nach dem Betragsoptimum (BO), siehe [85, S. 46ff]. Dadurch wird die verzögernde Wirkung der langsameren von beiden Zeitkonstanten (für die meisten Antriebsanordnungen gilt $T_A > T_\sigma$) eliminiert. Der Stromregler mit PI-Verhalten wird beschrieben durch

$$F_R(s) = V_R \cdot \frac{1 + sT_n}{sT_n} \quad (2.47)$$

wobei sich nach BO die Reglerverstärkung V_R und Nachstellzeit T_n zu

$$V_R = \frac{T_A}{2T_\sigma V_S} \quad \text{und} \quad T_n = T_A \quad (2.48)$$

ergeben. In der Serienschaltung von Regler und Strecke erfolgt eine Kürzung des Faktors $1 + sT_A$, so dass ausschließlich die Zeitkonstante des Stromrichters die Dynamik des geschlossenen Regelkreises

$$F_i(s) = \frac{i(s)}{i_{soll}(s)} = \frac{1}{2T_\sigma^2 s^2 + 2T_\sigma s + 1} \quad (2.49)$$

bestimmt. Das ursprüngliche Verhalten der Maschine ist durch die gewählte Reglereinstellung nicht mehr sichtbar. Die Optimierung nach BO bedingt die Polstellen

$$-\frac{1}{2T_\sigma} \pm j \frac{1}{2T_\sigma} \quad (2.50)$$

für den Regelkreis, deren Real- und Imaginärteil stets gleich groß sind. Daraus bestimmt sich der Dämpfungsgrad zu $1/\sqrt{2}$, der zu einer schnellstmöglichen Einschwingzeit und zu einem Überschwingen von 4.3% führt. Bei dieser Reglereinstellung ist ausschließlich die Zeitkonstante des Stromrichters maßgeblich für die Zeitskala der Regelkreisdynamik.

2.3.2 Drehzahlregelkreis

Abhängig von der Abweichung zwischen Soll- und Istdrehzahl beaufschlagt der überlagerte Drehzahlregler den inneren Stromregelkreis mit einem entsprechenden Sollwert. Damit stellt der Stromregelkreis zusammen mit dem elektromagnetisch-mechanischen Subsystem, das den Strom der Maschine zunächst in ein Drehmoment und schließlich in eine Drehzahl übersetzt, die Strecke aus Sicht des Drehzahlreglers dar. Für die Reglerauslegung ist zu unterscheiden, ob die Motor- oder die Lastdrehzahl als Regelgröße betrachtet wird und damit die Istdrehzahl darstellt. Ebenfalls beeinflusst die Art der Ankopplung (starr/hart/weich) die Einstellung der Reglerparameter.

Um einen Drehzahlregler ohne differenzierenden Anteil mittels Standard-Optimierung auslegen zu können, muss der Stromregelkreis – der durch die Auslegung nach BO bedingt PT_2 -Verhalten zeigt – durch ein System erster Ordnung approximiert werden. Eine geeignete Approximation ist ein PT_1 -Baustein, dessen Eckfrequenz mit der Resonanzfrequenz des PT_2 -Systems übereinstimmt. Aus den Polstellen (2.50) ergibt sich letztgenannte zu $1/(2T_\sigma)$, so dass eine Zeitkonstante von $2T_\sigma$ für die Approximation folgt.

Nachdem sich das schwingungsfähige Verhalten der elastischen Welle nicht durch die genannten Standard-Optimierungsverfahren beherrschen lässt, wird auch im mechanischen Subsystem eine Vereinfachung vorgenommen. Die elastische Kopplung bleibt unberücksichtigt, beide Massen werden als starr gekoppelt betrachtet. Eine starre Verbindung beider Massen liegt bei einer idealen, torsionssteifen Welle ($k_w = \infty$) vor und äußert sich in der Identität $\omega_1(\cdot) \equiv \omega_2(\cdot)$ beider Drehzahlen. Das starr gekoppelte Zweimassensystem ist daher äquivalent zum Einmassensystem mit dem Gesamtträgheitsmoment $J = J_1 + J_2$, das für die Reglerauslegung als Streckenmodell dient. Bei harter oder starrer Ankopplung sind folglich beide Trägheitsmomente J_1 und J_2 zu einem Gesamtträgheitsmoment zu summieren.

Demzufolge wird als Strecke des Drehzahlregelkreises das IT_1 -System

$$F_S(s) = \frac{\omega(s)}{i_{soll}(s)} = \frac{1}{(1 + 2T_\sigma s)sT_m} \quad (2.51)$$

angesetzt. Die Konstante $T_m > 0$ wird dabei von der Maschinenkonstanten $C_M > 0$, vom Erregerfluss $\Psi > 0$ und vom Massenträgheitsmoment $J > 0$ der zu beschleunigenden Masse gemäß

$$T_m := \frac{1}{C_M \Psi} \cdot J \quad (2.52)$$

bestimmt. Die Lastmasse wird bei der Reglerauslegung allerdings nicht berücksichtigt, wenn bei der Regelung der Motordrehzahl eine weiche Ankopplung vorliegt. In diesem Fall tritt die Rückwirkung der Welle auf die Motormasse aus Sicht des Reglers erst verzögert ein, so dass in Gleichung (2.52) lediglich das Trägheitsmoment $J = J_1$ eingesetzt wird. Die Rückwirkung der Welle wird dann als Störung aufgefasst, die vom Regler zu unterdrücken ist.

Für eine solche Strecke kann ein PI-Regler nach Standard-Optimierungsverfahren eingestellt werden. Es eignet sich zum Beispiel das symmetrische Optimum (SO) [85,

S. 60ff] als Einstellregel:

$$F_R(s) = V_R \frac{1 + sT_n}{sT_n} \quad \text{mit} \quad V_R = \frac{T_m}{2 \cdot 2T_\sigma} \quad \text{und} \quad T_n = 4 \cdot 2T_\sigma \quad (2.53)$$

Der daraus resultierende geöffnete Drehzahlregelkreis

$$F_o(s) = \underbrace{\frac{1 + 8T_\sigma s}{1 + 2T_\sigma s}}_{\text{PDT}_1} \cdot \underbrace{\frac{1}{32T_\sigma^2 s^2}}_{\text{I}_2} \quad (2.54)$$

setzt sich aus einem PDT₁-Glieder und einem I₂-Glieder zusammen, das im Betragsgang einen Abfall von 40dB/Dekade und einen Phasenwinkel von -180° verursacht. Die Eckfrequenz im Zähler des PDT₁-Gliederes liegt bei $\omega_Z = 1/(8T_\sigma)$ und ist damit stets kleiner als die Eckfrequenz $\omega_N = 1/(2T_\sigma)$ des Nenners. Folglich hebt der Zähler Betrag und Phase bereits an, bevor der Nenner erst bei höheren Frequenzen entgegen wirkt, wodurch das PDT₁-Glieder einen Betragsanstieg von 20dB/Dekade innerhalb der Eckfrequenzen $\omega_Z < \omega < \omega_N$ erhält. Außerhalb dieses Bereiches ($\omega < \omega_Z$ und $\omega > \omega_N$) entsteht ein konstanter Betragsverlauf. In Summe mit dem I₂-Glieder, das den Betrag pro Dekade um 40dB reduziert, ergibt sich daraus für den offenen Drehzahlregelkreis das in Abbildung 2.6 dargestellte Bodediagramm. Die Eckfrequenzen des PDT₁-Gliederes markieren die Knickfrequenzen, die symmetrisch zur Amplitudendurchtrittsfrequenz $\omega_d = 1/(4T_\sigma)$ liegen. Die Anhebung des Phasengangs in der Umgebung der Amplitudendurchtrittsfrequenz verschafft eine ausreichende Stabilitätsreserve von 36.9° .

Aus dieser Reglereinstellung resultiert die Übertragungsfunktion

$$F_R(s) = \frac{F_o}{1 + F_o} = \frac{1 + 8T_\sigma s}{64T_\sigma^3 s^3 + 32T_\sigma^2 s^2 + 8T_\sigma s + 1} \quad (2.55)$$

für den geschlossenen Drehzahlregelkreis, deren Polstellen bei

$$-\frac{1}{4T_\sigma} \quad \text{und} \quad -\frac{1}{8T_\sigma} \pm j\frac{\sqrt{3}}{8T_\sigma} \quad (2.56)$$

liegen. Der Dämpfungsgrad des konjugiert komplexen Polpaares ergibt sich zu 0.5, das Überschwingen beträgt 43% (siehe Abbildung 2.7). Mit zusätzlicher Sollwertglättung reduziert sich das Überschwingen auf 8%.

Die beschriebene Vorgehensweise legt bei der Einstellung des Drehzahlreglers ein Einmassensystem als Regelstrecke zugrunde, obwohl der Regler am Zweimassensystem eingesetzt wird und berücksichtigt dadurch die elastische Verbindung nicht. Diese Diskrepanz erfordert eine eingehende Untersuchung der Eigenschaften des derart geregelten Zweimassensystems, die in den folgenden Abschnitten detailliert durchgeführt wird. Dabei hängt das Verhalten stark von der gewählten Regelgröße ab, also davon, ob die Motor- oder die Lastdrehzahl geregelt wird.

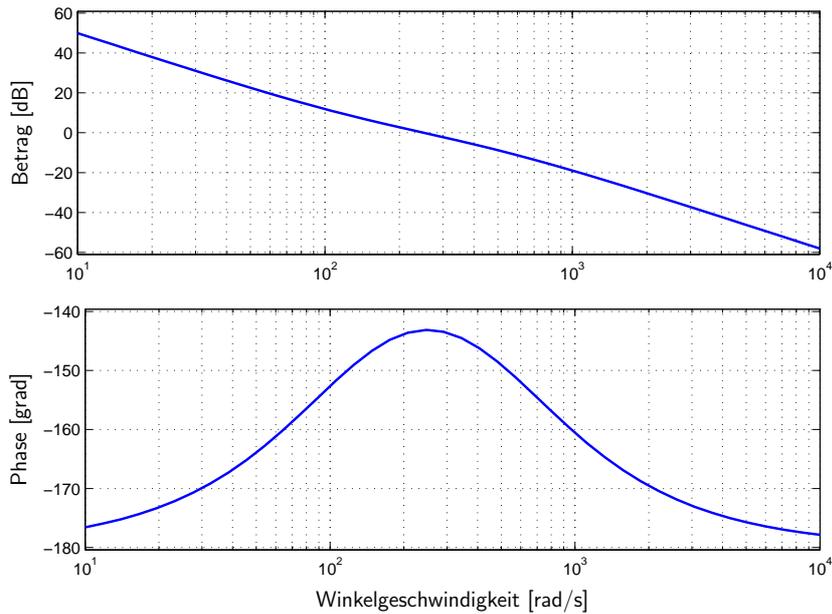


Abbildung 2.6: Bodediagramm des den offenen Kreises bei Regelung des Einmassensystems nach Standard-SO.

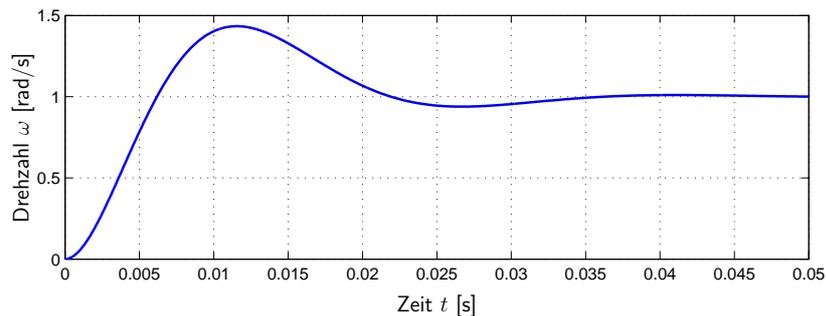


Abbildung 2.7: Sprungantwort des Drehzahlregelkreises (Einmassensystem), eingestellt nach Standard-SO.

2.3.3 Regelung der Antriebsmaschinendrehzahl

Bei der Regelung der Antriebsmaschinendrehzahl fokussiert sich der Drehzahlregler auf die Differenz

$$e(t) = w(t) - \omega_1(t) \quad (2.57)$$

von Sollwert $w(t)$ und Motordrehzahl $\omega_1(t)$. Damit unterscheidet sich im Allgemeinen die regelungstechnische von der prozesstechnischen Aufgabenstellung. Der Regler berücksichtigt einzig das Verhalten der Motordrehzahl $\omega_1(t)$ und vernachlässigt dabei vollständig das Verhalten der prozesstechnisch relevanten Größe, nämlich der Lastdrehzahl $\omega_2(t)$. Damit erfüllt die Regelung ihre Aufgabe bereits, sobald sich die Motordrehzahl wunschgemäß verhält. Dies gilt auch dann, wenn die Drehzahl der Lastmasse ihre Vorgabe nicht einhält, weswegen das Verhalten der Lastdrehzahl nicht zur Beurteilung der Regelgüte herangezogen werden darf. Folglich kann sich bei der Regelung der Antriebsmaschinendrehzahl das Gesamtsystem aus prozesstechnischer Sicht als un-

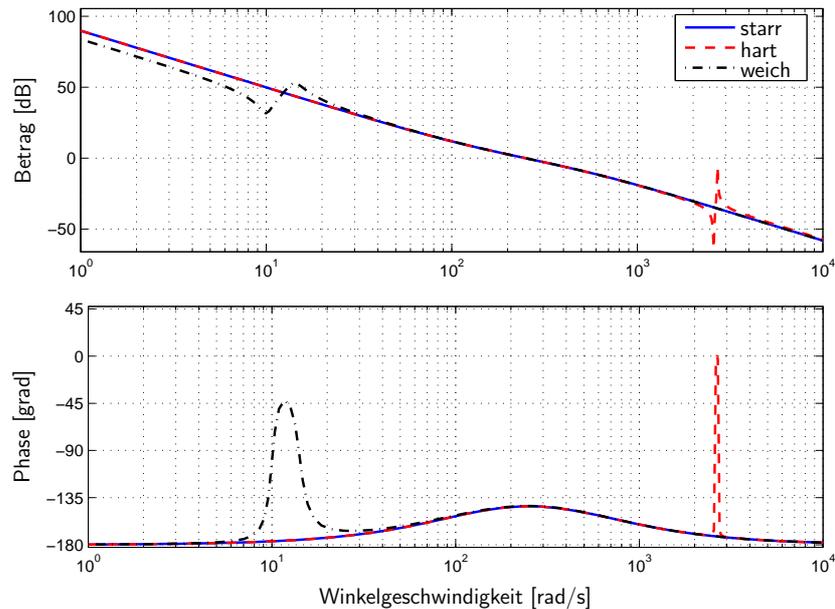


Abbildung 2.8: Bodediagramm für die Regelung des Einmassensystems und für die Regelung der Motordrehzahl des Zweimassensystems bei harter und weicher Ankopplung nach Standard-SO.

brauchbar erweisen, obwohl aus regelungstechnischer Sicht alle Anforderungen vom Regler erfüllt werden. Der Vorteil von diesem Konzept liegt in der Tatsache, dass für alle Situationen (starre/harte/weiche Ankopplung) ein stabiler Regelkreis¹⁾ entsteht. Allerdings treten bei einem solchen Vorgehen zwei Fragestellungen auf. Einerseits ist zu untersuchen, in wieweit das Verhalten der Motordrehzahl des Zweimassensystems vom Verhalten der Drehzahl des angenommenen Einmassensystems abweicht. Eine qualitative Aussage hierzu bietet Abbildung 2.8, in der exemplarisch die Bodediagramme für ein Einmassensystem und für ein hart bzw. weich angekoppeltes Zweimassensystem aufgetragen sind. Andererseits muss geklärt werden, ob das Verhalten der Lastdrehzahl den Anforderungen genügt, wenn sich die Motordrehzahl gemäß Vorgabe verhält.

Bevor beide Fragen in den Abschnitten 2.3.3.1 und 2.3.3.2 durch Simulationen beantwortet werden, soll zunächst diskutiert werden, ob durch eine Regelung der Antriebsdrehzahl die relevante Lastdrehzahl prinzipiell wunschgemäß beeinflusst werden kann.

Weil ein PI-Regler den Regelfehler $e(t)$ nur für einen konstanten Sollwert $w = \text{const.}$ asymptotisch zum Verschwinden bringt, wird ausschließlich der Fall einer konstanten Drehzahl untersucht. Wegen

$$\dot{\omega}_1(t) = \frac{u(t) - M_w(t)}{J_1} \rightarrow 0 \quad (2.58)$$

müssen sich Stellgröße des Reglers und Wellenmoment asymptotisch aufheben. Hierin liegt begründet, warum das Regelziel $e(t) \rightarrow 0$ nicht zu erreichen ist, solange die Welle Torsionsschwingungen ausführt. Es ist nicht möglich, dass eine Kombination aus

¹⁾ Das Nyquist-Kriterium erfordert von der Ortskurve des offenen Kreises eine globale Phasenwinkeländerung von π , was unabhängig von der Ankopplung stets erfüllt wird.

Idealverhalten an der Motormasse

$$\omega_1(t) = w = \text{const.} \quad (2.59)$$

und schwingender Lastmasse

$$\omega_2(t) = A \sin(\omega t) + w, \quad A, \omega \geq 0 \quad (2.60)$$

auftritt. Dies würde eine sinoidal verlaufende Differenzgeschwindigkeit

$$\omega_1(t) - \omega_2(t) = -A \sin(\omega t) \quad (2.61)$$

zwischen beiden Massen bedingen, deren Integral den Torsionswinkel ergibt:

$$\varphi_w(t) = \int_0^t -A \sin(\omega \tau) d\tau + \varphi_w(0) = \frac{A}{\omega} (\cos(\omega t) - 1) + \varphi_w(0) \quad (2.62)$$

Über die Beziehung (2.2) entsteht daraus ein ebenfalls sinusförmiges Wellenmoment, das als Störgröße in den Drehzahlregelkreis einwirkt:

$$M_w(t) = k_w \varphi_w(t) + d_w \dot{\varphi}_w(t) = k_w \frac{A}{\omega} \cos(\omega t) - k_w \frac{A}{\omega} + k_w \varphi_w(0) - d_w A \sin(\omega t) \quad (2.63)$$

Für eine konstante Motordrehzahl muss der Regler darauf mit einer schwingenden Stellgröße reagieren. Die konstante Regelgröße $\omega_1(t) \rightarrow w = \text{const.}$ führt allerdings zu einem konstanten Regelfehler, auf den der PI-Regler mit einer konstanten oder linear ansteigenden Stellgröße antwortet, nicht jedoch mit einer sinoidalen. Dieser Widerspruch schließt die angenommene Konstellation aus, bei der die Motordrehzahl exakt konstant verläuft, in der Welle jedoch Torsionsschwingungen auftreten.

Um den stationären Zustand $\omega_1(t) = w = \text{const.}$ exakt zu erreichen, muss daher die Schwingung in der Welle zunächst vom Regler bedämpft werden oder von selbst durch die Eigendämpfung abklingen, so dass ein *konstantes* Wellenmoment als Störung wirkt. Ein solches kann von einem PI-Regler vollständig kompensiert werden. Andernfalls lässt sich das Regelziel nicht exakt erreichen.

Allerdings wird der stationäre Zustand und damit das Regelziel näherungsweise erreicht, wenn die Amplitudendurchtrittsfrequenz des Reglers sehr viel höher liegt als die Eigenfrequenz der Wellenschwingung und damit ein verhältnismäßig schneller Regler verwendet wird. Dieser Fall tritt bei weicher Ankopplung auf (siehe Abschnitt 2.3.3.2). Nachdem ein schneller Regler einer hohen Reglerverstärkung entspricht,²⁾ erzeugen bereits kleine Regelfehler wirksame Stellgrößen. Sobald das störende Wellenmoment die Motordrehzahl geringfügig vom Arbeitspunkt auslenkt, antwortet ein schneller Regler bereits auf den kleinen Fehler mit einem entsprechend hohen Gegenmoment, um

²⁾ Eine Vergrößerung der Reglerverstärkung in Gleichung (2.53) äußert sich als zusätzliches Proportionalglied im offenen Regelkreis (2.54). Der Betragsgang eines Proportionalgliedes verläuft konstant, die Phase ist identisch 0° . Folglich ändert sich durch eine zusätzliche Verstärkung der Phasenverlauf von $F_o(j\omega)$ nicht, der Betrag erfährt eine Parallelverschiebung zu höheren Werten. Das streng monotone Fallen der Betragskurve (siehe Abbildung 2.6) führt dadurch zu einer Erhöhung der Amplitudendurchtrittsfrequenz.

die Drehzahlabweichung gering zu halten. Aus diesem Grund kann das niederfrequente Wellenmoment aus Sicht der schnellen Reglerdynamik näherungsweise als konstant betrachtet werden.

Je langsamer der Regler im Vergleich zur Torsionsschwingung eingestellt ist, desto stärker tritt die zeitliche Veränderung des schwingenden Wellenmomentes in den Vordergrund, so dass die Störung umso schlechter unterdrückt werden kann. Im Falle harter Ankopplung liegt die Torsionsfrequenz oberhalb der Amplitudendurchtrittsfrequenz des offenen Regelkreises, weshalb der Effekt der Eigenschwingung vom Regler nicht verringert werden kann. Die Motordrehzahl ist demnach vollständig vom oszillierenden Wellenmoment beeinflusst, das lediglich aufgrund der Eigendämpfung $d_w > 0$ abklingt.

Die angestellte Diskussion lässt erwarten, dass der beschriebene PI-Regler bei harter Kopplung das Regelziel nicht erreicht, bei weicher Kopplung zwar das Regelziel erreicht ($\omega_1(t)$ verhält sich wunschgemäß), jedoch eine stark oszillierende Lastdrehzahl toleriert. In den folgenden beiden Abschnitten wird diese Aussage durch Simulationen verifiziert.

2.3.3.1 Harte Ankopplung

Bei einer harten Verbindung beider Massen übersteigt die Eigenfrequenz ω_{oZ} die Amplitudendurchtrittsfrequenz ω_d des offenen Regelkreises deutlich, der Regler ist daher langsam im Vergleich zur Wellenschwingung. Dies schlägt sich in der Ungleichung

$$\omega_o > \omega_{oZ} \gg \omega_d = \frac{1}{4T_\sigma} \quad (2.64)$$

nieder. In der Praxis tritt dieser Fall näherungsweise dann ein, wenn eine verhältnismäßig kurze Welle mit großem Durchmesser aus einem steifem Material eingesetzt wird. Weiterhin trägt ein vergleichsweise großes Massenträgheitsmoment der Antriebsmaschine im Verhältnis zu dem der Arbeitsmaschine zu einer Versteifung der Kopplung bei. Der langsame Regler kann im geschlossenen Kreis durch seine Stellgröße die Eigenschwingung weder anregen, noch bedämpfen. Der Regler wird wie bei starrer Kopplung optimiert, d.h., als Strecke wird wiederum das Einmassensystem mit dem Trägheitsmoment $J = J_1 + J_2$ zugrunde gelegt. Somit ergibt sich die selbe Reglereinstellung wie bei starrer Ankopplung. Der Einfluss der schwingungsfähigen Kopplung wird als hochfrequente Störung betrachtet, die beispielsweise durch Lastsprünge angeregt werden kann.

Damit ist der Regler einer zeitveränderlichen Störung (Torsionsschwingung der Welle) ausgesetzt, für deren Bedämpfung zu wenig Bandbreite vorhanden ist. Im groben Verlauf bewegen sich die Drehzahlen zwar ähnlich wie beim Einmassensystem nach Standard-SO. Allerdings zeigt Abbildung 2.9 deutlich die Überlagerung hochfrequenter Oszillationen, deren Frequenz mit der Eigenschwingung der Strecke übereinstimmt. Eine aktive Bedämpfung durch den Regler ist nicht erkennbar, einzig die Eigendämpfung des Systems führt zu einer Verringerung der Torsionsschwingung. Dieser Sachverhalt wird in [85, S. 817] durch die Aussage

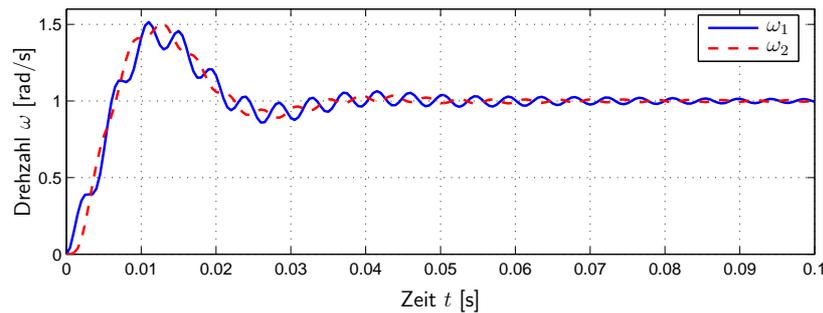


Abbildung 2.9: Sprungantwort für den Drehzahlregelkreis (Zweimassensystem, harte Kopplung), eingestellt nach Standard-SO.

„Eine einfache Kaskadenstruktur ist nicht in der Lage, diese Schwingungen zu dämpfen.“

zusammengefasst. Dennoch wird diese Regelstrategie in [85] mit der Begründung als brauchbar beschrieben, dass bei harter Ankopplung die Amplitude der Eigenschwingung wegen der hohen Wellensteifigkeit gewöhnlich vernachlässigbar klein ist. Je härter die Welle, desto geringer die Amplitude. Allerdings bleibt dann die Frage, ob nicht die Modellierung der Regelstrecke als Einmassensystem einer Modellierung als Zweimassensystem vorzuziehen ist.

Obwohl damit das Regelziel als verfehlt zu werten ist, kann diese Methode dennoch für gewisse Anwendungszwecke eingesetzt werden.

2.3.3.2 Weiche Ankopplung

Wenn dagegen der Regler im Vergleich zur Eigenschwingung der Welle sehr schnell eingestellt ist ($\omega_d \gg \omega_o > \omega_{oZ}$) oder im Bereich der Eigenfrequenz liegt ($\omega_d \approx \omega_o$), besteht eine weiche Ankopplung. Aus Sicht der Reglerdynamik setzt die Rückwirkung der Welle erst mit deutlicher Verzögerung ein, so dass in Gleichung (2.52) als Trägheitsmoment lediglich $J = J_1$ einzusetzen ist. Daraus resultiert eine geringere Reglerverstärkung als bei harter oder starrer Ankopplung.

Für eine Untersuchung des Systemverhaltens bietet sich eine Fallunterscheidung an.

Fall a: $\omega_d \gg \omega_o$

Das obere Diagramm in Abbildung 2.10 zeigt, dass die Regelgröße $\omega_1(t)$ ähnlich wie beim Einmassensystem verläuft, eine überlagerte Schwingung ist nicht sichtbar. Ein Regler mit ausreichender Bandbreite ($\omega_d \gg \omega_o$) kann demzufolge die Auswirkung eines niederfrequenten Wellenmomentes nahezu vollständig eliminieren, der Effekt der Störung auf die Regelgröße $\omega_1(t)$ bleibt vernachlässigbar gering. Nachdem die Lastdrehzahl mit hoher Amplitude schwingt, erfüllt die Regelung die gestellten Anforderungen selbst unter starker Störeinwirkung und ist daher als gut zu beurteilen. Allerdings ist eine solche Regelungsstrategie ungeeignet, um die Prozessgröße $\omega_2(t)$ zu beherrschen, die Torsionsschwingung in der Welle wird nicht aktiv bedämpft.

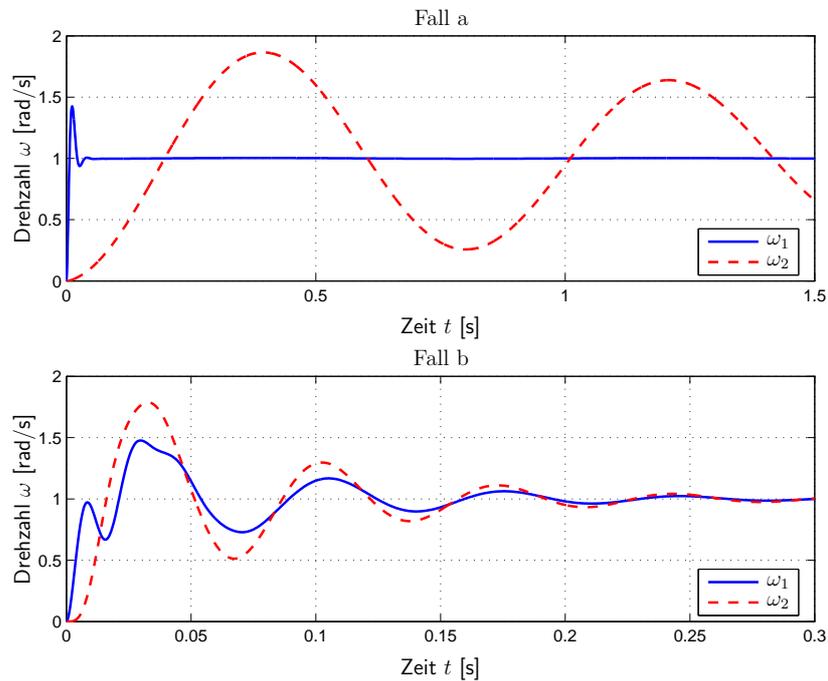


Abbildung 2.10: Sprungantworten für den Drehzahlregelkreis (Zweimassensystem, weiche Kopplung), eingestellt nach Standard-SO. Oberes Diagramm für Fall a: $\omega_d \gg \omega_o$. Unteres Diagramm für Fall b: $\omega_d \approx \omega_o$.

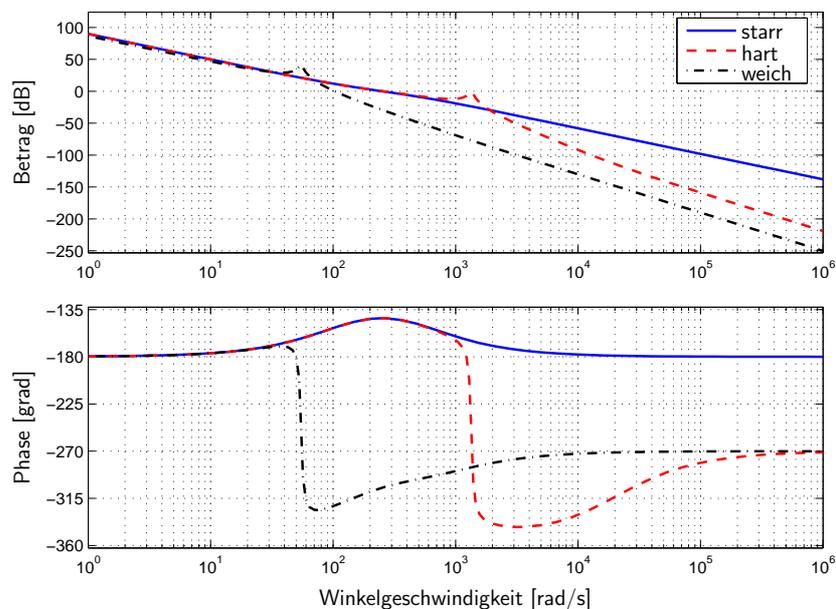


Abbildung 2.11: Bodediagramm für die Regelung des Einmassensystems und für die Regelung der Lastdrehzahl des Zweimassensystems bei harter und weicher Ankopplung nach Standard-SO.

Fall b: $\omega_d \approx \omega_o$ (Ankopplung im Bereich der Reglerdynamik)

Hier bewegt sich die Reglerdynamik im Bereich der Eigenfrequenz der Welle. Somit besitzt der Regler nicht genügend Bandbreite, um die Störung auszuregeln. Der Einfluss der elastischen Welle tritt jedoch deutlich in Erscheinung, da die Steifigkeit der Welle zu gering ist, um die Amplitude der Eigenschwingung vernachlässigbar klein zu halten. Aus diesem Grund erfüllt weder der Regler die gestellte Aufgabe, noch verhält sich die Prozessgröße wunschgemäß. Das untere Diagramm in Abbildung 2.10 verdeutlicht, dass dieser Fall für eine Anwendung gänzlich ungeeignet ist.

2.3.4 Regelung der Arbeitsmaschinendrehzahl

Eine Unterscheidung zwischen Regel- und Prozessgröße entfällt, wenn anstelle der Motordrehzahl $\omega_1(t)$ die Arbeitsmaschinendrehzahl $\omega_2(t)$ rückgeführt wird. Damit reagiert der Drehzahlregler direkt auf die maßgebliche Differenz

$$e(t) = w(t) - \omega_2(t) \quad (2.65)$$

von Sollwert $w(t)$ und Lastdrehzahl $\omega_2(t)$.

Wie allerdings die Systemanalyse in Abschnitt 2.2.2 erwarten lässt, liegt hierin die kritischere beider Strategien. Die Regelstrecke besitzt bei der Wahl $y(t) = \omega_2(t)$ den Relativgrad 2, so dass nicht in jedem Falle stabiles Verhalten gewährleistet ist.

Weiterhin bleibt die Annahme einer starren Verbindung bei der Reglerauslegung bestehen. Der Drehzahlregler wird gemäß der beschriebenen Vorgehensweise nach SO für ein Einmassensystem eingestellt. Damit bleibt die nominelle Amplitudendurchtrittsfrequenz ω_d für den offenen Kreis erhalten.

2.3.4.1 Harte Ankopplung

Bei harter Ankopplung wird das Verhalten des Regelkreises nicht wesentlich davon beeinflusst, welche Regelgröße gewählt wird. Wegen $\omega_o \gg \omega_d$ verhält sich der Regler langsam im Verhältnis zur Systemdynamik und kann die Eigenschwingung daher weder dämpfen noch anregen. Das Bodediagramm des offenen Kreises wird im Vergleich zu jenem des Standard-SO (EMS) im kritischen Bereich der Amplitudendurchtrittsfrequenz kaum verändert. Dadurch bleibt die Phasenreserve erhalten, es bestehen keine Stabilitätsprobleme.

2.3.4.2 Weiche Ankopplung

Besitzt der Regler eine große Bandbreite gegenüber der Systemdynamik ($\omega_o \leq \omega_d$ oder $\omega_o \approx \omega_d$), tritt die Abweichung vom Standard-SO (EMS) bereits deutlich unterhalb der Amplitudendurchtrittsfrequenz ein. Am Bodediagramm des offenen Kreises in Abbildung 2.11 wird offensichtlich, dass bei weicher Ankopplung nicht auf die Lastdrehzahl

geregelt werden kann. Die Phase senkt sich bereits bei niedrigen Frequenzen unter -180° ab, weswegen ein instabiler Regelkreis folgt.

Stabilität kann ausschließlich unter den Verhältnissen der harten Ankopplung gegeben sein. Diese werden erreicht, wenn die Parametrierung des Reglers nach SO aufgegeben und stattdessen eine erheblich langsamere Einstellung gewählt wird. Dadurch geht dem Regelkreis Dynamik verloren, was sich in einer verlängerten Anregelzeit niederschlägt. Gleichzeitig verhindert die reduzierte Bandbreite das ungewollte Anregen der Eigenschwingung durch den Regler und verschafft der Ungleichung $\omega_d \ll \omega_o$ wieder Gültigkeit, so dass zumindest die erforderliche Eigenschaft der Stabilität gewährleistet ist (siehe Abbildung 2.11).

2.3.5 Bewertung der Kaskadenstruktur

Die beschriebenen Regelungskonzepte erlauben im Falle harter Kopplung die gewünschte Beeinflussung der Drehzahlen im niedrigen Frequenzbereich. Allerdings kann die Eigenfrequenz der Regelstrecke, die in höheren Frequenzbereichen liegt, nicht gedämpft werden. Einzig durch die gutmütige Systemeigenschaft, dass bei ausreichender Torsionssteifigkeit der Welle die Amplitude der Eigenschwingung gering bleibt, bestehen Einsatzmöglichkeiten, wenn die verbleibenden Vibrationen der Welle hingenommen werden können.

Liegt eine weiche Ankopplung vor, scheitert die Kaskadenregelung gänzlich. Obwohl die Motordrehzahl wunschgemäß geregelt werden kann, zeigt die Lastdrehzahl unbefriedigendes Verhalten. Wird dagegen die Lastdrehzahl geregelt, entsteht ein instabiler Regelkreis.

Da die einfache Struktur dieser Konzepte nur einen einzigen Drehzahlsensor erfordert, der bei modernen Drehfeldmaschinen häufig bereits integriert ist, wird die Drehzahlregelung in Kaskadenform vor allem bei Antrieben mit nahezu starrer Ankopplung eingesetzt.

Trotz einiger Anwendungsgebiete für die Kaskadenregelung muss zur Zustandsregelung übergegangen werden, wenn gleichzeitig große Bandbreite des Reglers und gute Bedämpfung der Eigenschwingung gefordert ist. Die Bewertung der konventionellen Kaskadenregelung in [85, S. 819] kommt zu dem Schluss, dass die Verwendung des Torsionswinkels unabdingbar ist, um sämtliche Fälle zufriedenstellend beherrschen zu können. Diese Tatsache rechtfertigt den Einsatz einer Zustandsauskopplung zur Relativgradreduzierung, was für hochverstärkungsbasierte Regler grundlegend ist.

2.3.6 Zustandsregelung

Bekanntlich lässt sich durch eine entsprechend ausgelegte Zustandsregelung einer steuerbaren Regelstrecke beliebiges Verhalten aufprägen, die Pole des geschlossenen Regelkreises können vollkommen frei in der s -Ebene positioniert werden. Insbesondere ist die Nähe zur reellen Achse gewünscht, weil dadurch eine beliebig hohe Dämpfung der

Torsionsschwingung erreicht werden kann.

Hierfür muss jedoch eine Linearkombination aller drei Systemzustände zurückgeführt werden, die auch den Verdrehwinkel $\varphi_w(t)$ mit einschließt. Daraus ergibt sich der Nachteil eines erhöhten messtechnischen Aufwandes oder eines zusätzlichen Beobachters. Letzterer stützt sich im Falle des Luenberger-Beobachters jedoch auf vollständige Systemkenntnis und ist damit im Kontext von Systemen mit unbekanntem Parametern unbrauchbar. Aus diesem Grund stellt sich die Frage, ob für Zweimassensysteme die Messbarkeit der Zustände aus technischer Sicht gegeben und zu rechtfertigen ist.

In modernen Industrieantrieben werden überwiegend Drehfeldmaschinen, d.h. umrichter-gespeiste Asynchronmaschinen verwendet. Die intern im Umrichter implementierte Stromregelung für die Maschine verlangt das Drehzahlsignal, weshalb ein Sensor für die Drehzahl der Antriebsseite unumgänglich ist. Die gemessene Motordrehzahl kann dann nicht nur dem Umrichter, sondern auch einer übergeordneten Regelung zur Verfügung gestellt werden. Wenn die regelungstechnische Aufgabenstellung die Lastdrehzahl als zu regelnde Ausgangsgröße vorsieht, ist eine Drehzahlmessung auf der Lastseite erforderlich. Unter diesen Umständen dürfen beide Drehzahlen als messbar angenommen werden.

Die dritte Zustandsgröße, der Verdrehwinkel der Welle, kann ohne zusätzlichen Messaufwand aus beiden Drehzahlsignalen bestimmt werden, indem die Differenz beider Drehzahlen über der Zeit integriert wird. Hier besteht allerdings die Problematik, dass sich minimale Fehler im Laufe der Zeit zu unannehmbar hohen Abweichungen aufsummieren können. Das Zweimassensystem des Lehrstuhls ist anstelle von Tachogeneratoren zur Drehzahlmessung mit Positionssensoren ausgestattet. Die Drehzahlen errechnen sich aus einer Differentiation beider Positionssignale, der Verdrehwinkel wird aus einer einfachen Subtraktion beider Positionen gewonnen. Daher ist die Voraussetzung vollständiger Zustandsmessbarkeit im Falle rotatorischer Antriebssysteme im Allgemeinen keine restriktive Einschränkung, sondern wird hier als erfüllt angenommen.

Schwieriger gestaltet sich die Zustandsmessung, wenn das Feder-Dämpfer-Element beispielsweise wie bei Roboterarmen realisiert ist. Hier wird die Drehzahl der Antriebsmaschine in eine Geschwindigkeit entlang einer Trajektorie umgesetzt, die gewöhnlich nicht gemessen wird. Deshalb kann aus der Geschwindigkeitsdifferenz der benötigte Zustand nicht berechnet werden. In einer solchen Anwendung ist ein zusätzlicher Sensor (z.B. Dehnmessstreifen) für die Dehnung der Feder meist unumgänglich.

Falls nicht der vollständige Zustandsvektor gemessen werden kann, ist eine Regelung dennoch möglich, siehe Abschnitt 2.3.3 und 2.3.4. Allerdings kann dann, wie beschrieben, keine zufriedenstellende Schwingungsdämpfung erfolgen. Insofern ist zur Bedämpfung der Torsionsschwingung die Messbarkeit aller Zustände vorausgesetzt und anzunehmen.

Für die Auslegung des Rückführvektors $c \in \mathbb{R}^{1 \times 3}$ im Zustandsregelgesetz

$$u(t) = -cx(t) + Vw(t), \quad V \neq 0 \quad (2.66)$$

bestehen grundsätzlich zwei unterschiedliche Möglichkeiten. Zum einen ist die Minimierung verschiedener Gütefunktionale bzw. Kostenfunktionen geeignet, um geeignete

Rückführungen zu berechnen. Hier bieten sich vor allem LQ-Entwurf (in der Literatur auch unter der Bezeichnung Riccati-Entwurf bekannt) [23, S. 479ff] oder Kriterien wie ITAE (Integral Time Multiplied Absolute Error) oder ISE (Integral Square Error) [85, S. 104] an. Zum anderen kann, basierend auf der gewünschten Dynamik des geschlossenen Regelkreises, die anzustrebende Verteilung der Eigenwerte direkt vorgegeben werden (Polvorgabe). Daraus lässt sich ein Wunschpolynom erstellen, aus dem mittels Koeffizientenvergleich (Vergleich mit dem charakteristischen Polynom des Regelkreises in Abhängigkeit von den noch unbestimmten Rückführkoeffizienten) ein Rückführvektor zu berechnen ist, der dem Regelkreis schließlich die gewünschte Dynamik aufprägt. Zur Wahl eines geeigneten Sollpolynomes eignet sich beispielsweise das Dämpfungsoptimum (DO) [85, S. 84ff], das für eine Strecke dritter Ordnung das Polynom

$$N_{Norm}(s) = s^3 + \frac{4}{T}s^2 + \frac{8}{T^2}s + \frac{8}{T^3}, \quad T > 0 \quad (2.67)$$

mit den Wurzeln

$$-\frac{2}{T}, \quad -\frac{1}{T} + j\frac{\sqrt{3}}{T} \quad \text{und} \quad -\frac{1}{T} - j\frac{\sqrt{3}}{T} \quad (2.68)$$

ergibt. Mit Hilfe des freien Parameters $T > 0$ lässt sich die gewünschte Bandbreite des Regelkreises festlegen. Dabei sind vor allem Restriktionen durch die maximale Stellgröße oder durch die Belastbarkeit der Regelstrecke zu berücksichtigen.

Für das Zweimassensystem (2.6) lässt sich ein Zustandsregler nach DO auslegen, indem ein Vektor $c = (c_1, c_2, c_3) \in \mathbb{R}^{1 \times 3}$ gewählt wird, der dem geschlossenen Regelkreis

$$\begin{aligned} \dot{x}(t) &= (A - bc)x(t) + bVw(t), & x(0) &\in \mathbb{R}^3 \\ y(t) &= cx(t) \end{aligned} \quad (2.69)$$

ein charakteristisches Polynom

$$\begin{aligned} \det(sI_3 - (A - bc)) &= \\ &= s^3 + \frac{(J_1 + J_2)d_w + J_2c_3}{J_1J_2}s^2 + \frac{J_2c_2 + (c_1 + c_3)d_w + (J_1 + J_2)k_w}{J_1J_2}s + \frac{(c_1 + c_3)k_w}{J_1J_2} \end{aligned} \quad (2.70)$$

zuweist, das mit dem Normpolynom $N_{Norm}(s)$ übereinstimmt. Ein Koeffizientenvergleich führt daher auf einen Rückführvektor mit den Elementen

$$c_3 = \frac{4}{T}J_1 - \frac{J_1 + J_2}{J_2}d_w \quad (2.71)$$

$$c_2 = \frac{8}{T^2}J_1 - \frac{8}{T^3}J_1\frac{d_w}{k_w} - (J_1 + J_2)\frac{k_w}{J_2} \quad (2.72)$$

$$c_1 = \frac{8}{T^3}\frac{J_1J_2}{k_w} - c_3, \quad (2.73)$$

mit dem der geschlossene Regelkreis die vom Sollpolynom vorgegebene Dynamik erhält.

Unabhängig von der gewählten Auslegung des Zustandsreglers muss ein Vorfilter zur Sollwertskalierung verwendet werden, um die stationäre Verstärkung des Regelkreises anzupassen. Dadurch wird der Tatsache Rechnung getragen, dass die Zustandsregelung nicht die relevante Prozessgröße, sondern eine Linearkombination der Zustände regelt. Der benötigte Verstärkungsfaktor V ergibt sich aus dem Verhalten der geregelten Strecke im Gleichgewichtspunkt. Stationär gilt $\dot{x}(t) = 0$, so dass aus (2.69) und $\omega_2(t) = (1, 0, 0) x(t)$ der Zusammenhang

$$\omega_2(t) = - \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} (A - bc)^{-1} bVw(t) \quad (2.74)$$

resultiert. Aus der Forderung $\omega_2(t) \stackrel{!}{=} w(t)$ folgt daraus allgemein die benötigte Vorverstärkung

$$V = \frac{-1}{(1, 0, 0)(A - bc)^{-1}b}. \quad (2.75)$$

Angepasst auf den speziellen Fall des Zweimassensystems lautet Gleichung (2.75):

$$V = c_1 + c_3 \quad (2.76)$$

Der Vorfilter ist damit unabhängig von den Systemparametern auszulegen und hängt allein von der gewählten Rückführung ab.

Eine Zustandsregelung gestattet prinzipiell, für lineare Strecken eine beliebige Polverteilung zu erreichen und damit das Verhalten der geregelten Strecke frei zu beeinflussen. In diesem Sinne ist die Art der Ankopplung unerheblich, es lässt sich (theoretisch) ein beliebig schnelles Einschwingen erzielen bei gleichzeitig vollständiger Bedämpfung der Vibrationen in der Welle. In der praktischen Anwendung werden allerdings Grenzen durch Stellgrößenbeschränkungen und mechanische Randbedingungen gesetzt. Einerseits existiert ein maximal zulässiger Grenzwert bzw. ein erreichbarer Höchstwert für den Strom in der Antriebsmaschine. Andererseits darf eine maximale Torsion der Welle nicht überschritten werden, um eine Zerstörung zu vermeiden. Daraus folgt, dass die Beschleunigung des Prozessverhaltens nicht in beliebigem Maße gesteigert werden kann.

Um die genannten Restriktionen beim Reglerentwurf berücksichtigen zu können, bietet sich zum Beispiel für Zustandsregler die beschriebene Auslegung nach DO an. Daraus resultiert ein gut gedämpftes Einschwingverhalten (siehe Abbildung 2.12), wobei die Einschwingzeit und die Einhaltung der bestehenden Restriktionen von der Wahl der Zeitkonstanten T abhängt.

Da der Zustandsregler (2.66) proportionales Verhalten aufweist, treten unter Last bleibende Abweichungen auf. Abbildung 2.13 zeigt die Auswirkung eines Lastsprunges auf die Drehzahlen des Systems.

Um stationäre Regelfehler zu vermeiden, wird üblicherweise in der linearen Regelungstechnik ein Führungsintegrator eingesetzt. Im Sinne einer Kaskadenstruktur wird hierbei der Zustandsregelung (2.66) eine weitere Schleife überlagert. Diese bildet die Differenz aus Solldrehzahl $\omega_2^*(t)$ und Motordrehzahl und bestimmt daraus mittels I-Regler

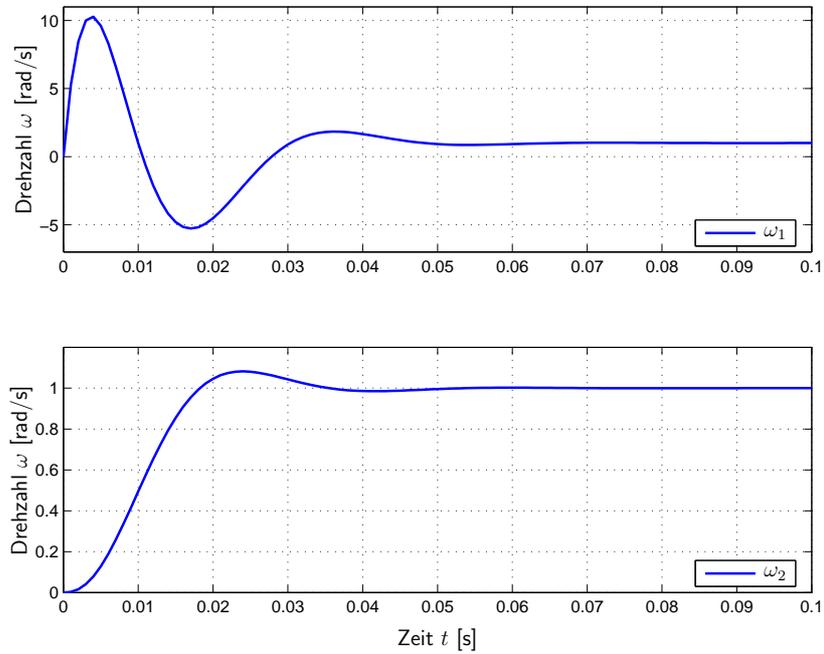


Abbildung 2.12: Sprungantwort des Drehzahlregelkreises (Zweimassensystem) mit Zustandsregler, eingestellt nach DO.

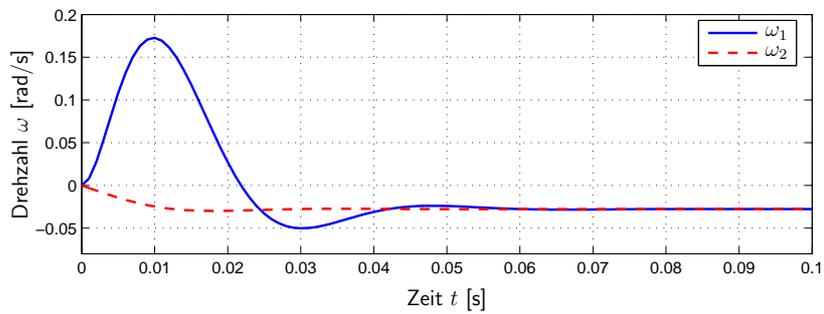


Abbildung 2.13: Antwort des Drehzahlregelkreises (Zweimassensystem) mit Zustandsregler auf einen Störgrößensprung, eingestellt nach DO.

(Führungsintegrator)

$$\begin{aligned} \dot{x}_I(t) &= \omega_2^*(t) - \omega_1(t), & x_I(0) &\in \mathbb{R} \\ w(t) &= c_4 x_I(t), & c_4 &\in \mathbb{R} \end{aligned} \quad (2.77)$$

einen geeigneten Sollwertverlauf $w(\cdot)$ für den inneren, zustandsgeregelten Kreis. Durch den zusätzlichen Zustand im I-Regler steigt die Ordnung des Regelkreises von 3 auf 4 an. Um die Optimierung nach DO beizubehalten, muss demnach die bisherige Einstellung des Zustandsreglers verworfen werden und sämtliche Auskopplungen unter Berücksichtigung des Führungsintegrators neu berechnet werden. Ergänzt um den Zustand $x_I(t)$ lautet mit $c = (c_1, c_2, c_3)$ die Beschreibung des geschlossenen Regelkreises:

$$\begin{pmatrix} \dot{x}(t) \\ \dot{x}_I(t) \end{pmatrix} = \underbrace{\begin{bmatrix} A - bc & bc_4 \\ -\tilde{c} & 0 \end{bmatrix}}_{A_{ZR}} \begin{pmatrix} x(t) \\ x_I(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \omega_2^*(t) + \begin{pmatrix} b_L \\ 0 \end{pmatrix} M_2(t) \quad (2.78)$$

mit $\tilde{c} = (0, 0, 1)$. Mittels Rückführvektor muss das charakteristische Polynom der Zustandsmatrix A_{ZR}

$$\begin{aligned} \det(sI_4 - A_{ZR}) &= s^4 + \frac{J_2 c_3 + (J_1 + J_2) d_w}{J_1 J_2} s^3 + \\ &+ \frac{J_2(c_2 + c_4) + (c_1 + c_3) d_w + (J_1 + J_2) k_w}{J_1 J_2} s^2 + \frac{d_w c_4 + (c_1 + c_3) k_w}{J_1 J_2} s + \frac{k_w c_4}{J_1 J_2} \end{aligned} \quad (2.79)$$

wiederum so eingestellt werden, dass es mit dem Normpolynom

$$N_{Norm}(s) = s^4 + \frac{8}{T} s^3 + \frac{32}{T^2} s^2 + \frac{64}{T^3} s + \frac{64}{T^4} = \left(s + \frac{2}{T} + j \frac{2}{T} \right)^2 \left(s + \frac{2}{T} - j \frac{2}{T} \right)^2 \quad (2.80)$$

nach DO für Systeme vierter Ordnung übereinstimmt. Gleichsetzen der Polynomkoeffizienten führt auf:

$$c_4 = \frac{64}{T^4} \frac{J_1 J_2}{k_w} \quad (2.81)$$

$$c_3 = \frac{8}{T} J_1 - \frac{J_1 + J_2}{J_2} d_w \quad (2.82)$$

$$c_2 = \frac{32}{T^2} J_1 - (J_1 + J_2) \frac{k_w}{J_2} - (c_1 + c_3) \frac{d_w}{J_2} - c_4 \quad (2.83)$$

$$c_1 = \frac{64}{T^3} \frac{J_1 J_2}{k_w} - \frac{d_w}{k_w} c_4 - c_3 \quad (2.84)$$

Aufgrund des Führungsintegrators ist ein Vorfilter nicht nötig, da die stationäre Verstärkung des Regelkreises

$$V_\infty = \frac{-1}{(1, 0, 0, 0) A_{ZR}^{-1} (0, 0, 0, 1)^T} = 1 \quad (2.85)$$

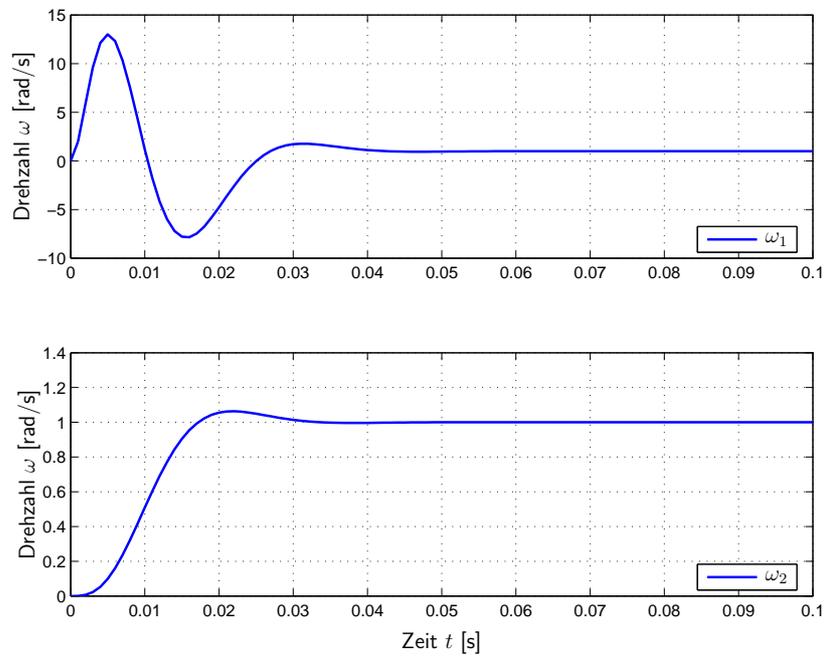


Abbildung 2.14: Sprungantwort des Drehzahlregelkreises (Zweimassensystem) mit Zustandsregler mit Führungsintegrator, eingestellt nach DO.

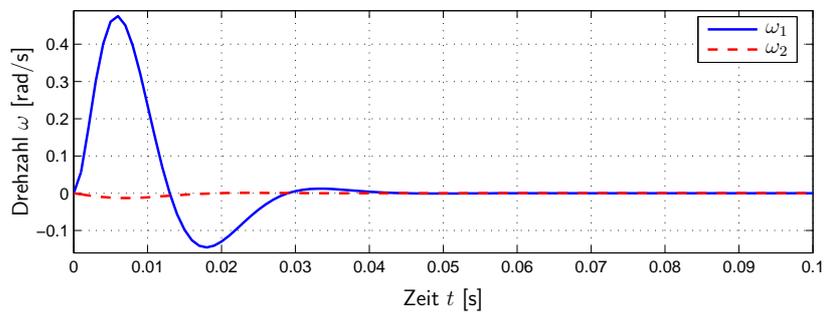


Abbildung 2.15: Antwort des Drehzahlregelkreises (Zweimassensystem) mit Zustandsregler mit Führungsintegrator auf einen Störgrößensprung, eingestellt nach DO.

bei 1 liegt. Eine Sollwertanpassung entfällt daher.

Die Ausführungen zum proportionalen Zustandsregler treffen auch bei Verwendung eines überlagerten I-Reglers zu. Abbildung 2.14 zeigt ein zu Abbildung 2.12 näherungsweise identisches Systemverhalten. Abweichend ist lediglich die in Abbildung 2.15 dargestellte Tatsache, dass der Führungsintegrator den Effekt einer konstanten Störgröße eliminiert und die Regelstrecke in ihren ursprünglichen Gleichgewichtszustand zurückkehren lässt.

2.4 Nichtlineares Modell

Die dominante Dynamik vieler Antriebssysteme lässt sich durch ein lineares Zweimassensystem modellieren und ist eingehend in Abschnitt 2.2 dargelegt. Eine lineare Beschreibung (LTI-Modell) erfasst das Verhalten eines realen Antriebssystems nicht immer in ausreichendem Maße, weil zusätzlich Effekte wie Reibung oder Lose in Betracht kommen, die im mechanischen System die wesentlichen Nichtlinearitäten darstellen. In mehrgelenkigen Roboterarmen zum Beispiel ändert sich das Massenträgheitsmoment in Abhängigkeit von den Gelenkwinkeln, weil durch eine Bewegung die Geometrie der Anordnung verändert wird und somit der Abstand zwischen Massenelement und jeweiliger Drehachse variiert. Desweiteren sind Anordnungen denkbar, deren Feder- oder Dämpferelement einer nichtlinearen Charakteristik gehorcht. Daraus motiviert sich die Erweiterung des linearen Modelles um nichtlineare Elemente.

2.4.1 Reibmodell

Im Kontext hochverstärkungsbasierter Regler fokussiert sich die Untersuchung der Reibung weniger auf deren physikalisch und quantitativ exakte Modellierung, sondern konzentriert sich viel mehr auf deren strukturelle und qualitative Eigenheiten. Damit ist es vorrangigstes Ziel dieses Abschnittes, etablierte und antriebstechnisch relevante Reibmodelle zu beleuchten und deren Einfluss auf die Regelung mechatronischer Systeme zu diskutieren.

Aus Sicht der Antriebstechnik ist die Konstellation von besonderem Interesse, bei der zwei Festkörper flächig bzw. formschlüssig gegeneinander bewegt werden.³⁾ Durch mechanische Verklammerung mikroskopischer Unebenheiten auf den Gleitflächen oder durch Adhäsion (molekulare Anziehung) können zwischen beiden Festkörperflächen in tangentialer Richtung Kraftwirkungen (Scherkräfte) übertragen werden, die sich als Reibkraft auswirken. Abhängig davon, ob ein Schmiermittel (i. allg. Mineralöle und

³⁾ Obwohl bei elektrischen Antrieben überwiegend rotatorische Bewegungen auftreten, beschränkt sich dieser Abschnitt auf translatorische Bewegungen. Die grundsätzlichen Eigenschaften sämtlicher Modelle können direkt auf rotatorische Bewegungen übertragen werden. Ein Übergang von *Reibkraft* zu *Reibmoment* ist möglich, wenn zusätzliche geometrische Gegebenheiten, wie z.B. Wellenradius, berücksichtigt werden. Weiterhin kann in rotatorischen Lagern die Anpresskraft über den Umfang variieren, was eine Integration der Reibanteile entlang des Umfangs erfordert.

synthetische Öle bzw. Fette) die Reibpartner trennt, werden folgende Arten der Reibung unterschieden:

(i) Festkörperreibung (trockene Reibung, Coulomb-Reibung):

Liegt trockene Reibung vor, so gleiten die Festkörper ohne Schmiermittel aufeinander ab. Dadurch kommt es zur Übertragung der Scherkräfte durch gegenseitiges Verhaken der Unebenheiten auf den Kontaktflächen. Die resultierende Reibkraft ist dabei unabhängig von der Gleitgeschwindigkeit und verursacht den größtmöglichen Abrieb.

(ii) Flüssigkeitsreibung (viskose Reibung):

Bei dieser Art der Reibung sind beide Festkörperflächen durch einen Schmierstoff vollständig voneinander getrennt. Eine direkte Wechselwirkung zwischen beiden Gleitflächen besteht daher nicht. Die reibungsverursachenden Scherkräfte können ausschließlich durch den Schmiermittelfilm übertragen werden, Reibung und Verschleiß reduzieren sich daher im Vergleich zur trockenen Reibung. Die vom Gleitfilm übertragbare Scherkraft hängt u.a. von der Viskosität des Schmiermittels und von der Schergeschwindigkeit ab. Entstehen kann der Gleitfilm entweder auf hydrostatische Weise (durch andauerndes Einpressen des Schmiermittels zwischen die Gleitflächen) oder durch hydrodynamischen Auftrieb. Dieser Effekt setzt eine ausreichend schnelle Bewegung voraus und lässt sich durch die Reynolds-Differentialgleichung [76, S. 233] beschreiben. Unterhalb einer Mindestgeschwindigkeit (Ausklink-Geschwindigkeit) kann sich kein tragfähiger Gleitfilm entwickeln, so dass keine vollständige Trennung der Gleitflächen gewährleistet ist. Die viskose Reibung steigt proportional zur Bewegungsgeschwindigkeit an.

(iii) Mischreibung:

Diese Form der Reibung liegt vor, wenn das Schmiermittel für eine unvollständige Trennung beider Gleitflächen sorgt und stellenweise eine direkte Berührung verbleibt. Die Dicke des Schmiermittelfilms ist hier zu gering, um die Höhe der Oberflächenunebenheiten vollständig auszugleichen. Dieser Fall tritt vor allem bei Mangel an Schmiermittel ein oder bei sehr langsamen Bewegungen (z.B. beim Anlaufen). Die niedrige Gleitgeschwindigkeit ist nicht ausreichend, um durch hydrodynamische Effekte genügend Schmiermittel zwischen die Gleitflächen zu bringen. Charakteristisch für die Mischreibung ist die Tatsache, dass eine Reibkraft geringer als bei Flüssigkeits- oder Festkörperreibung entsteht. Dennoch tritt ein im Vergleich zur Flüssigkeitsreibung höherer Abrieb auf.

Eine Übertragung von Scherkräften findet allerdings nicht nur bei gleitenden Oberflächen statt, sondern kann auch zwischen unbewegten Kontaktflächen auftreten. In diesem Falle spricht man von Haftung bzw. von Haftreibung (static friction). Im Zustand des Haftens bleiben die Kontaktstellen zwischen den mikroskopischen Oberflächenunebenheiten im Eingriff und lösen sich nicht. Dadurch kann sich eine Scherkraft aufbauen und die Wirkung einer möglichen externen Kraft eliminieren, weswegen eine Verschiebung zwischen den Festkörpern verhindert wird. Reibkraft und externe Kraft stehen damit im statischen Gleichgewicht. Sobald die externe Kraft die maximal

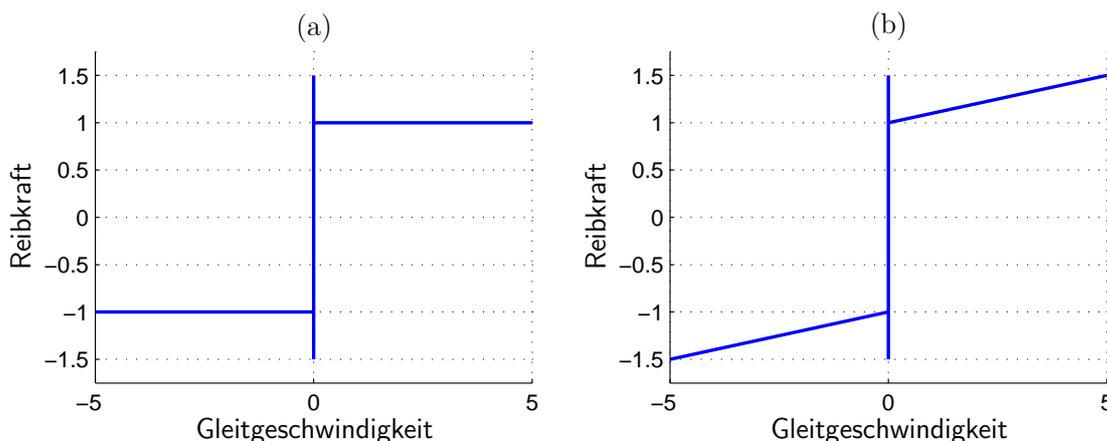


Abbildung 2.16: Klassische Reibmodelle (statisch).

mögliche Scherkraft übersteigt, lösen sich die Verbindungen zwischen den Kontaktflächen (Losbrechen) und es kommt zur Gleitbewegung.

Gewöhnlich ist zu beobachten, dass die Haftkraft die Gleitreibung übersteigt. Die Rauheiten der Kontaktflächen stehen im Ruhezustand in größtmöglichem Eingriff. Beim Losbrechen heben sich die Gleitflächen geringfügig voneinander ab, so dass die Tiefe der Verhakungen abnimmt. Daraus resultiert im Allgemeinen eine kleinere Reibkraft beim Gleiten.

Aus dem Blickwinkel der Regelungstechnik treten die physikalischen Effekte der Reibung in den Hintergrund, das Interesse gilt vor allem dem Zusammenhang zwischen Gleitgeschwindigkeit und Reibkraft, der mit dem Begriff „Reibmodell“ bezeichnet wird. Der umfassende Überblicksaufsatz [7] stellt nicht nur die wesentlichen Erkenntnisse der Tribologie ausführlich dar, sondern leitet daraus detaillierte Reibmodelle ab.

2.4.1.1 Statische Reibkraftmodelle

Statische Reibkraftmodelle besitzen keine gedächtnisbehafteten Elemente und nehmen einen statischen Zusammenhang zwischen Gleitgeschwindigkeit und Reibkraft an. Abgesehen vom Zustand der Haftung, bei dem die Reibkraft von externen Kräften beeinflusst wird, bestimmt der momentane Wert der Gleitgeschwindigkeit eindeutig den Betrag der Reibung. Der zeitliche Verlauf der Gleitgeschwindigkeit in der Vergangenheit oder die momentane Änderung der Gleitgeschwindigkeit sind dabei unerheblich.

Klassische Reibmodelle sind einfach, damit naturgemäß grob und vernachlässigen die Mischreibung. In Abbildung 2.16 sind zwei derartige Modelle dargestellt. Nachdem die Reibung der Bewegungsrichtung entgegen wirkt, wird die Orientierung des Reibkraftvektors antiparallel zum Vektor der Bewegungsgeschwindigkeit bzw. zum Vektor der extern angreifenden Kräfte ausgerichtet. Mit dieser Konvention ergibt sich daher für positive Geschwindigkeiten/Kräfte eine positive Reibkraft. Die Betrachtung der Reibung beschränkt sich in diesem Kapitel auf eindimensionale Bewegungsvorgänge,

weshalb Reibkraft und Geschwindigkeit als skalare Größen angesetzt werden können. Diagramm (a) zeigt die Kennlinie, wenn ausschließlich Haftreibung (für $v = 0$) und Coulomb'sche Reibung (für $v \neq 0$) berücksichtigt werden. Als mathematisches Modell zur Beschreibung der Coulomb'schen Reibung eignet sich zum Beispiel die Signum-Funktion.

Definition 2.1 (Signum-Funktion)

Die an der Stelle $v = 0$ unstetige Funktion

$$v \mapsto \text{sign}(v) = \begin{cases} -1 & \forall v < 0 \\ 0 & \text{für } v = 0 \\ 1 & \forall v > 0 \end{cases} \quad (2.86)$$

heißt Signum-Funktion.

Mit der maximal möglichen Haftkraft $F_H \geq 0$, dem dimensionslosen und material-abhängigen Reibungskoeffizienten $\mu_r \geq 0$, einer senkrecht zur Kontaktfläche wirkenden Anpresskraft $F_N \geq 0$ und einer externen, zur Kontaktfläche tangentialen Kraft $F_{ext} \in \mathbb{R}$, kann die Reibkraft $F_R : \mathbb{R} \rightarrow \mathbb{R}$ durch

$$v \mapsto F_R(v) = \begin{cases} F_{ext} & \text{für } v = 0 \text{ und } |F_{ext}| \leq F_H \quad (\text{Haften}) \\ F_H \cdot \text{sign}(F_{ext}) & \text{für } v = 0 \text{ und } |F_{ext}| > F_H \quad (\text{Losbrechen}) \\ \mu_r \cdot F_N \cdot \text{sign}(v) & \text{für } |v| > 0 \quad (\text{Gleiten}) \end{cases} \quad (2.87)$$

beschrieben werden. Im Stillstand ($v = 0$) haften die Gleitflächen, es kann jede beliebige Kraft aus dem Intervall $[-F_H, F_H]$ aufgebracht werden. Übersteigt der Betrag von F_{ext} den Wert F_H , kommt es zum Gleiten. Der Betrag der Gleitgeschwindigkeit ist dabei für die Berechnung der Reibkraft unerheblich, lediglich deren Richtung ist relevant.

Als zusätzliche Erweiterung ist in Diagramm (b) der Einfluss viskoser Reibung mit eingeschlossen. Die einfachste Modellierung modifiziert den Gleiterterm durch einen zusätzlichen linearen Summanden, so dass der Anstieg der Reibkraft im Bereich hoher Geschwindigkeiten wiedergegeben wird. Mit $\mu_g \geq 0$ lässt sich das Reibmodell durch

$$v \mapsto F_R(v) = \begin{cases} F_{ext} & \text{für } v = 0 \text{ und } |F_{ext}| \leq F_H \quad (\text{Haften}) \\ F_H \cdot \text{sign}(F_{ext}) & \text{für } v = 0 \text{ und } |F_{ext}| > F_H \quad (\text{Losbrechen}) \\ \mu_r \cdot F_N \cdot \text{sign}(v) + \mu_g v & \text{für } |v| > 0 \quad (\text{Gleiten}) \end{cases} \quad (2.88)$$

darstellen. Die Addition von Coulomb'scher und viskoser Reibung entspricht zwar außerhalb des Mischreibbereiches nicht der physikalischen Realität. Unterhalb der Ausklink-Geschwindigkeit besteht kein Gleitfilm, es tritt ausschließlich Trockenreibung auf. Darüber trennt die Schmierschicht die Gleitflächen, so dass anstelle der Trockenreibung nun viskose Reibung wirkt. Dennoch bietet dieses Modell eine Näherung an die realen Verhältnisse, die für viele Anwendungszwecke ausreichend genau ist. Ein Beispiel hierfür ist in [30] zu finden; die gemessene Reibkennlinie des dort untersuchten Positionierantriebes zeigt vor allem Coulomb'sche und viskose Reibanteile. Wegen ihrer

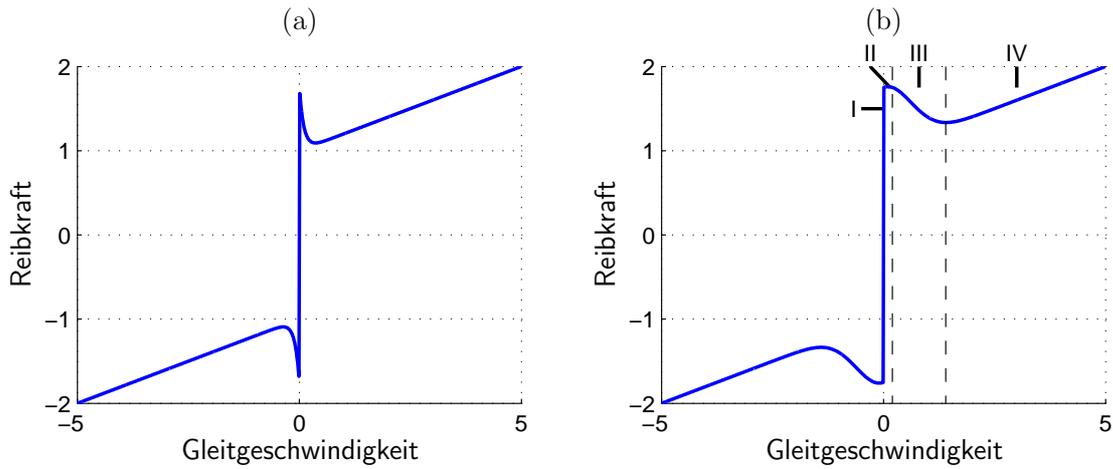


Abbildung 2.17: Präzisierte Reibmodelle mit Stribeck-Effekt (statisch).

Einfachheit und dennoch oft ausreichender Genauigkeit finden solche Reibmodelle breite Verwendung.

Allerdings versagen diese einfachen Modelle bei der Beschreibung einiger Phänomene. So lassen sich damit z.B. weder Stribeck- noch Sick-Slip Effekt erfassen.

Experimentelle Ergebnisse zeigen, dass in einigen mechanischen Anordnungen unmittelbar nach Verlassen des Haftbereiches (Stillstand) für langsame Geschwindigkeiten eine Verminderung der Reibung eintritt, was in der Tribologie als Stribeck-Effekt [7] bekannt ist. Ursache hierfür ist Mischreibung, bei der sich mangels ausreichender hydrodynamischer Hilfe der Gleitfilm nicht vollständig aufbauen kann. Mit wachsender Geschwindigkeit nimmt die Mächtigkeit der Schmierschicht zu und bewirkt einen deutlichen Rückgang der Trockenreibung sowie gleichzeitig einen geringfügigen Zuwachs viskoser Reibanteile. In Summe vermindert sich die Reibkraft dadurch. Oberhalb der Ausklink-Geschwindigkeit steigt die Reibkraft mit dem für Flüssigkeitsreibung typischen linearen Anstieg. Präzisere Modelle erfordern daher den Einschluss von Mischreibung. Eine mögliche Formulierung eines solchen Reibmodelles ist durch die Abwandlung des Terms im Gleitbereich

$$v \mapsto F_R(v) = \begin{cases} F_{ext} & \text{für } v = 0 \text{ und } |F_{ext}| \leq F_H \quad (\text{Haften}) \\ F_H \cdot \text{sign}(F_{ext}) & \text{für } v = 0 \text{ und } |F_{ext}| > F_H \quad (\text{Losbrechen}) \\ F_G & \text{für } |v| > 0 \quad (\text{Gleiten}) \end{cases} \quad (2.89)$$

mit

$$F_G = r_c \cdot \text{sign}(v) + r_v v + (r_h - r_c) e^{-\frac{|v|}{v_0}} \cdot \text{sign}(v), \quad r_c, r_h, r_v, v_0 \geq 0 \quad (2.90)$$

gegeben [7] und ist in Diagramm (a) der Abbildung 2.17 dargestellt. Der Parameter r_c beschreibt die Coulomb'sche Reibung, r_h modelliert die Haftkraft und r_v gibt die Steigung der viskosen Reibung wider.

Besonders im Bereich extrem langsamer Geschwindigkeiten kann eine weitere Verbesserung des Modelles erzielt werden, wenn auch Grenzschichtreibung (boundary lubrication) berücksichtigt wird. Eine exemplarische Kennlinie eines solchen Modells ist in

Abbildung 2.17 (b) eingezeichnet und gliedert sich in vier Gleitgeschwindigkeitsbereiche:

- **Haftung (I):**
Im Stillstand wird Haftreibung angenommen, die abhängig von den externen Kräften einen beliebigen Wert zwischen $-F_H$ und $+F_H$ annehmen kann.

- **Grenzschichtreibung (II):**
Bei der Grenzschichtreibung wird die Gleitfläche nicht von den Reibkörpern selbst, sondern durch Oberflächenbeläge (Grenzschichten) gebildet, mit Dicken von wenigen Atomschichten bis hin zu einigen Mikrometern. Zum einen können solche Grenzschichten durch Oxidreaktionsschichten (z.B. Korrosion auf Metalloberflächen) selbsttätig entstehen. In geschmierten Gleitlagern dagegen sorgen meist Additive im Schmierstoff für die bewusste Ausprägung einer solchen Adsorptionsschicht auf der Gleitfläche. Insofern tritt Grenzschichtreibung unabhängig vom Vorhandensein eines Schmiermittels auf. Nachdem es sich bei den Grenzschichten stets um Festkörperstrukturen⁴⁾ handelt, liegt hier Trockenreibung vor.

Da die Reibpartner durch die Oberflächenbeläge voneinander getrennt sind, wird das Reibverhalten erheblich von den Grenzschichten bestimmt. Gewöhnlicherweise besitzen solche Beläge kleinere Scherfestigkeiten als die Reibkörper selbst, weshalb dort die größte Scherung auftritt und damit die Reibkraft reduziert. Aus diesem Grund können Additive über den gezielten Aufbau von Grenzbelägen die Reibung beträchtlich verändern. Wird die Haftung unter die Coulomb'sche Reibung reduziert, kann dadurch das unerwünschte Auftreten des Stick-Slip Effektes vermieden werden.

Vor allem in hochpräzisen Positionsregelungen macht sich der Einfluss der Grenzschichtreibung bemerkbar. Mechanische Anordnungen mit starker Pressung der Reibflächen und geringer Reibgeschwindigkeit (z.B. Zahnflanken von Zahnrädern in Getrieben) können vollständig im Bereich der Grenzschichtreibung arbeiten [7].

- **Mischreibung (III):**
Mit wachsender Geschwindigkeit setzen verstärkt hydrodynamische Effekte ein, so dass Schmiermittel bereits teilweise in den Reibspalt gezogen wird. Allerdings bildet sich noch kein tragfähiger Gleitfilm in ausreichender Stärke aus, so dass ein Teil der Reibkraft weiterhin durch trockene Reibung verursacht wird. Mit steigender Geschwindigkeit wächst die Stärke des Schmierfilmes an, d.h. der Anteil der Trockenreibung reduziert sich. Aus diesem Grund verringert sich im Bereich der Mischreibung die Reibkraft mit zunehmender Geschwindigkeit. Dieses Phänomen ist unter dem Namen Stribeck-Effekt bekannt.
- **Viskose Reibung (IV):**
Oberhalb der Ausklink-Geschwindigkeit trennt der Schmiermittelfilm beide Gleit-

⁴⁾ Korrosionsbeläge sind in diesem Zusammenhang den Festkörpern zuzurechnen. Werden die Beläge durch Additive im Schmiermittel gebildet, so müssen auch hier Festkörperstrukturen vorliegen, um den auftretenden Belastungen widerstehen zu können.

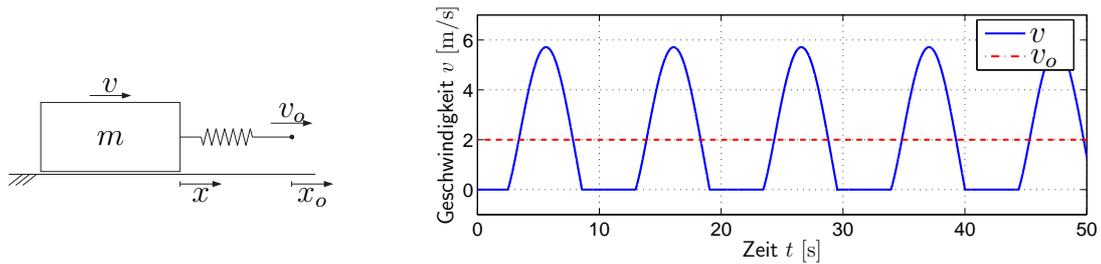


Abbildung 2.18: Schematische Versuchsanordnung zum Stick-Slip Effekt und resultierender Verlauf der Gleitgeschwindigkeit.

schichten vollständig und es dominiert viskose Reibung mit einem linearen Anstieg über der Geschwindigkeit.

Als geeignete Beschreibung dieses Modells im Gleitbereich wird in der Literatur [16] die Funktion

$$v \mapsto F_R(v) = \begin{cases} F_{ext} & \text{für } v = 0 \text{ und } |F_{ext}| \leq F_H \quad (\text{Haften}) \\ F_H \text{sign}(F_{ext}) & \text{für } v = 0 \text{ und } |F_{ext}| > F_H \quad (\text{Losbrechen}) \\ r_c \cdot \text{sign}(v) + r_v v + \\ + (r_h - r_c) e^{-\left(\frac{v}{v_o}\right)^2} \cdot \text{sign}(v) & \text{für } |v| > 0 \quad (\text{Gleiten}) \end{cases} \quad (2.91)$$

angegeben, die sich von (2.89) durch das Quadrat im Exponenten unterscheidet.

2.4.1.2 Stick-Slip Effekt

In mechanischen Systemen mit Stribeck-Effekt im Reibverhalten kann der Stick-Slip Effekt auftreten. Eine mögliche Anordnung ist in Abbildung 2.18 gegeben. Es wird ein Reibkörper mit der Masse $m = 2 \text{ kg}$ mittels elastischer Kopplung (Federhärte $k = 1 \text{ N/m}$) über eine Gleitfläche gezogen. Dabei sei angenommen, dass die masseabgewandte Seite der Feder (Zugpunkt) mit konstanter Geschwindigkeit $v_o = 2 \text{ m/s}$ bewegt wird. Die Position der Masse wird durch die Koordinate x beschrieben, die des Zugpunktes durch x_o . Die Reibcharakteristik mit der maximalen Haftkraft $F_H = 5 \text{ N}$ wird durch die Funktion

$$v \mapsto F_R(v) = \begin{cases} k(x_o - x) = F_{ext} & \text{für } v = 0 \text{ und } |F_{ext}| \leq F_H \quad (\text{Haften}) \\ F_H \cdot \text{sign}(F_{ext}) & \text{für } v = 0 \text{ und } |F_{ext}| > F_H \quad (\text{Losbrechen}) \\ (e^{-|v|} + 0.7) \text{sign}(v) & \text{für } |v| > 0 \quad (\text{Gleiten}) \end{cases} \quad (2.92)$$

festgelegt. Solange die Feder noch nicht ausreichend gespannt ist, haftet die Masse. Erreicht die Federkraft den kritischen Wert, löst sich der Reibkörper und beschleunigt. Der Stribeck-Effekt vermindert die Gleitreibung stark, so dass der Körper einer nur geringen Reibkraft und einer hohen Zugkraft ausgesetzt ist. Wegen der daraus resultierenden Beschleunigung holt der Reibkörper den Zugpunkt ein und reduziert folglich die Dehnung der Feder. Die Federkraft nimmt solange ab, bis die Reibkraft überwiegt und

den Reibkörper verlangsamt. Der Stribeck-Effekt verursacht schließlich einen raschen Anstieg der Reibung und bremst dadurch den Reibkörper bis zum Stillstand ab. Von diesem Zustand aus wiederholt sich der Zyklus erneut. Der Reibkörper nimmt bei auftretendem Stick-Slip Effekt keine konstante Geschwindigkeit an, sondern bewegt sich schubweise, wie das t - v -Diagramm in Abbildung 2.18 darstellt.

2.4.1.3 Dynamische Reibkraftmodelle

Dynamische Reibkraftmodelle geben einen Zusammenhang von Geschwindigkeitsverlauf und Reibkraft wieder. Im Gegensatz zu statischen Modellen ist hier nicht nur der aktuelle Wert der Geschwindigkeit von Interesse, sondern auch vergangene Werte, ebenso wie die aktuelle Geschwindigkeitsänderung.

Besondere Bedeutung kommt dynamischen Modellen zu, wenn die physikalischen Gegebenheiten präzise modelliert werden müssen. Phänomene wie Hysterese-Charakteristik bei Geschwindigkeitsänderungen, Dahl-Effekt oder variierende (verweilzeitabhängige/kraftänderungsabhängige) Losbrechkraft können durch den Einschluss von dynamischem Verhalten vom Reibmodell nachgebildet werden.

- Dahl-Effekt:

In statischen Reibmodellen werden die hervorstehenden Unebenheiten, welche die Kontaktkräfte übertragen, als vollkommen starr angenommen. In der Haftung ist keine Verschiebung möglich, beide Reibpartner sind fest gegeneinander fixiert. Anders gestaltet sich der Sachverhalt, wenn die Unebenheiten der Oberflächen als elastische Biegebalken (bzw. als Federn mit hoher Steifigkeit) modelliert werden, die im gegenseitigen Eingriff stehen. Zwar ist auch dann ein Gleiten zwischen den Berührungspunkten der Biegebalken ausgeschlossen. Allerdings gestattet hier die angenommene Verformbarkeit der Unebenheiten eine Verschiebung der Reibkörper gegeneinander (presliding displacement) [7], [16].

Gemäß des Hooke'schen Gesetzes hängt die entstehende Verschiebung linear von der äußeren Kraft ab. Bei Stahl beispielsweise liegt die im Haftbereich maximal mögliche Verschiebung in der Größenordnung weniger Mikrometer [7]. Oberhalb einer kritischen Kraft wird die Verschiebung zu groß, so dass die haftende Verbindung losbricht. Der Dahl-Effekt spielt vor allem bei hochpräzisen Positionierungsaufgaben eine Rolle. Es ist festzustellen, dass sich der Übergang von elastischer Haftung zu Gleiten nicht schlagartig vollzieht und daher eine aufwändige Modellierung erfordert.

- Variierende Losbrechkraft:

Es ist zu beobachten, dass die erforderliche Losbrechkraft, um aus der Haftung in die Gleitbewegung überzugehen, von der Kraftänderung abhängt. Je schneller die äußere Kraft ansteigt, desto geringer ist die Losbrechkraft [16].

- Hysterese:

Wenn eine unidirektionale Bewegung vorliegt ($\text{sign}(v) = \text{const.}$), können in geschmierten Gleitlagern bei Geschwindigkeitsänderungen Hysterese-Effekte auf-

treten. Die Reibkraft hängt demzufolge nicht allein von der Gleitgeschwindigkeit ab, sondern zusätzlich auch von der Geschwindigkeitsänderung. Grund hierfür ist eine zeitliche Verzögerung in der Änderung der Schichtstärke des Schmierfilms (frictional memory bzw. time lag). Bei Beschleunigungsvorgängen ergibt das zeitlich verzögerte Anwachsen des Gleitfilms im Mittel geringere Filmstärken als beim Bremsen; hier sorgt die Trägheit im Abbau des Gleitfilms für vergleichsweise höhere Schichtstärken. In [7] und [16] ist dieses Phänomen experimentell untersucht und die resultierenden Hystereseschleifen ermittelt. Die dargestellten Messergebnisse zeigen, dass die Weite der Hystereseschleifen zunimmt, wenn die Geschwindigkeitsänderung gesteigert wird.

2.4.1.4 Reibkennlinie des Prüfstandes

Die Dissertation [78] befasst sich mit modellbasierter prädiktiver Regelung des beschriebenen Zweimassensystems und ist daher auf ein hinreichend genaues Reibmodell angewiesen. Die in [78] experimentell ermittelten Kennlinien für beide Maschinen des Prüfstandes sind in Abbildung 2.19 aufgetragen. Eine geeignete Beschreibung der Reib-

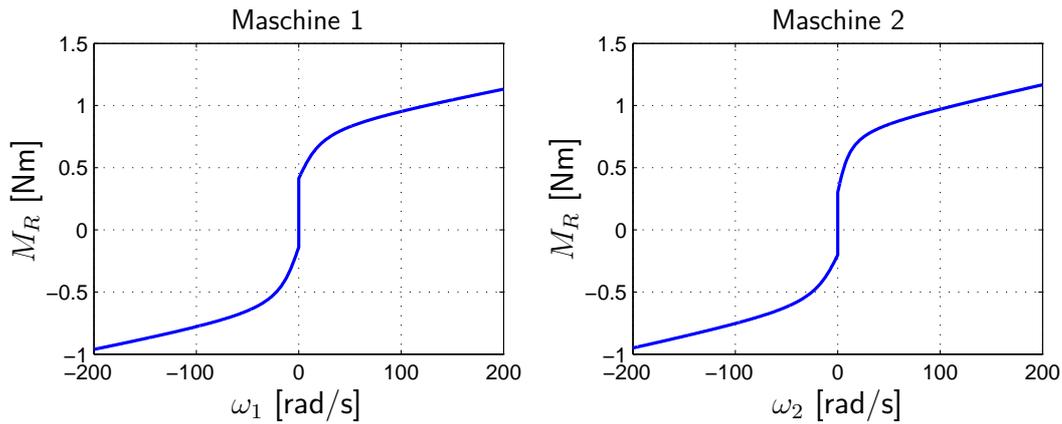


Abbildung 2.19: Am Zweimassensystem experimentell bestimmte Reibkennlinien.

kennlinie des Motors und der Lastmaschine ist durch

$$\omega_1 \mapsto \begin{cases} 0.4131 + 0.273 \cdot \arctan(0.0581 \cdot \omega_1) + 0.0016 \cdot \omega_1 & \text{für } \omega_1 > 0 \\ [-0.1394, 0.4131] & \text{für } \omega_1 = 0 \\ -0.1394 + 0.3398 \cdot \arctan(0.0652 \cdot \omega_1) + 0.0016 \cdot \omega_1 & \text{für } \omega_1 < 0 \end{cases} \quad (2.93)$$

bzw. durch

$$\omega_2 \mapsto \begin{cases} 0.3029 + 0.3293 \cdot \arctan(0.1048 \cdot \omega_2) + 0.0018 \cdot \omega_2 & \text{für } \omega_2 > 0 \\ [-0.2025, 0.3029] & \text{für } \omega_2 = 0 \\ -0.2025 + 0.2735 \cdot \arctan(0.0532 \cdot \omega_2) + 0.0017 \cdot \omega_2 & \text{für } \omega_2 < 0 \end{cases} \quad (2.94)$$

gegeben. Bedingt durch die Haftreibung im Stillstand ist eine Unstetigkeitsstelle in der Kennlinie enthalten. Dadurch ergeben sich Probleme bei der mathematischen Analyse des Systemverhaltens. Abhilfe kann geschaffen werden, indem eine stetige Kennlinie als Näherung angesetzt wird.

2.4.1.5 Reibmodell im Hinblick auf hochverstärkungsbasierte Regelung

Reibung ist bei der Regelung mechanischer Systeme meist unerwünscht, da beispielsweise Stick-Slip-Bewegungen oder bleibende Regelfehler im Falle proportionaler Regler eintreten können. Werden zur Beseitigung der Regelfehler integrale Anteile im Regler eingesetzt, können in Positionierantrieben Grenzyklen (hunting) entstehen [16]. Neben diesen negativen Auswirkungen, die sich durch geeignete Regelungstechnik beseitigen lassen, führt Reibung auch zu erhöhtem Energieverbrauch.

Die gängigsten Methoden, um den Einfluss von Reibung durch regelungstechnische Maßnahmen zu eliminieren, stützen sich auf präzise Reibmodelle. Sind solche Modelle nicht vorhanden, können Identifikationsverfahren (z.B. mit Hilfe Neuronaler Netze) eingesetzt werden, um die vorliegende Reibkennlinie zu lernen. Das Modell schätzt bzw. sagt den Einfluss der Reibung voraus, so dass mittels Vorsteuerung oder Störgrößenaufschaltung eine Kompensation erreicht wird [30]. Alternativ zu dieser aufwändigen Vorgehensweise bietet sich der Einsatz hochverstärkungsbasierter Regler an [16]. Hierfür ist allerdings eine Regelstrecke mit stabiler Nulldynamik erforderlich.

Als Endergebnis sei festgehalten, dass die hier vorgestellten Reibmodelle in zwei Komponenten aufgespaltet werden können. Ein linearer Anteil beschreibt die viskose Reibung und ermöglicht ein unbeschränktes Anwachsen der Reibkraft über der Gleitgeschwindigkeit. Alle weiteren Effekte summieren sich in einem nichtlinearen Operator

$$\mathbf{N} : \mathcal{C}([0, \infty[; \mathbb{R}) \rightarrow \mathcal{L}^\infty([0, \infty[; \mathbb{R}), \quad (2.95)$$

der allerdings global beschränkt ist und deshalb die Bedingung

$$\sup \left\{ |(\mathbf{N}\zeta)(t)| \mid t \geq 0, \zeta \in \mathcal{C}([0, \infty[; \mathbb{R}) \right\} < \infty \quad (2.96)$$

erfüllt. Für das betrachtete Zweimassensystem lässt sich daher das Reibmoment durch

$$M_R(\cdot) = \mu_g \omega_2(\cdot) + (\mathbf{N}\omega_2)(\cdot), \quad \mu_g \geq 0 \quad (2.97)$$

beschreiben. Mit dieser Modellierung erfüllt die Reibung die notwendigen Eigenschaften, damit eine hochverstärkungsbasierte Regelung des Zweimassensystems möglich ist.

2.4.2 Integration von Reibung in das Modell des ZMS

Eine nichtlineare Reibcharakteristik an der Arbeitsmaschine wird in das lineare Modell (2.6) eingeschlossen, indem die Momentenbilanz an der Lastmasse abgeändert und Gleichung (2.4) als

$$J_2 \ddot{\varphi}_2(\cdot) = M_w(\cdot) - M_2(\cdot) - \mu_g \omega_2(\cdot) - (\mathbf{N}\omega_2)(\cdot) \quad (2.98)$$

geschrieben wird. Das Reibmoment ist abhängig von der Lastdrehzahl und wirkt zusammen mit dem Lastmoment bremsend auf die Lastmasse ein. Daraus folgt das nichtlineare Zustandsmodell

$$\begin{aligned} \dot{x}(t) &= Ax(t) + b_L [\mu_g \omega_2(t) + (\mathbf{N}\omega_2)(t)] + b [u(t) + \hat{d}(t)], & x(0) \in \mathbb{R}^3 \\ y(t) &= cx(t) \end{aligned} \quad (2.99)$$

mit dem Gleitreibungskoeffizienten $\mu_g \geq 0$ (Einheit: [Nms/rad]) und der zusätzlichen Störung $\hat{d} \in \mathcal{L}^\infty(\mathbb{R}; \mathbb{R})$ am Eingang der Strecke. Eine solche Störung ist realistisch, da der Regler die gewünschte Stellgröße (antreibendes Moment $M_1(t)$) nicht direkt einprägt, sondern über einen Umrichter aufbringt. Die hierdurch bedingten Verfälschungen sind im Modell nicht explizit berücksichtigt und werden in $\hat{d}(\cdot)$ zusammengefasst.

Kapitel 3

Hochverstärkungsbasierte adaptive Regler

Hochverstärkungsbasierte Regler bringen eine bemerkenswerte Neuerung im Gebiet der adaptiven Regelung, da auf die Identifikation unbekannter Systemdynamik verzichtet werden kann. In der Folge reduziert sich die Komplexität von Adaptionsgesetz und Regler, so dass mit vergleichsweise geringen Anforderungen an die notwendige Rechenkapazität eine Stabilisierung von unbekanntem Regelstrecken erreichbar wird.

Kennzeichnend für alle hochverstärkungsbasierten Regler ist die zugrunde liegende Annahme, dass der Regelkreis mit steigender Kreisverstärkung zunehmend verbesserte Stabilitätseigenschaften erhält. Vor diesem Hintergrund wird die Verstärkung eines proportionalen Reglers hinreichend hoch gewählt bzw. adaptiv eingestellt. Eine solche Voraussetzung schränkt den Anwendungsbereich hochverstärkungsbasierter Regler allerdings stark ein und stellt harte Anforderungen an die Regelstrecke. Der Einsatz hochverstärkungsbasierter Regler bleibt für Prozesse aus der Systemklasse \mathcal{S} vorbehalten, die sich durch folgende Eigenschaften auszeichnen:

Definition 3.1 (Systemklasse \mathcal{S})

Die Systemklasse \mathcal{S} enthält alle linearen Regelstrecken

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), & x(0) &= x_o \in \mathbb{R}^n \\ y(t) &= cx(t) \end{aligned} \quad (3.1)$$

der Ordnung $n \in \mathbb{N}$, $A \in \mathbb{R}^{n \times n}$ und $b, c^T \in \mathbb{R}^n$ mit

- (i) Relativgrad $r = 1$, d.h. $cb \neq 0$,
- (ii) stabiler Nulldynamik, bzw. minimalphasigem Verhalten, d.h.

$$\det \begin{bmatrix} sI_n - A & -b \\ c & 0 \end{bmatrix} \neq 0 \quad \forall s \in \mathbb{C} \text{ mit } \operatorname{Re}\{s\} \geq 0 \quad (3.2)$$

- (iii) und einer instantanen Verstärkung mit strikt positivem Vorzeichen, d.h. $cb > 0$.

Anmerkung

- (i) Jede Regelstrecke aus der Systemklasse \mathcal{S} kann durch Anwendung der in [35, S. 11] vorgestellten Ähnlichkeitstransformation

$$\begin{pmatrix} y(t) \\ z(t) \end{pmatrix} = Sx(t), \quad S \in \mathbf{GL}_n(\mathbb{R}), \quad y(t) \in \mathbb{R}, \quad z(t) \in \mathbb{R}^{n-1} \quad (3.3)$$

in BINF dargestellt werden. In dieser Systembeschreibung ist die Regelstrecke in „Vorwärtszweig“ und „Nullodynamik“ partitioniert und wird demnach durch die Differentialgleichung

$$\begin{pmatrix} \dot{y}(t) \\ \dot{z}(t) \end{pmatrix} = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{pmatrix} y(t) \\ z(t) \end{pmatrix} + \begin{pmatrix} cb \\ 0 \end{pmatrix} u(t), \quad \begin{pmatrix} y(0) \\ z(0) \end{pmatrix} = \begin{pmatrix} y_o \\ z_o \end{pmatrix} \in \mathbb{R}^n \quad (3.4)$$

wiedergegeben, wobei der Relative Grad von 1 eine skalare Größe $A_1 \in \mathbb{R}$ und eine Matrix $A_4 \in \mathbb{R}^{(n-1) \times (n-1)}$ bedingt. Durch $A_2 \in \mathbb{R}^{1 \times (n-1)}$ und $A_3 \in \mathbb{R}^{(n-1) \times 1}$ ist die wechselseitige Verkopplung von Vorwärtszweig und Nullodynamik festgelegt.

- (ii) Aus Abschnitt A.3 wird die Berechnung und die physikalische Bedeutung der instantanen Verstärkung cb offensichtlich. Da dieser Wert für Systeme der Klasse \mathcal{S} der zeitlichen Änderung des Ausgangs unmittelbar nach Sprunganregung entspricht, kann die Kenntnis des Vorzeichens als Forderung nach einer bekannten Wirkungsrichtung der Stellgröße angesehen werden. Dies bedeutet im Allgemeinen keine besondere Einschränkung und kann für technisch relevante Systeme häufig bereits ohne genaue Kenntnis der Streckenparameter angegeben werden.
- (iii) Die Forderung nach einem *positiven* Vorzeichen der instantanen Verstärkung bedeutet keine wesentliche Restriktion, weil durch eine Multiplikation der Stellgröße mit -1 das Vorzeichen invertiert werden kann. Sofern die Bedingungen an den Relativgrad und die Nullodynamik erfüllt werden, genügt daher ein zusätzlicher Vorfaktor -1 , um eine Strecke mit negativer instantaner Verstärkung in die Systemklasse \mathcal{S} einzuschließen. Der Vorfaktor beeinflusst dabei weder den Relativgrad, noch die Stabilität der Nullodynamik.
- (iv) Die benötigte Stabilität der Nullodynamik schließt hingegen bereits eine große Klasse von Regelstrecken aus, wird allerdings von den Prozessen der Antriebstechnik meist erfüllt. Im Spezialfall linearer Systeme ist die Stabilität der Nullodynamik (das Spektrum der Matrix A_4 weist ausschließlich negative Realteile auf) gleichbedeutend mit Minimalphasigkeit, sämtliche Wurzeln des Zählerpolynoms müssen einen negativen Realteil besitzen. Die Forderung nach minimalphasigem Systemverhalten impliziert Stabilisierbarkeit und Detektierbarkeit, siehe Theorem A.42.
- (v) Große Probleme bereitet jedoch die Anforderung an den Relativgrad. Bei Prozessen mit einem Relativgrad von 1 beeinflusst der Systemeingang direkt die erste Ableitung des Ausgangs, siehe Gleichung (3.4) und Abschnitt A.6. Demnach kann durch eine entsprechend hohe Stellgröße ohne Verzögerung eine beliebig schnelle Änderung des Ausgangssignales hervorgerufen werden. Ein niedriger

Relativgrad ist daher gleichbedeutend mit schnell reagierenden Prozessen. Die überwiegende Mehrzahl industrieller Regelstrecken weist allerdings eine deutlich ausgeprägte Trägheit in der Systemantwort auf, die sich durch einen entsprechend hohen Relativgrad äußert. So ist zum Beispiel eine elastische Kopplung zwischen Antrieb und Arbeitsmaschine bei Zweimassensystemen bereits ausreichend, um die Relativgradbedingung zu verletzen. Wird zusätzlich ein Umrichter als Stellglied berücksichtigt, erhöht sich der Relativgrad nochmals. Deswegen können hochverstärkungsbasierte Regelungsverfahren nicht ohne weitere Modifikation zur Regelung von elektrischen Antrieben eingesetzt werden. Eine anwendungstaugliche Möglichkeit zur Relativgradreduzierung wird in Kapitel 5 vorgestellt, um die Regelung des vorliegenden Zweimassensystems mit hochverstärkungsbasierten Methoden zu ermöglichen. □

Bevor in ausführlicher Weise hochverstärkungsbasierte Regler mit den zugehörigen Adaptionsgesetzen dargestellt werden, sollen zunächst in den Abschnitten 3.1 bis 3.3 die Auswirkungen hoher aber konstanter Verstärkungen auf den Regelkreis beleuchtet werden.

3.1 Auswirkung hoher Kreisverstärkungen

Bekanntermaßen existieren Regelsysteme, die sich durch zu hohe Kreisverstärkungen destabilisieren lassen. Für lineare Regelkreise ist diese Tatsache anhand von Wurzelortskurven direkt ersichtlich. Falls die Regelstrecke jedoch der Systemklasse \mathcal{S} angehört, entstehen für hohe Verstärkungen keine Stabilitätsprobleme. Die Zugehörigkeit zur Systemklasse \mathcal{S} ist hinreichend, um hochverstärkungsbasierte Regler erfolgreich anwenden zu können. In [54, S. 660f] wird ausgeführt, dass die geregelten Strecken aufgrund hoher Kreisverstärkungen folgende Eigenschaften erhalten:

- (i) Eine wichtige Folge hoher Verstärkungen im Vorwärtszweig eines Regelkreises ist, dass die Abbildung vom Sollwert (Eingang) auf den Istwert (Ausgang) näherungsweise mit der inversen Abbildung des Rückwärtszweiges übereinstimmt, was durch die Bezeichnung „inverser Regler“ ausgedrückt wird. Der Baustein im Rückwärtspfad gibt damit die Eigenschaften des Regelkreises vor. Im Spezialfall der Einheitsrückführung sind Sollwert und Istwert daher gleich.
- (ii) Für hohe Verstärkungen entsteht der inverse Regler auch dann, wenn die Strecke nur unzureichend bekannt ist, oder ungünstige dynamische Eigenschaften besitzt. Diese Eigenschaft kann unter dem Stichwort „hohe Robustheit“ eingeordnet werden.
- (iii) Sowohl die Linearität des Regelkreises, als auch die Bandbreite lassen sich durch hohe Verstärkungen verbessern.

Allerdings stehen diesen Vorteilen auch Nachteile gegenüber, mit welchen die positiven Effekte erkauft werden müssen:

- (i) Aufgrund der hohen Verstärkung führen bereits kleine Regelfehler zu extrem hohen Stellgrößen. Reale Stellglieder weisen jedoch stets Stellgrößenbeschränkungen auf. Insofern ist zu klären, ob das Stellglied genügend Stellgröße liefern kann (Kostenfrage) und ob die Anlage mit solch hohen Stellenergien schadlos beaufschlagt werden darf.
- (ii) Messfehler und vor allem Messrauschen verschlechtern bei hohen Verstärkungen das Regelergebnis erheblich. In der Praxis muss daher auf hochwertige Sensoren Wert gelegt werden, was entsprechend höhere Kosten mit sich bringt.

Punkt (i) in der Vorteilsliste deutet darauf hin, dass bei Einheitsrückführung alle Strecken aus \mathcal{S} für hohe Verstärkungen das selbe Verhalten aufweisen (inverser Regler). Unabhängig von der Ordnung der Regelstrecke, nähert sich das dynamische Verhalten des Regelkreises im Grenzfall, d.h. für hinreichend hohe Verstärkung, einem beliebig schnellen System erster Ordnung an. Die beiden folgenden Abschnitte gehen darauf näher ein, wobei zuerst nur lineare Systeme untersucht werden. Abschnitt 3.3 erweitert die Betrachtungen schließlich auf nichtlineare Systeme.

3.2 Hohe Verstärkung im linearen Regelkreis

Aussagen über die Auswirkung hoher Verstärkungen im linearen Regelkreis werden in der Literatur meist mittels Wurzelortskurven abgeleitet [23, S. 200ff], [59, S. 415ff], [79, S. 157ff]. Hierbei wird eine Systemdarstellung im Laplace-Bereich genutzt, um mit Hilfe der Eigenschaften des offenen Regelkreises das Verhalten des geschlossenen Regelkreises zu bestimmen.

Davon abweichend bietet sich eine Analyse im Zeitbereich an, die für den geschlossenen Kreis die Eigenwerte der Systemmatrix untersucht. Den folgenden Betrachtungen wird eine Strecke (3.4) der Ordnung $n \in \mathbb{N}$ zugrunde gelegt und deren Zugehörigkeit zu \mathcal{S} vorausgesetzt.

Das Spektrum der Matrix $A_4 \in \mathbb{R}^{(n-1) \times (n-1)}$ ist entscheidend für das Verhalten der Strecke unter hohen Verstärkungen im Regelkreis. Aus diesem Grund kommt den $n - 1$ Eigenwerten $\lambda_i \in \mathbb{C}$, $i = 1, 2, \dots, n - 1$, der Matrix A_4 besondere Bedeutung zu, wenn das Stabilisationsgesetz

$$u(t) = -ky(t) \tag{3.5}$$

auf die Regelstrecke (3.4) angewendet wird. Es lässt sich zeigen, dass ein stabiles Spektrum von A_4 notwendig ist, damit bei Überschreiten einer Mindestverstärkung $k^* \in \mathbb{R}$ für alle $k \geq k^*$ ein stabiler Regelkreis resultiert. Die Bedingung $k \geq k^*$ ist hinreichend, damit $n - 1$ Eigenwerte des Regelkreises in einer ε -Umgebung um das Spektrum von A_4 liegen.

Um eine derartige Umgebung präzise beschreiben zu können, wird in der komplexen Ebene zunächst ein offener ε -Ball mit Zentrum $\lambda \in \mathbb{C}$ durch

$$\mathcal{B}_\varepsilon(\lambda) := \{s \in \mathbb{C} \mid |\lambda - s| < \varepsilon\}, \quad \varepsilon > 0 \tag{3.6}$$

definiert. Wird eine solche Kreisscheibe um jeden Eigenwert $\{\lambda_1, \dots, \lambda_{n-1}\} = \sigma(A_4)$ gelegt, ergibt die Vereinigungsmenge der Kreisscheiben

$$G_\varepsilon = \mathcal{B}_\varepsilon(\lambda_1) \cup \mathcal{B}_\varepsilon(\lambda_2) \cup \dots \cup \mathcal{B}_\varepsilon(\lambda_{n-1}) \quad (3.7)$$

ein Gebiet G_ε , das eine ε -Umgebung um das Spektrum von A_4 bildet und hilfreich ist, um folgendes Theorem zu formulieren.

Theorem 3.2

(i) Zu jedem beliebigen $\varepsilon > 0$ existiert eine Konstante $E > 0$, so dass das Stabilisierungsgesetz (3.5) zusammen mit der Regelstrecke (3.4) zu einem geschlossenen Regelkreis führt, dessen Eigenwerte ausschließlich im Gebiet $G_\varepsilon \cup \mathcal{B}_E(A_1 - kcb)$ liegen.

(ii) Zusätzlich existiert zu jedem beliebigen $\varepsilon > 0$ eine Konstante $k^* > 0$, so dass sich für alle $k \geq k^*$ die Eigenwerte des Regelkreises in $G_\varepsilon \cup \mathcal{B}_\varepsilon(A_1 - kcb)$ befinden, wobei G_ε und $\mathcal{B}_\varepsilon(A_1 - kcb)$ disjunkte Gebiete darstellen, d.h. $G_\varepsilon \cap \mathcal{B}_\varepsilon(A_1 - kcb) = \emptyset$.

Beweis

Durch Einsetzen von (3.5) in (3.4) ergibt sich die homogene lineare Differentialgleichung

$$\begin{pmatrix} \dot{y}(t) \\ \dot{z}(t) \end{pmatrix} = \begin{bmatrix} A_1 - kcb & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{pmatrix} y(t) \\ z(t) \end{pmatrix}, \quad \begin{pmatrix} y(0) \\ z(0) \end{pmatrix} = \begin{pmatrix} y_o \\ z_o \end{pmatrix}, \quad (3.8)$$

die den geschlossenen Regelkreis beschreibt. Analog zum Beweis von Theorem 2.2 in [36] wird die Systemmatrix zunächst in geeignete Form transformiert, um anschließend die Lage der Eigenwerte mittels Geršgorin-Theorem abzuschätzen.

Nachdem zu jeder quadratischen Matrix $A_4 \in \mathbb{C}^{(n-1) \times (n-1)}$ die zugehörige Jordanform existiert (siehe z.B. Folgesatz 1 in [55, S. 237]), ist mit einer geeignet gewählten Matrix $S \in \mathbf{GL}_{n-1}(\mathbb{C})$ die Ähnlichkeitstransformation

$$A_4^J := SA_4S^{-1} = \begin{bmatrix} \lambda_1 & \delta_1 & & \\ & \ddots & \ddots & \\ & & \lambda_{n-2} & \delta_{n-2} \\ & & & \lambda_{n-1} \end{bmatrix}, \quad \delta_j \in \{0, 1\}, \quad j = 1, \dots, n-2 \quad (3.9)$$

möglich. Sofern $\lambda_p \neq \lambda_q$ für alle $p, q = 1, 2, \dots, n-1$ mit $p \neq q$, liegen ausschließlich einfache Eigenwerte vor. In diesem Falle gilt $\delta_j = 0$ für alle $j = 1, \dots, n-2$, so dass A_4^J Diagonalstruktur besitzt. Bei mehrfachen Eigenwerten verliert die Matrix A_4^J ihre Diagonalgestalt, da die entsprechenden Stellen in der ersten Nebendiagonalen den Eintrag 1 enthalten (Jordan-Blöcke). Damit erzeugt die Transformation

$$\begin{bmatrix} 1 & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} A_1 - kcb & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & S^{-1} \end{bmatrix} = \begin{bmatrix} A_1 - kcb & A_2S^{-1} \\ SA_3 & A_4^J \end{bmatrix} \quad (3.10)$$

ein Zustandsraummodell, dessen Nulldynamik in Jordanform vorliegt. Dem Vorgehen in [36] folgend, wird eine zweite Ähnlichkeitstransformation durchgeführt, vornehmlich

um einen Eingriff auf den Spaltenvektor SA_3 und die Nebendiagonale von A_4^J zu erhalten und dadurch die Geršgorin-Radien der Zeilen 2 bis n beeinflussen zu können. Mit einer Konstanten $K > 0$ lässt sich die invertierbare Diagonalmatrix

$$T = \text{diag}(K^{-(n-1)}, K^{-(n-2)}, \dots, K^{-1}) \in \mathbf{Gl}_{n-1}(\mathbb{R}) \quad (3.11)$$

aufbauen, mit deren Hilfe die Transformation

$$\begin{bmatrix} 1 & 0 \\ 0 & T \end{bmatrix} \begin{bmatrix} A_1 - kcb & A_2 S^{-1} \\ SA_3 & A_4^J \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & T^{-1} \end{bmatrix} = \begin{bmatrix} A_1 - kcb & A_2 (TS)^{-1} \\ TSA_3 & TA_4^J T^{-1} \end{bmatrix} \quad (3.12)$$

erfolgen kann. Wird das Matrizenprodukt $TA_4^J T^{-1}$ ausgeschrieben, lautet die Systemmatrix nach der Transformation:

$$\begin{bmatrix} A_1 - kcb & A_2 (TS)^{-1} \\ TSA_3 & TA_4^J T^{-1} \end{bmatrix} = \left[\begin{array}{c|ccc} A_1 - kcb & & & A_2 (TS)^{-1} \\ \hline & \lambda_1 & \delta_1/K & \\ & & \ddots & \ddots \\ TSA_3 & & & \ddots & \delta_{n-2}/K \\ & & & & \lambda_{n-1} \end{array} \right] \quad (3.13)$$

Um die Eigenwerte der transformierten Systemmatrix abzuschätzen, lässt sich das in Abschnitt A.12.4 angegebene Geršgorin-Theorem anwenden. Die Radien der Geršgorin-Scheiben berechnen sich dabei durch die jeweilige Zeilensumme der Beträge ohne Diagonalelement. Für die Zeilen 2 bis n sind mit der Definition $\delta_{n-1} := 0$ und den Einheitsvektoren e_i^{n-1} , $i = 1, 2, \dots, n-1$, die Geršgorin-Radien durch

$$R_q = |e_{q-1}^T SA_3| \cdot \frac{K^q}{K^{n+1}} + \frac{\delta_{q-1}}{K} \Rightarrow R_q = \mathcal{O}\left(\frac{1}{K}\right), \quad q = 2, 3, \dots, n \quad (3.14)$$

gegeben. Die Auswertung der ersten Zeile resultiert abweichend in

$$R_1 = K^{n-1} |e_1^T A_2 S^{-1}| + K^{n-2} |e_2^T A_2 S^{-1}| + \dots + K |e_{n-1}^T A_2 S^{-1}|. \quad (3.15)$$

Offensichtlich existiert eine hinreichend große Konstante K , so dass die Ungleichungen

$$R_q < \varepsilon \quad \text{für alle } q = 2, 3, \dots, n \quad (3.16)$$

erfüllt sind. Damit können mit Hilfe von K die Geršgorin-Radien um die Eigenwerte der Nulldynamik beliebig klein eingestellt werden. Allerdings gewichtet der hohe Betrag von K , der in den Potenzen von 1 bis $n-1$ auftritt, die Summanden in (3.15) und weitet dadurch den Radius R_1 auf. Um (3.16) für jedes beliebige $\varepsilon > 0$ zu erfüllen, muss folglich ein entsprechend großer Radius R_1 in Kauf genommen werden. Jede Konstante $E \geq R_1$ definiert ein Gebiet $G_\varepsilon \cup \mathcal{B}_E(A_1 - kcb)$ in der Weise, dass darin sämtliche Eigenwerte der Systemmatrix enthalten sind. Aussage (i) in Theorem 3.2 ist daher zwar für alle $k \in \mathbb{R}$ verifiziert, beinhaltet allerdings nicht, dass $n-1$ Eigenwerte des geschlossenen Kreises in G_ε liegen. Letzteres trifft erst zu, wenn zusätzliche Einschränkungen an k gestellt werden.

Nachdem k Einfluss auf die Position der Geršgorin-Scheibe um das Zentrum $A_1 - kcb$ nimmt, gewährleistet eine hinreichend große Verstärkung

$$\forall k \geq k^* : G_\varepsilon \cap \mathcal{B}_E(A_1 - kcb) = \emptyset. \quad (3.17)$$

Daraus resultiert die Folgerung, dass sich $n - 1$ Eigenwerte der Systemmatrix im ε -Ball um das Spektrum der Nulldynamik befinden.

Für den verbleibenden Eigenwert $\lambda_n \in \mathcal{B}_E(A_1 - kcb)$ wird eine verbesserte Abschätzung erreicht, wenn das Geršgorin-Theorem auf die transponierte Systemmatrix angewendet wird. Nach Transponieren von Gleichung (3.13) ergibt sich für die Radien:

$$\bar{R}_1 = \sum_{q=1}^{n-1} \frac{|e_q^T S A_3|}{K^{n-q}} \Rightarrow \bar{R}_1 = \mathcal{O}\left(\frac{1}{K}\right) \quad (3.18)$$

$$\bar{R}_q = |A_2(TS)^{-1}e_{q-1}^T| + \frac{\delta_{q-2}}{K}, \quad q = 2, 3, \dots, n \quad (3.19)$$

Wegen $T^{-1} = \text{diag}(K^{n-1}, K^{n-2}, \dots, K)$ wird der Zeilenvektor A_2S^{-1} mit K in den Potenzen von 1 bis $n - 1$ multipliziert. Dadurch errechnen sich zwar große Radien um $\lambda_i, i = 1, 2, \dots, n - 1$, der Radius \bar{R}_1 dagegen verringert sich mit wachsendem K . Unter Ausnutzung obiger Argumentation kann durch hinreichend große k der Mittelpunkt $A_1 - kcb$ der Geršgorin-Scheibe mit dem Radius $\bar{R}_1 \leq \varepsilon$ genügend weit entfernt positioniert werden, so dass für hinreichend großes k die Schnittmenge

$$\forall k \geq k^* : G_\varepsilon \cap \mathcal{B}_\varepsilon(A_1 - kcb) = \emptyset \quad (3.20)$$

leer bleibt. Weil $\mathcal{B}_\varepsilon(A_1 - kcb)$ disjunkt ist zu allen restlichen Scheiben enthält dieser ε -Ball genau den einen verbleibenden Eigenwert λ_n . ■

Theorem 3.3

Unter der Voraussetzung der Minimalphasigkeit, die zu $\sigma(A_4) \subset \mathbb{C}_-$ äquivalent ist, zeigt der geschlossene Regelkreis, bestehend aus Strecke (3.4) und Stabilisationsgesetz (3.5) für hinreichend große Verstärkungen exponentiell stabiles Verhalten.

Beweis

Minimalphasigkeit bedingt, dass alle $n - 1$ Eigenwerte $\lambda_i \in \mathbb{C}$, $i = 1, 2, \dots, n - 1$, der Matrix A_4 negativen Realteil aufweisen. Werden ohne Einschränkung der Allgemeinheit die Eigenwerte derart bezeichnet, dass

$$\text{Re}\{\lambda_{n-1}\} \leq \text{Re}\{\lambda_{n-2}\} \leq \dots \leq \text{Re}\{\lambda_1\} < 0 \quad (3.21)$$

gilt, folgt aus der Wahl $0 < \varepsilon < -\text{Re}\{\lambda_1\}$, dass sich G_ε vollständig in der linken offenen Halbebene befindet. Für hinreichend große Werte k liegt $\mathcal{B}_\varepsilon(A_1 - kcb)$ ebenfalls vollständig in der linken offenen Halbebene, d.h.

$$\text{Re}\{s\} < 0 \quad \forall s \in (G_\varepsilon \cup \mathcal{B}_\varepsilon(A_1 - kcb)). \quad (3.22)$$

Theorem 3.2 gewährleistet die Existenz einer Konstanten $k^* > 0$, so dass für alle $k \geq k^*$ die Eigenwerte des geschlossenen Regelkreises in $G_\varepsilon \cup \mathcal{B}_\varepsilon(A_1 - kcb)$ liegen und damit ausschließlich negative Realteile besitzen. Wegen der Linearität des Regelkreises folgt daraus die exponentielle Stabilität der Ruhelage im Ursprung. ■

Abschnitt A.8.2 zeigt, dass die Übertragungsnullstellen einer linearen Regelstrecke mit den Polstellen ihrer Nulldynamik übereinstimmen. Letztere decken sich, unter der Annahme einer minimalen Realisierung, mit den Eigenwerten der Zustandsmatrix A_4 . Unter der Voraussetzung einer Strecke mit Relativgrad $r = 1$ impliziert Theorem 3.2 folglich die Aussage, dass $m = n - 1$ Pole der geregelten Strecke bei monoton anwachsender Verstärkung gegen die m Nullstellen der ungeregelten Strecke konvergieren und deshalb für $k \rightarrow \infty$ eine m -fache Pol-Nullstellen-Kompensation verursachen. Dieses Ergebnis ist in Übereinstimmung mit dem bekannten und in der linearen Regelungstechnik häufig verwendeten Wurzelortskurvenverfahren [23, S. 200ff], [59, S. 415ff], [79, S. 157ff], dessen Herleitung im Frequenzbereich erfolgt.

Als Konsequenz geht aus Theorem 3.2 hervor, dass jede Nullstelle mit positivem Realteil für hinreichend hohe Verstärkungen eine Polstelle in der rechten, instabilen Halbebene erzwingt. Sofern dagegen sämtliche Nullstellen einen negativen Realteil besitzen, werden im Falle $r = 1$ alle Pole bei Überschreiten einer Mindestverstärkung in die linke, stabile Halbebene verschoben. Eine destabilisierende Wirkung durch zu hohe Verstärkungen ist daher ausgeschlossen. Darin liegt die zur Anwendung hochverstärkungsbasierter Regler erforderliche Minimalphasigkeit der Strecke begründet.

Minimalphasige Systeme mit Relativgrad $r = 1$ erfahren im Grenzfall $k \rightarrow \infty$ eine $n - 1$ -fache Pol-Nullstellen-Kompensation in der stabilen Halbebene. Dadurch wird die charakteristische Dynamik der ungeregelten Strecke gekürzt und es verbleibt lediglich die Dynamik der ungekürzten Polstelle bei $A_1 - kcb$ im E/A-Verhalten des Regelkreises sichtbar. Unabhängig von der Streckenordnung lässt sich ein hochverstärkungsbasierter Regelkreis im Grenzfall durch eine Übertragungsfunktion erster Ordnung beschreiben, die eine verschwindend kleine Zeitkonstante aufweist. Daran ist abzulesen, dass sich das Folgeverhalten des Regelkreises verbessert, je höher die Verstärkung eingestellt ist. Fundiert wird diese Aussage in den beiden nachfolgenden Theoremen.

Theorem 3.4

Gegeben seien eine reelle Konstante $g \neq 0$ und zwei Polynome $Z, N \in \mathbb{R}[s]$, wobei $Z(\cdot)$ der Forderung

$$Z(j\omega) \neq 0 \quad \forall \omega \in \mathbb{R} \quad (3.23)$$

genüge. Zu jedem vorgegebenen $\omega_g > 0$ und zu jedem beliebigen $\varepsilon \in]0, 1[$ existiert eine Schranke $k^* > 0$, so dass für alle $\omega \in \Omega := [-\omega_g, \omega_g]$ und für alle $k \geq k^*$ die Ungleichungen

$$1 - \varepsilon \leq \left| \frac{N(j\omega) + kgZ(j\omega)}{kgZ(j\omega)} \right| \leq 1 + \varepsilon \quad (3.24)$$

und

$$-\varepsilon \leq \arg \left(\frac{N(j\omega) + kgZ(j\omega)}{kgZ(j\omega)} \right) \leq \varepsilon \quad (3.25)$$

erfüllt sind.

Beweis

Weil Polynome stetig sind, existiert eine Schranke $M \in \mathbb{R}$, die $M > |N(j\omega)|$ für alle $\omega \in \Omega$ erfüllt. Sei

$$\kappa := \sup_{\omega \in \Omega} \left| \frac{N(j\omega)}{gZ(j\omega)} \right| \in \mathbb{R}, \quad (3.26)$$

so ist für alle $k > \kappa/\varepsilon > 0$ und $\omega \in \Omega$ die Behauptung (3.24) gezeigt. Wird $k > \kappa$ gewählt, folgt $|N(j\omega)|/|kgZ(j\omega)| < 1$ für alle $\omega \in \Omega$. Daher liegt der Phasenwinkel von $1 + N(j\omega)/[kgZ(j\omega)]$ im Intervall $[-\gamma, \gamma]$, mit

$$\gamma = \arctan \left(\sup_{\omega \in \Omega} \left| \frac{N(j\omega)}{kgZ(j\omega)} \right| \right) = \arctan \left(\frac{\kappa}{k} \right). \quad (3.27)$$

Durch $k \geq \kappa/\tan \varepsilon > 0$ wird für alle $\omega \in \Omega$ die Behauptung (3.25) sichergestellt. Weil $\tan \varepsilon \geq \varepsilon$ für alle $\varepsilon \in]0, 1[$, sind beide Aussagen des Theorems simultan erfüllt, wenn $k > k^* := \kappa/\varepsilon > 0$ gesetzt wird. ■

Theorem 3.5

Gegeben sei eine Regelstrecke (3.1) aus der Systemklasse \mathcal{S} mit Anfangswert $x(0) = 0$, deren E/A-Verhalten im Frequenzbereich durch die Übertragungsfunktion

$$F(s) = \frac{y(s)}{u(s)} = g \frac{Z(s)}{N(s)}, \quad g > 0 \quad (3.28)$$

mit den monischen Polynomen $Z, N \in \mathbb{R}[s]$, $N(\cdot) \not\equiv 0$, beschrieben werde. Wird ein geeignetes Sollwertsignal $y_{ref}(\cdot) \stackrel{\mathcal{L}}{\bullet} w(\cdot)$ vorgegeben und das proportionale Regelgesetz

$$u(s) = -k[y(s) - w(s)] \quad \text{mit} \quad k > 0 \quad (3.29)$$

verwendet, entsteht ein geschlossener Regelkreis mit dem Frequenzgang

$$F_{RK}(s) \Big|_{s=j\omega} = \frac{y(s)}{w(s)} \Big|_{s=j\omega} = \frac{kF(j\omega)}{1 + kF(j\omega)} = \frac{kgZ(j\omega)}{N(j\omega) + kgZ(j\omega)}, \quad (3.30)$$

für den gilt: Zu jedem vorgegebenen $\omega_g > 0$ und zu jedem beliebigen $\delta \in]0, 1[$ existiert eine Schranke $k^* > 0$, so dass für alle $\omega \in \Omega := [-\omega_g, \omega_g]$ und für alle $k \geq k^*$ die Ungleichungen

$$1 - \delta \leq \left| \frac{kgZ(j\omega)}{N(j\omega) + kgZ(j\omega)} \right| \leq 1 + \delta \quad (3.31)$$

und

$$-\delta \leq \arg \left(\frac{kgZ(j\omega)}{N(j\omega) + kgZ(j\omega)} \right) \leq \delta \quad (3.32)$$

erfüllt sind. Darüberhinaus lässt sich k^* so festlegen, dass $N(s) + kgZ(s) \neq 0$ für alle $s \in \mathbb{C}_+$ und der Regelkreis daher exponentiell stabil ist.

Beweis

Die Zugehörigkeit zur Systemklasse \mathcal{S} verlangt Minimalphasigkeit, was $Z(j\omega) \neq 0$ für

alle $\omega \in \mathbb{R}$ bedingt. Zusammen mit $g > 0$ sind die Voraussetzungen für die Anwendung von Theorem 3.4 erfüllt, aus dem für $\varepsilon := \delta/(1 + \delta) \in]0, 1/2[$ die Aussage

$$\frac{1}{1 + \delta} = 1 - \varepsilon \leq \left| \frac{N(j\omega) + kgZ(j\omega)}{kgZ(j\omega)} \right| \leq 1 + \varepsilon = \frac{1 + 2\delta}{1 + \delta} \quad \forall \omega \in \Omega, k > k^* \quad (3.33)$$

resultiert. Umstellen ergibt:

$$1 - \delta \leq \frac{1 + \delta}{1 + 2\delta} \leq \left| \frac{kgZ(j\omega)}{N(j\omega) + kgZ(j\omega)} \right| \leq 1 + \delta \quad \forall \omega \in \Omega, k > k^* \quad (3.34)$$

Weil das Invertieren einer komplexen Zahl den Phasenwinkel mit -1 multipliziert und gleichzeitig $\varepsilon := \delta/(1 + \delta) < \delta$ gilt, geht aus (3.25) unmittelbar die Behauptung (3.32) hervor. Die exponentielle Stabilität ist das Ergebnis von Folgesatz 3.3. ■

Ein Proportionalregler wird als inverser Regler bezeichnet, wenn der geschlossene Regelkreis auf einem vorgegebenen Frequenzintervall $[-\omega_g, \omega_g]$ die Eigenschaften (3.31) und (3.32) besitzt. Dadurch ist gewährleistet, dass der Streckenausgang y (Istwert) einem Sollwertsignal $w : t \mapsto A \sin(\omega_w t)$, $A \in \mathbb{R}$, $\omega_w \leq \omega_g$, mit beliebig geringer Abweichung folgt. Für hinreichend hohe Verstärkungen $k > k^*$ verliert die Dynamik der unregelmäßigten Strecke mit der Übertragungsfunktion $F(s)$ ihren Einfluss. Die geregelten Strecken weisen daher umso geringere Unterschiede auf, je höher die Kreisverstärkung eingestellt wird. Dies zeigt, dass proportionale, hochverstärkungsbasierte Regler in den inversen Regler übergehen. Der Grenzfall $\omega_g = \infty$ und $\delta = 0$ entspricht dem Verhalten eines idealen Regelkreises, da frequenzunabhängig Sollwert und Istwert übereinstimmen und kein Regelfehler auftritt. Offensichtlich sind (3.31) und (3.32) für $\omega_g = \infty$ und $\delta = 0$ nicht erfüllbar, da das Nennerpolynom der Strecke nicht für alle Frequenzen identisch verschwindet. Wegen

$$\deg N > \deg Z \Rightarrow \lim_{\omega \rightarrow \infty} \frac{|N(j\omega)|}{|Z(j\omega)|} = \infty \quad (3.35)$$

überwiegt für hohe Frequenzen $\omega \rightarrow \infty$ das Nennerpolynom, so dass aus (3.26) die Aussage $\kappa \rightarrow \infty$ resultiert. Daher lässt sich ein Frequenzgang mit $F_{RK}(j\omega) \approx 1$ (und damit näherungsweise ideales Folgeverhalten) nur auf einem Kompaktum $\Omega = [-\omega_g, \omega_g]$, $\omega_g \in \mathbb{R}$, erreichen.

3.3 Hohe Verstärkung im nichtlinearen Regelkreis

Die Überlegungen des vorangegangenen Abschnitts, die sich auf lineare Regelstrecken beschränken, werden hier auf nichtlineare Regelstrecken erweitert. Dazu muss die bisher verwendete Systembeschreibung mittels Eigenwerten, Übertragungsfunktionen und Wurzelortskurven aufgegeben und stattdessen eine allgemeinere Systemdarstellung gewählt werden. In Anlehnung an die Ausführungen und die Vorgehensweise in [54, S. 659f] wird das dynamische Verhalten eines nichtlinearen Systems als E/A-Abbildung dargestellt. Somit wird der nichtlineare Prozess als kausaler Operator (oder Funktional)

$$\mathbf{P} : \mathcal{C}([0, \infty]; \mathbb{R}) \rightarrow \mathcal{L}_{lok}^\infty([0, \infty]; \mathbb{R}) \quad (3.36)$$

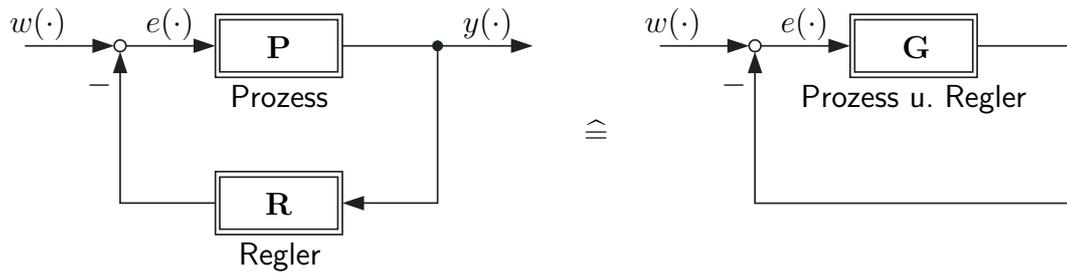


Abbildung 3.1: Betrachteter Regelkreis, bestehend aus Prozess **P** und Regler **R** (links) und äquivalente Darstellung (rechts), bei der die Serienschaltung aus Regler und Prozess zu einem Gesamtsystem **G** zusammengefasst ist.

angesehen, der eine Eingangsfunktion $e(\cdot)$ in eine Ausgangsfunktion $y(\cdot)$ durch die Vorschrift

$$y(\cdot) = (\mathbf{P}e)(\cdot) \quad (3.37)$$

abbildet. In analoger Weise ist der Regler durch einen Operator

$$\mathbf{R} : \mathcal{C}([0, \infty]; \mathbb{R}) \rightarrow \mathcal{L}_{lok}^{\infty}([0, \infty]; \mathbb{R}) \quad (3.38)$$

beschrieben. Der zugrundeliegende Regelkreis wird durch den Prozess **P** und den Regler **R** gebildet und ist in Abbildung 3.1 dargestellt. Die Beschreibung des geschlossenen Regelkreises setzt sich aus den beiden Gleichungen

$$y(\cdot) = (\mathbf{P}e)(\cdot) \quad \text{und} \quad e(\cdot) = w(\cdot) - (\mathbf{R}y)(\cdot) \quad (3.39)$$

zusammen, wenn $w(\cdot) \in \mathcal{L}^{\infty}([0, \infty]; \mathbb{R})$ als Sollwertverlauf vorgegeben wird. Mit der Verknüpfung $\mathbf{G} := \mathbf{R} \circ \mathbf{P}$, die als Abbildung des aufgeschnittenen Kreises angesehen werden kann, ergibt sich

$$e(\cdot) + (\mathbf{G}e)(\cdot) = w(\cdot). \quad (3.40)$$

Es sei angenommen, dass für diese Gleichung eine beschränkte Lösung $e(\cdot) \in \mathcal{L}^{\infty}$ bei beliebigem, aber beschränktem $w(\cdot) \in \mathcal{L}^{\infty}$ existiert. Weiterhin besteht die Annahme, dass eine reelle Konstante $k > 0$ existiert, welche die „Verstärkung“ der Abbildung des aufgeschnittenen Kreises nach unten hin begrenzt:

$$\|(\mathbf{G}e)(\cdot)\|_{\infty} \geq k \cdot \|e(\cdot)\|_{\infty} \quad \text{für alle } e(\cdot) \in \mathcal{L}^{\infty} \quad (3.41)$$

Der Wert für k , für welchen in (3.41) die Gleichheit gilt, wird als die Verstärkung des Operators **G** bezeichnet. Aus den Gleichungen (3.40) und (3.41) folgt:

$$\|w(\cdot) - e(\cdot)\|_{\infty} = \|(\mathbf{G}e)(\cdot)\|_{\infty} \geq k \cdot \|e(\cdot)\|_{\infty} \quad (3.42)$$

Mit Hilfe der Dreiecksungleichung kann die Norm $\|w(\cdot) - e(\cdot)\|_{\infty}$ in die Summe zweier Normen aufgetrennt werden, was die Abschätzung

$$\|w(\cdot)\|_{\infty} + \|e(\cdot)\|_{\infty} \geq \|w(\cdot) - e(\cdot)\|_{\infty} \geq k \cdot \|e(\cdot)\|_{\infty} \quad (3.43)$$

ermöglicht. Daraus resultiert durch Umstellen die Ungleichung

$$\|e(\cdot)\|_{\infty} \leq \frac{1}{k-1} \cdot \|w(\cdot)\|_{\infty}, \quad (3.44)$$

welche eine obere Schranke für die Norm des Regelfehlers angibt. Mittels einer ausreichend hohen Kreisverstärkung k kann die Norm $\|e(\cdot)\|_\infty$ beliebig klein gehalten werden. Weil für hinreichend große Verstärkungen k näherungsweise

$$k \gg 1 \quad \Rightarrow \quad \frac{1}{k-1} \approx \frac{1}{k} \quad (3.45)$$

gilt, lässt sich für große Verstärkungen

$$\|e(\cdot)\|_\infty \leq \frac{1}{k-1} \cdot \|w(\cdot)\|_\infty \approx \frac{1}{k} \cdot \|w(\cdot)\|_\infty \quad (3.46)$$

schreiben. Die Gewichtung von $\|w(\cdot)\|_\infty$ mit einem verschwindend kleinen Faktor $1/k$ führt zusammen mit dem Ungleichheitszeichen in (3.46) auf

$$\|e(\cdot)\|_\infty \ll \|w(\cdot)\|_\infty. \quad (3.47)$$

Damit ist gezeigt, dass eine hohe Verstärkung auch im Falle nichtlinearer Regelstrecken auf einen kleinen Regelfehler führt. Dabei ist „klein“ bezogen auf die Sollwertvorgabe. Je höher die Kreisverstärkung, desto kleiner der Regelfehler („large gain – small error“). Allerdings gelten diese Aussagen nur bei Existenz einer beschränkten Lösung für Gleichung (3.40). Eine solche Lösung für $e(\cdot)$ muss nicht notwendigerweise immer und für alle beschränkten Eingangssignale $w(\cdot)$ existieren. Falls dies jedoch zutrifft, ist der Regelkreis BIBO-stabil mit beliebig keinem Regelfehler.

Um den Schritt zum inversen Regler zu vollziehen, muss

$$e(\cdot) \approx 0 \quad \Leftrightarrow \quad w(\cdot) \approx (\mathbf{R}y)(\cdot) \quad \Leftrightarrow \quad y(\cdot) \approx (\mathbf{R}^{-1}w)(\cdot) \quad (3.48)$$

gezeigt werden. Aus Einsetzen von Gleichung (3.39) in Gleichung (3.47) folgt die Ungleichung:

$$\|w(\cdot)\|_\infty \gg \|e(\cdot)\|_\infty = \|w(\cdot) - (\mathbf{R}y)(\cdot)\|_\infty \quad (3.49)$$

Verglichen mit $\|w(\cdot)\|_\infty$ kann die Norm von $\|w(\cdot) - (\mathbf{R}y)(\cdot)\|_\infty$ wegen der starken Ungleichheit vernachlässigt werden, d.h die Näherung

$$\|w(\cdot) - (\mathbf{R}y)(\cdot)\|_\infty \approx 0 \quad (3.50)$$

ist zulässig. Es ist sofort ersichtlich, dass $w(\cdot) \approx (\mathbf{R}y)(\cdot)$ gilt, die Annahme (3.48) also zutreffend ist. Nachdem der Regler frei gewählt werden kann, darf für den Operator \mathbf{R} die Einheitsrückführung verwendet werden. Folglich ist die Inverse \mathbf{R}^{-1} die Einheitsabbildung, Gleichung (3.48) zeigt damit die Gleichheit zwischen Sollwert und Istwert, was die charakteristische Eigenschaft des inversen Reglers ist. Daraus folgt, dass die E/A-Abbildung eines hochverstärkungsbasierten Regelkreises mit Einheitsrückführung näherungsweise die Einheitsabbildung selbst ist.

3.4 Adaptive Stabilisierung

In den vorangegangenen Abschnitten wird die Auswirkung konstanter, hoher Verstärkungen auf den geregelten Prozess diskutiert. Praktische Anwendbarkeit ist in diesen

Fällen jedoch nur dann gegeben, wenn die Regelstrecke bekannt ist, oder wenn zumindest Abschätzungen über die Systemparameter vorliegen, um eine benötigte Mindestverstärkung zu ermitteln. Falls jedoch die Parameter der Regelstrecke als gänzlich unbekannt angenommen werden müssen, ist eine Adaption der Reglerverstärkung ein möglicher Ausweg. Aufgrund der vorausgesetzten Eigenschaften der Regelstrecke ist der Einsatz einer monoton zunehmenden Kreisverstärkung denkbar. Dieses Konzept soll im Folgenden zusammen mit dem zugehörigen Stabilitätsbeweis vorgestellt werden, die Regelstrecke ist dabei zunächst auf die erste Ordnung beschränkt.

3.4.1 Ordnung $n = 1$

Die grundlegende Idee adaptiver, hochverstärkungsbasierter Regler ist an einem Beispielsystem erster Ordnung am einfachsten zu verdeutlichen. Deshalb wird in einem ersten Schritt zunächst nur eine Regelstrecke erster Ordnung

$$\begin{aligned} \dot{x}(t) &= ax(t) + bu(t), & x(0) &= x_o \in \mathbb{R} \\ y(t) &= cx(t), & a, b, c &\in \mathbb{R}, \quad cb > 0 \end{aligned} \quad (3.51)$$

betrachtet, die den unbekanntem Eigenwert a besitzt, der sowohl positive, als auch negative Werte annehmen kann. Damit sind stabile und instabile Prozesse zugelassen, auch ein grenzstabiles System mit dem Eigenwert $a = 0$ ist eingeschlossen. Ziel ist der Entwurf eines stabilisierenden Regelgesetzes. Zu diesem Zweck kann, wie in 3.2 ausführlich dargestellt, der lineare, zeitinvariante Stabilisator (3.5) eingesetzt werden. Dieser ergibt zusammen mit der Regelstrecke (3.51) ein Gesamtsystem, das beschrieben wird durch die homogene Differentialgleichung

$$\dot{x}(t) = (a - kcb)x(t), \quad x(0) = x_o \in \mathbb{R} \quad (3.52)$$

und exponentiell stabil ist, falls $a - kcb < 0$ zutrifft. Aufgrund unbekannter Systemparameter a , b und c kann diese Bedingung weder überprüft noch sichergestellt werden, so dass eine Berechnung der nötigen Verstärkung $k > a/(cb)$ nicht möglich ist. Da $cb > 0$ wird durch Vergrößerung der Rückführverstärkung k eine Verschiebung des ursprünglichen Eigenwertes a entlang der negativen reellen Achse vorgenommen. Es ist offensichtlich, dass eine anwachsende Verstärkung k den Regelkreis immer stärker stabilisiert. Genau darin besteht die Grundidee adaptiver hochverstärkungsbasierter Regler. Weil die benötigte Verstärkung nicht berechnet werden kann, wird als Ausweg demzufolge ein zeitvarianter Stabilisator

$$u(t) = -k(t)y(t) \quad (3.53)$$

eingesetzt, dessen Verstärkung mit dem Adaptionsgesetz

$$\dot{k}(t) = |y(t)|^p, \quad k(0) = k_o \in \mathbb{R}, \quad p \geq 1 \quad (3.54)$$

eingestellt wird. Durch die Betragsbildung resultiert ein monotones (nicht notwendig streng monoton) Wachstum der Funktion $k(\cdot)$. Aus der Tatsache, dass die Verstärkung nicht abnehmen kann, sondern gegen „große“ Werte strebt, leitet sich die Bezeichnung „hochverstärkungsbasiert“ ab.

Theorem 3.6

Die erste Komponente der Menge der Gleichgewichtslagen $\{0\} \times \mathbb{R}$ des zeitvarianten bzw. nichtlinearen Gesamtsystems mit $p \geq 1$

$$\begin{aligned} \dot{x}(t) &= (a - k(t)cb) \cdot x(t), & x(0) &= x_o \in \mathbb{R} \\ \dot{k}(t) &= |c \cdot x(t)|^p, & k(0) &= k_o \in \mathbb{R} \end{aligned} \quad (3.55)$$

bestehend aus Strecke (3.51), Stabilisator (3.53) und Adaptionsgesetz (3.54) ist attraktiv, d.h.

$$\lim_{t \rightarrow \infty} x(t) = 0. \quad (3.56)$$

Für jeden Anfangswert $(x_o, k_o) \in \mathbb{R} \times \mathbb{R}$ existiert ein beschränkter Grenzwert

$$\lim_{t \rightarrow \infty} k(t) = k_\infty < \infty \quad (3.57)$$

für die Verstärkung. Falls $x_o \neq 0$, ist die erste Komponente der Gleichgewichtslagen $\{0\} \times \mathbb{R}$ asymptotisch stabil.

Beweis

Es sei $(x_o, k_o) \in \mathbb{R} \times \mathbb{R}$. Dann existiert für das nichtlineare Anfangswertproblem (3.55) eine eindeutige Lösung (x, k) auf einem maximalen Intervall $[0, T[$ mit $T \in]0, \infty[$.

Schritt 1:

Aus der Annahme

$$k(\cdot) \notin \mathcal{L}^\infty([0, T[; \mathbb{R}) \quad (3.58)$$

folgt wegen $\dot{k}(t) \geq 0$, dass für $t \rightarrow T$ die Funktion $k(\cdot)$ monoton gegen unendlich strebt. Damit existiert ein Zeitpunkt $0 < t^* < T$, so dass die Ungleichung

$$a - k(t)cb < -1 \quad \text{für alle } t \in [t^*, T[\quad (3.59)$$

formuliert werden kann. Zusätzlich führt die Monotonie wegen $t^* < T$ auf

$$\sup_{t \in [0, t^*]} k(t) = k(t^*) < \infty \quad \text{und} \quad k(\cdot) \in \mathcal{L}^\infty([0, t^*]; \mathbb{R}). \quad (3.60)$$

Wird $k(\cdot)$ nicht als Systemzustand, sondern als exogene Funktion interpretiert, reduziert sich die Ordnung des Gesamtsystems (3.55) von 2 auf 1. Die niedrige Ordnung der Regelstrecke gestattet – trotz zeitvariantem Regler – die Lösung

$$x(t) = x_o \cdot e^{\int_0^{t^*} (a - k(\tau)cb) d\tau} \cdot e^{\int_{t^*}^t (a - k(\tau)cb) d\tau} \quad \forall t \in [0, T[\quad (3.61)$$

der zeitvarianten Differentialgleichung (3.55) explizit anzugeben. Wegen (3.60) und $t^* < \infty$ ist das Integral

$$\int_0^t (a - k(\tau)cb) d\tau < \infty \quad \forall t \in [0, t^*] \quad (3.62)$$

im Intervall $[0, t^*]$ beschränkt, weshalb eine Konstante $M \in \mathbb{R}$ existiert, mit der die Abschätzung

$$e^{\int_0^t (a - k(\tau)cb) d\tau} \leq M < \infty \quad \forall t \in [0, t^*] \quad (3.63)$$

möglich wird. Desweiteren lässt sich mit (3.59)

$$e^{\int_{t^*}^t (a-k(\tau)cb)d\tau} < e^{-\int_{t^*}^t d\tau} = e^{-(t-t^*)} \leq 1 \quad \forall t \in [t^*, T[\quad (3.64)$$

angeben. Folglich ist im gesamten Intervall $[0, T[$ der Betrag der Lösung

$$|x(t)| \leq \begin{cases} |x_o| \cdot M & \forall t \in [0, t^*] \\ |x_o| \cdot M \cdot e^{\int_{t^*}^t (a-k(\tau)cb)d\tau} & \forall t \in]t^*, T[\end{cases} \quad (3.65)$$

beschränkt und es gilt $x(\cdot) \in \mathcal{L}^\infty([0, T[; \mathbb{R})$. Um den gewünschten Widerspruch zu erhalten, wird $k(t)$ aus der Integration des Adaptionsgesetzes (3.54) berechnet. Zum Zeitpunkt t^* errechnet sich:

$$\begin{aligned} k(t^*) &= k_o + \int_0^{t^*} \dot{k}(\tau)d\tau = k_o + |c|^p \int_0^{t^*} |x(\tau)|^p d\tau \leq \\ &\leq k_o + |c|^p \int_0^{t^*} |x_o|^p M^p d\tau = k_o + |c|^p |x_o|^p M^p \cdot t^* =: k^* \in \mathbb{R} \end{aligned} \quad (3.66)$$

Aus (3.59) folgt

$$e^{p \int_{t^*}^\tau (a-k(s)cb)ds} < e^{p \int_{t^*}^\tau -1 ds} = e^{-p(\tau-t^*)} \quad \forall \tau \in]t^*, T[, \quad (3.67)$$

so dass im Intervall $]t^*, T[$ gilt:

$$\begin{aligned} k(t) &= k_o + |c|^p \int_0^{t^*} |x(\tau)|^p d\tau + |c|^p \int_{t^*}^t |x(\tau)|^p d\tau \leq k^* + |c|^p \int_{t^*}^t |x(\tau)|^p d\tau \leq \\ &\leq k^* + |c|^p |x_o|^p M^p \int_{t^*}^t e^{p \int_{t^*}^\tau (a-k(s)cb)ds} d\tau < \\ &< k^* + |c|^p |x_o|^p M^p \int_{t^*}^t e^{-p(\tau-t^*)} d\tau = k^* + |c|^p |x_o|^p M^p \frac{1}{p} (1 - e^{-p(t-t^*)}) \leq \\ &\leq N \in \mathbb{R} \quad \Rightarrow \quad k(\cdot) \in \mathcal{L}^\infty([0, T[; \mathbb{R}) \end{aligned} \quad (3.68)$$

Letzteres widerspricht der Annahme, mit der Folge, dass k im Intervall $[0, T[$ nicht unbegrenzt wächst.

Schritt 2:

Solange die Lösung $k(\cdot)$ auf dem Intervall $[0, T[$ beschränkt bleibt, lässt sich eine stetige Fortsetzung finden und T ist daher nicht maximal gewählt. Weil dieses Argument für alle endlichen $T \in \mathbb{R}$ zutrifft, folgt $T = \infty$, woraus sich

$$k(\cdot) \in \mathcal{L}^\infty([0, \infty]; \mathbb{R}) \quad (3.69)$$

ergibt.

Schritt 3:

Aussage (3.69) ist in Kombination mit $\dot{k}(t) = |c \cdot x(t)|^p$ äquivalent zu

$$x(\cdot) \in \mathcal{L}^p([0, \infty]; \mathbb{R}). \quad (3.70)$$

Da in (3.55) die Zustandstrajektorie $x(\cdot)$ mit einer beschränkten Funktion multipliziert wird, folgt auch

$$\dot{x}(\cdot) \in \mathcal{L}^p([0, \infty]; \mathbb{R}). \quad (3.71)$$

Mit Hilfssatz A.41 ergibt sich daraus $\lim_{t \rightarrow \infty} x(t) = 0$.

Schritt 4:

Die Monotonie von $k(\cdot)$ führt auf

$$k(s) \leq k_\infty \quad \forall s \geq 0 \quad \Rightarrow \quad \int_0^\tau (a - k(s)cb) ds \geq \int_0^\tau (a - k_\infty cb) ds \quad (3.72)$$

und ermöglicht dadurch mit (3.61) für alle $\tau \in [0, \infty]$ die Abschätzung

$$|x(\tau)|^p = |x_o|^p \cdot e^{p \int_0^\tau (a - k(s)cb) ds} \geq |x_o|^p \cdot e^{p \int_0^\tau (a - k_\infty cb) ds} = |x_o|^p \cdot e^{p(a - k_\infty cb)\tau}. \quad (3.73)$$

Daraus lässt sich eine untere Schranke für $k(\cdot)$ gewinnen:

$$\begin{aligned} k(t) &= k_o + |c|^p \int_0^t |x(\tau)|^p d\tau \geq k_o + |c|^p \int_0^t |x_o|^p \cdot e^{p(a - k_\infty cb)\tau} d\tau = \\ &= k_o + |c|^p |x_o|^p \frac{1}{p(a - k_\infty cb)} \left[e^{p(a - k_\infty cb)\tau} \right]_0^t = \\ &= k_o + |c|^p |x_o|^p \frac{1}{p(a - k_\infty cb)} (e^{p(a - k_\infty cb)t} - 1) =: \tilde{k}(t) \quad \forall t \in [0, \infty] \end{aligned} \quad (3.74)$$

Aufgrund (3.69) und $k(\cdot) \geq \tilde{k}(\cdot)$ muss notwendigerweise $\tilde{k}(\cdot) \in \mathcal{L}^\infty([0, \infty]; \mathbb{R})$ gelten, was $a - k_\infty cb > 0$ unter der Voraussetzung $x_o \neq 0$ ausschließt. Das lineare, zeitinvariante Limit-System

$$\dot{x}(t) = (a - k_\infty cb) \cdot x(t), \quad x(0) = x_o \in \mathbb{R} \quad (3.75)$$

kann daher keinen positiven Eigenwert aufweisen.

Um auch die Konstellation $a - k_\infty cb = 0$ auszuschließen, wird eine Betrachtung des Phasenportraits von (3.55) herangezogen. Es zeigt sich, dass ausschließlich die triviale Trajektorie $(x(\cdot) = 0, k(\cdot) = a/(cb))$ in der Gleichgewichtslage $G = (0, a/(cb))$ „endet“. Sämtliche Trajektorien mit der Eigenschaft $x_o \neq 0$ können nicht in diese Gleichgewichtslage münden. Um dies zu zeigen, ist eine Aufteilung der Phasenebene in 5 disjunkte Teilbereiche dienlich:

1. $x = 0, k \in \mathbb{R}$:

Nachdem durch $x = 0$ die Ruhelagen von (3.55) gekennzeichnet sind, findet entlang der k -Achse keine Bewegung statt, es existiert daher keine Trajektorie, die mit der Nebenbedingung $x(\cdot) = \dot{x}(\cdot) = 0$ den Punkt G erreicht.

2. $x \in \mathbb{R} \setminus \{0\}, k = a/(cb)$:

Gleichung (3.55) ergibt $\dot{x} = 0$, so dass der Zustand x unverändert bleibt und nicht gegen 0 strebt. Von diesem Gebiet aus kann G nicht erreicht werden.

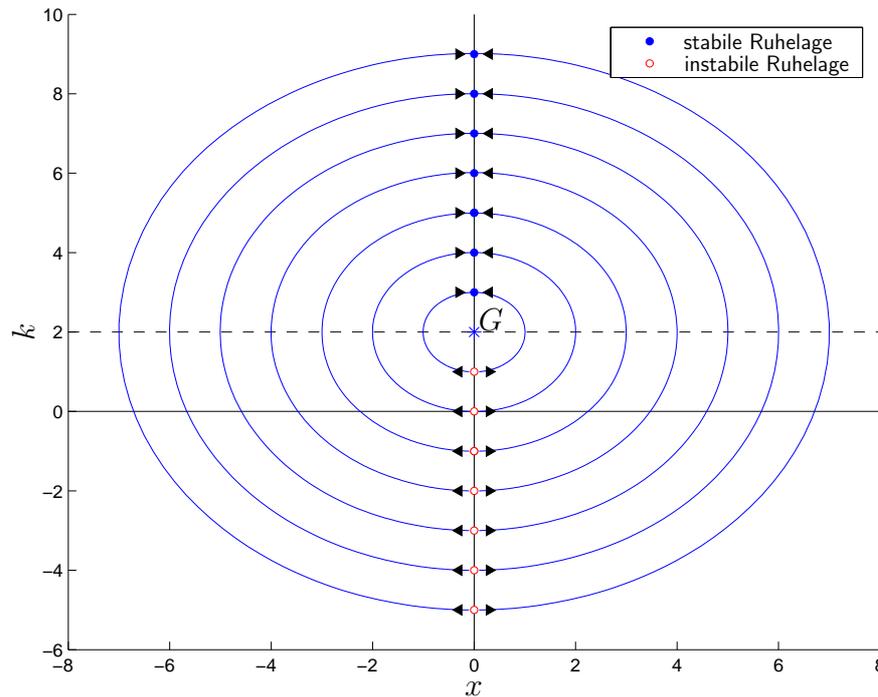


Abbildung 3.2: Phasenportrait für den Regelkreis (3.76)

3. $x > 0, k < a/(cb)$:

Gleichung (3.55) ergibt $\dot{x} > 0$, so dass der Zustand x nicht gegen 0 strebt. Von diesem Gebiet aus kann G nicht erreicht werden.

4. $x < 0, k < a/(cb)$:

Gleichung (3.55) ergibt $\dot{x} < 0$, so dass der Zustand x nicht gegen 0 strebt. Von diesem Gebiet aus kann G nicht erreicht werden.

5. $x \neq 0, k > a/(cb)$:

Gleichung (3.55) ergibt $\dot{k} > 0$, so dass der Zustand k zunimmt und nicht gegen $a/(cb)$ strebt. Von diesem Gebiet aus kann G nicht erreicht werden.

Dadurch verbleibt als einzige Möglichkeit $a - k_\infty cb < 0$, weswegen das Limit-System (3.75) asymptotisch stabil ist. ■

Anmerkung

Ein instabiles Limit-System ergibt sich jedoch, wenn als Anfangswert eine instabile Ruhelage gewählt wird. Die triviale Lösung $x(\cdot) = 0$ und $\dot{k}(\cdot) = 0$ verhindert eine Adaption der Verstärkung, so dass $k(t) = k_o$ für alle $t \geq 0$ gilt. Für jeden Anfangswert $(0, k_o)$, der die Bedingung $a - k_o cb > 0$ erfüllt, befindet sich die geregelte Strecke im instabilen Gleichgewicht und behält wegen $k(t) = k_o$ für alle $t \geq 0$ diesen Zustand bei. Deshalb darf nicht allein aus (3.56) auf die Stabilität der geregelten Strecke geschlossen werden. Beispielhaft wird dieser Sachverhalt durch das Phasenportrait für den Regelkreis

$$\begin{aligned} \dot{x}(t) &= (2 - k(t))x(t), & x(0) &= x_o \in \mathbb{R} \\ \dot{k}(t) &= x(t)^2, & k(0) &= k_o \in \mathbb{R} \end{aligned} \quad (3.76)$$

in Abbildung 3.2 verdeutlicht. Dabei sind die Trajektorien für diejenigen Anfangswerte eingezeichnet, die durch die Kombinationen von $x_o = \pm 0.1$ und $k_o \in \{-5, -4, \dots, 1\}$ festgelegt sind. Die Ruhelagen des geregelten Systems sind durch die notwendige und hinreichende Bedingung $x = 0$ gegeben. Die Menge

$$\{(x, k) \mid x = 0, k < 2\} \quad (3.77)$$

beinhaltet alle instabilen Ruhelagen. Im Gegensatz hierzu bilden die Zustände

$$\{(x, k) \mid x = 0, k > 2\} \quad (3.78)$$

die stabilen Ruhelagen. Am Phasenportrait ist ersichtlich, dass stets ein stabiles Limit-System entsteht, wenn eine Bewegung in den Systemzuständen erfolgt. \square

Die in Theorem 3.6 getroffene Aussage $\lim_{t \rightarrow \infty} k(t) < \infty$ lässt sich für $p = 2$ tiefergehend formulieren, weil der Grenzwert k_∞ explizit berechnet werden kann [34]. Wird $p = 2$ und $y(t) = cx(t)$ gesetzt, lautet das Adaptionsgesetz:

$$\dot{k}(t) = |y(t)|^p = y(t)^2, \quad k(0) = k_o \in \mathbb{R} \quad (3.79)$$

Ausgehend von einer positiv definiten Funktion

$$t \mapsto V(y(t)) = \frac{1}{2}y(t)^2 \geq 0 \quad (3.80)$$

wird deren zeitliche Änderung entlang einer Trajektorie $y(\cdot)$ bestimmt:

$$\begin{aligned} \dot{V}(t) &= \frac{dV}{dy} \frac{dy}{dt} = y(t)\dot{y}(t) = c^2x(t)\dot{x}(t) = c^2x^2(t)(a - k(t)cb) = \\ &= (a - k(t)cb)y^2(t) = (a - k(t)cb)\dot{k}(t) \end{aligned} \quad (3.81)$$

Durch eine Integration auf beiden Seiten

$$\int_0^t \dot{V}(\tau) d\tau = \int_0^t (a - k(\tau)cb)\dot{k}(\tau) d\tau \quad (3.82)$$

wird wieder die ursprüngliche Funktion $V(\cdot)$ gewonnen, die nun jedoch das Argument k anstelle von y aufweist:

$$V(t) - V(0) = a(k(t) - k_o) - \frac{cb}{2}(k(t)^2 - k_o^2) \quad (3.83)$$

Wegen (3.56) strebt $y(t)$ asymptotisch gegen 0, so dass aus (3.80) $\lim_{t \rightarrow \infty} V(y(t)) = 0$ folgt. Damit ergibt ein Grenzübergang $t \rightarrow \infty$ für Gleichung (3.83) den quadratischen Zusammenhang

$$-\frac{1}{2}y_o^2 = ak_\infty - ak_o - \frac{cb}{2}(k_\infty^2 - k_o^2) \quad \text{für } t \rightarrow \infty, \quad (3.84)$$

dessen zwei Lösungen sich zu

$$k_\infty = \frac{a}{cb} \pm \sqrt{\left(\frac{a}{cb} - k_o\right)^2 + \frac{y_o^2}{cb}} \quad (3.85)$$

bestimmen. Für die Anfangsbedingung $k_o = 0$ und $y_o = cx_o \neq 0$ entsteht eine negative Lösung, wenn das Minuszeichen vor der Wurzel verwendet wird. Da jedoch die Funktion $k(\cdot)$ wegen der Wahl des Adaptionsgesetzes monoton wächst und damit $k_\infty \geq k_o$ gelten muss, ist diese Lösung ausgeschlossen. Deshalb errechnet sich als Endwert

$$k_\infty = \frac{a}{cb} + \sqrt{\left(\frac{a}{cb} - k_o\right)^2 + \frac{y_o^2}{cb}} < \infty, \quad (3.86)$$

gegen den die Reglerverstärkung konvergiert.

Der hier beschriebene adaptive Stabilisator hat allerdings lediglich einen untergeordneten Nutzwert in der Praxis, da das notwendige Mindestmaß an Robustheit gänzlich fehlt. Wird am Streckeneingang eine additive Störgröße $d \in \mathbb{R} \setminus \{0\}$ berücksichtigt, ergibt sich das Gesamtsystem

$$\begin{aligned} \dot{x}(t) &= (a - k(t)cb) \cdot x(t) + d, & x(0) &= x_o \in \mathbb{R} \\ \dot{k}(t) &= |c \cdot x(t)|^p, & k(0) &= k_o \in \mathbb{R} \end{aligned} \quad (3.87)$$

mit $p \geq 1$, für das die Zustände $(x = 0, k \in \mathbb{R})$ offensichtlich keine Ruhelage mehr sind. Um $\dot{x}(t) = 0$ und $\dot{k}(t) = 0$ zu erfüllen, müssen die beiden Bedingungen

$$x(t) = -\frac{d}{a - k(t)cb} \quad \text{und} \quad x = 0 \quad (3.88)$$

simultan gelten. Diese schließen sich jedoch wechselseitig aus, sofern für $k(t)$ endliche Werte angesetzt werden. Folglich gehen durch die Störung sämtliche Gleichgewichtslagen mit der Eigenschaft $k < \infty$ verloren, einzig die Konstellation $(x = 0, k = \infty)$ führt zum Verschwinden der beiden Ableitungen in Gleichung (3.87).

3.4.2 Ordnung $n > 1$

Die Ergebnisse des vorangegangenen Abschnittes werden hier verallgemeinert, indem Strecken der Ordnung $n > 1$ zugelassen werden. Die Zugehörigkeit der Strecke zur Systemklasse \mathcal{S} sei jedoch weiterhin beibehalten.

Das Problem erschwert sich für $n > 1$ erheblich, da sich die Lösung der zeitvarianten Differentialgleichung nicht mehr explizit angeben lässt. In [35, S. 59f] wird auf ein Theorem zurückgegriffen, das in der Literatur unter dem Namen „High-gain Lemma“ [35, S. 23f] bekannt ist, um den entsprechenden Satz für beliebige Ordnung $n \in \mathbb{N}$ zu beweisen.

Theorem 3.7

Die in Gleichung (3.4) angegebene Regelstrecke mit $(A, b, c) \in \mathcal{S}$, $n \in \mathbb{N}$ und

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}, \quad A_1 \in \mathbb{R}, \quad A_2 \in \mathbb{R}^{1 \times (n-1)}, \quad A_3 \in \mathbb{R}^{(n-1) \times 1}, \quad A_4 \in \mathbb{R}^{(n-1) \times (n-1)} \quad (3.89)$$

bildet zusammen mit dem Stabilisator $u(t) = -k(t)y(t)$ und dem Adaptionsgesetz (3.54) ein Gesamtsystem, das durch

$$\begin{pmatrix} \dot{y}(t) \\ \dot{z}(t) \end{pmatrix} = \begin{bmatrix} A_1 - k(t)cb & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{pmatrix} y(t) \\ z(t) \end{pmatrix}, \quad \begin{pmatrix} y(0) \\ z(0) \end{pmatrix} = \begin{pmatrix} y_o \\ z_o \end{pmatrix} \in \mathbb{R}^n \quad (3.90)$$

$$\dot{k}(t) = |y(t)|^p, \quad p \geq 1, \quad k(0) = k_o \in \mathbb{R}$$

beschrieben wird. Das zeitvariante bzw. nichtlineare Anfangswertproblem (3.90) besitzt eine eindeutige maximale Lösung $(y(\cdot), z(\cdot), k(\cdot)) : [0, \infty[\rightarrow \mathbb{R}^{n+1}$ mit den Eigenschaften:

- (i) $\lim_{t \rightarrow \infty} y(t) = 0$ und $\lim_{t \rightarrow \infty} z(t) = 0$
- (ii) $(y(\cdot), z(\cdot))^T =: x(\cdot)^T \in \mathcal{L}^p \cap \mathcal{L}^\infty$
- (iii) $\lim_{t \rightarrow \infty} k(t) = k_\infty < \infty$

Aus Abschnitt 3.2 geht hervor, dass es zwei Konstante $\varepsilon > 0$ und $k^* \in \mathbb{R}$ gibt, so dass für alle $k(t) \geq k^*$ die Ungleichung

$$\operatorname{Re} \left\{ \sigma \left(\begin{bmatrix} A_1 - k(t)cb & A_2 \\ A_3 & A_4 \end{bmatrix} \right) \right\} < -\varepsilon < 0 \quad (3.91)$$

formuliert werden kann. Im Falle einer konstanten Verstärkung $k \geq k^*$ lässt sich daraus asymptotische Stabilität des linearen Regelkreises folgern. Durch eine zeitvariable Verstärkung $\dot{k} \neq 0$ geht das lineare Verhalten und damit auch die Aussagekraft von (3.91) bezüglich der Stabilität verloren.

Beweis

Es sei $(x_o, k_o) \in \mathbb{R}^n \times \mathbb{R}$. Das nichtlineare Anfangswertproblem (3.90) weist eine in x und k stetige rechte Seite auf und besitzt daher eine eindeutige maximale Lösung $(x, k) : [0, T[\rightarrow \mathbb{R}^n \times \mathbb{R}$ mit $T \in]0, \infty[$. Mit einer Lyapunov-Funktion, deren zeitliche Ableitung unter der Annahme $k(\cdot) \notin \mathcal{L}^\infty([0, T[; \mathbb{R})$ negativ definit wird, lässt sich ein Widerspruch konstruieren und somit $k(\cdot) \in \mathcal{L}^\infty([0, T[; \mathbb{R})$ folgern. Anschließend erfolgt die Ausweitung des Intervalles auf $T = \infty$ und der Beweis von Aussage (i).

Aus der Zugehörigkeit zur Systemklasse \mathcal{S} folgt $cb > 0$ und $\sigma(A_4) \subset \mathbb{C}_-$, weshalb die Gleichung $A_4^T P + P A_4 = -2I_{n-1}$, $P \in \mathbb{R}^{(n-1) \times (n-1)}$, eine symmetrische und positiv definite Lösung $P = P^T > 0$ besitzt. Mit dieser Matrix wird eine positiv definite Lyapunov-Funktion

$$V(y(t), z(t)) = \frac{1}{2} \frac{1}{cb} y(t)^2 + \frac{1}{2} z(t)^T P z(t) \geq 0 \quad \forall t \in [0, T[\quad (3.92)$$

angesetzt. Deren zeitliche Ableitung ist durch

$$\dot{V}(y(t), z(t)) = \frac{1}{cb} y(t) \dot{y}(t) + \frac{1}{2} [\dot{z}(t)^T P z(t) + z(t)^T P \dot{z}(t)] \quad (3.93)$$

gegeben und lautet nach Einsetzen von (3.90):

$$\dot{V}(y(t), z(t)) = \left(\frac{A_1}{cb} - k(t) \right) y(t)^2 - \|z(t)\|_2^2 + \left(\frac{1}{cb} A_2 + A_3^T P \right) y(t) z(t) \quad (3.94)$$

Mit $M := |A_1|/(cb) + \|A_2\|/(cb) + \|PA_3\|$ resultiert:

$$\begin{aligned} \dot{V}(y(t), z(t)) &\leq \left(\frac{|A_1|}{cb} - k(t) + \frac{\|A_2\|}{cb} + \|PA_3\| \right) y(t)^2 - \|z(t)\|_2^2 + \\ &\quad + \left(\frac{\|A_2\|}{cb} + \|PA_3\| + \frac{|A_1|}{cb} \right) \|y(t)z(t)\|_2 \leq \\ &\leq (M - k(t))y(t)^2 - \|z(t)\|_2^2 + \sqrt{2}M|y(t)| \cdot \frac{1}{\sqrt{2}}\|z(t)\|_2 \leq \\ &\leq (M - k(t))y(t)^2 - \|z(t)\|_2^2 + 2M^2y(t)^2 + \frac{1}{2}\|z(t)\|_2^2 = \\ &= (M + 2M^2 - k(t))y(t)^2 - \frac{1}{2}\|z(t)\|_2^2 \end{aligned} \quad (3.95)$$

Aus der Annahme $k(\cdot) \notin \mathcal{L}^\infty([0, T[; \mathbb{R})$ folgt wegen $\dot{k}(t) \geq 0$, dass für $t \rightarrow T$ die Funktion $k(\cdot)$ monoton gegen unendlich strebt. Dies garantiert die Existenz eines Zeitpunktes $0 < t^* < T$, so dass für alle $t \in [t^*, T[$ die Funktion $t \mapsto k(t)$ die Bedingung $k(t) \geq k^* \in \mathbb{R}$ erfüllt. Wird die Konstante k^* mit der Eigenschaft $k^* \geq M + 2M^2 + 1/2$ angesetzt, legt dies eine ausreichend hohe Verstärkung fest, um für alle $k(t) \geq k^*$ eine negativ definite Ableitung der Lyapunov-Funktion zu erreichen. Demzufolge gilt ab einem Zeitpunkt t^* :

$$\dot{V}(y(t), z(t)) \leq -\frac{1}{2}y(t)^2 - \frac{1}{2}\|z(t)\|_2^2 \leq 0 \quad \forall t \in [t^*, T[. \quad (3.96)$$

Mit $\alpha := \min \{cb, 1/\|P\|\} > 0$ ergibt sich

$$\left. \begin{array}{l} \text{entweder } \alpha = cb \quad \Rightarrow \alpha\|P\| \leq 1 \\ \text{oder } \alpha = 1/\|P\| \quad \Rightarrow \alpha/(cb) \leq 1 \end{array} \right\} \Rightarrow \frac{\alpha}{cb} \leq 1, \quad \alpha\|P\| \leq 1. \quad (3.97)$$

Daraus folgt die Abschätzung

$$-\alpha V(t) \geq -\frac{1}{2} \frac{\alpha}{cb} y(t)^2 - \frac{1}{2} \alpha \|P\| \cdot \|z(t)\|_2^2 \geq -\frac{1}{2} y(t)^2 - \frac{1}{2} \|z(t)\|_2^2 \quad \forall t \geq 0, \quad (3.98)$$

mit der sich die Differentialgleichung erster Ordnung

$$\dot{V}(t) \leq -\frac{1}{2} y(t)^2 - \frac{1}{2} \|z(t)\|_2^2 \leq -\alpha V(t) \quad \forall t \in [t^*, T[\quad (3.99)$$

formulieren lässt, deren Lösung nach oben durch eine abklingende Exponentialfunktion

$$V(t) \leq e^{-\alpha(t-t_o)} V(t_o) \quad \forall t \in [t_o, T[, \quad t_o \geq t^* \quad (3.100)$$

beschränkt wird. Dabei ist der Anfangswert $V(t_o)$ durch

$$\begin{aligned} V(t_o) &= V(y_o, z_o) = \frac{1}{2} \frac{1}{cb} y(t_o)^2 + \frac{1}{2} z(t_o)^T P z(t_o) \leq \\ &\leq \frac{1}{2} (1/(cb) + \|P\|) y(t_o)^2 + \frac{1}{2} (1/(cb) + \|P\|) \|z(t_o)\|_2^2 \end{aligned} \quad (3.101)$$

gegeben und kann mit $\gamma_1 := (1/(cb) + \|P\|)/2$ abkürzend durch

$$V(t_o) \leq \gamma_1 (y(t_o)^2 + \|z(t_o)\|_2^2) = \gamma_1 \|x(t_o)\|_2^2 \quad (3.102)$$

dargestellt werden. Eine untere Schranke für $V(t)$ wird wegen

$$\frac{1}{2} z(t)^T P z(t) \geq \frac{1}{2} \lambda_{\min}(P) \|z(t)\|_2^2 \quad (3.103)$$

aus (3.92) zusammen mit der Definition $\gamma_2 := 1/2 \cdot \min \{1/(cb), \lambda_{\min}(P)\}$ durch

$$V(t) \geq \frac{1}{2} \frac{1}{cb} \|y(t)\|_2^2 + \frac{1}{2} \lambda_{\min}(P) \|z(t)\|_2^2 \geq \gamma_2 \|y(t)\|_2^2 + \gamma_2 \|z(t)\|_2^2 = \gamma_2 \|x(t)\|_2^2 \quad (3.104)$$

festgelegt. Daraus folgt

$$\gamma_2 \|x(t)\|_2^2 \leq V(t) \leq e^{-\alpha(t-t_o)} V(t_o) \leq e^{-\alpha(t-t_o)} \gamma_1 \|x(t_o)\|_2^2 \quad \forall t \in [t_o, T[, \quad (3.105)$$

bzw. nach Umstellen

$$\|x(t)\|_2^2 \leq \frac{\gamma_1}{\gamma_2} e^{-\alpha(t-t_o)} \|x(t_o)\|_2^2 \quad \forall t \in [t_o, T[. \quad (3.106)$$

Der Widerspruch zur Annahme ergibt sich, wenn die Verstärkung aus der Integration der Gleichung

$$\dot{k}(t) = |y(t)|^p \leq \|x(t)\|_2^p \quad \forall t \geq 0 \quad (3.107)$$

bestimmt wird. Wegen $t_o < T$ und der Monotonie von $k(\cdot)$ trifft die Aussage $k(t_o) < \infty$ zu, so dass $k(\cdot)$ im Intervall $[0, T[$ beschränkt ist, wenn das Integral

$$k(t) = \int_{t_o}^t \dot{k}(\tau) d\tau + k(t_o) \quad (3.108)$$

für alle $t \in [t_o, T[$ beschränkt bleibt. Letzteres lässt sich abschätzen durch:

$$\begin{aligned} \int_{t_o}^t \dot{k}(\tau) d\tau &\leq \int_{t_o}^t \|x(\tau)\|_2^p d\tau \leq \int_{t_o}^t \left(\frac{\gamma_1}{\gamma_2} \right)^{\frac{p}{2}} e^{-\alpha \frac{p}{2} (\tau-t_o)} \|x(t_o)\|_2^p d\tau = \\ &= \left(\frac{\gamma_1}{\gamma_2} \right)^{\frac{p}{2}} \|x(t_o)\|_2^p \cdot \frac{-2}{\alpha p} \left[e^{-\alpha \frac{p}{2} (\tau-t_o)} \right]_{t_o}^t = \\ &= \left(\frac{\gamma_1}{\gamma_2} \right)^{\frac{p}{2}} \|x(t_o)\|_2^p \cdot \frac{-2}{\alpha p} \left(e^{-\alpha \frac{p}{2} (t-t_o)} - 1 \right) < \infty \quad \forall t \in [t_o, T[. \end{aligned} \quad (3.109)$$

Daran ist $k(t) < \infty$ für alle $t \in [t_o, T[$ abzulesen, was der Annahme widerspricht. Aus diesem Grund resultiert die Folgerung

$$k(\cdot) \in \mathcal{L}^\infty([0, T[; \mathbb{R}). \quad (3.110)$$

Mit dem Argument der stetigen Fortsetzbarkeit der Lösung, das für alle endlichen $T \in \mathbb{R}$ greift, folgt $T = \infty$, woraus sich

$$k(\cdot) \in \mathcal{L}^\infty([0, \infty]; \mathbb{R}) \quad (3.111)$$

ergibt. Diese Aussage ist in Kombination mit $\dot{k}(t) = |y(t)|^p$ äquivalent zu

$$y(\cdot) \in \mathcal{L}^p([0, \infty]; \mathbb{R}). \quad (3.112)$$

Nachdem für A_4 eine Hurwitz-Matrix vorausgesetzt ist, beschreibt die zweite Gleichung in (3.90) ein asymptotisch stabiles lineares Teilsystem, auf das die Eingangsgröße $y(\cdot) \in \mathcal{L}^p([0, \infty]; \mathbb{R})$ einwirkt. Wie in [50, S. 265ff] dargestellt, entsteht damit ein Ausgangssignal $z(\cdot) \in \mathcal{L}^p([0, \infty]; \mathbb{R}^{n-1})$. Weil $k(\cdot) \in \mathcal{L}^\infty([0, \infty]; \mathbb{R})$ eine essentiell beschränkte Funktion ist, lässt sich aus (3.90) weiterhin auch $\dot{y}(\cdot) \in \mathcal{L}^p([0, \infty]; \mathbb{R})$ und $\dot{z}(\cdot) \in \mathcal{L}^p([0, \infty]; \mathbb{R}^{n-1})$ ablesen. Anwenden von Hilfssatz A.41 führt auf $\lim_{t \rightarrow \infty} y(t) = 0$ und $\lim_{t \rightarrow \infty} z(t) = 0$. ■

3.5 Adaptive Regelung

Abschnitt 3.4 benutzt adaptive, hochverstärkungsbasierte Proportionalregler für die Stabilisierung unbekannter Strecken und beschränkt sich damit auf den Sonderfall der Festwertregelung mit Sollwert $w(\cdot) = 0$. Eine Erweiterung auf einen Festwertregler mit beliebigem Sollwert $w(\cdot) = \text{const} \in \mathbb{R}$ erfordert die Abänderung des Reglers (3.53) und des Adaptionsgesetzes (3.54) in

$$e(t) = w(t) - y(t), \quad u(t) = k(t)e(t), \quad \dot{k}(t) = e(t)^2, \quad k(0) = k_o \in \mathbb{R}, \quad (3.113)$$

ist jedoch nicht in allen Fällen ohne Einschränkungen möglich. Die Aussage

$$\lim_{t \rightarrow \infty} k(t) = k_\infty < \infty \quad (3.114)$$

von Theorem 3.6 ist zum Beispiel nicht auf die Festwertregelung aller Strecken aus der Systemklasse \mathcal{S} übertragbar. Um eine geeignete Klassifikation der Regelstrecken vorzunehmen, wird die Strecke in einer minimalen Realisierung

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), & x(0) &= x_o \in \mathbb{R}^n \\ y(t) &= cx(t), & (A, b, c) &\in \mathcal{S} \end{aligned} \quad (3.115)$$

angenommen.

Definition 3.8 (globales Verhalten)

Das minimal realisierte LTI-System (3.115) hat genau dann (global) integrales Verhalten, wenn

$$0 \in \sigma(A) \quad (3.116)$$

gilt. Andernfalls, d.h. wenn $0 \notin \sigma(A)$, liegt (global) proportionales Verhalten vor.

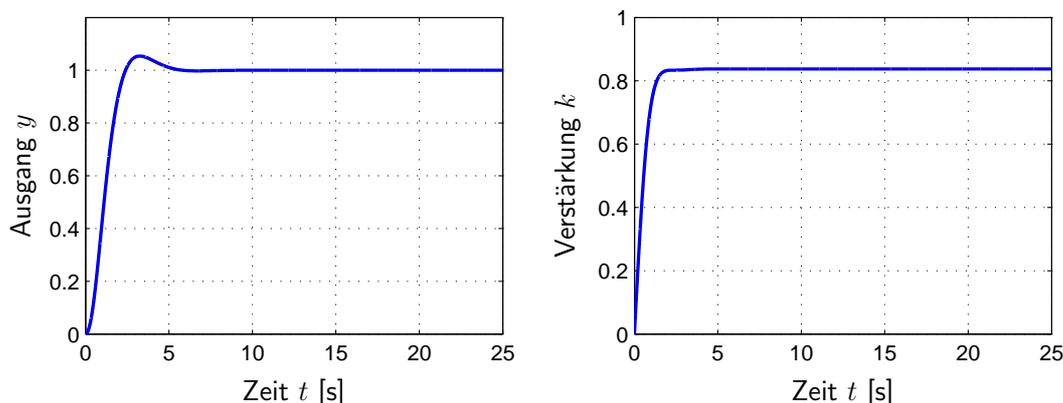


Abbildung 3.3: Hochverstärkungsbasierte Festwertregelung für einen global integralen Prozess.

Neben der Tatsache, dass bei Strecken mit global proportionalem Verhalten eine Festwertregelung mit dem Regler (3.113) und einem Sollwert $w(\cdot) \neq 0$ zu einer Verstärkung $k \notin \mathcal{L}^\infty$ führt, können ebenso Störgrößen ein unbeschränktes Anwachsen verursachen. Ähnliches gilt, wenn Folgeregelungen zu realisieren sind, also wenn die Strecke zeitlich veränderlichen Sollwerten folgen soll. Dieser Abschnitt widmet sich der Betrachtung der genannten Problematik und stellt geeignete Abhilfemaßnahmen dar.

3.5.1 Festwertregelung für Integrale Regelstrecken

Ohne Probleme hingegen lässt sich mit (3.113) eine Festwertregelung realisieren, wenn die Regelstrecke global integrales Verhalten zeigt. Für die Regelstrecke

$$\begin{aligned} \begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} &= \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} \cdot \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot u(t) - \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot d(t) \\ y(t) &= \begin{pmatrix} 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}, \quad x(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{aligned} \quad (3.117)$$

ist die entsprechende Simulation in Abbildung 3.3 für den Fall ohne Störung ($d(\cdot) = 0$) aufgezeichnet, wobei ein konstanter Sollwert $w(\cdot) = 1$ und $k_o = 0$ vorgegeben ist. Durch das integrale Verhalten der Regelstrecke kann der proportionale Regler hier stationäre Genauigkeit erreichen ($e \rightarrow 0$), die Verstärkung konvergiert auf einen endlichen Grenzwert.

3.5.2 Festwertregelung für Proportionale Regelstrecken

Ein gewichtiger Nachteil des Reglers (3.113) besteht darin, dass die Verstärkung divergiert, wenn die Regelstrecke (3.1) global proportionales Verhalten zeigt, siehe [41, Rem. 8]. Bei beschränkter Verstärkung $k < \infty$ kann der Regelfehler $e(t) = w(t) - y(t)$

nicht gegen 0 konvergieren, ohne dass dadurch auch die Stellgröße $u(t) = k(t)e(t)$ verschwindet. Eine proportionale Strecke erfordert jedoch eine konstante Stellgröße $u(\cdot) \neq 0$, damit der Ausgang auf einen beliebigen Sollwert $w(\cdot) \neq 0$ verharren kann. Daraus folgt im Umkehrschluss, dass ein Regelfehler bleiben muss, der mit wachsender Verstärkung allerdings abnimmt. Die monotone Adaption erhöht die Verstärkung, solange ein Fehler auftritt. Weil erst im Grenzfall $k \rightarrow \infty$ der Sollwert exakt erreicht wird, kann, bedingt durch das gewählte Adaptionsgesetz, die Verstärkungsfunktion nicht auf einen endlichen Wert konvergieren.

Der in Abschnitt 3.4 angegebene Stabilitätsbeweis, der eine endliche Verstärkung garantiert, widerspricht dieser Überlegung nicht, da dort lediglich das Problem der Stabilisierung behandelt ist, also eine Festwertregelung auf den Sollwert $w(\cdot) = 0$ erfolgt. Dies stellt insofern einen Sonderfall dar, weil der Arbeitspunkt $y = 0$ eine Gleichgewichtslage bildet, die ohne Stellgröße (also mit $u \equiv 0$) beibehalten werden kann. Der Regelfehler kann in diesem Falle bereits für eine endliche Verstärkung gegen 0 konvergieren. Bedingt durch die Adaption liegt ein zeitvariantes und damit nichtlineares Gesamtsystem vor und es kann nicht, wie bei linearen Systemen üblich, durch eine Verschiebung der Gleichgewichtslage die Aufgabe der Stabilisierung in eine beliebige Festwertregelung überführt werden. Dies gelingt nur, wenn die Regelstrecke global integrierendes Verhalten aufweist und deshalb eine Ruhelage $y = w \neq 0$ auch ohne Stellgröße beibehalten kann. Dadurch wird ermöglicht, dass der Regelfehler bei endlicher Verstärkung verschwindet. In allen anderen Fällen ist das Regelgesetz (3.53) mit der Adaption (3.54) lediglich als Stabilisator zu gebrauchen, solange keine Modifikationen vorgenommen werden.

Die linken beiden Diagramme in Abbildung 3.4 enthalten das Simulationsergebnis für die Stabilisierung ($w(\cdot) = 0$) der instabilen Strecke

$$\dot{y}(t) = 2y(t) + u(t), \quad y(0) = y_o = 1, \quad (3.118)$$

wobei der adaptive Regler (3.113) mit $k_o = 0$ angesetzt ist. Der Systemzustand y startet bei 1 und wächst zunächst an, da der unregelte Prozess instabil ist. Währenddessen wird die Verstärkung $k(t)$ monoton erhöht. Solange $k(t) < 2$ bleibt, überwiegt die instabile Eigendynamik und das Anwachsen von y besteht fort. Sobald die Verstärkung den Wert $k = 2$ erreicht, heben sich Eigendynamik und Stellgröße auf, so dass der Systemausgang y keine zeitliche Änderung erfährt. Zu diesem Zeitpunkt erreicht das Ausgangssignal das Maximum. Erst für Werte $k(t) > 2$ wird eine stabilisierende Wirkung erzielt, das Ausgangssignal strebt asymptotisch gegen 0. Dadurch wird die Adaptionsschwindigkeit kontinuierlich verlangsamt, bis die Verstärkung gegen einen konstanten, beschränkten Wert konvergiert.

Anders verhält sich dagegen der selbe Regelkreis, wenn anstelle von $w(\cdot) = 0$ ein Sollwert $w(\cdot) = 1$ vorgegeben wird, wie in den rechten beiden Diagrammen von Abbildung 3.4 aufgetragen ist. Auch hier wachsen Ausgang und Verstärkung aus den selben Gründen zunächst an. Sobald die kritische Verstärkung $k = 2$ überschritten wird, strebt der Systemausgang gegen den Sollwert. Allerdings kann der Sollwert für eine endliche Verstärkung nicht exakt erreicht werden. Deshalb steigt die Verstärkung unbegrenzt an, um damit den Regelfehler zu reduzieren. Dieser Fall ist in der Praxis unerwünscht,

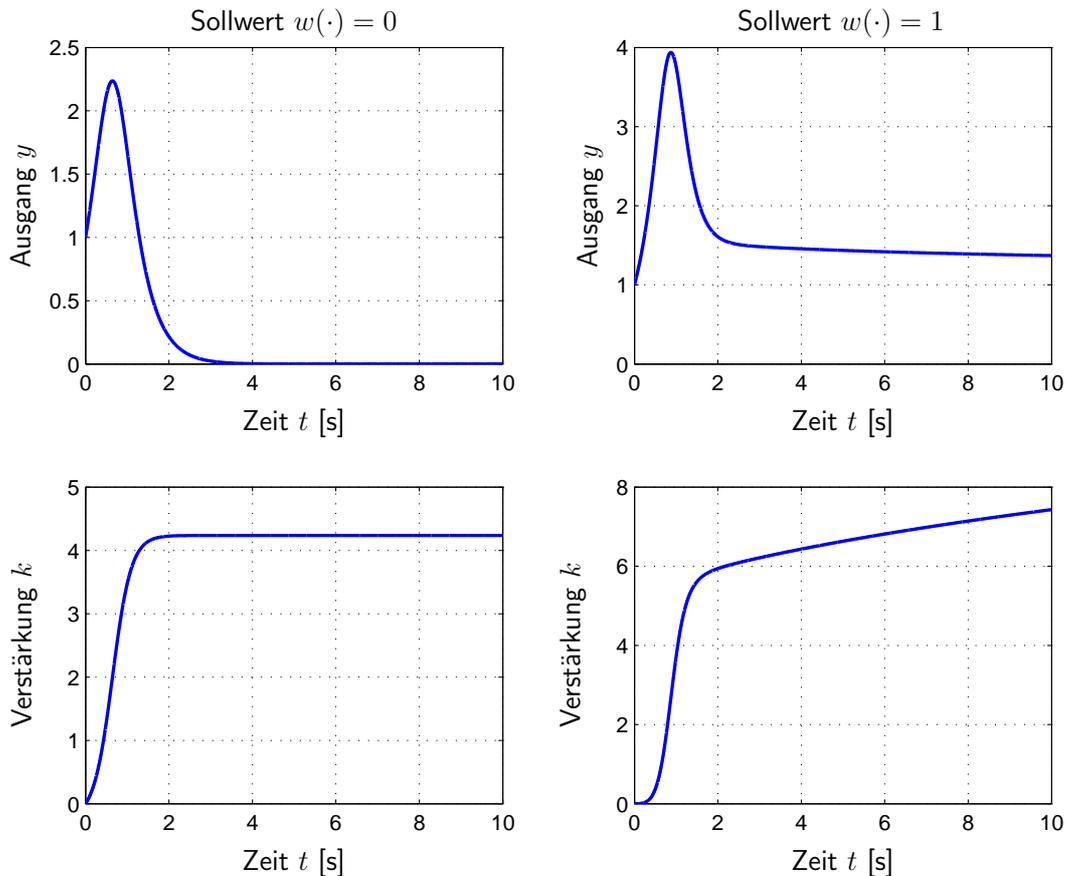


Abbildung 3.4: Hochverstärkungsbasierte Regelung für einen global proportionalen Prozess. Linke Spalte: Sollwert $w(\cdot) = 0$. Rechte Spalte: Sollwert $w(\cdot) = 1$.

weswegen diese Methode nicht für die Regelung, sondern nur zur Stabilisierung global proportionaler Strecken verwendet werden kann.

Die Anwendung proportionalen Regler bei proportionalen Strecken ist allgemein mit einem bleibenden Regelfehler in der Gleichgewichtslage verbunden. Das gewählte Adaptionsgesetz in Gleichung (3.113) generiert daraus eine unbeschränkte Verstärkung. Um die Verstärkung beschränkt zu halten, sind unterschiedliche Ansätze möglich. Zum einen kann die Adaptionsgeschwindigkeit auf 0 herabgesetzt werden, wenn der Betrag des Regelfehlers eine positive Konstante unterschreitet [35, Kap. 5.2], [37], [39]. Zum anderen genügt im speziellen Fall der Festwertregelung ein integraler Anteil im Regler, um dieses Problem zu beseitigen. Verallgemeinert auf beliebige Sollwertverläufe muss ein internes Modell im Regelkreis enthalten sein, dessen Eigendynamik der Sollwerttrajektorie folgen kann. Dadurch kann sich die Regelgröße asymptotisch an den Sollwertverlauf annähern.

Der Einsatz interner Modelle für lineare Strecken unter hochverstärkungsbasierter Regelung ist in [38] dargestellt. Die Linearität von Strecke und Modell gestattet eine Argumentation im Laplace-Bereich bzw. anhand von Übertragungsfunktionen, die auf direktem Wege die notwendigen Anforderungen an das interne Modell offenlegt.

Dagegen erfolgt die Eingliederung des internen Modelles in den Regelkreis hier im Zeitbereich und stützt sich auf Zustandsbeschreibungen.

Hilfssatz 3.9

Wird die Regelstrecke

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), & x(0) &= x_o \in \mathbb{R}^n \\ y(t) &= cx(t), & (A, b, c) &\in \mathcal{S} \end{aligned} \quad (3.119)$$

um einen integralen Anteil

$$\begin{aligned} \dot{x}_I(t) &= v(t), & x_I(0) &\in \mathbb{R} \\ u(t) &= k_I x_I(t) + v(t), & k_I &> 0 \end{aligned} \quad (3.120)$$

erweitert, der als internes Modell für konstante Führungsgrößen betrachtet werden kann, so gehört auch die Serienschaltung

$$\begin{aligned} \begin{pmatrix} \dot{x}(t) \\ \dot{x}_I(t) \end{pmatrix} &= \begin{bmatrix} A & k_I b \\ 0 & 0 \end{bmatrix} \begin{pmatrix} x(t) \\ x_I(t) \end{pmatrix} + \begin{pmatrix} b \\ 1 \end{pmatrix} v(t) \\ y(t) &= \begin{pmatrix} c & 0 \end{pmatrix} \begin{pmatrix} x(t) \\ x_I(t) \end{pmatrix} \end{aligned} \quad (3.121)$$

der Systemklasse \mathcal{S} an.

Beweis

Wegen $(A, b, c) \in \mathcal{S}$ ist

$$\det \begin{bmatrix} sI_n - A & -b \\ c & 0 \end{bmatrix} \neq 0 \quad \forall s \in \mathbb{C} \text{ mit } \operatorname{Re}\{s\} \geq 0 \quad (3.122)$$

und $cb > 0$. Für das Gesamtsystem (3.121) sind die drei Forderungen von Definition 3.1 zu prüfen.

Erste und dritte Forderung:

$$\begin{pmatrix} c & 0 \end{pmatrix} \begin{pmatrix} b \\ 1 \end{pmatrix} = cb > 0 \quad (3.123)$$

Zweite Forderung:

Durch Entwicklung der Determinante gemäß Laplace'schem Entwicklungssatz entlang

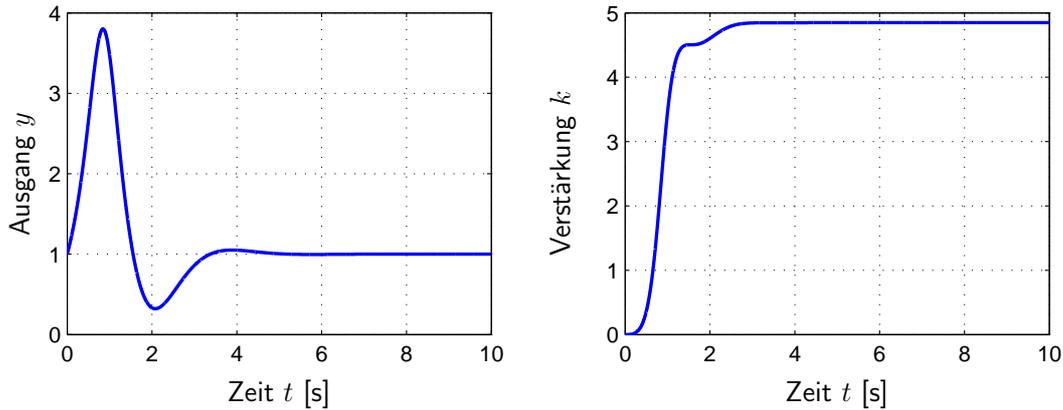


Abbildung 3.5: Hochverstärkungsbasierte Regelung mit integralem Anteil für einen global proportionalen Prozess mit Sollwert $w(\cdot) = 1$.

der vorletzten Zeile ergibt sich:

$$\begin{aligned}
 \det \left[\begin{array}{cc|c} sI_n - A & -k_I b & -b \\ 0 & s & -1 \\ \hline c & 0 & 0 \end{array} \right] &= s \det \left[\begin{array}{cc} sI_n - A & -b \\ c & 0 \end{array} \right] + 1 \det \left[\begin{array}{cc} sI_n - A & -k_I b \\ c & 0 \end{array} \right] = \\
 &= (s + k_I) \underbrace{\det \left[\begin{array}{cc} sI_n - A & -b \\ c & 0 \end{array} \right]}_{\neq 0 \quad \forall s \in \bar{\mathbb{C}}_+} \neq 0 \quad \forall s \in \bar{\mathbb{C}}_+
 \end{aligned} \tag{3.124}$$

Wegen $k_I > 0$ besitzt die Determinante ausschließlich Wurzeln mit negativem Realteil. ■

In Abbildung 3.5 ist im Gegensatz zu Abbildung 3.4 ein zusätzlicher Integralanteil eingefügt worden. Dieser wird beschrieben durch Gleichung (3.120) und ist mit $k_I = 1$ und $x_I(0) = 0$ parametrisiert. Der Integrator ermöglicht eine Stellgröße $u \neq 0$, auch wenn $e \rightarrow 0$. Dadurch kann die Verstärkung trotz stationärer Genauigkeit konvergieren.

3.5.3 Folgeregelung

Im Allgemeinen kann der Regler (3.113) mit dem dort angegebenen Adaptionsgesetz nicht für Folgeregelungen eingesetzt werden. Jede Sollwertänderung bewirkt (unter der wirklichkeitsnahen Annahme eines trägen Prozessverhaltens) einen Regelfehler, der zu einer Erhöhung der Verstärkung führt. Treten dauerhaft Sollwertänderungen auf (z.B. Sinus- oder Rechtecksignal), steigt die Verstärkung unbeschränkt an. Selbiges trifft bei Messrauschen zu. In diesem Zusammenhang ist verständlich, warum ein solches Konzept im industriellen Umfeld keine Verbreitung findet. Allerdings ist für Chargenprozesse ein Einsatz denkbar, wenn nach einem definierten, endlichen Zeitintervall die Regelung abgeschaltet und vor erneuter Inbetriebnahme auf die ursprünglichen

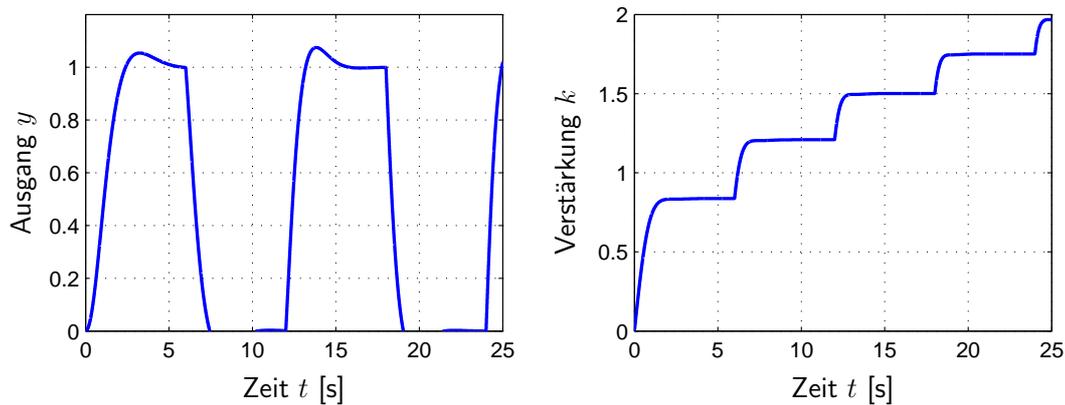


Abbildung 3.6: Hochverstärkungs-basierte Folgeregelung für einen global integralen Prozess (sprungartiger Sollwertverlauf).

Anfangswerte initialisiert werden kann. Bullinger [14] verwendet beispielsweise hochverstärkungs-basierte Regler, um die Dosierung von Narkosemitteln bei Patienten während einer Operation zu regeln. Dabei wird der Patient als „Prozess“ angesehen, der nichtlineares Verhalten aufweist und eine Vielzahl unbekannter Parameter enthält. Dennoch soll eine präzise Dosierung der Narkotika erfolgen. Nach jeder Operation wird die Reglerverstärkung wieder auf einen Ausgangswert zurückgesetzt und für jeden einzelnen Patienten erneut individuell adaptiert. Die bei Folgeregelungen entstehenden Konvergenzprobleme werden hier umgangen, weil ein Chargenprozess vorliegt und folglich die Reglerverstärkung nur über einem beschränkten Zeitintervall beschränkt bleiben muss.

Verdeutlicht wird die Problematik des Driftens durch das Simulationsergebnis, das in Abbildung 3.6 gezeigt ist. Hier ist ein sprunghafter (stückweise konstanter) Sollwert

$$w(t) = \begin{cases} 1 & \forall t \in [0, 6[\cup [12, 18[\cup [24, 30[\dots \\ 0 & \forall t \in [6, 12[\cup [18, 24[\dots \end{cases} \quad (3.125)$$

aufgeschaltet. Jede Sollwertänderung führt zu einem kurzzeitigen Regelfehler, der in der Folge die Verstärkung erhöht. Jene wächst dadurch stufenartig an, es kommt nicht zur Konvergenz der Verstärkung.

3.5.4 Einfluss von Störungen

Das monotone Adaptionsgesetz verursacht eine weitere Schwäche, die sich unter Störgrößeneinfluss zeigt [75]. Selbst im günstigen Falle einer integralen Strecke verschwindet der Regelfehler im gestörten System nicht, weshalb $k(\cdot)$ über alle Grenzen wächst.

Als Voraussetzung, damit die Adaption auf einen festen Wert konvergieren kann, muss der Regelfehler asymptotisch verschwinden. Daraus resultiert bei proportionalem Regelverhalten die Stellgröße $u \rightarrow 0$, so dass für $e \rightarrow 0$ der Prozess allein der Störung unterworfen ist. Diese treibt den Ausgang des unregulierten Prozesses aus dem Arbeitspunkt heraus und verursacht dadurch wegen $y \neq w$ bzw. $e \neq 0$ einen stetigen Anstieg

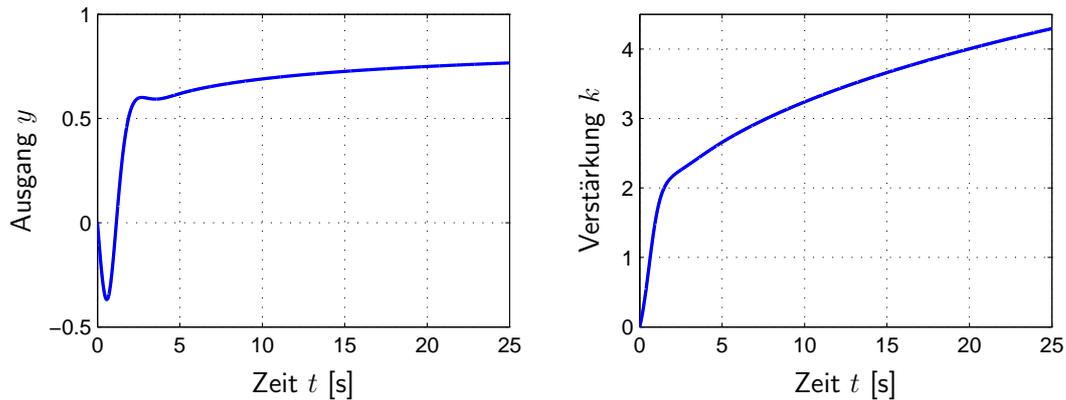


Abbildung 3.7: Hochverstärkungsbasierte Regelung für einen global integralen Prozess unter dem Einfluss einer konstanten Störgröße.

in der Verstärkung. Daher reagiert der betrachtete Regler in hohem Maße empfindlich auf Störeinträge.

Für die integrale Strecke (3.117) ist dieser Fall in Abbildung 3.7 simuliert, wobei eine konstante Störung $d(\cdot) = 1$ am Eingang der Strecke wirkt.

Um Einblick in die Problematik zu erhalten, die durch konstante Störungen verursacht wird, ist eine Umformung der Regelstrecke hilfreich, so dass deren Verhalten zwar nicht verändert wird, die Störung aber formal verschwindet. Die Störgröße wird gewissermaßen als Systemzustand interpretiert und dadurch die Ordnung des Systems erhöht.

Hilfssatz 3.10

Die Wirkung einer konstanten Störung $d \in \mathbb{R}^n$ am Eingang der Regelstrecke $(A, b, c^T) \in \mathbb{R}^{n \times n} \times \mathbb{R}^n \times \mathbb{R}^n$

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t) + d, & x(0) &= x_o \in \mathbb{R}^n \\ y(t) &= cx(t) \end{aligned} \quad (3.126)$$

mit differenzierbarem Stellsignal $u \in \mathcal{C}^1(\mathbb{R}_{\geq 0}; \mathbb{R})$ kann durch Erhöhung der Systemordnung und geeignete Wahl des Anfangszustandes berücksichtigt werden. Es gilt

$$q(t) = y(t), \quad \forall t \geq 0, \quad (3.127)$$

wenn $q(\cdot)$ als Ausgang des ungestörten, erweiterten Systems

$$\begin{aligned} \begin{pmatrix} \dot{x}_d(t) \\ \dot{q}(t) \end{pmatrix} &= \begin{bmatrix} A & 0 \\ c & 0 \end{bmatrix} \begin{pmatrix} x_d(t) \\ q(t) \end{pmatrix} + \begin{pmatrix} b \\ 0 \end{pmatrix} \dot{u}(t) \\ q(t) &= \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} x_d(t) \\ q(t) \end{pmatrix} \end{aligned} \quad (3.128)$$

$$\begin{pmatrix} x_d(0) \\ q(0) \end{pmatrix} = \begin{pmatrix} Ax_o + d + bu(0) \\ cx_o \end{pmatrix} \in \mathbb{R}^{n+1}$$

definiert ist.

Beweis

Für (3.126) wird die Zustandstrajektorie mittels Variation der Konstanten berechnet:

$$x(t) = e^{At}x_o + \int_0^t e^{A(t-\tau)}d\tau + \int_0^t e^{A(t-\tau)}bu(\tau)d\tau \quad \forall t \geq 0 \quad (3.129)$$

Differenzieren mit Hilfe der Leibniz-Formel

$$\frac{d}{dt} \int_0^t F(t, \tau)d\tau = F(t, t) + \int_0^t \frac{\partial}{\partial t} F(t, \tau)d\tau \quad (3.130)$$

und Einsetzen von

$$\begin{aligned} \int_0^t Ae^{-A\tau}d\tau &= \int_0^t \left(A - A^2\tau + \frac{A^3\tau^2}{2!} - \frac{A^4\tau^3}{3!} \pm \dots \right) d\tau = \\ &= I_n - I_n + At - \frac{(At)^2}{2!} + \frac{(At)^3}{3!} - \frac{(At)^4}{4!} \pm \dots = I_n - e^{-At} \end{aligned} \quad (3.131)$$

ergibt:

$$\dot{x}(t) = e^{At}(Ax_o + d) + bu(t) + \int_0^t Ae^{A(t-\tau)}bu(\tau)d\tau \quad (3.132)$$

Für das ungestörte System berechnet sich:

$$x_d(t) = e^{At}(Ax_o + d + bu(0)) + \int_0^t e^{A(t-\tau)}b\dot{u}(\tau)d\tau \quad (3.133)$$

Anwendung der partiellen Integration führt auf

$$\begin{aligned} \int_0^t e^{A(t-\tau)}b \cdot \dot{u}(\tau)d\tau &= \\ &= [e^{A(t-\tau)}b \cdot u(\tau)]_0^t + \int_0^t e^{A(t-\tau)}Ab \cdot u(\tau)d\tau = \\ &= bu(t) - e^{At}bu(0) + \int_0^t e^{A(t-\tau)}Ab \cdot u(\tau)d\tau, \end{aligned} \quad (3.134)$$

so dass

$$x_d(t) = e^{At}(Ax_o + d) + bu(t) + \int_0^t Ae^{A(t-\tau)}bu(\tau)d\tau = \dot{x}(t) \quad (3.135)$$

folgt. Wegen $\dot{q}(t) = cx_d(t) = c\dot{x}(t) = \dot{y}(t)$ und dem gewählten Anfangswert $q(0) = cx_o = y(0)$ folgt durch Integration (3.127). ■

Für weitere Überlegungen sei angenommen, dass die Regelstrecke (3.126) der Systemklasse \mathcal{S} angehöre, wofür

$$\det \begin{bmatrix} sI_n - A & -b \\ c & 0 \end{bmatrix} = b_{n-1}s^{n-1} + \dots + b_1s + b_0 \neq 0 \quad \forall s \in \mathbb{C} \text{ mit } \operatorname{Re}\{s\} \geq 0 \quad (3.136)$$

und

$$b_{n-1} \neq 0, \quad b_j \in \mathbb{R} \quad \text{für } j = 0, 1, \dots, n-1 \quad (3.137)$$

notwendige¹⁾ Voraussetzungen sind. Das charakteristische Polynom der Zustandsmatrix in (3.126) ist durch

$$\det(sI_n - A) = s^n + a_{n-1}s^{n-1} + \dots + a_1s + a_0, \quad a_j \in \mathbb{R} \quad \text{für } j = 0, 1, \dots, n-1 \quad (3.138)$$

gegeben. Daraus folgt, dass für (3.128)

$$\det \left[\begin{array}{cc|c} sI_n - A & 0 & -b \\ -c & s & 0 \\ \hline 0 & 1 & 0 \end{array} \right] = -1 \det \begin{bmatrix} sI_n - A & -b \\ -c & 0 \end{bmatrix} = \det \begin{bmatrix} sI_n - A & -b \\ c & 0 \end{bmatrix} \quad (3.139)$$

und

$$\det \begin{bmatrix} sI_n - A & 0 \\ -c & s \end{bmatrix} = s \det(sI_n - A) \quad (3.140)$$

gilt. Die Minimalphasigkeit bleibt erhalten, das charakteristische Polynom der Zustandsmatrix wird um eine zusätzliche Wurzel bei $s = 0$ erweitert, wodurch sich der Relativgrad von 1 auf 2 erhöht.

Die Zustandsdarstellung (3.128) verlangt als Eingangssignal die zeitliche Ableitung von $u(t)$. Um eine explizite Differentiation zu vermeiden, wird die Differentialgleichung (3.128) zunächst in Beobachtungsnormalform transformiert und anschließend in geeigneter Weise modifiziert.

Hilfssatz 3.11

Ist die Strecke (3.126) beobachtbar, so geht diese Eigenschaft auf die erweiterte Strecke (3.128) über.

Beweis

Sofern Beobachtbarkeit für (3.126) gegeben ist, gilt

$$\det \begin{bmatrix} c \\ cA \\ \vdots \\ cA^{n-1} \end{bmatrix} \neq 0. \quad (3.141)$$

Mit

$$\begin{pmatrix} 0 & 1 \end{pmatrix} \begin{bmatrix} A & 0 \\ c & 0 \end{bmatrix}^j = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{bmatrix} A^j & 0 \\ cA^{j-1} & 0 \end{bmatrix}^j = \begin{pmatrix} cA^{j-1} & 0 \end{pmatrix} \quad (3.142)$$

¹⁾ Für $b_{n-1} \neq 0$ weist die Regelstrecke (3.126) den Relativgrad $r = 1$ auf. Dies erklärt sich bei Betrachtung der zugehörigen Übertragungsfunktion $F(s) = y(s)/u(s) = (b_{n-1}s^{n-1} + \dots + b_1s + b_0)/(s^n + a_{n-1}s^{n-1} + \dots + a_1s + a_0)$, da der Relativgrad in Abschnitt A.6 als Differenz der Ordnungen von Nenner- und Zählerpolynom definiert ist.

errechnet sich die Determinante der Beobachtbarkeitsmatrix für (3.128) durch

$$\det \begin{bmatrix} (0 \ 1) \\ (0 \ 1) \begin{bmatrix} A & 0 \\ c & 0 \end{bmatrix} \\ \vdots \\ (0 \ 1) \begin{bmatrix} A & 0 \\ c & 0 \end{bmatrix}^n \end{bmatrix} = \det \left[\begin{array}{c|c} 0 & 1 \\ \hline c & 0 \\ cA & 0 \\ \vdots & \\ cA^{n-1} & 0 \end{array} \right] = (-1)^n \det \begin{bmatrix} c \\ cA \\ \vdots \\ cA^{n-1} \end{bmatrix} \neq 0. \quad (3.143)$$

Da die Determinanten beider Beobachtbarkeitsmatrizen identische Beträge besitzen, ist die Eigenschaft der Beobachtbarkeit für beide Strecken gleichermaßen gegeben. ■

Aus der Beobachtbarkeit folgt die Existenz [61, S. 594] einer invertierbaren Transformationsmatrix $S \in \mathbf{GI}_{n+1}(\mathbb{R})$, die den Koordinatenwechsel

$$\xi(t) = S \begin{pmatrix} x_d(t) \\ y(t) \end{pmatrix} \quad (3.144)$$

für (3.128) ermöglicht. Dadurch ergeben sich für das transformierte System

$$A_B = S \begin{bmatrix} A & 0 \\ c & 0 \end{bmatrix} S^{-1} = \begin{bmatrix} 0 & & 0 \\ 1 & \ddots & -a_0 \\ & \ddots & 0 \\ & & 1 & -a_{n-1} \end{bmatrix} \in \mathbb{R}^{(n+1) \times (n+1)}, \quad (3.145)$$

$$b_B = S \begin{pmatrix} b \\ 0 \end{pmatrix} = \begin{pmatrix} b_0 \\ \vdots \\ b_{n-1} \\ 0 \end{pmatrix} \in \mathbb{R}^{n+1}, \quad (3.146)$$

$$c_B = (0 \ 1) S^{-1} = (0 \ \dots \ 0 \ 1) \in \mathbb{R}^{1 \times (n+1)}, \quad (3.147)$$

wobei die Einträge a_j in der Systemmatrix A_B aus Gleichung (3.138) zu entnehmen sind und die Elemente des Einkoppelvektors b_B durch Gleichung (3.136) festgelegt werden. Aus

$$(0 \ 1) \begin{pmatrix} b \\ 0 \end{pmatrix} = 0, \quad (0 \ 1) \begin{bmatrix} A & 0 \\ c & 0 \end{bmatrix} \begin{pmatrix} b \\ 0 \end{pmatrix} = cb > 0 \quad (3.148)$$

zeigt sich, dass (3.128) den Relativen Grad 2 besitzt. Dadurch bedingt ist das letzte Element des Einkoppelvektors b_B zwangsweise 0.

Zusammen mit der Struktur der Beobachtungsnormalform ermöglicht der Relative Grad 2 die Differentiation des Eingangssignales in der Zustandsbeschreibung auf einfachem Wege zu berücksichtigen.

Hilfssatz 3.12

Die Wirkung einer expliziten Differentiation des Eingangssignales einer Regelstrecke in BNF mit Relativgrad 2

$$\begin{aligned} \dot{\xi}(t) &= A_B \xi(t) + b_B \dot{u}(t), & \xi(0) = \xi_o = S \begin{pmatrix} x_d(0)^T & y(0) \end{pmatrix}^T \in \mathbb{R}^{(n+1)} \\ y(t) &= c_B \xi(t), & c_B b_B = 0, \quad c_B A_B b_B = cb > 0 \quad \Rightarrow r = 2 \end{aligned} \quad (3.149)$$

kann durch Multiplikation des Einkoppelvektors mit der Zustandsmatrix und geeignete Wahl des Anfangszustandes berücksichtigt werden. Es gilt

$$q(t) = y(t), \quad \forall t \geq 0, \quad (3.150)$$

wenn $q(\cdot)$ als Ausgang des Systems ohne Differenzierer

$$\begin{aligned} \dot{\zeta}(t) &= A_B \zeta(t) + A_B b_B u(t), & \zeta(0) = \zeta_o = \xi_o - b_B u(0) \\ q(t) &= c_B \zeta(t), & c_B (A_B b_B) > 0 \quad \Rightarrow r = 1 \end{aligned} \quad (3.151)$$

definiert ist.

Beweis

Für das Anfangswertproblem (3.149) lautet für alle $t \geq 0$ die Zustandstrajektorie

$$\xi(t) = e^{A_B t} \xi_o + \int_0^t e^{A_B(t-\tau)} b_B \dot{u}(\tau) d\tau. \quad (3.152)$$

Mit Hilfe der partiellen Integration (3.134), wobei sinngemäß A durch A_B und b durch b_B zu ersetzen ist, errechnet sich

$$\xi(t) = e^{A_B t} (\xi_o - b_B u(0)) + b_B u(t) + \int_0^t e^{A_B(t-\tau)} A_B b_B \cdot u(\tau) d\tau \quad (3.153)$$

als Lösung des Systems mit expliziter Differentiation. Nach Einsetzen des Anfangswertes $\zeta_o = \xi_o - b_B u(0)$ folgt daraus für das Ausgangssignal

$$c_B \xi(t) = c_B e^{A_B t} \zeta_o + \underbrace{c_B b_B}_{=0} u(t) + c_B \int_0^t e^{A_B(t-\tau)} A_B b_B \cdot u(\tau) d\tau. \quad (3.154)$$

Für (3.151) errechnet sich für alle $t \geq 0$ das Ausgangssignal

$$c_B \zeta(t) = c_B e^{A_B t} \zeta_o + c_B \int_0^t e^{A_B(t-\tau)} A_B b_B \cdot u(\tau) d\tau, \quad (3.155)$$

so dass für alle $t \geq 0$ die Aussage $y(t) = c_B \xi(t) = c_B \zeta(t) = q(t)$ zutrifft. ■

Um die Auswirkung konstanter Störungen zu betrachten, wird im Folgenden eine Regelstrecke $(A, b, c^T) \in \mathbb{R}^{n \times n} \times \mathbb{R}^n \times \mathbb{R}^n$, $n \in \mathbb{N}$, zugrunde gelegt, die sich durch das Zustandsraummodell (3.126) beschreiben lässt, für $d = 0$ der Systemklasse \mathcal{S} angehört und die Kalman'sche Beobachtbarkeitsbedingung (3.141) erfüllt. Hilfssatz 3.11

überträgt die Beobachtbarkeit auf das erweiterte, ungestörte System (3.128), das mit einer geeigneten Matrix $S \in \mathbf{GI}_{n+1}(\mathbb{R})$ in BNF

$$\begin{aligned}\dot{\xi}(t) &= S \begin{bmatrix} A & 0 \\ c & 0 \end{bmatrix} S^{-1} \xi(t) + S \begin{pmatrix} b \\ 0 \end{pmatrix} \dot{u}(t), & \xi(0) &= S \begin{pmatrix} x_d(0) \\ y(0) \end{pmatrix} \in \mathbb{R}^{(n+1)} \\ y(t) &= \begin{pmatrix} 0 & 1 \end{pmatrix} S^{-1} \xi(t)\end{aligned}\quad (3.156)$$

transformiert wird. Anwenden von Hilfssatz 3.12 beseitigt die Differentiation des Eingangssignales und führt auf:

$$\begin{aligned}\dot{\zeta}(t) &= \underbrace{S \begin{bmatrix} A & 0 \\ c & 0 \end{bmatrix} S^{-1}}_{=A_B} \zeta(t) + S \begin{pmatrix} Ab \\ cb \end{pmatrix} \cdot u(t), & \zeta(0) &= S \begin{pmatrix} x_d(0) - bu(0) \\ y(0) \end{pmatrix} \\ y(t) &= \underbrace{\begin{pmatrix} 0 & 1 \end{pmatrix} S^{-1}}_{=c_B} \zeta(t)\end{aligned}\quad (3.157)$$

Die Voraussetzung $(A, b, c) \in \mathcal{S}$ bedingt $cb > 0$, was die Aussage

$$\begin{pmatrix} 0 & 1 \end{pmatrix} S^{-1} \cdot S \begin{pmatrix} Ab \\ cb \end{pmatrix} = cb > 0 \quad \Rightarrow \quad r = 1 \quad (3.158)$$

ermöglicht. Dies zeigt, dass die Differentialgleichung (3.157) Relativgrad 1 und eine positive instantane Verstärkung besitzt und damit den Bedingungen (i) und (iii) aus Definition 3.1 genügt. Dagegen ist Forderung (ii), die Minimalphasigkeit verlangt, nicht erfüllt. Wegen der Darstellung des Zustandsmodells (3.157) in BNF bestimmt ausschließlich der Einkoppelvektor

$$S \begin{pmatrix} Ab \\ cb \end{pmatrix} = S \begin{bmatrix} A & 0 \\ c & 0 \end{bmatrix} S^{-1} S \begin{pmatrix} b \\ 0 \end{pmatrix} = A_B b_B = \begin{pmatrix} 0 & b_0 & \dots & b_{n-1} \end{pmatrix}^T \quad (3.159)$$

das Zählerpolynom der zugehörigen Übertragungsfunktion. Folglich lässt sich

$$\begin{aligned}\det \begin{bmatrix} sI_{n+1} - A_B & -A_B b_B \\ c_B & 0 \end{bmatrix} &= s(b_{n-1}s^{n-1} + \dots + b_2s^2 + b_1s + b_0) = \\ &= s \cdot \det \begin{bmatrix} sI_n - A & -b \\ c & 0 \end{bmatrix} = 0 \quad \text{für } s = 0 \quad (3.160)\end{aligned}$$

ablesen und daraus $(A_B, A_B b_B, c_B) \notin \mathcal{S}$ entnehmen, weil (3.160) der Gleichung (3.2) in der Definition der Systemklasse \mathcal{S} entgegensteht. Durch eine konstante Störung $d \neq 0$, die sich als zusätzlicher Systemzustand modellieren lässt, verliert die ursprünglich minimalphasige Regelstrecke diese Eigenschaft.

Anmerkung

Da die erste Zeile der Matrix A_B ausschließlich mit den Einträgen 0 versehen ist, enthält

A_B^j für $j = 1, 2, \dots, n$ ebenfalls eine erste Zeile mit den Werten 0. Deshalb weist auch die Steuerbarkeitsmatrix

$$\begin{bmatrix} A_B b_B & A_B^2 b_B & \dots & A_B^n b_B \end{bmatrix} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ * & * & \dots & * \\ \vdots & \vdots & & \vdots \\ * & * & \dots & * \end{bmatrix} \in \mathbb{R}^{(n+1) \times (n+1)} \quad (3.161)$$

von (3.157) eine Nullzeile und damit Rangabfall auf, worin sich der Verlust der Steuerbarkeit zeigt. Jener ist durch die physikalische Tatsache begründet, dass Störgrößenverläufe nicht vom Stellsignal einer Regelstrecke abhängen. In der Beschreibung (3.157) wird die Störung durch einen Systemzustand modelliert, der folglich vom Eingangssignal $u(\cdot)$ aus nicht steuerbar sein kann. Nachdem die Störung d als konstante Größe angesetzt ist, verharrt der entsprechende Zustand auf seinem Anfangswert und konvergiert nicht asymptotisch gegen 0. Aus diesem Grund ist die Strecke (3.157) weder steuerbar noch stabilisierbar.

Matrizen, die ausschließlich mit dem Eintrag 0 in einer Zeile oder Spalte versehen sind, besitzen den Eigenwert $\lambda = 0$. Zusammen mit der Tatsache, dass Ähnlichkeitstransformationen die Eigenwerte einer Matrix beibehalten, folgt

$$0 \in \sigma \left(\begin{bmatrix} A & 0 \\ c & 0 \end{bmatrix} \right) = \sigma \left(S \begin{bmatrix} A & 0 \\ c & 0 \end{bmatrix} S^{-1} \right) = \sigma(A_B) = \{s \in \mathbb{C} \mid \det[sI_{n+1} - A_B] = 0\}. \quad (3.162)$$

Die beiden Gleichungen (3.160) und (3.162) zeigen, dass die Erweiterung der Strecke um eine Zustandsgröße, die der Berücksichtigung der Störung dient, gleichzeitig eine Nullstelle und eine Polstelle bei $s = 0$ hinzufügt. In der Übertragungsfunktion

$$F(s) = \frac{y(s)}{u(s)} = \frac{\det \begin{bmatrix} sI_{n+1} - A_B & -A_B b_B \\ c_B & 0 \end{bmatrix}}{\det[sI_{n+1} - A_B]} \quad (3.163)$$

von (3.157) bleibt dieser Effekt verborgen, weil eine Pol-Nullstellen-Kürzung auftritt. Der Einfluss der Störung $d \neq 0$ äußert sich in einem Anfangswert im nicht stabilisierbaren Subsystem, der von 0 verschieden ist. Die Darstellung eines linearen Systems als Übertragungsfunktion setzt jedoch grundsätzlich den Anfangszustand 0 voraus und erfasst dadurch das Systemverhalten für beliebige Anfangswerte nicht vollständig. Insofern ist das durch die Übertragungsfunktion (3.163) gegebene E/A-Verhalten unzureichend für die Beurteilung der Minimalphasigkeit nach Definition A.27, so dass aus

$$F(s) \neq 0 \quad \forall s \text{ mit } \operatorname{Re}\{s\} \geq 0 \quad (3.164)$$

nicht auf die Zugehörigkeit zur Systemklasse \mathcal{S} geschlossen werden darf. \square

Aus Simulationen gewonnene Beobachtungen legen die Vermutung nahe, dass der Streckenausgang bei Anwendung des Reglers (3.113) auch unter Störgrößeneinfluss beschränkt bleibt. Die Verstärkung dagegen konvergiert nicht, was vor dem Hintergrund

der eingangs dieses Abschnittes geführten Diskussion verständlich ist. Um trotz Störung eine beschränkte Verstärkungsfunktion zu erhalten, ergänzt [87] den Regler um einen integralen Anteil. Dadurch wird die Regelung des Zweimassensystems mit konstanter Laststörung möglich. Allerdings ist in [87] kein Stabilitätsbeweis enthalten, die Untersuchungen basieren ausschließlich auf Simulationen.

3.6 σ -Modifikation

Wie die Beispiele aus den vorigen Abschnitten zeigen, fehlt einem Regler mit monoton ansteigender Verstärkung die notwendige Robustheit gegenüber Störungen, Messfehlern und Rauschen. Um die Problematik des unbeschränkten Abdriftens der Verstärkung auszuschalten, die das Adaptionsgesetz

$$\dot{k}(t) = y(t)^2, \quad k(0) = k_o \in \mathbb{R} \quad (3.165)$$

mit sich bringt, eignet sich die in der Literatur [45], [49] unter dem Namen „ σ -Modifikation“ bekannte Abänderung. Dabei wird der Adaption eine Art Dämpfungsterm hinzugefügt, der im Falle einer hohen Verstärkung und eines kleinen Ausgangs zu einer Reduktion der Verstärkung führt. Mit der Konstanten $\sigma > 0$ lautet das modifizierte Adaptionsgesetz

$$\dot{k}(t) = -\sigma k(t) + y(t)^2, \quad k(0) = k_o \in \mathbb{R} \quad (3.166)$$

und garantiert eine beschränkte Funktion $k \in \mathcal{L}^\infty$, solange der Ausgang $y(\cdot)$ beschränkt bleibt. Ein unerwünschtes Driften der Verstärkung z.B. aufgrund einer Störeinwirkung wird dadurch verhindert. Neben diesem offensichtlichen Vorteil, der als Hauptgrund für die Verwendung der σ -Modifikation angeführt werden kann, bestehen allerdings auch schwerwiegende Nachteile. Fundierte Abhandlungen hierzu finden sich beispielsweise in [63], [74] und [75].

Der zusätzliche Term $-\sigma k(t)$ hindert $k(t)$ nicht nur am Divergieren, sondern ruft durch die Rücknahme der Verstärkung unter Umständen gleichzeitig auch einen destabilisierenden Effekt hervor, der die Ursache für Grenzzyklen bildet und damit die Regelgüte beeinträchtigt. Dieser Mechanismus lässt sich anhand einer instabilen Strecke erster Ordnung intuitiv erklären. Ausgehend von einem Anfangszustand, der hinreichend weit vom Ursprung entfernt liegt, darf zunächst $y(t)^2 > \sigma k(t)$ angenommen werden. Dadurch wächst anfänglich die Verstärkung an und erreicht bzw. überschreitet einen kritischen Wert k^* , der zur Stabilisierung der Strecke notwendig ist, weswegen ab diesem Zeitpunkt der Ausgang $y(t)$ abnimmt. Sobald der Term $-\sigma k(t)$ im Adaptionsgesetz gegenüber $y(t)^2$ überwiegt, gilt $\dot{k}(t) < 0$ und die Verstärkung vermindert sich. Der Regelkreis bewegt sich daher auf die Stabilitätsgrenze zu. Fällt die Verstärkung jedoch zu sehr ab und unterschreitet k^* , zeigt sich wiederum das instabile Eigenverhalten der Strecke und bedingt ein erneutes Anwachsen des Ausgangs, das zu $\dot{k}(t) > 0$ führt. In der Folge wird die kritische Verstärkung k^* wieder überschritten und der Zyklus beginnt erneut. Auf diese Weise entstehen Oszillationen im Ausgang y und in der Kreisverstärkung k .

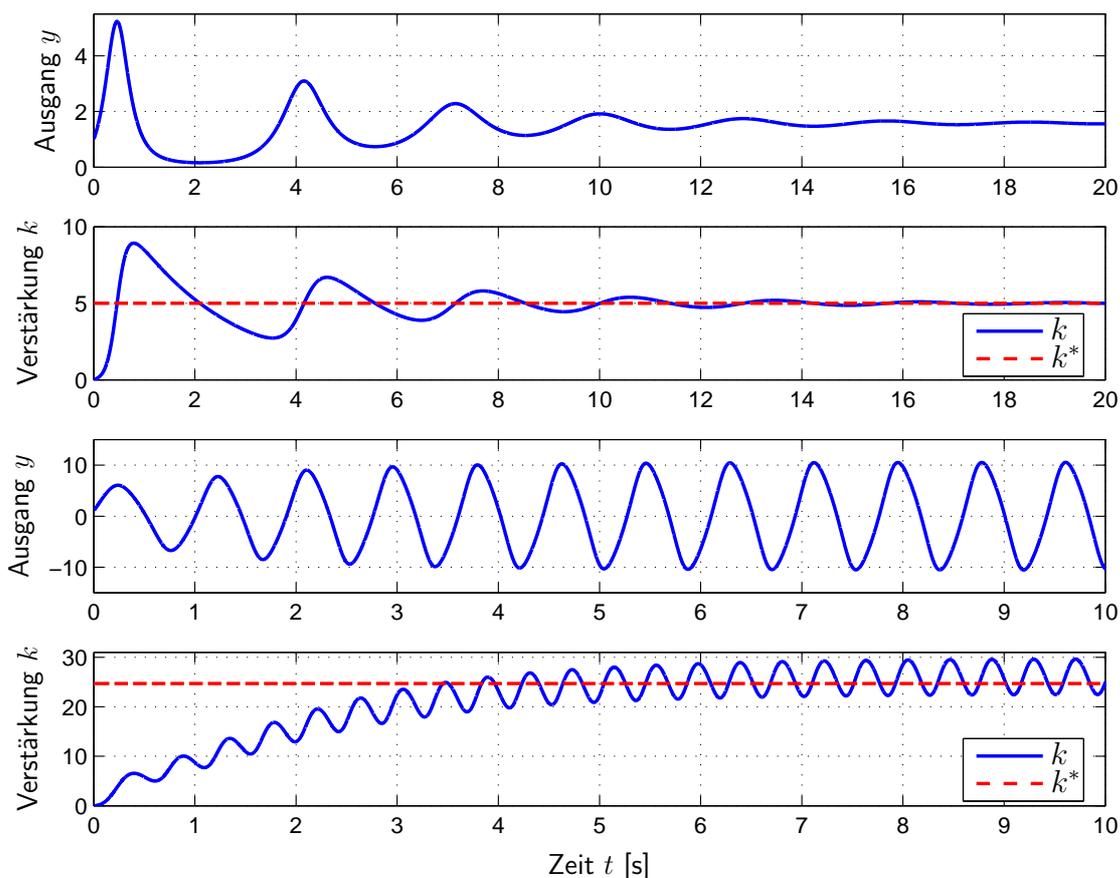


Abbildung 3.8: Diverse Regelstrecken unter hochverstärkungsbasierter Regelung mit σ -Modifikation.

In den ersten beiden Zeilen von Abbildung 3.8 ist das Verhalten des Regelkreises

$$\begin{aligned} \dot{y}(t) &= 5y(t) - k(t)y(t), & y(0) &= 1 \\ \dot{k}(t) &= -0.5k(t) + y(t)^2, & k(0) &= 0 \end{aligned} \quad (3.167)$$

aufgezeichnet, wobei $\sigma = 0.5$ gesetzt ist. Der oben beschriebene Effekt ist hier deutlich zu erkennen, die Verstärkung oszilliert um den kritischen Wert $k^* = 5$, der zur Stabilisierung der Strecke überschritten werden muss. Asymptotisch konvergiert die Verstärkung gegen k^* , so dass für $t \rightarrow \infty$ ein grenzstabiler Regelkreis entsteht. In der Gleichgewichtslage nimmt der Ausgang den Wert

$$\lim_{t \rightarrow \infty} y(t) = \lim_{t \rightarrow \infty} \sqrt{\sigma k(t)} = \sqrt{\frac{k^*}{2}} \approx 1.58 \neq 0 \quad (3.168)$$

an, die angestrebten Ruhelagen ($y = 0, k > k^*$) werden nicht erreicht. Damit ist das Ziel der Stabilisation, das $\lim_{t \rightarrow \infty} y(t) = 0$ verlangt, als verfehlt zu werten.

Die destabilisierende Wirkung, die von der σ -Modifikation ausgeht, tritt deutlicher in

Erscheinung, wenn die Regelstrecke höhere Ordnung aufweist. Für den Regelkreis

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} -4 & -6 & -4 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} x(t) - \begin{pmatrix} k(t) \\ 0 \\ 0 \end{pmatrix} y(t), & x(0) &= \begin{pmatrix} 0 \\ 0.5 \\ 0 \end{pmatrix} \\ y(t) &= (1 \ 2 \ 65) x(t) \\ \dot{k}(t) &= -2k(t) + y(t)^2, & k(0) &= 0, & \sigma &= 2 \end{aligned} \quad (3.169)$$

ist die Simulation in den unteren beiden Zeilen von Abbildung 3.8 dargestellt. Für kleine, konstante Verstärkungen $k \in [0, 0.398]$ verhält sich die Regelstrecke stabil. Durch Verstärkungen aus dem Intervall $k \in]0.398, 24.682[$ wird die Strecke destabilisiert. Alle konstanten Verstärkungen $k \geq k^* = 24.682$ erzeugen wiederum einen stabilen Regelkreis. Auch in diesem Beispiel sind Oszillationen um die kritische Verstärkung k^* zu beobachten. Allerdings konvergiert hier weder die Verstärkung, noch der Ausgang gegen einen Grenzwert, stattdessen stellt sich ein Grenzzyklus ein.

Daran ist ersichtlich, dass die σ -Modifikation (3.166) im Vergleich zum Adaptionsgesetz (3.165) unter Umständen erheblich komplexere Lösungstrajektorien verursacht.

Die Tatsache, dass der Ausgang $y(t)$ nicht asymptotisch gegen 0 konvergiert, lässt sich an folgender Überlegung einsehen. Aus dem Adaptionsgesetz (3.166) folgt, dass $y \rightarrow 0$ gleichzeitig auch $k \rightarrow 0$ bedingt und damit die Rückführung eliminiert. Wenn sich also die Strecke in der Ruhelage $y = 0$ befindet, ist der Regelkreis geöffnet. Liegt eine instabile Strecke vor, ist diese Gleichgewichtslage instabil und somit nicht attraktiv. Dagegen hat eine stabile Strecke naturgemäß eine stabile Ruhelage bei $y = 0$, ist dort allerdings wegen $k = 0$ ungerregelt. Im Simulationsbeispiel für die stabile Strecke konvergiert der Ausgang dennoch nicht gegen 0. Der Grund liegt darin, dass die Verstärkung einen Wert $k > k^*$ erreicht und nicht wieder verringert werden kann, ohne dass der Ausgang dabei anwächst. Für das gewählte $\sigma = 2$ kann die Verstärkung daher das Intervall $k \in]0.398, 24.682[$ nicht durchschreiten, um gegen 0 zu konvergieren, was jedoch Voraussetzung für das Erreichen der Gleichgewichtslage $y = 0$ ist. Diese Ruhelage ist, bedingt durch das instabile Verhalten der Strecke für $k \in]0.398, 24.682[$, lediglich lokal asymptotisch stabil und wird im Beispiel nicht erreicht. Wird dagegen ein höherer Wert für σ verwendet (z.B. $\sigma = 300$), so kann das instabile Intervall durchschritten werden und $y(t)$ konvergiert gegen 0, selbst wenn ein weit entfernter Anfangswert (z.B. $k(0) = 30$) eingestellt wird.

Im Gegensatz zur Adaption (3.166) führt die Verwendung von (3.165) in beiden Simulationsbeispielen zum Verschwinden des Ausgangs $y(t)$. Daran zeigt sich, dass die σ -Modifikation die Stabilitätseigenschaften der monotonen Adaption nicht notwendigerweise beibehält (es entsteht grenzstabiles Verhalten), Grenzzyklen erzeugen kann (komplexere Lösungstrajektorien) und eine Verschiebung der ursprünglichen Gleichgewichtslage bewirkt ($y \neq 0$). Diese Effekte sind verantwortlich dafür, dass bei Verwendung der σ -Modifikation die Aufgabenstellung der Regelung auf den Sollwert 0 nicht in jedem Falle erfüllt wird. Selbst in Abwesenheit einer Störung wird das Regelziel $y \rightarrow 0$ verfehlt, was somit als Preis für eine beschränkte, nicht monotone Verstärkungsfunktion

zu zahlen ist.

Polderman und Mareels lösen zwar die Probleme der σ -Modifikation jedoch auf Kosten einer wiederum monotonen Adaption [74], [75]. Durch die Verwendung von

$$\dot{k}(t) = \max \{ -\sigma k(t) + y(t)^2, 0 \} \quad (3.170)$$

wird die Funktion $k(\cdot)$ zwar beschränkt gehalten. Da jedoch $\dot{k}(t) \geq 0$ für alle $t > 0$, können transiente Vorgänge (z.B. verursacht durch Störungen oder Anfangszustände) eine Verstärkung größer als notwendig oder als erwünscht generieren. Ebenso wie im Falle von (3.165) ist eine Rücknahme der Verstärkung dabei nicht möglich. Aus diesem Grund soll dieser Ansatz hier nicht weiter verfolgt werden.

Nachdem die Nachteile einer monoton zunehmenden Verstärkungsfunktion als zu gewichtig erscheinen, um diese Art der adaptiven hochverstärkungsbasierten Regler erfolgreich in der Praxis anwenden zu können, soll im folgenden Abschnitt ein hochverstärkungsbasiertes Regelungsverfahren vorgestellt werden, das eine Rücknahme der Reglerverstärkung erlaubt.

Kapitel 4

Funnel-Control

Kapitel 3 widmet sich in vertiefter Form hochverstärkungsbasierten Reglern und diskutiert dabei ausführlich die stabilisierenden Auswirkungen hoher Verstärkungen im Regelkreis. Dadurch ist bereits ein Teil der theoretischen Grundlage für Funnel-Control geschaffen, weil auch hier eine hochverstärkungsbasierte Regelstrategie eingesetzt wird.

Nachdem bei Betrachtung eines mathematischen Modelles vor einem rein theoretischen Hintergrund hohe Verstärkungen die Eigenschaften des Regelkreises positiv beeinflussen, indem die Stabilität verbessert und die Regelung beschleunigt wird, braucht eine Reduktion der Verstärkung nicht in Erwägung gezogen zu werden, solange ausschließlich das Verhalten der beschreibenden Differentialgleichung untersucht wird.

In der praktischen Anwendung dagegen treten zusätzlich zum modellierten Idealverhalten weitere unmodellerte Effekte auf, welche für die erfolgreiche Anwendung der Regelung eine Rücknahme der Kreisverstärkung erzwingen. So wird beispielsweise vorhandenes Messrauschen hochfrequente Anteile in die Stellgröße einbringen, deren Amplituden wegen der großen Verstärkungen nicht zu vernachlässigen sind. Dadurch wird der Prozess in einem unerwünschten Frequenzband angeregt, sofern die Aktorik in der Lage ist, eine solche Stellgröße zu erzeugen. Hierin begründet sich der zweite Aspekt, weswegen sehr hohe Verstärkungen vermieden werden müssen: jedes physikalische Stellglied ist Beschränkungen unterworfen und kann daher nicht beliebig hohe Stellgrößen generieren. Weiterhin existieren Konstellationen (wie eingehend in Kapitel 3 dargestellt), bei welchen eine stationäre Regeldifferenz verbleibt, solange die Reglerverstärkung endliche Werte annimmt. In diesem Fall wird durch das monotone Adaptionsgesetz die Verstärkung unbeschränkt anwachsen. Hier kann zwar durch die Einführung einer Toleranzgrenze (λ -Band) Abhilfe geschaffen werden [35, Kap. 5.2], [39]. Dennoch bleibt die Monotonie in der Adaption erhalten und schließt damit eine Verringerung der Verstärkung aus. Dies verhindert eine situationsgerechte Einstellung der Verstärkung, die idealerweise geringe Werte annimmt und nur kurzzeitig ansteigt, falls für das Erreichen des Regelzieles notwendig.

Die genannten Gründe motivieren daher die Verwendung einer nicht-monotonen Verstärkungsfunktion für praxisrelevante Anwendungen. Deshalb wird hier eine tragfähige Lösung betrachtet, um hohe Verstärkungen ohne Verlust der Stabilität wieder verrin-

gern zu können. Hierfür besteht eine geeignete Möglichkeit darin, die aktuell notwendige Verstärkung direkt aus dem gegenwärtigen Prozessverhalten zu berechnen und dabei das bisherige Adaptionsgesetz zu verwerfen, das durch Integration eines Fehlersignales eine zusätzliche Zustandsgröße erzeugt. Trotz dieser Modifikation beruht ein solcher Regler weiterhin auf den Grundgedanken des hochverstärkungs-basierten Ansatzes und stellt daher ähnliche Anforderungen an die Regelstrecken wie die in Kapitel 3 beschriebenen Konzepte. Im Vergleich zur Systemklasse \mathcal{S} werden hier jedoch zusätzlich nichtlineare Effekte und zeitvariante Parameter erlaubt.

Die mathematische Grundlage wird von Ilchmann et. al. gelegt, indem in [41] zwei essentielle Theoreme zu dieser Regelstrategie aufgestellt und bewiesen werden. Beide Theoreme sollen in diesem Kapitel zusammen mit ihren Beweisen ausführlich dargestellt und eingehend beleuchtet werden.

Ein besonderer Vorteil liegt darin, dass im Gegensatz zu anderen adaptiven Verfahren bei Funnel-Control nicht nur für $t \rightarrow \infty$ ein kleiner Regelfehler gewährleistet ist, sondern auch das transiente Verhalten des Regelkreises vorgegeben werden kann. Diese Besonderheit ist von wichtiger Bedeutung bei praktischen Anwendungen. Es ist für die Regelung realer Anlagen wenig relevant, welchen Zustand der Prozess nach *unendlich* langer Zeit (also asymptotisch) einnehmen wird. Vielmehr von Interesse ist das Prozessverhalten während der gesamten Betriebsdauer der Anlage, die sich vom Einschalten der Regelung bis zu einem Zeitpunkt in *endlicher* Zukunft erstreckt. Dabei müssen sowohl während des Einschwingvorganges vorgegebene Beschränkungen eingehalten werden, als auch asymptotisch ein stationärer Zustand oder Zustandsverlauf mit vorgegebener Genauigkeit erreicht werden. Nahezu alle adaptiven Regler garantieren jedoch ausschließlich asymptotische Stabilität, d.h. nach unendlich langer Zeit wird die entsprechende Gleichgewichtslage erreicht, wobei die Systemzustände stets beschränkt bleiben. Damit ist jedoch keinerlei Aussage getroffen über die tatsächliche Größe der Schranke für den maximal auftretenden Betrag des Zustandsvektors. Es besteht durchaus die Gefahr, während des transienten Vorganges kritische Werte zu überschreiten und dadurch Defekte an der Anlage zu verursachen, obwohl asymptotisch stabiles Verhalten vorliegt. Damit sind Stabilitätsbeweise für derartige Regler rein von akademischem Interesse und aus Sicht der praktischen Anwendung nahezu bedeutungslos. Miller und Davison schreiben zu diesem Thema in [66]:

„In most results on MRAC [. . .], little can be said about the system’s closed-loop transient response, even though it is an important property.“

Dem gegenüber steht Funnel-Control, das einen Fehlerverlauf innerhalb eines vorab wählbaren Trichters garantiert. Auf diese Weise können kritische Grenzen für den Prozessausgang berücksichtigt werden, sowie die gegebene Spezifikation, die Einschwingzeiten oder auch maximal zulässiges Überschwingen festlegt. Eine vorab gewählte Grenze, der Trichterrand, wird vom Regelfehler eingehalten, sofern die maximale Stellgröße dafür ausreicht.

Nachdem Funnel-Control in der Lage ist, sowohl nichtlineare als auch zeitvariante Regelstrecken zu beherrschen, ist die Berücksichtigung nichtlinearer Reibkennlinien bei

Antriebssystemen keine ernsthafte Schwierigkeit. Auch die additive Einwirkung einer beschränkten Störung auf die Stellgröße kann den Regelkreis ebenso wenig destabilisieren wie Messrauschen. Dies ist ein klarer Vorteil gegenüber vielen anderen adaptiven Regelungskonzepten, die nur lineare Regelstrecken behandeln können, oder sich durch beständig einwirkende Störungen destabilisieren lassen.

4.1 Regelstrategie

Das kennzeichnende und auffälligste Merkmal dieser Regelungsstrategie besteht in der Besonderheit, dass nicht nur ein gewünschter Sollwert $w(t)$ vorgegeben wird, sondern zusätzlich auch eine trichterförmig zulaufende Funktion, die den Betrag des aktuellen Regelfehlers $|e(t)| = |w(t) - y(t)|$ beschränkt. Durch eine geeignete Anpassung der Reglerverstärkung wird das Einhalten dieser Grenze garantiert. Der Fehler kann den frei wählbaren Trichter zu keinem Zeitpunkt verlassen, wodurch ein beliebig schnelles Einschwingen und ein beliebig kleiner stationärer Fehler vorab festgelegt werden können.

Die Schranke für den maximal zulässigen Regelfehler wird durch eine Funktion beschrieben, die in folgendem Funktionenraum angesiedelt sein muss:

Definition 4.1 (Funktionsraum \mathcal{F})

Der Funktionsraum \mathcal{F} beinhaltet alle Funktionen $x : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ mit den Eigenschaften:

- (i) Die Funktion $x : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ ist auf jedem Kompaktum $I \subset \mathbb{R}_{\geq 0}$ absolut stetig, formal: $x \in \mathcal{AC}_{lok}(\mathbb{R}_{\geq 0}; \mathbb{R}_{\geq 0})$
- (ii) Die Funktion und ihre Ableitung sind essentiell beschränkt. Es gilt: $x, \dot{x} \in \mathcal{L}^\infty$.
- (iii) Die Funktion besitzt auf der positiven reellen Achse einen streng positiven Wertebereich und ist von 0 wegbeschränkt. Dies drückt sich durch $x(t) > 0$ für alle $t > 0$ und $\liminf_{t \rightarrow \infty} x(t) > 0$ aus.

Für die Vorgabe des Trichters wird aus dem Funktionsraum \mathcal{F} mit $\psi \in \mathcal{F}$ eine beschränkte und stetige Funktion $\psi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ mit der Eigenschaft

$$\forall t > 0 : \psi(t) > 0 \quad (4.1)$$

gewählt, deren Kehrwert

$$\partial F_\psi(t) := \frac{1}{\psi(t)} > 0 \quad \forall t > 0 \quad (4.2)$$

den Trichterrand definiert. Die Verwendung des ∂ -Zeichens betont dabei, dass es sich bei den beiden Funktionen $\pm \partial F_\psi(\cdot)$ um den Rand des Trichters handelt. Die Wahl der Trichterrandfunktion $1/\psi(\cdot)$ orientiert sich an den Anforderungen, die an die Dynamik

der geregelten Strecke gestellt werden. Die Vorgabe eines exponentiell abklingenden Trichterrandes ist selbst im Falle nichtlinearer Regelstrecken zulässig [40]. Ein mögliches Beispiel für die Funktion $\psi(\cdot)$ ist daher durch

$$\psi : t \mapsto (a_F \cdot e^{-\lambda_F t^2} + \Delta_F)^{-1}, \quad a_F, \Delta_F > 0, \quad \lambda_F \geq 0 \quad (4.3)$$

gegeben. Der maximal zulässige Fehler für $t \rightarrow \infty$ kann durch eine geeignete Wahl der Konstanten Δ_F festgelegt werden. Die Geschwindigkeit der Abnahme des Trichterrandes hängt von λ_F ab, so dass mit dieser Größe ein Freiheitsgrad vorhanden ist, um die gewünschte Abklingrate für den Fehler vorzugeben und auf diesem Wege das transiente Verhalten des Regelkreises zu beeinflussen. Damit das Fehlersignal für alle Zeiten im Trichter verläuft, muss auch der anfängliche Fehler zum Zeitpunkt $t = 0$ im Trichter enthalten sein. Dieser Forderung kann auf zwei unterschiedlichen Arten Rechnung getragen werden. Einerseits ist eine entsprechende Anpassung der Konstanten a_F möglich, so dass der Trichter den (bekannten) Fehler $e(0) \in \mathbb{R}$ beinhaltet. Die in Abschnitt 4.4 erzielten Resultate treffen in diesem Falle lediglich für $|e(0)| < 1/\psi(0)$ zu und sind daher semi-globaler Natur. Die Alternative besteht darin, eine Funktion $\psi(\cdot)$ mit der Eigenschaft $\psi(0) = 0$ zu wählen. Dadurch beginnt der Trichterrand im Unendlichen, der initiale Fehler ist somit prinzipiell vom Trichter umschlossen – folglich gelten die Aussagen in Abschnitt 4.4 global. Allerdings bleibt anfänglich der Prozess unregelt, weil die Reglerverstärkung bei 0 beginnt.

Sofern das Fehlersignal $e(\cdot)$ im Trichter verläuft und der Regelfehler unterhalb des Trichterrandes bleibt, ist die Ungleichung

$$|e(t)| < \partial F_\psi(t) = 1/\psi(t) \quad \forall t > 0 \quad (4.4)$$

erfüllt, aus der durch Umstellen der Zusammenhang $\psi(t) \cdot |e(t)| < 1$ folgt. Der Trichter selbst wird definiert als die Menge aller Punkte zwischen den beiden Funktionen $\pm \partial F_\psi(\cdot)$, was durch

$$F_\psi : t \mapsto \{e \in \mathbb{R} \mid \psi(t) \cdot |e| < 1\} \quad (4.5)$$

beschrieben wird. In Abbildung 4.1 ist diese Menge (der Trichter) als graue Fläche dargestellt. Die vertikale Distanz

$$d_v(t) = \partial F_\psi(t) - |e(t)| \quad (4.6)$$

zwischen Regelfehler und Trichterrand bildet die Grundlage für die Einstellung der aktuellen Reglerverstärkung $k(t)$, die aus

$$k(t) := \frac{1}{d_v(t)} = \frac{1}{\partial F_\psi(t) - |e(t)|} = \frac{1}{\frac{1}{\psi(t)} - |e(t)|} = \frac{\psi(t)}{1 - \psi(t)|e(t)|} \quad (4.7)$$

berechnet wird. Um einen (nichtlinearen) Prozess mit unbekanntem Parametern zu regeln, wird auf dieser Basis ein zeitvariantes, proportionales Regelgesetz

$$u(t) = -k(t) e(t) \quad (4.8)$$

festgelegt. In [41] ist gezeigt, dass ein solcher Regler alle Prozesse aus der in Definition 4.3 notierten Systemklasse \mathcal{P} stabil regeln kann. Die in [41] vorgestellten Beweise sind in den Abschnitten 4.3 und 4.4 wiedergegeben.

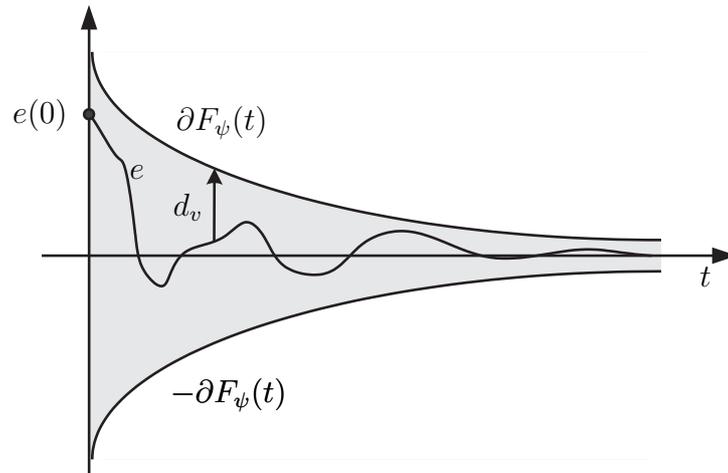


Abbildung 4.1: Trichter und Fehlersignal über der Zeit. Der Trichterrand wird gebildet von den Funktionen $\pm\partial F_\psi(\cdot)$.

Der Hintergrund dieser Festlegung für die Reglerverstärkung ist folgendes Wunschverhalten: nur wenn erforderlich, soll die Verstärkung ansteigen, um den Regelkreis zu stabilisieren oder zu beschleunigen. Bei Erreichen des Regelzieles dagegen ist eine niedrige Verstärkung vorzuziehen, die Verstärkung soll demnach wieder fallen, den Regelkreis beruhigen und damit robuster einstellen gegenüber Störeinflüssen wie z.B. Messrauschen. Aufgrund der Vorgabe der Reglerverstärkung in Gleichung (4.7) wird das Wunschverhalten erreicht: genau dann, wenn sich der Regelfehler dem Trichterrand nähert, erfährt die Reglerverstärkung wegen der reduzierten Distanz $d_v(t)$ eine Erhöhung und vergrößert folglich das Stellsignal. Das, zusammen mit dem in Abschnitt 4.3 dargelegten Charakteristikum der Regelstrecke, hindert das Fehlersignal am Verlassen des Trichters. Bei einem Kontakt des Fehlers mit dem Trichterrand reduziert sich der Abstand auf 0 und generiert eine unendlich große Verstärkung. Diese potentielle Singularität in der rechten Seite der Differentialgleichung des Regelkreises bedarf einer intensiven mathematischen Untersuchung. Es lässt sich jedoch zeigen, dass für die vorgegebene Klasse von Regelstrecken dieser Fall ausgeschlossen ist. Die Reglerverstärkung bleibt stets beschränkt, der Fehlerverlauf ist vom Trichterrand wegbeschränkt und kann daher den Trichter zu keiner Zeit verlassen.

Falls dagegen der Fehler abnimmt (bzw. zutreffender, wenn die vertikale Distanz zunimmt), verringert sich die Verstärkung wunschgemäß. Damit entsteht eine weniger aggressive Regelung, sobald die momentane Situation dies zulässt. Der hauptsächliche Nachteil der in Kapitel 3 diskutierten Regler, nämlich die monotone Zunahme der Verstärkung, ist damit eliminiert.

Wie Definition 4.1 verlangt, muss die Funktion $\psi(\cdot) \in \mathcal{F}$ beschränkt bleiben. Dadurch ist deren Kehrwert strikt positiv, ein Absinken der Trichterränder auf den Wert 0 wird verhindert und es gilt $\partial F_\psi(t) > 0$ für alle $t > 0$. Es ist von besonderer Bedeutung, dass sich der Trichter nicht auf eine verschwindende Breite verschmälert, da ansonsten die vertikale Distanz gegen 0 konvergiert und damit die Verstärkung unendlich stark anwächst. Allerdings darf sich der Trichter theoretisch auf beliebig kleine, aber po-

sitive Werte verengen. Dies führt dazu, dass der stationäre Regelfehler beliebig klein gehalten werden kann. Es kann jedoch nicht erwartet werden, dass der Regelfehler vollständig verschwindet, eine (kleine) Abweichung muss stets zugelassen werden. Dieser verbleibende Regelfehler ist kein spezifisches Merkmal von Funnel-Control, sondern ist vielmehr typisch für alle proportionalen Regler. Diesbezüglich lässt sich bekanntlich eine Verbesserung erzielen, wenn ein integraler Anteil in das Regelgesetz eingebracht wird, ohne dabei die Stabilität des Regelkreises zu gefährden. Diesem Aspekt widmet sich Abschnitt 4.7.

4.2 Regelbare Systemklasse

Analog zur adaptiven hochverstärkungsbasierten Regelung beziehen sich die wesentlichen Eigenschaften der regelbaren Systemklasse auf Relativgrad, instantane Verstärkung und Nulldynamik. Um letztere zu beschreiben, wird zunächst eine Operatorklasse \mathcal{Z} mit allen erforderlichen Attributen definiert. Weil die Ordnung der Nulldynamik in diesem Zusammenhang keine Bedeutung besitzt, erweisen sich E/A-basierte Systembeschreibungen als vorteilhaft. Daher wird zur Spezifikation der benötigten Eigenschaften eine Darstellung als (nicht-)linearer Operator herangezogen.

Definition 4.2 (Operatorklasse \mathcal{Z})

Die Operatorklasse \mathcal{Z} fasst sämtliche Operatoren $\mathbf{Z} : \mathcal{C}([0, \infty]; \mathbb{R}) \rightarrow \mathcal{L}_{lok}^{\infty}([0, \infty]; \mathbb{R})$ zusammen, die durch folgende Eigenschaften charakterisiert werden:

- (i) Zu jeder beliebigen Schranke $\delta > 0$ existiert eine zugehörige Schranke $\Delta > 0$, so dass alle beschränkten Eingangssignale $\zeta(\cdot) \in \mathcal{C}([0, \infty]; \mathbb{R})$ durch \mathbf{Z} gemäß

$$\sup_{t \geq 0} |\zeta(t)| \leq \delta \quad \Rightarrow \quad |(\mathbf{Z}\zeta)(\tau)| \leq \Delta \quad \text{für fast alle } \tau \geq 0 \quad (4.9)$$

in ein beschränktes Ausgangssignal abgebildet werden.

- (ii) Für alle beliebigen Signale $\zeta(\cdot), \xi(\cdot) \in \mathcal{C}([0, \infty]; \mathbb{R})$ und für alle Zeitpunkte $T \geq 0$ gilt:

$$\zeta(t) = \xi(t) \quad \forall t \in [0, T] \quad \Rightarrow \quad (\mathbf{Z}\zeta)(\tau) = (\mathbf{Z}\xi)(\tau) \quad \text{für fast alle } \tau \in [0, T] \quad (4.10)$$

- (iii) Für alle Zeitpunkte $T \geq 0$ und zu jeder stetigen Funktion $\beta(\cdot) \in \mathcal{C}([0, T]; \mathbb{R})$ existieren passende Konstante $\tau > 0$, $\delta > 0$ und $c > 0$, so dass für alle Signale $\zeta(\cdot), \xi(\cdot) \in \mathcal{C}([0, \infty]; \mathbb{R})$ mit den beiden Eigenschaften

$$\zeta(t) = \beta(t) = \xi(t) \quad \forall t \in [0, T]$$

und

$$\zeta(s), \xi(s) \in [\beta(T) - \delta, \beta(T) + \delta] \quad \forall s \in [T, T + \tau]$$

die Ungleichung

$$\operatorname{ess-sup}_{s \in [T, T+\tau]} |(\mathbf{Z}\zeta)(s) - (\mathbf{Z}\xi)(s)| \leq c \cdot \sup_{s \in [T, T+\tau]} |\zeta(s) - \xi(s)| \quad (4.11)$$

erfüllt ist.

Anmerkung

- (i) Der Operator $\mathbf{Z} \in \mathcal{Z}$ bildet eine stetige Funktion in den Raum der (Lebesgue-) messbaren, lokal essentiell beschränkten Funktionen ab.
- (ii) Der Operator $\mathbf{Z} \in \mathcal{Z}$ beschreibt ein E/A-stabiles dynamisches Übertragungssystem, weil ein Eingangssignal mit beschränkter Norm ein Ausgangssignal mit ebenfalls beschränkter Norm generiert.
- (iii) Der Operator $\mathbf{Z} \in \mathcal{Z}$ besitzt kausales Verhalten, da der Ausgang ausschließlich von vergangenen und aktuellen Werten des Eingangs abhängt, nicht jedoch vom zukünftigen Verlauf des Eingangssignales. Sind zwei verschiedene Eingangssignale im Intervall $[0, T]$ identisch, ergeben sie innerhalb dieses Zeitraumes die selben Ausgangssignale, unbeeinflusst von einem späteren, möglicherweise unterschiedlichen Verlauf beider Eingänge. Nachdem kein reales System in die Zukunft blicken kann, ist diese Bedingung keine Einschränkung, sondern wird automatisch von allen praxisrelevanten Regelstrecken eingehalten.
- (iv) Der Operator $\mathbf{Z} \in \mathcal{Z}$ erfüllt eine Art lokale Lipschitz-Bedingung. Die maximale Abweichung zweier Ausgangssignale, die von zwei unterschiedlichen Eingangssignalen erzeugt werden, ist durch die Abweichung zwischen den Eingangssignalen beschränkt. Je ähnlicher also die Eingänge, desto ähnlicher auch die Ausgänge. \square

Mit Hilfe der Operatorklasse \mathcal{Z} lässt sich schließlich die zulässige Systemklasse definieren.

Definition 4.3 (Systemklasse \mathcal{P})

Eine Regelstrecke mit dem Eingangssignal $u(\cdot)$ und dem Ausgangssignal $y(\cdot)$ gehört zur Systemklasse \mathcal{P} , wenn die beschreibende Differentialgleichung

$$\dot{y}(t) = f(p(t), (\mathbf{Z}y)(t), u(t)), \quad y(0) = y_0 \in \mathbb{R} \quad (4.12)$$

folgende Voraussetzungen erfüllt:

- (i) Die (nicht-)lineare Funktion $f : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ist stetig.
- (ii) Der Operator $\mathbf{Z} : C([0, \infty]; \mathbb{R}) \rightarrow \mathcal{L}_{lok}^{\infty}(\mathbb{R}_{\geq 0}; \mathbb{R})$ gehört der Operatorklasse \mathcal{Z} an.
- (iii) Die (externe) Störung $p(\cdot) \in \mathcal{L}^{\infty}(\mathbb{R}; \mathbb{R})$ ist essentiell beschränkt.
- (iv) Für jede nicht-leere kompakte Menge $C \subset \mathbb{R} \times \mathbb{R}$ und für jede Folge $(u_n) \subset \mathbb{R} \setminus \{0\}$, $n \in \mathbb{N}$ muss gelten:

$$|u_n| \rightarrow \infty \text{ wenn } n \rightarrow \infty \Rightarrow \min_{(\rho, \zeta) \in C} (\operatorname{sign}(u_n) \cdot f(\rho, \zeta, u_n)) \rightarrow \infty \text{ wenn } n \rightarrow \infty$$

Bedingung (iv) in Definition 4.3 fordert eine instantane Verstärkung mit konstantem Vorzeichen und impliziert dadurch zusammen mit Gleichung (4.12), dass mit einem entsprechend gewählten Stelleingriff die Ableitung des Ausgangs zu jedem Zeitpunkt frei einstellbar ist. Unter der in Definition 4.3 getroffenen Annahme $|u_n| \rightarrow \infty$ wenn $n \rightarrow \infty$, $(u_n) \subset \mathbb{R} \setminus \{0\}$ und $C \subset \mathbb{R} \times \mathbb{R}$ darf (für hinreichend große $n \in \mathbb{N}$) $\text{sign}(u_n) = 1$ gesetzt werden. Daraus resultiert

$$\min_{(\rho, \zeta) \in C} f(\rho, \zeta, u) = \min_{(\rho, \zeta) \in C} \dot{y}(\rho, \zeta, u) \rightarrow \infty \text{ wenn } u \rightarrow \infty \quad (4.13)$$

und zeigt, dass unbeeinflusst vom Verlauf der Nulldynamik und der Einwirkung der Störung instantan eine beliebig starke Änderung des Ausgangssignales möglich ist. Damit – unabhängig von der Rate, mit der sich der vorgegebene Trichter verjüngt – der Fehler den Trichter nicht verlässt, muss die zeitliche Änderung des Fehlersignales durch die Stellgröße vorgegeben werden können. Weil das Fehlersignal aus der Differenz von Ist- und Sollwert gebildet wird, muss notwendigerweise auch die Ableitung des Ausgangs beliebig große Werte annehmen können, was die Aussage von Gleichung (4.13) erklärt. Durch die instantane Verstärkung, die bei Darstellung in BINF durch den Einkoppelfaktor $\alpha(z(t))$ gegeben ist (siehe Abschnitt A.8), muss daher dauerhaft eine Eingriffsmöglichkeit gewährleistet sein. Insofern ist $|\alpha(z(t))| > 0$ für alle $z(t) \in \mathbb{R}^n$ eine notwendige Voraussetzung zu Bedingung (iv).

Da in Gleichung (4.12) der Eingang zu jedem Zeitpunkt unmittelbar die erste Ableitung des Ausgangs beeinflusst, besitzen sämtliche Prozesse der Systemklasse \mathcal{P} strikt einen Relativgrad von 1. Dadurch wird die Anwendbarkeit von Funnel-Control stark eingeschränkt, besonders im Bereich der Elektrischen Antriebstechnik und der mechatronischen Systeme. Deshalb werden in Kapitel 5 Abhilfemaßnahmen und Lösungsmöglichkeiten für dieses Problem vorgestellt.

Eindeutig lässt sich kein Relativgrad bestimmen, wenn nichtlineare dynamische Systeme nicht für alle beliebigen Werte der Eingangsgröße eine direkte Zugriffsmöglichkeit auf die Ableitung des Ausgangs bieten. Obwohl solche Strecken nicht in die Systemklasse \mathcal{P} fallen, ist dennoch deren Regelung mit Funnel-Control möglich. Die notwendigen Erweiterungen sind in [40] dargestellt.

Der nichtlineare Operator $(\mathbf{Z}y)(\cdot)$ kann als dynamisches Subsystem beliebiger Ordnung betrachtet werden und beschreibt die Nulldynamik. Die erforderliche E/A-Stabilität des Operators ist gleichbedeutend mit der Stabilität der Nulldynamik, die im Spezialfall linearer Systeme mit der Minimalphasigkeit übereinstimmt.

Mit dem Signal $p(\cdot)$ ist in der Systemklasse \mathcal{P} eine beschränkte Störung explizit zugelassen. Daraus resultiert ein wesentlicher Vorteil gegenüber identifizierenden Verfahren, die sich bei Untermodellierung durch solche Störeinträge meist destabilisieren lassen. Rohrs et. al. schreiben in einer renommierten und vielbeachteten Veröffentlichung [80]:

„[...] adaptive algorithms as published in the literature are likely to produce unstable control systems if they are implemented on physical systems directly as they appear in the literature.“

Nachdem Störungen stets präsent sind, müssen identifizierende Regler durch zusätzliche Maßnahmen robustifiziert werden. Eine Möglichkeit hierfür ist der Einsatz von Störmodellen [20], die allerdings in vielen Anwendungen von geringem Nutzen sind, da keine beliebigen Verläufe des Störsignales zugelassen sind. Es werden lediglich Störungen eliminiert, die als Linearkombination von Sinussignalen unterschiedlicher Frequenzen dargestellt werden können, wobei die einzelnen Frequenzen im Vorhinein als bekannt vorausgesetzt sind.

Bei der Überprüfung, ob ein gegebener Prozess der Systemklasse \mathcal{P} angehört, sind vor allem strukturelle und qualitative, nicht jedoch quantitative Eigenschaften von Interesse. Beispielsweise spielt nicht die instantane Verstärkung selbst, sondern nur deren Vorzeichen eine Rolle: relevant ist die Wirkrichtung, also ob bei einer Vergrößerung der Stellgröße das Ausgangssignal sinkt oder steigt. Wie stark diese Veränderung ausfällt, ist dagegen unerheblich. Aus diesem Grund können die maßgeblichen Eigenschaften der Systemklasse \mathcal{P} oftmals ohne Kenntnis der exakten Streckenparameter geprüft werden. Ein grobes „Prozessmodell“ mit unsicheren Parametern, bzw. reine Strukturkenntnis ist dann ausreichend. Der Regler selbst verwendet kein parametrisiertes Modell (nicht-parameterbasiert), so dass keine Identifikation notwendig ist, um das Strukturmodell mit korrekten Werten zu versehen.

Bei der Bildung des Strukturmodells, dessen Analyse über die Zugehörigkeit der Regelstrecke zur Systemklasse \mathcal{P} entscheidet, ist wegen der zulässigen Störung eine gewisse Untermodellierung akzeptabel. Solange die wesentlichen Eigenschaften unverändert bleiben (Relativgrad, Stabilität der Nulldynamik und Vorzeichen der instantanen Verstärkung), sind demzufolge keine Stabilitätsprobleme zu erwarten, wenn parasitäre Effekte bei der Modellierung unberücksichtigt bleiben. So darf die Nulldynamik nach bekannten Regeln im Sinne einer Ordnungsreduktion um stabile Anteile gekürzt werden. Das Störsignal $p(\cdot)$ dient in diesem Falle dazu, die vernachlässigten inneren Effekte zu berücksichtigen. Neben unmodellierter Dynamik können sich im Signal $p(\cdot)$ ebenso auch externe Einflüsse widerspiegeln. Dadurch lassen sich Störungen wie Lastmomente, Reibung, o.ä. in das Modell einschließen.

Die Differentialgleichung (4.12) beschreibt damit ein dynamisches System beliebiger Ordnung mit E/A-stabiler Nulldynamik und Relativgrad 1. Zusätzlich muss die Wirkungsrichtung der Stellgröße bekannt und konstant sein. Folglich darf der Einkoppelkoeffizient $\alpha(z(t))$ zu keinem Zeitpunkt verschwinden, so dass durch eine entsprechend hohe Stellgröße das Ausgangssignal beliebig schnell verstellt werden kann.

4.3 Charakteristische Besonderheit der regelbaren Systemklasse

Ohne die zugrundeliegende Systemklasse zu spezifizieren, stützt sich Abschnitt 3.3 auf die Annahme, dass eine beschränkte Lösung existiert und zeigt auf dieser Basis, dass eine hohe, konstante Verstärkung auf einen kleinen Regelfehler führt. Diese Aussage wird hier erneut aufgegriffen, präzisiert und konkret für die nichtlineare Systemklasse \mathcal{P}

gezeigt, dass hohe Verstärkungen eine stabilisierende Wirkung besitzen und daher eine hochverstärkungs-basierte Regelung möglich ist. Dabei ist gegenüber Abschnitt 3.3 insofern eine Erweiterung enthalten, als nicht nur die Norm des Fehlersignales betrachtet wird, sondern eine Schranke für den Zeitverlauf des Regelfehlers angegeben wird.

Für die folgenden Betrachtungen wird nicht der allgemeine Trichterrand $\psi(\cdot) \in \mathcal{F}$ verwendet, sondern zusätzlich monotonen Wachstum $\dot{\psi}(t) \geq 0$ für alle $t \geq 0$ und $\psi(0) > 0$ gefordert. Dadurch ist insbesondere ausgeschlossen, dass anfänglich (für $t = 0$) der Trichterrand im Unendlichen liegt. Desweiteren werde die Reglerverstärkung nicht über die vertikale Distanz eingestellt, sondern als konstant angesetzt.

Die kennzeichnende Eigenschaft der Systemklasse \mathcal{P} wird durch [41, Prop. 6] verdeutlicht, die den positiven Effekt hoher Verstärkungen darstellt. Da der zugehörige Beweis auf [41, Thm. 5] zurückgreift, ist letzteres hier als Hilfssatz wiedergegeben.

Hilfssatz 4.4 (Existenz einer maximalen Lösung) [41]

Bezeichne $N \in \mathbb{N}$ eine natürliche Zahl, $\mathcal{D} \subset \mathbb{R}^N$ eine nicht-leere, offene Menge, $\widehat{\mathbf{Z}}$ einen kausalen Operator der Klasse \mathcal{Z} und $F : [0, \infty[\times \mathcal{D} \times \mathbb{R} \rightarrow \mathbb{R}^N$ eine Carathéodori-Funktion. Eine Lösung des Anfangswertproblems

$$\dot{x}(t) = F(t, x(t), (\widehat{\mathbf{Z}}x)(t)), \quad x(0) = x_o \in \mathcal{D} \quad (4.14)$$

auf dem Intervall $[0, T[$, $T \in]0, \infty]$, ist gegeben durch eine absolut stetige Funktion $x(\cdot) : [0, T[\rightarrow \mathbb{R}^N$, die der Differentialgleichung (4.14) für fast alle $t \in [0, T[$ genügt, wobei der Wertebereich der Lösung auf die Menge \mathcal{D} beschränkt ist. Es gilt folglich im gesamten Definitionsbereich der Lösung $t \in [0, T[$ die Aussage $x(t) \in \mathcal{D}$. Die Funktion $x(\cdot)$ heißt maximal, wenn keine Rechtserweiterung ebenfalls eine Lösung darstellt.

Es existiert eine Lösung $x : [0, T[\rightarrow \mathbb{R}^N$ des Anfangswertproblems (4.14) mit der Eigenschaft $x([0, T[) \subset \mathcal{D}$. Jede Lösung kann zu einer maximalen Lösung erweitert werden. Wenn $F(\cdot, x(t), (\widehat{\mathbf{Z}}x)(t))$ im ersten Argument lokal essentiell beschränkt ist und $x : [0, T[\rightarrow \mathbb{R}^N$ mit $x([0, T[) \subset \mathcal{D}$ eine maximale Lösung ist, wobei $T < \infty$, dann existiert für jede kompakte Menge $C \subset \mathcal{D}$ ein $t' \in [0, T[$, so dass $x(t') \notin C$.

Beweis

siehe [41, Thm. 5]. ■

Theorem 4.5

Unter der Voraussetzung, dass die Funktion $\psi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_+ \in \mathcal{F}$ monoton wächst, gegen einen positiven, beschränkten Grenzwert $\lim_{t \rightarrow \infty} \psi(t) = \psi_\infty \in \mathbb{R}_+$ konvergiert und der Funktionswert $\psi(0) > 0$ die Forderung $1/\psi(0) > |y(0) - r(0)|$ erfüllt, existiert für alle Sollwerte $w(\cdot) \in \mathcal{F}$, für jede Strecke aus der Systemklasse \mathcal{P} und für jeden Anfangswert $y(0) = y_o \in \mathbb{R}$ eine untere Schranke $k^* \in \mathbb{R}$ für die Verstärkung, so dass das zeitinvariante, proportionale Regelgesetz

$$u(t) = -k \cdot [y(t) - w(t)], \quad \text{mit } k \geq k^* \quad (4.15)$$

auf einen Regelkreis führt, der durch das Anfangswertproblem

$$\dot{y}(t) = f(p(t), (\mathbf{Z}y)(t), -k[y(t) - w(t)]), \quad y(0) = y_o \in \mathbb{R}, \quad \psi(0)|y_o - w(0)| < 1 \quad (4.16)$$

beschrieben wird, das eine Lösung auf dem Zeitintervall $[0, \infty[$ besitzt und wegen

$$\psi(t)|y(t) - w(t)| < 1 \quad \forall t \geq 0 \quad (4.17)$$

einen Zeitverlauf des Fehlersignales $e(\cdot) = y(\cdot) - w(\cdot)$ hervorruft, das sich stets innerhalb des durch die Funktion $1/\psi(\cdot)$ festgelegten Trichters bewegt.

Beweis

Per Voraussetzung ist der anfängliche Regelfehler vom Trichter umschlossen, weswegen durch entsprechende Wahl des Funktionswertes $\psi(0) \in \mathbb{R}$ die Bedingung

$$|e(0)| = |y(0) - w(0)| < 1/\psi(0) \quad (4.18)$$

sichergestellt ist. Werden $k \geq k^*$ und das Gebiet \mathcal{D} in Hilfssatz 4.4 zu $\mathcal{D} = \mathbb{R}$ festgelegt, so ist dadurch eine Lösung für das Anfangswertproblem (4.16) garantiert, wobei jede Lösung eine maximale Fortsetzung besitzt, die sich über das Intervall $[0, T[$, $T > 0$ erstreckt. Zusätzlich sei angenommen, dass die Funktion $y : [0, T[\rightarrow \mathbb{R}$ mit $T > 0$ eine maximale Lösung darstellt.

Im weiteren Verlauf gliedert sich der Beweis in fünf Teilschritte, von welchen drei durch Widerspruch bewiesen werden. Die Konstruktion der gesuchten Widersprüche verlangt die zeitliche Ableitung der Funktionen $|e(\cdot)|^2$ und $(\psi(\cdot)|e(\cdot)|)^2$ bzw. $\psi(\cdot)|e(\cdot)|$. Aus diesem Grund wird als vorbereitende Maßnahme zunächst entlang der Lösung differenziert. Es ergibt sich

$$\begin{aligned} \frac{d}{dt}e(t)^2 &= 2e(t)\dot{e}(t) = 2e(t)[\dot{y}(t) - \dot{w}(t)] = \\ &= 2e(t)[f(p(t), (\mathbf{Z}y)(t), -ke(t)) - \dot{w}(t)] \quad \forall t \in [0, T[\end{aligned} \quad (4.19)$$

und ist im Folgenden abzuschätzen.

Mit der kompakten Menge $C \subset \mathbb{R} \times \mathbb{R}$ und der Menge $U := \{-1; +1\}$ lässt sich eine stetige Funktion $\gamma_C : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ durch

$$\gamma_C(s) := \min_{(\rho, \zeta, u) \in C \times U} u \cdot f(\rho, \zeta, su) \quad (4.20)$$

definieren. Wird $s = k|e|$ und $u = -e/|e|$ gesetzt, kann

$$\gamma_C(k|e|) = \min_{(\rho, \zeta, u) \in C \times U} u \cdot f(\rho, \zeta, k|e|u) \leq -\frac{e}{|e|} \cdot f\left(\rho, \zeta, -k|e|\frac{e}{|e|}\right) \quad \forall (\rho, \zeta) \in C \quad (4.21)$$

geschrieben werden, woraus die Ungleichung

$$e \cdot f(\rho, \zeta, -ke) \leq -\gamma_C(k|e|) \cdot |e| \quad \forall (\rho, \zeta) \in C \quad (4.22)$$

gewonnen wird. Zusammen mit $|\dot{w}(t)| \leq \text{ess-sup}_{t \in [0, T[} |\dot{w}(t)| = \|\dot{w}(\cdot)\|_\infty$ für fast alle $t \in [0, T[$ ergibt sich:

$$\begin{aligned} \frac{d}{dt} e(t)^2 &\leq -2\gamma_C(k|e(t)|)|e(t)| - 2e(t)\dot{w}(t) \leq \\ &\leq -2\gamma_C(k|e(t)|)|e(t)| + 2|e(t)\dot{w}(t)| = \\ &= -2[\gamma_C(k|e(t)|) - |\dot{w}(t)|] \cdot |e(t)| \tag{4.23} \\ \Rightarrow \frac{d}{dt} e(t)^2 &\leq -2 \left[\gamma_C(k|e(t)|) - \|\dot{w}(\cdot)\|_\infty \right] \cdot |e(t)| \quad \text{für fast alle } t \in [0, T[\end{aligned}$$

Mit den Abkürzungen $\Lambda := 1/\psi(0) < \infty$ und $\lambda := \lim_{t \rightarrow \infty} (1/\psi(t)) = 1/\psi_\infty > 0$ lässt sich ein Wertebereich

$$\mathcal{A} := \{e \in \mathbb{R} \mid \lambda/2 \leq |e| \leq \Lambda\} \tag{4.24}$$

für den Fehler definieren, in dem $|e(\cdot)|$ beschränkt und gleichzeitig von 0 wegbeschränkt ist. Wird in Gleichung (4.20) $u = \text{sign}(u_n) \in U$ und $s = |u_n|$ gesetzt, folgt wegen $su = u_n$ und Bedingung (iv) in Definition 4.3 unmittelbar

$$s \rightarrow \infty \quad \Rightarrow \quad \gamma_C(s) \rightarrow \infty. \tag{4.25}$$

Dadurch bedingt findet sich eine hinreichend große Konstante $k^* > 0$, so dass

$$\gamma_C(k|e|) \geq 1 + \|\dot{w}(\cdot)\|_\infty + \Lambda^2 \|\dot{\psi}(\cdot)\|_\infty \quad \forall k \geq k^*, \quad \forall e \in \mathcal{A} \tag{4.26}$$

erfüllt ist. Ein abgeschlossener Δ -Ball mit Zentrum 0

$$\bar{\mathcal{B}}_\Delta(0) := \{s \in \mathbb{R} \mid |s| < \Delta\}, \quad \Delta > 0 \tag{4.27}$$

und ein Kompaktum

$$P \subset \mathbb{R} \quad \text{mit der Eigenschaft} \quad p(t) \in P \quad \text{für fast alle } t \geq 0 \tag{4.28}$$

dienen zur Definition einer Menge $C := P \times \bar{\mathcal{B}}_\Delta(0)$, mit deren Hilfe

$$I := \{t \in [0, T[\mid (e(t), p(t), (\mathbf{Z}y)(t)) \in \mathcal{A} \times C\} \tag{4.29}$$

festgelegt wird. Aus (4.24) folgt $-2|e| \leq -\lambda$ für alle $e \in \mathcal{A}$ und aus (4.26) folgt $\gamma_C(k|e|) - \|\dot{w}(\cdot)\|_\infty \geq 1 + \Lambda^2 \|\dot{\psi}(\cdot)\|_\infty \geq 1$ für alle $e \in \mathcal{A}$ und $k \geq k^*$. Damit lässt sich schließlich

$$\frac{d}{dt} e(t)^2 \leq - \underbrace{[\gamma_C(k|e(t)|) - \|\dot{w}(\cdot)\|_\infty]}_{\geq 1} \cdot |e(t)| \leq -\lambda \quad \text{für fast alle } t \in I \tag{4.30}$$

als Abschätzung für die Steigung von $e(\cdot)^2$ schreiben. Die ebenfalls zu berechnende Ableitung

$$\frac{d}{dt} (\psi(t)|e(t)|)^2 = \frac{d}{dt} (\psi(t)^2 e(t)^2) = \psi(t)^2 \left(\frac{d}{dt} e(t)^2 + 2 \frac{\dot{\psi}(t)}{\psi(t)} e(t)^2 \right) \tag{4.31}$$

enthält die Ableitung des quadrierten Fehlers, die sich durch (4.23) substituieren lässt. Da für fast alle t die Beziehung $\dot{\psi}(t) \leq \|\dot{\psi}(\cdot)\|_\infty$ gilt, ergibt sich

$$\frac{d}{dt} (\psi(t)|e(t)|)^2 \leq 2\psi(t)^2 \left(\left[-\gamma_C(k|e(t)|) + \|\dot{w}(\cdot)\|_\infty \right] + \frac{\|\dot{\psi}(\cdot)\|_\infty}{\psi(t)} |e(t)| \right) |e(t)| \quad (4.32)$$

für fast alle $t \in [0, T[$.

Bedingt durch das monotone Wachsen der Funktion $\psi(\cdot)$ lässt sich die Ungleichung

$$\Lambda = \frac{1}{\psi(0)} \geq \frac{1}{\psi(t)} \quad \forall t \geq 0 \quad (4.33)$$

aufstellen. Von Interesse ist das Verhalten der Funktion $\psi(\cdot)|e(\cdot)|$ in I . Aus (4.24) und (4.29) lässt sich

$$|e(t)| \leq \Lambda \quad \forall t \in I \quad (4.34)$$

ablesen, woraus die Abschätzung

$$\frac{d}{dt} (\psi(t)|e(t)|)^2 \leq -2\psi(t)^2 \left[\gamma_C(k|e(t)|) - \|\dot{w}(\cdot)\|_\infty - \|\dot{\psi}(\cdot)\|_\infty \Lambda^2 \right] |e(t)| \quad \text{für fast alle } t \in I \quad (4.35)$$

resultiert, die durch (4.26) in

$$\frac{d}{dt} (\psi(t)|e(t)|)^2 \leq -2\psi(t)^2 |e(t)| \quad \text{für fast alle } t \in I \quad (4.36)$$

übergeht und wegen $t \in I \Rightarrow |e(t)| \geq \lambda/2$ durch

$$\frac{d}{dt} (\psi(t)|e(t)|)^2 \leq -2\psi(t)^2 \frac{\lambda}{2} = -\psi(t)^2 \lambda < 0 \quad \text{für fast alle } t \in I \quad (4.37)$$

ausgedrückt werden kann.

Schritt 1: $|e(t)| < \Lambda$ für alle $t \in [0, T[$

Annahme: es existieren Zeitpunkte $t \in [0, T[$, für die $|e(t)| \geq \Lambda$ erfüllt ist.

Aufgrund der Annahme, der Stetigkeit von $e(\cdot)$ und der Tatsache, dass sich der anfängliche Regelfehler im Trichter befindet ($|e(0)| < \Lambda$), ist der Zeitpunkt

$$\hat{t} := \min \{t \in [0, T[\mid |e(t)| = \Lambda\} > 0 \quad (4.38)$$

wohldefiniert. Darüberhinaus kann ein zweiter Zeitpunkt $\hat{s} \in [0, \hat{t}[$ festgelegt werden, der die Forderung $\lambda/2 < |e(\hat{s})| < |e(\hat{t})| = \Lambda$ erfüllt. Da $\psi(\cdot)$ einen positiven Wertebereich besitzt und monoton wächst, gilt $\psi(\hat{t}) > 0$ und $\psi(\hat{s}) \leq \psi(\hat{t})$, weshalb sich die Ungleichung

$$0 < \psi(\hat{t})(|e(\hat{t})| - |e(\hat{s})|) \leq \psi(\hat{t})|e(\hat{t})| - \psi(\hat{s})|e(\hat{s})| \quad (4.39)$$

aufstellen lässt. Im Zeitintervall $[\hat{s}, \hat{t}] \subset I$ darf die Abschätzung (4.37) angesetzt werden, aus der sich $\psi(\hat{t})|e(\hat{t})| < \psi(\hat{s})|e(\hat{s})|$ ergibt. Daraus entsteht schließlich der Widerspruch

$$0 < \psi(\hat{t})|e(\hat{t})| - \psi(\hat{s})|e(\hat{s})| < 0, \quad (4.40)$$

aus dem $|e(t)| < \Lambda$ für alle $t \in [0, T[$ zu folgern ist.

Schritt 2: $T = \infty$

Nachdem die Funktionen $e(\cdot)$ und $w(\cdot)$ im Intervall $[0, T[$ beschränkt sind, ist auch die Summe $y(\cdot) = e(\cdot) + w(\cdot)$ in diesem Intervall beschränkt. Solange die Lösung des Anfangswertproblems (4.16) beschränkt bleibt ist eine stetige Fortsetzung möglich, woraus Hilfssatz 4.4 die Aussage $T = \infty$ ermittelt.

Dadurch ist zwar sichergestellt, dass eine Lösung für alle Zeiten existiert. Allerdings wird wegen $\mathcal{D} = \mathbb{R}$ ein Verlauf des Fehlersignales auch dann als Lösung bezeichnet, wenn dieses den Trichter verlässt. Im Folgenden bleibt daher zu zeigen, dass Gleichung (4.17) Gültigkeit besitzt und der Fehler im vorgegebenen Trichter verläuft.

Schritt 3: Es existiert ein Zeitpunkt $s \geq 0$, so dass $|e(s)| < \lambda/2$

Annahme: dieser Zeitpunkt existiert nicht, also $|e(s)| \geq \lambda/2$ für alle $s \geq 0$.

Zusammen mit $|e(t)| < \Lambda$ für alle $t \in \mathbb{R}_{\geq 0}$ führt die Annahme auf $e(t) \in \mathcal{A}$ für alle $t \in \mathbb{R}_{\geq 0}$. Daher ist $\mathbb{R}_{\geq 0} \subset I$ und (4.30) ist für alle $t \in \mathbb{R}_{\geq 0}$ gültig, woraus der Widerspruch

$$|e(t)|^2 \leq |e(0)|^2 - \lambda t \quad \forall t \geq 0 \quad (4.41)$$

resultiert. Ab dem Zeitpunkt $t = |e(0)|^2/\lambda$ ist Gleichung (4.41) nicht erfüllbar, was die Annahme widerlegt.

Schritt 4: $|e(t)| < \lambda$ für alle $t \geq s^* \in \mathbb{R}$

Annahme: es existieren Zeitpunkte $t \geq s^*$, für die $|e(t)| \geq \lambda$ erfüllt ist.

Aufgrund der Annahme, der Stetigkeit von $e(\cdot)$ und der in Schritt 3 bewiesenen Tatsache $|e(s)| < \lambda/2$ ist der Zeitpunkt

$$s^* := \min \{s \geq 0 \mid |e(s)| \leq \lambda/2\} \quad (4.42)$$

wohldefiniert. Werden, wie in Abbildung 4.2 markiert, zwei weitere Zeitpunkte durch

$$t^* := \min \{t \geq s^* \mid |e(t)| = \lambda\}, \quad t_* := \max \{t \in [s^*, t^*] \mid |e(t)| = \lambda/2\} \quad (4.43)$$

festgelegt, gilt $[t_*, t^*] \subset I$ und aus Anwendung von (4.30) resultiert der Widerspruch

$$\lambda = |e(t^*)| \leq |e(t_*)| = \lambda/2. \quad (4.44)$$

Daraus folgt $|e(t)| < \lambda$ für alle $t \geq s^*$, der Fehler ist ab dem Zeitpunkt s^* stets im Trichter, da $\psi(t)|e(t)| \leq \lambda^{-1}|e(t)| < 1$ für alle $t \geq s^*$.

Schritt 5: $\psi(t)|e(t)| < 1$ für alle $t \in [0, s^*]$

Im Falle $s^* = 0$ ist die Beweisführung trivial, da per Voraussetzung $|e(0)| < 1/\psi(0)$ gilt. Daher sei im Folgenden $s^* > 0$ angenommen. Die Aussagen von Schritt 1 und 2 führen auf $[0, s^*] \subset I$, so dass aus (4.37) direkt

$$\psi(t)|e(t)| \leq \psi(0)|e(0)| < 1 \quad \forall t \in [0, s^*] \quad (4.45)$$

folgt. ■

Die geregelte Strecke zeigt somit das gewünschte dynamische Verhalten, das durch den Trichter bzw. die Funktion $\psi(\cdot)$ beschrieben wird, wenn eine konstante Verstärkung

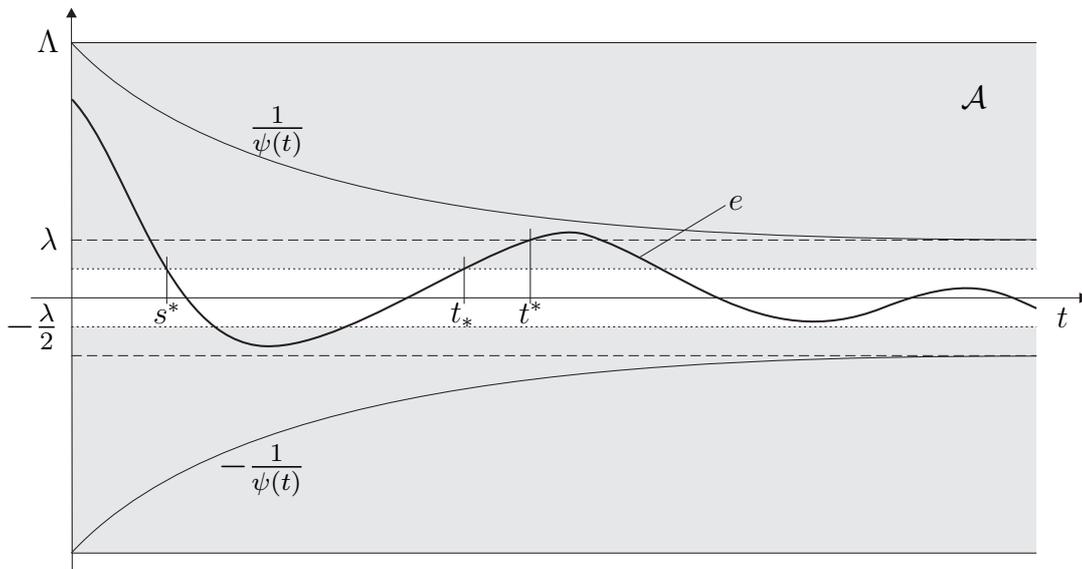


Abbildung 4.2: Skizzierter Verlauf von Fehlersignal und Trichter zur Illustration der Zeitpunkte s^* , t_* und t^* .

$k \geq k^* \in \mathbb{R}$ verwendet wird. Allerdings ist eine Realisierung dieses Reglers nicht möglich, da die Schranke k^* nicht ausschließlich von der gegebenen Funktion $\psi(\cdot)$, sondern auch vom Verhalten der unregelmäßigten Strecke abhängt und daher für Prozesse mit unbekanntem Parametern nicht berechnet werden kann. Aus diesem Grund ist eine zeitvariable Verstärkung notwendig, die auf geeignete Weise dem momentanen Systemverhalten angepasst wird.

4.4 Hochverstärkungsbasierte Regelung mit nicht-monotoner Verstärkung

Die Anpassung der Verstärkung erfolgt, wie in Abschnitt 4.1 beschrieben, über die vertikale Distanz zwischen Fehlersignal und Trichterrand. Die Regelstrategie auf Grundlage einer solchen Einstellung der Kreisverstärkung ist erstmals in [41, Thm. 7] angegeben und bewiesen.

Theorem 4.6

Wird das zeitvariable Regelgesetz

$$u(t) = -k(t)e(t) \quad \text{mit} \quad k(t) = \frac{1}{1 - \psi(t)|e(t)|} \quad \text{und} \quad e(t) = y(t) - w(t) \quad (4.46)$$

auf eine beliebige Regelstrecke aus der Systemklasse \mathcal{P} angewandt, ergibt sich ein Anfangswertproblem, welches das Verhalten des geschlossenen Regelkreises durch

$$\dot{y}(t) = f(p(t), (\mathbf{Z}y)(t), -k(t) \cdot e(t)), \quad y(0) = y_o \in \mathbb{R} \quad (4.47)$$

beschreibt. Unter den Voraussetzungen

- (i) die Funktion $\psi(\cdot) \in \mathcal{F}$ besitzt die Eigenschaften $\psi(t) > 0$ für alle $t > 0$ und $\liminf_{t \rightarrow \infty} \psi(t) > 0$
- (ii) das Sollsignal $w(\cdot) \in \mathcal{F}$ stammt aus der selben Klasse wie $\psi(\cdot)$
- (iii) der Wert $\psi(0) \geq 0$ ist so gewählt, dass zusammen mit dem gegebenen Anfangswert $y(0)$ und dem gegebenen Sollwert $w(0)$ die Bedingung $\psi(0)|y(0) - w(0)| < 1$ erfüllt ist, d.h. der anfängliche Fehler zum Zeitpunkt $t = 0$ liegt innerhalb des Trichters

besitzt das Anfangswertproblem (4.47) eine Lösung, wobei jede Lösung maximal erweiterbar ist. Mit $T > 0$ hat jede maximale Lösung $y : [0, T[\rightarrow \mathbb{R}$ die Eigenschaften:

- (i) Die Lösung ist für alle Zeiten $t \geq 0$ definiert, d.h. $T = \infty$.
- (ii) Es existiert ein $\varepsilon \in]0, 1[$, so dass $\psi(t)|e(t)| \leq 1 - \varepsilon$ für alle $t \geq 0$. Der Fehler bewegt sich nicht nur für alle Zeiten im Trichter, sondern hält einen Abstand $\varepsilon/\psi(t)$ zum Trichterrand ein.
- (iii) Die Stellgröße $u : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ und die Verstärkung $k : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ sind stetig und beschränkt.

Beweis

Der in [41] angegebene Beweis gliedert sich in vier Schritte auf. Zuerst wird gezeigt, dass eine maximale Lösung existiert, wobei der Regelfehler $e(\cdot)$ beschränkt bleibt. Hintergrund dieser Überlegung ist die Tatsache, dass die potentielle Singularität im Regelgesetz (4.46) zu einer Differentialgleichung führt, welche die Lipschitz-Bedingung nicht erfüllt. Daher kann der bekannte Existenz- und Eindeutigkeitsatz [65, S. 97] nicht angewendet werden, was zunächst die Frage aufwirft, ob das zu betrachtende Anfangswertproblem eine eindeutige Lösung besitzt. In einem zweiten Schritt wird bewiesen, dass auch die Verstärkung $k(\cdot)$ beschränkt bleibt, woraus auf einen beschränkten Stellgrößenverlauf $u(\cdot)$ geschlossen werden kann. Zuletzt wird das Existenzintervall der Lösung bis unendlich ausgedehnt und der Mindestabstand zum Trichterrand nachgewiesen.

Schritt 1:

Zunächst wird die Zustandsgröße y in Gleichung (4.47) durch den Regelfehler e ersetzt und die Größe $z(t) := t$ eingeführt. Daraus ergibt sich die Beschreibung

$$\begin{aligned}
 \dot{e}(t) &= f(p(t), (\mathbf{Z}(e + w))(t), -k(t) \cdot e(t)) - \dot{w}(t) \\
 \dot{z}(t) &= 1 \\
 x(t) &:= (e(t), z(t)) \in \mathcal{D} := \{(e, z) \in \mathbb{R} \times \mathbb{R} \mid \psi(z)|e| < 1\} \\
 e(0) &= e_o \in \mathbb{R}, \quad z(0) = 0, \quad \psi(0)\|x(0)\| < 1,
 \end{aligned} \tag{4.48}$$

die für $(\widehat{\mathbf{Z}}x)(t) = (\widehat{\mathbf{Z}}(e, z))(t) := (\mathbf{Z}(e + w))(t)$ strukturell mit dem Anfangswertproblem (4.14) übereinstimmt, wenn dort $N = 2$ gesetzt wird. Die lokal essentiell beschränkte Funktion $F : [0, T[\times \mathcal{D} \times \mathbb{R} \rightarrow \mathbb{R}^2$ ist durch

$$(t, x, \zeta) \mapsto F(t, x, \zeta) = F(t, (e, z), \zeta) := \left(f(p(t), \zeta, -k(t) \cdot e) - \dot{w}(t), 1 \right) \tag{4.49}$$

gegeben. Aus Hilfssatz 4.4 folgt, dass eine Lösung existiert und jede Lösung maximal erweitert werden kann. Eine solche Lösung sei durch $(e, z) : [0, T[\rightarrow \mathbb{R}^2$ gegeben, wobei $T \in]0, \infty]$ unbekannt und maximal gewählt ist. Durch die Definition von \mathcal{D} in Gleichung (4.48) und aus $(e(t), z(t)) \in \mathcal{D}$ für alle $t \in [0, T[$ folgt $\psi(t)|e(t)| < 1$ für alle $t \in]0, T[$, woraus die Stetigkeit des Regelfehlers auf eine beschränkte Funktion $e(\cdot)$ führt. Zusammen mit einem beschränkten Sollwert ergibt sich eine beschränkte Ausgangsgröße y .

Schritt 2:

Innerhalb des Zeitintervalles $[0, T[$, in dem die Lösung der Differentialgleichung (4.47) definiert ist, wird zunächst ein beliebiger Zeitpunkt $\delta \in]0, T[$ festgelegt. Nachdem $k : [0, T[\rightarrow \mathbb{R}_{\geq 0}$ eine stetige Funktion ist, kann der Schrankensatz (siehe Theorem A.2) angewendet werden, aus dem direkt

$$k(\cdot) \in \mathcal{L}^\infty([0, \delta]; \mathbb{R}_{\geq 0}) \quad (4.50)$$

folgt. Per Widerspruch lässt sich beweisen, dass die Verstärkung $k(\cdot)$ auch für alle Zeiten $t > \delta$ beschränkt bleibt. Unter der Annahme einer unbeschränkt wachsenden Verstärkung wird dazu die Existenz eines Zeitintervalles nachgewiesen, auf dem die Funktion $k(\cdot)$ zwar eine strikt negative Ableitung besitzt, gleichzeitig aber anwächst. Aus diesen unvereinbaren Charakteristika darf auf eine beschränkte Funktion $k(\cdot) \in \mathcal{L}^\infty$ geschlossen werden. Der hier beschriebene Grundgedanke des Beweises wird nachfolgend ausführlich dargestellt.

Analog zu Abschnitt 4.3 stützt sich der Entwurf des Widerspruches auf die Ableitung der Funktion $(\psi(\cdot)|e(\cdot)|)^2$, die vorausgehend unter Ausnutzung der dort vorgenommenen Berechnungen differenziert wird, wobei k durch $k(t)$ zu ersetzen ist. Ausgehend von Gleichung (4.31) ergibt sich durch Einsetzen von (4.23) die Abschätzung:

$$\frac{d}{dt} \left(\psi(t)|e(t)| \right)^2 \leq -2\psi(t)^2 \left[\gamma_C(k(t)|e(t)|) - \|\dot{w}(\cdot)\|_\infty - \frac{\dot{\psi}(t)}{\psi(t)}|e(t)| \right] \cdot |e(t)| \quad (4.51)$$

für fast alle $t \in [0, T[$

Da aus Hilfssatz 4.4 hervorgeht, dass für $t \in [0, T[$ der Fehler im Trichter verläuft und dessen Rand nicht berühren kann, lässt sich die Ungleichung $\psi(t)|e(t)| < 1$ ansetzen. Wegen $\psi(t) > 0$ für alle $t > 0$ darf diese im Intervall $]0, T[$ durch $\psi(t)^2$ geteilt werden. Mit $|e(t)|/\psi(t) < 1/\psi(t)^2$ und $\dot{\psi}(t) \leq |\dot{\psi}(t)|$ ergibt sich

$$\frac{d}{dt} \left(\psi(t)|e(t)| \right)^2 \leq -2\psi(t)^2 \left[\gamma_C(k(t)|e(t)|) - \|\dot{w}(\cdot)\|_\infty - \frac{|\dot{\psi}(t)|}{\psi(t)^2} \right] \cdot |e(t)| \quad (4.52)$$

für fast alle $t \in]0, T[$.

Im Hinblick auf den gewünschten Widerspruch ist ein Zeitintervall festzulegen, in welchem die Funktion $(\psi(\cdot)|e(\cdot)|)^2$ streng monoton fällt. Hierzu ist das Kriterium

$$\gamma_C(k(t)|e(t)|) > \|\dot{w}(\cdot)\|_\infty + \frac{|\dot{\psi}(t)|}{\psi^2(t)} \quad (4.53)$$

zu erfüllen. Aus diesem Grund ist der Einfluss von $\|\dot{w}(\cdot)\|_\infty + |\dot{\psi}(t)|/\psi^2(t)$ abzuschätzen. Durch die Voraussetzungen an die Funktion $\psi(\cdot)$ ist gewährleistet, dass eine Konstante $c_1 > 1$ gefunden werden kann, für die

$$\frac{1}{c_1} \leq \psi(t) \leq c_1 \quad \Rightarrow \quad \frac{1}{\psi(t)^2} \leq c_1^2 \quad \forall t \in [\delta, T[\quad (4.54)$$

zutrifft. Weil die Ableitung $\dot{\psi}(\cdot)$ essentiell beschränkt ist, kann darüberhinaus c_1 so gewählt werden, dass zusätzlich auch

$$|\dot{\psi}(t)| \leq c_1 \quad \text{für fast alle } t \in [\delta, T[\quad (4.55)$$

gilt. Da außerdem der Sollwert $w(\cdot)$ der Funktionsklasse \mathcal{F} angehört, ist dessen Signalnorm $\|\dot{w}(\cdot)\|_\infty < \infty$ beschränkt. Wird $c_2 := \|\dot{w}(\cdot)\|_\infty + c_1^3 < \infty$ definiert, lässt sich schließlich

$$\begin{aligned} \frac{d}{dt} \left(\psi(t)|e(t)| \right)^2 &\leq -2\psi(t)^2 \left[\gamma_C(k(t)|e(t)) - \|\dot{w}(\cdot)\|_\infty - c_1^3 \right] \cdot |e(t)| = \\ &= -2\psi(t)^2 \left[\gamma_C(k(t)|e(t)) - c_2 \right] \cdot |e(t)| \quad \text{für fast alle } t \in [\delta, T[\end{aligned} \quad (4.56)$$

angeben.

Der Formulierung des angestrebten Widerspruches wird die Annahme

$$k(\cdot) \notin \mathcal{L}^\infty([\delta, T[; \mathbb{R}) \quad (4.57)$$

zugrunde gelegt. Desweiteren werden für jedes $j \in \mathbb{N}$ zwei konstante Zeitpunkte definiert, um damit geeignete Grenzen

$$\tau_j := \inf \{ t \in [\delta, T[\mid k(t) = k(\delta) + j + 1 \} \quad (4.58)$$

$$\sigma_j := \sup \{ t \in [\delta, \tau_j[\mid k(t) = k(\delta) + j \} \quad (4.59)$$

für das gesuchte Intervall festlegen zu können. Bedingt durch (4.57) sind die Größen σ_j und τ_j für alle $j \geq 0$ wohldefiniert. Die durchschnittliche Steigung von $k(\cdot)$ im Intervall $[\sigma_j, \tau_j]$ ist durch

$$\frac{k(\tau_j) - k(\sigma_j)}{\tau_j - \sigma_j} = \frac{1}{\tau_j - \sigma_j} > 0 \quad (4.60)$$

gegeben. Offensichtlich ist die Definition des Zeitintervalles $[\sigma_j, \tau_j]$ so gewählt, dass $k(\cdot)$ darin im Durchschnitt anwächst. Es handelt sich jedoch nicht notwendigerweise um monotonen Wachstum, eine zwischenzeitliche Verringerung ist zulässig. Abbildung 4.3 enthält beispielhaft einen möglichen Verlauf der Funktion $k(\cdot)$. Eingetragen sind auch die Zeitpunkte τ_j und σ_j , zu denen k zum ersten Mal nach δ den Wert $k(\delta) + j + 1$ erreicht, bzw. zum letzten Mal vor dem Zeitpunkt τ_j den Wert $k(\delta) + j$ annimmt. Im Intervall $[\sigma_j, \tau_j]$ ist $k(\cdot)$ durch die beiden Schranken $k(\delta) + j$ und $k(\delta) + j + 1$ eingegrenzt, was sich in der Beziehung

$$k(\sigma_j) = k(\delta) + j \leq k(t) \leq k(\delta) + j + 1 = k(\tau_j) \quad \forall t \in [\sigma_j, \tau_j] \quad (4.61)$$

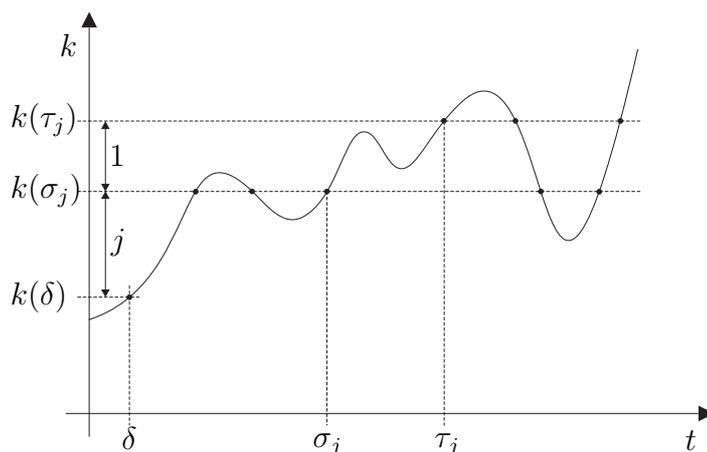


Abbildung 4.3: Exemplarischer Verlauf der Verstärkung.

niederschlägt.

Um die zu (4.60) konträre Aussage aufzustellen, ist lediglich das Vorzeichen der zeitlichen Ableitung von $k(\cdot)$ im Intervall $[\sigma_j, \tau_j]$ relevant. Die Monotonie der Funktion $\psi(t)|e(t)| \mapsto 1/(1 - \psi(t)|e(t)|) = k(t)$ aus der die Verstärkung gemäß (4.46) berechnet wird, gestattet hierzu alternativ eine Untersuchung ihres Argumentes, d.h. der Funktion $\psi(\cdot)|e(\cdot)|$ bzw. auch von $(\psi(\cdot)|e(\cdot)|)^2$. Aus der vorbereitenden Abschätzung (4.56) ist zu entnehmen, dass innerhalb sämtlicher Zeitintervalle, in welchen $\gamma_C(k(t)|e(t)|)$ die Konstante $c_2 < \infty$ übersteigt, eine negative Ableitung vorliegt, weshalb dort die gewünschte strenge Monotonie besteht. Daher muss im Folgenden das Verhalten der Funktion $\gamma_C(\cdot)$ abhängig von ihrem Argument $k(t)|e(t)|$ analysiert werden, so dass daraus ein geeignetes Zeitintervall festgelegt werden kann. Aus (4.25) ist abzulesen, dass $\gamma_C(k(t)|e(t)|) > c_2$ erzwungen wird, wenn das Produkt $k(t)|e(t)|$ entsprechend anwächst. Falls $|e(\cdot)|$ von 0 wegbeschränkt ist, wird dies durch eine ausreichend hohe Verstärkung $k(t)$ garantiert.

Unmittelbar kann eine hinreichend große Schranke $K \in \mathbb{N}$ gewählt werden, die $K + k(\delta) \geq 1$ erfüllt und daher für jede Zahl $j > K$ den Zusammenhang

$$k(t) = \frac{1}{1 - \psi(t)|e(t)|} \geq j + k(\delta) > K + k(\delta) \geq 1 > 0 \quad \forall t \in [\sigma_j, \tau_j] \quad (4.62)$$

hervorrufen. Umstellen ergibt

$$\psi(t)|e(t)| \geq 1 - \frac{1}{j + k(\delta)} > 0 \quad \forall t \in [\sigma_j, \tau_j], \quad (4.63)$$

woraus mit $\psi(t) \leq c_1$ für alle $t \in [\delta, T[$ die Beziehung

$$|e(t)| \geq \frac{1}{c_1} \left(1 - \frac{1}{j + k(\delta)} \right) \geq \frac{1}{c_1} \left(1 - \frac{1}{k(\delta)} \right) =: c_3 \quad \forall t \in [\sigma_j, \tau_j] \quad (4.64)$$

resultiert. Die gesuchte untere Schranke für $|e(t)|$ ist demzufolge durch c_3 gegeben, weswegen Annahme (4.57) direkt die Existenz eines Zeitintervalles mit hinreichend

großer Verstärkung $k(t)$ bedingt, um die Ungleichung

$$\gamma_C(k(t)|e(t)|) > c_2 \quad \forall t \in [\sigma_j, \tau_j] \quad (4.65)$$

zu erfüllen. In der Folge entsteht eine negative Ableitung

$$\frac{d}{dt}(\psi(t)|e(t)|)^2 \leq -2\psi^2(t)(\gamma_C(k(t)|e(t)|) - c_2)|e(t)| < 0 \quad (4.66)$$

für fast alle $t \in [\sigma_j, \tau_j]$,

die ermöglicht, auf eine streng monoton fallende Funktion $\psi(\cdot)|e(\cdot)|$ zu schließen. Da $\sigma_j < \tau_j$ lässt sich daraus $\psi(\tau_j)|e(\tau_j)| < \psi(\sigma_j)|e(\sigma_j)|$ bzw. die Aussage

$$\frac{1}{1 - \psi(\tau_j)|e(\tau_j)|} < \frac{1}{1 - \psi(\sigma_j)|e(\sigma_j)|} \quad \text{für fast alle } t \in [\sigma_j, \tau_j] \quad (4.67)$$

ansetzen, mit der sich schlussendlich der gewünschte Widerspruch

$$1 = k(\tau_j) - k(\sigma_j) = \frac{1}{1 - \psi(\tau_j)|e(\tau_j)|} - \frac{1}{1 - \psi(\sigma_j)|e(\sigma_j)|} < 0 \quad (4.68)$$

für fast alle $t \in [\sigma_j, \tau_j]$

aufstellen lässt. Aus diesem Grund ist die Annahme unzutreffend und $k(\cdot) \in \mathcal{L}^\infty$ ist zu folgern. Zusammen mit $e(\cdot) \in \mathcal{L}^\infty$ ergibt sich darüberhinaus auch eine beschränkte Stellgröße $u(\cdot) = -k(\cdot)e(\cdot)$.

Schritt 3:

Ausgehend von der Annahme $T < \infty$ zeigt Hilfssatz 4.4, dass ab einem Zeitpunkt $t' \in [0, T[$ die Lösung $x(t') = (e(t'), z(t')) \notin C$ jede kompakte Menge $C \subset \mathcal{D}$ verlässt. Wegen $e(\cdot) \in \mathcal{L}^\infty$ gibt es eine Konstante $E > 0$, die $|e(t)| \leq E$ für alle $t \in [0, T[$ erfüllt. Weil sich C jedoch durch

$$C := \{(e, z) \in \mathbb{R} \times \mathbb{R}_{\geq 0} \mid \psi(z)|e| \leq 1 - \varepsilon, \quad |e| \leq E, \quad z \in [0, T]\} \quad (4.69)$$

als kompakte Teilmenge von \mathcal{D} mit der Eigenschaft $x(t) = (e, z) \in C$ für alle $t \in [0, T[$ definieren lässt, widerspricht dies der Aussage von Hilfssatz 4.4 und führt daher auf das Resultat $T = \infty$.

Schritt 4:

Abschließend bleibt zu zeigen, dass die Funktion $e(\cdot)$ durch $\varepsilon > 0$ vom Trichterrand wegbeschränkt ist. Einzig im Falle, dass der Regelfehler $e(t)$ den Trichterrand erreicht, gilt $\psi(t)|e(t)| = 1$, woraus sich eine Verstärkung $k(t) = \infty$ errechnet. Weil $k(\cdot)$ jedoch erwiesenermaßen beschränkt bleibt, muss offenkundig eine Konstante $\varepsilon > 0$ existieren, so dass

$$\psi(t)|e(t)| \leq 1 - \varepsilon \quad \forall t \in [0, \infty[\quad (4.70)$$

gilt. Die Aussage $k(\cdot) \in \mathcal{L}^\infty$ bedeutet daher nicht nur, dass der Fehler im Trichter verbleibt, sondern darüberhinaus einen Mindestabstand zum Trichterrand einhält. ■

Nachdem sich aus der theoretischen Untersuchung ableitet, dass der Regler sämtliche Strecken aus der Systemklasse \mathcal{P} in ihrem transienten und asymptotischen Verhalten wunschgemäß beeinflusst, wird die Verwendbarkeit der Regelstrategie in einer Simulationsstudie evaluiert.

4.5 Anwendung am Einmassensystem

Eine mögliche Anwendung für Funnel-Control ist die Regelung von Einmassensystemen. Wird die Masse der antreibenden Maschine und die Masse der Last als ideal (starr) gekoppelt angenommen, liegt ein Einmassensystem vor, und die einzelnen Trägheitsmomente von Motor und Last lassen sich zu einem Gesamtträgheitsmoment $J > 0$ zusammenfassen. Damit resultiert eine Zustandsbeschreibung erster Ordnung

$$\dot{\omega}(t) = \frac{1}{J} \cdot u(t) - \frac{1}{J} \cdot p(t), \quad \omega(0) = \omega_o \in \mathbb{R}, \quad y(t) = \omega(t) \quad (4.71)$$

für diese Regelstrecke. Der Zustand ist durch die Drehgeschwindigkeit $\omega(t)$ gegeben, die gleichzeitig auch den Ausgang $y(t)$ bildet. Der Eingang des Prozesses ist das antreibende Drehmoment $u(t)$, eine Last wird durch die Störung $p(t)$ eingebracht. Ein solches System gehört zur Systemklasse \mathcal{P} , weil das Vorzeichen der instantanen Verstärkung $1/J$ aus physikalischen Gründen positiv und damit bekannt ist, der Relativgrad bei 1 liegt, und da keine Nulldynamik existiert, liegt folglich auch keine instabile Rückkopplung vor. Für ein Gesamtträgheitsmoment $J = 0.2 \text{ kgm}^2$ sind entsprechende Simulationsergebnisse in den Abbildungen 4.4 und 4.5 dargestellt.

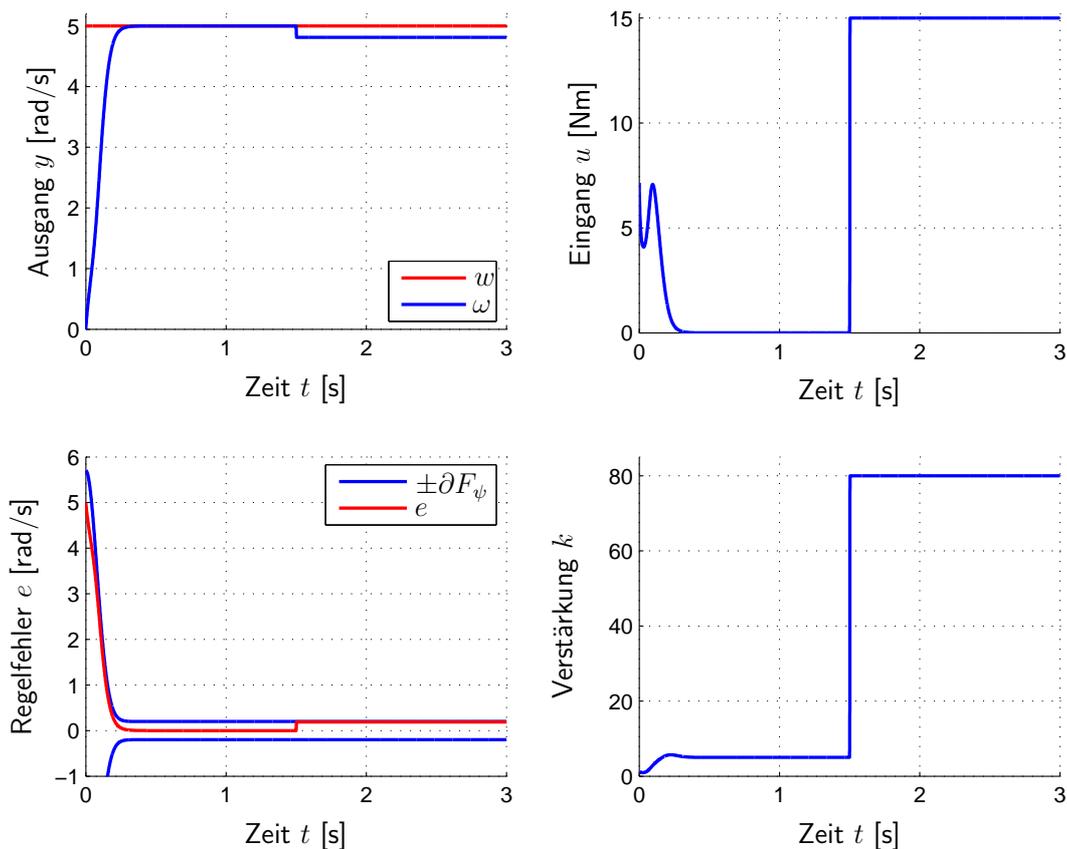


Abbildung 4.4: Hochverstärkungs-basierte Regelung mit nicht-monotoner Verstärkungsfunktion für das Einmassensystem (4.71) für die Trichterbreite $\Delta_F = 0.2 \text{ rad/s}$.

Die beiden Simulationen unterscheiden sich im gewählten Trichterrand. Während sich

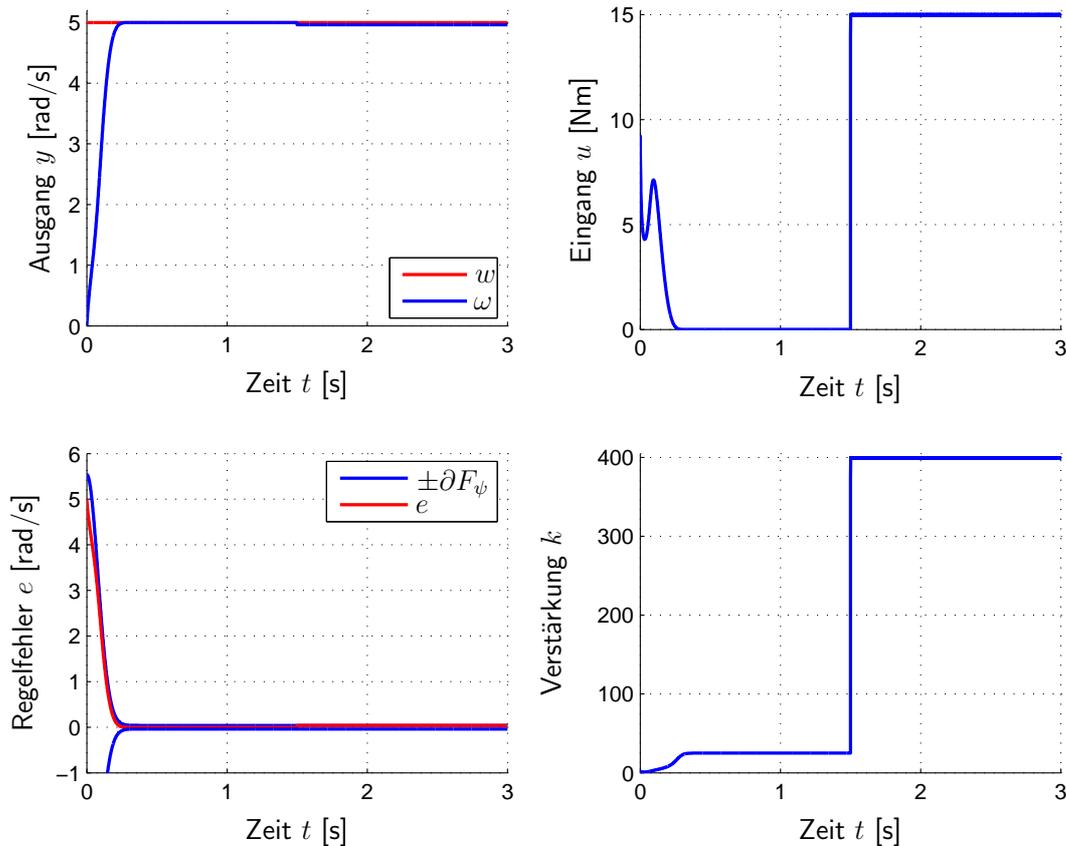


Abbildung 4.5: Hochverstärkungs-basierte Regelung mit nicht-monotoner Verstärkungsfunktion für das Einmassensystem (4.71) für die Trichterbreite $\Delta_F = 0.04$ rad/s.

in Abbildung 4.4 der Trichter asymptotisch auf eine Breite $\Delta_F = 0.2$ rad/s verengt, ist in Abbildung 4.5 eine Endweite von $\Delta_F = 0.04$ rad/s angesetzt. In beiden Fällen wird der Trichterverlauf durch die Funktion

$$\partial F_\psi : t \mapsto a_F \cdot e^{-\lambda_F t^2} + \Delta_F \quad \text{mit} \quad \Delta_F > 0, \quad a_F = 5.5 \quad \text{und} \quad \lambda_F = 80 \quad (4.72)$$

beschrieben. Nachdem die Regelstrecke aus der Systemklasse \mathcal{P} stammt, kann der beschriebene Regler diese erwartungsgemäß regeln. Die Strecke befindet sich in der Ruhelage $\omega = 0$ rad/s, wenn zum Zeitpunkt $t = 0$ s ein Sollwertsprung von 0 auf 5 rad/s erfolgt. Die Trichterweite $\partial F_\psi(0) = a_F + \Delta_F > 5$ rad/s ist dem Regelfehler $e(0) = 5$ rad/s so angepasst, dass der anfängliche Fehler vom Trichter umschlossen wird. Der Parameter λ_F legt die Rate fest, mit der sich der Trichterrand zusammenzieht. Dieser Vorgabe folgt der Regelfehler, der durch den Trichter begrenzt wird und daher ebenfalls mit dieser Mindestrate abnimmt. Dadurch kann über λ_F die gewünschte Mindestbeschleunigung des Systems eingestellt werden. Weil die Regelstrecke global integrales Verhalten zeigt, kann der proportionale Regler (unabhängig von dessen Verstärkung) stationär genaues Verhalten erzielen, der Sollwert wird asymptotisch exakt erreicht, wobei keine Stellgröße mehr erforderlich ist. Daher reduziert sich die Stellgröße von anfänglich positiven Werten, die für die gewünschte Beschleunigung benötigt werden, auf 0. Zum Zeitpunkt $t = 1.5$ s wird ein Lastsprung von 15 Nm aufgeschaltet. Bekanntlich kann ein proportionaler Regler unter Störgrößeneinwirkung die stationäre Genauigkeit nicht

beibehalten. Die Stellgröße steigt auf den Wert der Last, um die Störgröße zu kompensieren. Dennoch verursacht das Lastmoment einen Drehzahleinbruch, der jedoch durch den Wert Δ_F begrenzt werden kann. Trotz Störung verbleibt der Fehler im Trichter, so dass sich die Drehzahl höchstens auf 4.8 rad/s (für $\Delta_F = 0.2$ rad/s) bzw. auf 4.96 rad/s (für $\Delta_F = 0.04$ rad/s) reduzieren kann. An dieser Stelle ist ersichtlich, dass über eine hinreichend starke Verengung des Trichters der verbleibende Regelfehler beliebig klein gehalten werden kann. Der für proportionale Regler typische Nachteil eines permanenten Fehlers fällt nicht ins Gewicht, solange ein ausreichend schmaler Trichter angesetzt werden darf. In der Anwendung allerdings verlangt Messrauschen stets eine Mindestbreite. Diese fällt meist größer aus als die maximale Breite, die aus Gesichtspunkten der Genauigkeit wünschenswert ist. Diesem Problem widmen sich die Abschnitte 4.6 und 4.7, die den proportionalen Regler um ein internes Modell erweitern.

4.6 Folgeregelung und Störgrößenunterdrückung

Generell teilt sich die Anforderung an eine Regelung in zwei unterschiedliche Aufgabenstellungen auf. Zum einen wird die Stabilität des Regelkreises verlangt, der Regler muss beschränkte Zustandsgrößen garantieren. Besonders bei der Anwendung an realen Anlagen darf die Stabilität nicht verloren gehen, wenn die Regelstrecke nicht exakt ihre nominalen Parameter, sondern davon abweichende Werte besitzt (robuste Stabilität). Andererseits soll der Prozessausgang einem vorgegebenen Sollwert folgen oder sich diesem asymptotisch annähern, so dass ein gewünschtes Regelziel erreicht wird. Idealerweise sollen beide Vorgaben auch dann eingehalten werden, wenn eine Störgröße einwirkt.

Aus den vorangegangenen Abschnitten wird deutlich, dass Funnel-Control zwar jede Regelstrecke aus der Systemklasse \mathcal{P} stabilisiert und den Regelfehler dabei durch den Trichter begrenzt. Die Aufgabe der (robusten) Stabilisierung wird dadurch erfüllt. Nachdem der Regelfehler aber nicht in jedem Falle asymptotisch gegen 0 strebt (vor allem nicht unter Störeinwirkung), ist das Ziel der asymptotischen Folgeregelung und der Störgrößenunterdrückung als verfehlt zu betrachten. Aus diesem Grund müssen zusätzliche Maßnahmen erfolgen, deren theoretische Grundlagen im Folgenden dargelegt werden.

Francis und Wonham widmen Mitte der siebziger Jahre eine Reihe prominenter Publikationen [24], [25], [99], [100] der Fragestellung, welche Struktur ein Regler benötigt, damit ein stabiler, linearer Regelkreis asymptotisches Folgeverhalten aufweist und asymptotisch die Wirkung einer Störgröße unterdrückt. Die Antwort darauf wird von Francis und Wonham im „*Internal Model Principle*“ zusammengefasst. Sontag erweitert diese Resultate in [96] auf nichtlineare Systeme. Demnach muss im Regelkreis notwendigerweise ein Subsystem – das interne Modell – vorhanden sein, das den Sollwert bzw. das Störsignal nachbilden kann. In [100, S. 210] stellt Wonham diesen Sachverhalt folgendermaßen dar:

„A regulator is structurally stable only if the controller utilizes feedback

of the regulated variable, and incorporates in the feedback loop a suitably reduplicated model of the dynamic structure of the exogenous signals which the regulator is required to process.

In plain terms, every good regulator must incorporate a model of the outside world.“

Zur Herleitung dieser fundamentalen Aussage wird eine lineare Regelstrecke

$$\begin{aligned}\dot{x}_s(t) &= A_s x_s(t) + b_s u(t), & x_s(0) &\in \mathbb{R}^{n_s} \\ y_s(t) &= c_s x_s(t), & A_s \in \mathbb{R}^{n_s \times n_s}, b_s, c_s^T &\in \mathbb{R}^{n_s}\end{aligned}\quad (4.73)$$

der Ordnung $n_s \in \mathbb{N}$ betrachtet. Zusammen mit dem Regler

$$\begin{aligned}\dot{x}_r(t) &= A_r x_r(t) + b_r e(t), & x_r(0) &\in \mathbb{R}^{n_r}, \quad n_r \in \mathbb{N} \\ y_r(t) &= c_r x_r(t) + d_r e(t), & A_r \in \mathbb{R}^{n_r \times n_r}, b_r, c_r^T &\in \mathbb{R}^{n_r}, \quad d_r \in \mathbb{R},\end{aligned}\quad (4.74)$$

dem Sollwertsignal $w(\cdot)$ und den Festlegungen

$$e(t) = y_s(t) - w(t) \quad \text{und} \quad u(t) = y_r(t) \quad (4.75)$$

ergibt sich ein Regelkreis, der durch

$$\begin{aligned}\begin{pmatrix} \dot{x}_s(t) \\ \dot{x}_r(t) \end{pmatrix} &= \begin{bmatrix} A_s + d_r b_s c_s & b_s c_r \\ b_r c_s & A_r \end{bmatrix} \begin{pmatrix} x_s(t) \\ x_r(t) \end{pmatrix} - \begin{pmatrix} b_s d_r \\ b_r \end{pmatrix} w(t), & \begin{pmatrix} x_s(0) \\ x_r(0) \end{pmatrix} &\in \mathbb{R}^{n_s + n_r} \\ e(t) &= \begin{pmatrix} c_s & 0 \end{pmatrix} \begin{pmatrix} x_s(t) \\ x_r(t) \end{pmatrix} - w(t)\end{aligned}\quad (4.76)$$

beschrieben wird. Durch die Festlegung des Regelfehlers $e(\cdot) = y_s(\cdot) - w(\cdot)$ als Ausgang des Regelkreises wird die Aufgabe der asymptotischen Folgeregelung in das Problem der asymptotischen Stabilisierung überführt.

Zum Zwecke der Abkürzung wird

$$A := \begin{bmatrix} A_s + d_r b_s c_s & b_s c_r \\ b_r c_s & A_r \end{bmatrix}, \quad b := \begin{pmatrix} -b_s d_r \\ -b_r \end{pmatrix}, \quad c := \begin{pmatrix} c_s & 0 \end{pmatrix}, \quad x(t) := \begin{pmatrix} x_s(t) \\ x_r(t) \end{pmatrix} \quad (4.77)$$

festgelegt. Essentiell für asymptotisches Auslöchen des Fehlersignales ist die asymptotische Stabilität des Regelkreises, weshalb in den nachfolgenden Ausführungen stets $\sigma(A) \subset \mathbb{C}_-$ vorausgesetzt wird. Diese Forderung ist bei der Auswahl eines geeigneten Reglers zu berücksichtigen.

Durch die Koeffizienten $\gamma_0, \dots, \gamma_{l-1} \in \mathbb{R}$, $l \in \mathbb{N}$, und $\gamma_l = 1$ wird die Signalklasse

$$\mathcal{W} := \left\{ w \in \mathcal{C}^l(\mathbb{R}_{\geq 0}; \mathbb{R}) \mid \sum_{i=0}^l \gamma_i w^{(i)}(\cdot) = 0 \right\} \quad (4.78)$$

festgelegt, der das Sollwertsignal $w(\cdot)$ angehöre. Mittels Variablensubstitution

$$\eta_1 := w, \quad \eta_2 := \dot{\eta}_1 = \dot{w}, \quad \eta_3 := \dot{\eta}_2 = \ddot{w}, \quad \dots \quad \eta_l := \dot{\eta}_{l-1} = w^{(l-1)} \quad (4.79)$$

lässt sich die Differentialgleichung der Ordnung l

$$\gamma_0 w(\cdot) + \gamma_1 \dot{w}(\cdot) + \dots + \gamma_{l-1} w^{(l-1)}(\cdot) + w^{(l)}(\cdot) = 0 \quad (4.80)$$

in l verkoppelte Differentialgleichungen erster Ordnung

$$\begin{aligned} \begin{pmatrix} \dot{\eta}_1(t) \\ \dot{\eta}_2(t) \\ \vdots \\ \dot{\eta}_l(t) \end{pmatrix} &= \underbrace{\begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -\gamma_0 & \dots & \dots & -\gamma_{l-1} \end{bmatrix}}_{=: A_G \in \mathbb{R}^{l \times l}} \begin{pmatrix} \eta_1(t) \\ \eta_2(t) \\ \vdots \\ \eta_l(t) \end{pmatrix}, \quad \eta(0) = \begin{pmatrix} \eta_1(0) \\ \eta_2(0) \\ \vdots \\ \eta_l(0) \end{pmatrix} \in \mathbb{R}^l \\ v(t) &= \underbrace{\begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}}_{=: c_G \in \mathbb{R}^{1 \times l}} \begin{pmatrix} \eta_1(t) & \eta_2(t) & \dots & \eta_l(t) \end{pmatrix}^T \end{aligned} \quad (4.81)$$

umwandeln. Wird der Anfangswert

$$\eta(0) = \eta_o^* := \begin{pmatrix} w(0) \\ \dot{w}(0) \\ \vdots \\ w^{(l-1)}(0) \end{pmatrix} \in \mathbb{R}^l \quad (4.82)$$

gewählt, ist durch das Zustandsmodell (4.81) eine Realisierung der Differentialgleichung in (4.78) gegeben, die jede Funktion der Klasse \mathcal{W} erzeugen kann, weil $v(\cdot) = w(\cdot)$ für alle $w \in \mathcal{W}$ gilt. Es soll die Voraussetzung gelten, dass $l \in \mathbb{N}$ minimal gewählt ist und daher der Generator (A_G, c_G) als vollständig zustandsbeobachtbar angenommen werden kann.

Einblicke in den Zusammenhang zwischen der homogenen Eigendynamik des Regelkreises (4.76) und den Eigenschaften der Signalklasse \mathcal{W} eröffnen sich, wenn der Regelkreis umgeformt wird.

Wird in Gleichung (4.76) der Summand $bcx(t) - bcx(t) = 0$ hinzugefügt und anschließend $cx(t) = e(t) + w(t)$ eingesetzt, folgt:

$$\begin{aligned} \dot{x}(t) &= [A + bc]x(t) - be(t), \quad x(0) \in \mathbb{R}^{n_s + n_r} \\ y_s(t) &= cx(t) \\ e(t) &= y_s(t) - w(t) \end{aligned} \quad (4.83)$$

Dadurch wird offensichtlich, dass sich die beiden in Abbildung 4.6 dargestellten Systeme ineinander überführen lassen.

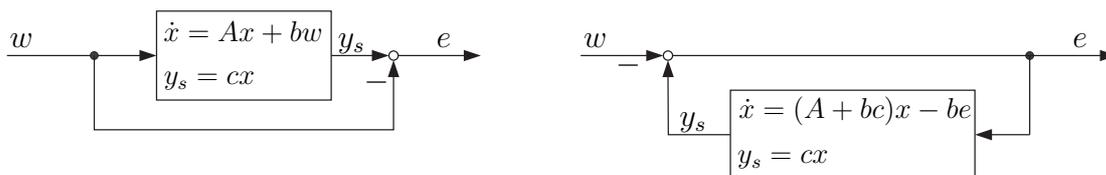


Abbildung 4.6: Regelkreis (4.76) in verschiedener Darstellung.

Das Problem der exakten Nullung des Ausganges wird gelöst, wenn für alle Zeitpunkte $t \geq 0$ die Gleichung $y_s(t) = cx(t) = w(t)$ erfüllt wird. Aus $e(t) = y_s(t) - w(t) = 0$ für alle $t \geq 0$ folgt, dass in der Differentialgleichung (4.83) das Eingangssignal entfällt. Deshalb muss das homogene System

$$\begin{aligned} \dot{x}(t) &= [A + bc]x(t), & x(0) &\in \mathbb{R}^{n_s+n_r} \\ y_s(t) &= cx(t) \end{aligned} \quad (4.84)$$

aufgrund seiner Eigendynamik

$$x(t) = \exp\{(A + bc)t\}x(0) \quad \forall t \geq 0 \quad (4.85)$$

am Ausgang den Sollwert $w(\cdot)$ generieren können. Hierzu ist in der Differentialgleichung (4.84) ein Subsystem erforderlich, das als internes Modell bezeichnet wird und im Ähnlichkeitsorbit des Generators (4.81) liegt. Demzufolge ist bei der Wahl des Reglers (4.74) die Forderung

$$\sigma(A_G) \subset \sigma(A + bc) \quad (4.86)$$

zu berücksichtigen. Diese Überlegungen sind in Theorem 4.8 zusammengefasst, dessen Beweis auf Hilfssatz 4.7 gestützt ist. Jener zeigt, dass für ein beobachtbares System in Jordan'scher Normalform eine Ähnlichkeitstransformation existiert, welche die Jordanform für die Systemmatrix beibehält, dem Auskoppelvektor aber eine beliebige Struktur zuweist.

Hilfssatz 4.7

Sei $n \in \mathbb{N}$, $c^T \in \mathbb{C}^n$ und $A \in \mathbb{C}^{n \times n}$ eine Matrix in Jordan'scher Normalform. Diese besitze ein Spektrum mit $1 \leq \bar{n} \leq n$ verschiedenen Eigenwerten $\lambda_i \in \mathbb{C}$, $i = 1, 2, \dots, \bar{n}$, deren Vielfachheiten durch $1 \leq m_i \leq n$ bezeichnet werden und $\sum_{i=1}^{\bar{n}} m_i = n$ erfüllen. Die Indizierung der Eigenwerte und der zugehörigen Jordan-Blöcke

$$J_i := \lambda_i \in \mathbb{C} \quad \text{für } m_i = 1 \quad \text{bzw.} \quad J_i := \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix} \in \mathbb{C}^{m_i \times m_i} \quad \text{für } m_i > 1 \quad (4.87)$$

sei derart vorgenommen, dass $A = \text{diag}(J_1, J_2, \dots, J_{\bar{n}})$ gelte. Falls das Tuple (A, c)

die Kalman'sche Beobachtbarkeitsbedingung

$$\det \begin{bmatrix} c \\ cA \\ \vdots \\ cA^{n-1} \end{bmatrix} \neq 0 \quad (4.88)$$

erfüllt, existiert eine Matrix $S \in \mathbf{GL}_n(\mathbb{C})$, die simultan den beiden Gleichungen

$$SAS^{-1} = A \quad \text{und} \quad cS^{-1} = \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{1 \times n} \quad (4.89)$$

genügt.

Beweis

Für jedes $i = 1, 2, \dots, \bar{n}$ wird eine Matrix D_i und eine Matrix $S_i \in \mathbf{GL}_{m_i}(\mathbb{C})$ mit Hilfe der komplexen Zahlen $s_{i,1}, s_{i,2}, \dots, s_{i,m_i} \in \mathbb{C}$, $s_{i,1} \neq 0$, durch

$$S_i^{-1} := \begin{bmatrix} s_{i,1} & s_{i,2} & \dots & \dots & s_{i,m_i} \\ & s_{i,1} & s_{i,2} & \ddots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & s_{i,2} \\ & & & & s_{i,1} \end{bmatrix} \quad \text{und} \quad D_i := \begin{bmatrix} 0 & 1 & 0 & & \\ & 0 & 1 & \ddots & \\ & & 0 & \ddots & 0 \\ & & & \ddots & 1 \\ & & & & 0 \end{bmatrix} \in \mathbb{R}^{m_i \times m_i} \quad (4.90)$$

festgelegt, um $S^{-1} = \text{diag}(S_1^{-1}, S_2^{-1}, \dots, S_{\bar{n}}^{-1})$ aufzubauen. Weil A in Blockdiagonalform gegeben ist, kann $SAS^{-1} = \text{diag}(S_1 J_1 S_1^{-1}, \dots, S_{\bar{n}} J_{\bar{n}} S_{\bar{n}}^{-1})$ geschrieben werden. Wegen $J_i = \lambda_i I_i + D_i$ gilt $S_i J_i S_i^{-1} = \lambda_i S_i S_i^{-1} + S_i D_i S_i^{-1}$. Die Struktur von D_i führt auf $S_i D_i S_i^{-1} = D_i$, woraus sich $S_i J_i S_i^{-1} = J_i$ und damit $SAS^{-1} = A$ ablesen lässt.

Die Jordan-Blöcke J_i besetzen auf der Diagonalen von A jeweils die Positionen p_i bis $p_i + m_i - 1$, die durch

$$p_1 = 1 \quad \text{und} \quad p_i = p_{i-1} + m_{i-1} \quad \text{für } i = 2, 3, \dots, \bar{n} \quad (4.91)$$

gegeben sind. Mittels Vektor-Matrix-Multiplikation

$$\begin{pmatrix} c_{p_i} & c_{p_i+1} & \dots & c_{p_i+m_i-1} \end{pmatrix} := c \begin{bmatrix} O_{(p_i-1) \times m_i} \\ I_{m_i \times m_i} \\ O_{q_i \times m_i} \end{bmatrix}, \quad q_i := n - p_i + 1 - m_i \quad (4.92)$$

werden aus dem Vektor c diejenigen Elemente selektiert, die dem Jordan-Block J_i entsprechen. Die vorausgesetzte Beobachtbarkeit fordert

$$\det \begin{bmatrix} (c_{p_i} \dots c_{p_i+m_i-1}) & J_i^0 \\ (c_{p_i} \dots c_{p_i+m_i-1}) & J_i^1 \\ \vdots & \vdots \\ (c_{p_i} \dots c_{p_i+m_i-1}) & J_i^{m_i-1} \end{bmatrix} = \det \begin{bmatrix} c_{p_i} \lambda_i^0 & * & \dots & * \\ c_{p_i} \lambda_i^1 & * & \dots & * \\ \vdots & \vdots & \dots & \vdots \\ c_{p_i} \lambda_i^{m_i-1} & * & \dots & * \end{bmatrix} \neq 0 \quad (4.93)$$

und verlangt dadurch $c_{p_i} \neq 0$. Deshalb bietet S_i^{-1} die notwendigen Freiheitsgrade, um

$$\begin{aligned} \begin{pmatrix} c_{p_i} & c_{p_i+1} & \dots & c_{p_i+m_i-1} \end{pmatrix} S_i^{-1} &= \\ \begin{pmatrix} c_{p_i} s_{i,1} & c_{p_i} s_{i,2} + c_{p_i+1} s_{i,1} & c_{p_i} s_{i,3} + c_{p_i+1} s_{i,2} + c_{p_i+2} s_{i,1} & \dots \end{pmatrix} &= \\ \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{1 \times m_i} & \quad (4.94) \end{aligned}$$

zu erzielen. ■

Theorem 4.8 (Exakte Auslöschung des Regelfehlers)

Gegeben sei der Regelkreis (4.76), (4.77), dessen Sollwert durch den Generator (4.81) mit dem Anfangswert (4.82) erzeugt werde. Zu jedem beliebigen Sollwertsignal $w \in \mathcal{W}$ existiert unter der Voraussetzung

$$\sigma(A_G) \subset \sigma(A + bc) \quad (4.95)$$

ein Vektor $x_o^* \in \mathbb{R}^{n_s+n_r}$, so dass der Regelkreis (4.76) mit dem Anfangszustand $x(0) = x_o^*$ auf eine Anregung $w(\cdot)$ mit der Antwort

$$e(t) = 0 \quad \forall t \geq 0 \quad (4.96)$$

reagiert und damit den Regelfehler vollständig auslöscht.

Beweis

Wie Folgesatz 1 in [55, S. 237] zeigt, lässt sich jede quadratische Matrix mit komplexen Einträgen in Jordanform darstellen. Mit Hilfe einer komplexen Matrix $S_G \in \mathbf{GL}_l(\mathbb{C})$ wird der Signalgenerator (4.81) durch die Abbildung $\zeta(t) = S_G \eta(t)$ in Jordanform

$$\begin{aligned} \dot{\zeta}(t) &= S_G A_G S_G^{-1} \zeta(t), & \zeta(0) &= S_G \eta(0) \in \mathbb{R}^l \\ v(t) &= c_G S_G^{-1} \zeta(t) \end{aligned} \quad (4.97)$$

transformiert. Wegen Hilfssatz 4.7 bestehen bei der Wahl von S_G ausreichend Freiheitsgrade, um neben der Forderung, dass $S_G A_G S_G^{-1}$ in Jordanform vorliegt, auch die Bedingung

$$c_G S_G^{-1} = \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{1 \times l} \quad (4.98)$$

zu erfüllen.

Durch $l \in \mathbb{N}$ ist die Anzahl der Eigenwerte $\tilde{\lambda}_i \in \mathbb{C}$, $i = 1, 2, \dots, l$, gegeben, so dass die Zustandsmatrix das Spektrum

$$\sigma(A_G) = \sigma(S_G A_G S_G^{-1}) = \{\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_l\} \quad (4.99)$$

besitzt. Falls zwei natürliche Zahlen $i, j \in \{1, 2, \dots, l\}$ mit $i \neq j$ existieren, für die $\tilde{\lambda}_i = \tilde{\lambda}_j$ gilt, liegen mehrfache Eigenwerte vor. Sei durch $1 \leq \bar{l} \leq l$ die Anzahl

der verschiedenen Eigenwerte gegeben. Um für jeden verschiedenen Eigenwert einen Repräsentanten zu erhalten, werden die komplexen Zahlen

$$\tilde{\mu}_1, \tilde{\mu}_2, \dots, \tilde{\mu}_{\bar{l}} \in \sigma(A_G) = \{\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_l\} \quad (4.100)$$

so gewählt, dass für alle Tuple (i, j) mit $i, j \in \{1, 2, \dots, \bar{l}\}$ und $i \neq j$ die Aussage $\tilde{\mu}_i \neq \tilde{\mu}_j$ gilt. Die Vielfachheiten der Repräsentanten $\tilde{\mu}_i$ werden durch $m(\tilde{\mu}_i) \leq l$ bezeichnet und erfüllen $\sum_{i=1}^{\bar{l}} m(\tilde{\mu}_i) = l$.

Darüberhinaus wird für jede Zahl $i = 1, 2, \dots, \bar{l}$ ein Jordan-Block durch

$$\tilde{J}_i := \tilde{\mu}_i \in \mathbb{C} \quad \text{wenn } m(\tilde{\mu}_i) = 1 \quad (4.101)$$

beziehungsweise durch

$$\tilde{J}_i := \begin{bmatrix} \tilde{\mu}_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \tilde{\mu}_i \end{bmatrix} \in \mathbb{C}^{m(\tilde{\mu}_i) \times m(\tilde{\mu}_i)} \quad \text{wenn } m(\tilde{\mu}_i) > 1 \quad (4.102)$$

eingeführt. Die Indizierung für $\tilde{\mu}_i$ und damit gleichzeitig auch für \tilde{J}_i sei so angesetzt, dass sich

$$S_G A_G S_G^{-1} = \begin{bmatrix} \tilde{J}_1 & & \\ & \ddots & \\ & & \tilde{J}_{\bar{l}} \end{bmatrix} \quad (4.103)$$

schreiben lässt. Dabei besetzen die Jordan-Blöcke \tilde{J}_i auf der Diagonalen von $S_G A_G S_G^{-1}$ jeweils die Positionen \tilde{p}_i bis $\tilde{p}_i + m(\tilde{\mu}_i) - 1$, die durch

$$\tilde{p}_1 = 1 \quad \text{und} \quad \tilde{p}_i = \tilde{p}_{i-1} + m(\tilde{\mu}_{i-1}) \quad \text{für } i = 2, 3, \dots, \bar{l} \quad (4.104)$$

gegeben sind. Dieses Zustandsraummodell ist ein homogener Generator für den Sollwert $w(\cdot)$, weil für $\zeta(0) = \zeta_o^* := S_G \eta_o^*$

$$v(t) = w(t) \quad \text{für alle } t \geq 0 \quad \text{und } w \in \mathcal{W} \quad (4.105)$$

gilt und daher jede Funktion aus der Klasse \mathcal{W} erzeugt werden kann.

Mit einer geeignet gewählten Matrix $S \in \mathbf{G}\mathbf{I}_{n_s+n_r}(\mathbb{C})$ und der Transformation $\xi(t) = Sx(t)$ wird ebenfalls das System $(A + bc, -b, c)$ in Jordanform überführt:

$$\begin{aligned} \dot{\xi}(t) &= S(A + bc)S^{-1}\xi(t) - Sbe(t), & \xi(0) &= Sx(0) \\ y_s(t) &= cS^{-1}\xi(t) \end{aligned} \quad (4.106)$$

Im Spektrum

$$\sigma(A + bc) = \sigma(S(A + bc)S^{-1}) = \{\lambda_1, \lambda_2, \dots, \lambda_{n_s+n_r}\} \quad (4.107)$$

befinden sich $1 \leq \bar{n} \leq n_s + n_r$ verschiedene Eigenwerte, für die wiederum die Repräsentanten

$$\mu_1, \mu_2, \dots, \mu_{\bar{n}} \in \sigma(A + bc) \quad (4.108)$$

so gewählt werden, dass für alle Tuple (i, j) mit $i, j \in \{1, 2, \dots, \bar{n}\}$ und $i \neq j$ die Aussage $\mu_i \neq \mu_j$ gilt. Deren Vielfachheiten sind durch $m(\mu_i) \leq n_s + n_r$ bezeichnet, mit $\sum_{i=1}^{\bar{n}} m(\mu_i) = n_s + n_r$.

Analog zur Partitionierung der Matrix $S_G A_G S_G^{-1}$ wird ebenso $A_J := S(A + bc)S^{-1}$ in Jordan-Blöcke eingeteilt. Diese sind für alle $i = 1, 2, \dots, \bar{n}$ durch

$$J_i := \mu_i \in \mathbb{C} \quad \text{wenn } m(\mu_i) = 1 \quad (4.109)$$

beziehungsweise durch

$$J_i := \begin{bmatrix} \mu_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \mu_i \end{bmatrix} \in \mathbb{C}^{m(\mu_i) \times m(\mu_i)} \quad \text{wenn } m(\mu_i) > 1 \quad (4.110)$$

definiert und besitzen eine Indizierung in geeigneter Reihenfolge, weswegen sich

$$A_J := S(A + bc)S^{-1} = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_{\bar{n}} \end{bmatrix} \quad (4.111)$$

angeben lässt. Die Jordan-Blöcke J_i decken sich mit der Diagonalen von A_J auf den Positionen p_i bis $p_i + m(\mu_i) - 1$ mit

$$p_1 = 1 \quad \text{und} \quad p_i = p_{i-1} + m(\mu_{i-1}) \quad \text{für } i = 2, 3, \dots, \bar{n}. \quad (4.112)$$

Unter der geforderten Voraussetzung $\sigma(A_G) \subset \sigma(A_J)$ gilt für alle $j = 1, 2, \dots, \bar{l}$:

$$\tilde{\lambda}_j \in \{\lambda_1, \lambda_2, \dots, \lambda_{n_s+n_r}\} \quad (4.113)$$

Nachdem das Spektrum von A_G in demjenigen von A_J enthalten ist, finden sich sämtliche Eigenwerte $\tilde{\mu}_i$ für $i = 1, 2, \dots, \bar{l}$ in der Menge $\{\mu_1, \mu_2, \dots, \mu_{\bar{n}}\}$ wieder. Daher lässt sich festhalten, dass für jedes $i = 1, 2, \dots, \bar{l}$ ein Index $j = 1, 2, \dots, \bar{n}$ existiert, so dass $\tilde{\mu}_i = \mu_j$ mit $m(\tilde{\mu}_i) \leq m(\mu_j)$. Der Anfangszustand von (4.106) sei durch

$$\xi(0) = \xi_o^* = \begin{pmatrix} \xi_1^* \\ \vdots \\ \xi_{n_s+n_r}^* \end{pmatrix} \in \mathbb{C}^{n_s+n_r} \quad (4.114)$$

dargestellt. Für alle Indizes $i = 1, 2, \dots, \bar{l}$ werden dessen Elemente durch die Festlegung

$$\begin{aligned} \xi_{p_j}^* &:= \zeta_{\tilde{p}_i}^* \\ \xi_{p_j+1}^* &:= \zeta_{\tilde{p}_i+1}^* \\ &\vdots \\ \xi_{p_j+m(\tilde{\mu}_i)-1}^* &:= \zeta_{\tilde{p}_i+m(\tilde{\mu}_i)-1}^* \end{aligned} \quad \text{mit} \quad \begin{pmatrix} \zeta_1^* \\ \vdots \\ \zeta_{\bar{l}}^* \end{pmatrix} := \zeta_o^* = S_G \eta_o^* \quad (4.115)$$

in Übereinstimmung mit denjenigen des Anfangswertes ζ_o^* gebracht, wobei die nicht allozierten Elemente von ξ_o^* den Eintrag 0 erhalten. Die bestehende Wahlmöglichkeit für $S \in \mathbf{Gl}_{n_s+n_r}(\mathbb{C})$ soll ausgenutzt werden, damit nicht nur $A_J = S(A + bc)S^{-1}$ Jordanform erhält, sondern damit darüberhinaus auch der Vektor cS^{-1} in denjenigen Elementen mit 1 besetzt ist, für deren Positionen $p_j, p_j + 1, \dots, p_j + m(\tilde{\mu}_i) - 1$ in (4.115) eine Zuweisung erfolgt.

Ausgehend von diesem Anfangswert ξ_o^* erzeugt die Eigendynamik des homogenen Systems (4.106) eine Zustandstrajektorie $\xi(\cdot)$, die im Skalarprodukt $y_s(\cdot) = cS^{-1}\xi(\cdot)$ mit $v(\cdot) = w(\cdot)$ übereinstimmt, weshalb $y_s(\cdot) = w(\cdot)$ gilt. Durch Rücktransformation $x_o^* = S^{-1}\xi_o^*$ errechnet sich der Anfangswert, der die Aussage des Theorems erfüllt. ■

In der Praxis ist jedoch die Zielsetzung meist nicht die exakte, sondern vielmehr die asymptotische Unterdrückung des Regelfehlers. Die Fragestellung, ob ein gewählter Regler den Ausgang einer Regelstrecke mit beliebigem (und möglicherweise unbekanntem) Anfangszustand asymptotisch an ein gegebenes Sollwertsignal annähert, wird durch Theorem 4.9 gleichgesetzt mit der Frage, ob ein Anfangswert existiert, der eine exakte Unterdrückung des Regelfehlers ermöglicht. Eine Antwort hierauf bietet Theorem 4.8.

Theorem 4.9 (Asymptotische Auslöschung des Regelfehlers)

Der Regelkreis (4.76), (4.77) sei mit einem beliebigen Anfangswert $x(0) \in \mathbb{R}^{n_s+n_r}$ initialisiert und werde durch den Generator (4.81), (4.82) mit einem Sollwertsignal $w \in \mathcal{W}$ angeregt. Unter der Voraussetzung

$$\sigma(A_G) \subset \sigma(A + bc) \quad (4.116)$$

konvergiert der Ausgang des Regelkreises asymptotisch gegen 0, es gilt $e(t) \rightarrow 0$.

Beweis

Durch Variation der Konstanten lässt sich für das Anfangswertproblem (4.76) die Lösung

$$e(t) = c \exp\{At\}x(0) + c \int_0^t \exp\{A(t-\tau)\}bw(\tau)d\tau - w(t) \quad \forall t \geq 0 \quad (4.117)$$

ermitteln. Theorem 4.8 garantiert unter den Voraussetzungen $\sigma(A_G) \subset \sigma(A + bc)$ und $w \in \mathcal{W}$ die Existenz eines Vektors $x_o^* \in \mathbb{R}^{n_s+n_r}$, so dass im Sonderfall $x(0) = x_o^*$ die Aussage

$$e(t) = c \exp\{At\}x_o^* + c \int_0^t \exp\{A(t-\tau)\}bw(\tau)d\tau - w(t) = 0 \quad \forall t \geq 0 \quad (4.118)$$

gilt. Mit einem geeigneten Vektor $\Delta x_o \in \mathbb{R}^{n_s+n_r}$ kann jeder beliebige Anfangswert $x(0) \in \mathbb{R}^{n_s+n_r}$ in

$$x(0) = x_o^* + \Delta x_o \quad (4.119)$$

zerlegt werden, so dass sich die Lösung (4.117) zu

$$e(t) = \underbrace{c \exp\{At\}x_o^* + c \int_0^t \exp\{A(t-\tau)\}bw(\tau)d\tau - w(t) + c \exp\{At\}\Delta x_o}_{=0 \quad \forall t \geq 0} = c \exp\{At\}\Delta x_o \quad \forall t \geq 0 \quad (4.120)$$

ergibt. Bedingt durch $\sigma(A) \subset \mathbb{C}_-$ lässt sich $c \exp\{At\}\Delta x_o \rightarrow 0$ und damit $e(t) \rightarrow 0$ entnehmen. ■

Hinsichtlich der Fragestellung, ob die Unterdrückung einer beliebigen Störung $z(\cdot) \in \mathcal{W}$ möglich ist, wird die Strecke

$$\begin{aligned} \dot{x}_s(t) &= A_s x_s(t) + b_s u(t) + \bar{b}_s z(t), & x_s(0) &\in \mathbb{R}^{n_s} \\ y_s(t) &= c_s x_s(t) + d_s z(t), & A_s \in \mathbb{R}^{n_s \times n_s}, b_s, \bar{b}_s, c_s^T &\in \mathbb{R}^{n_s}, d_s \in \mathbb{R} \end{aligned} \quad (4.121)$$

betrachtet. Zusammen mit dem Regler (4.74), $e(t) = y_s(t)$ und $u(t) = y_r(t)$ ergibt sich der Regelkreis

$$\begin{aligned} \dot{x}(t) &= Ax(t) + \bar{b}z(t), & x(0) &\in \mathbb{R}^{n_s+n_r} \\ e(t) &= cx(t) + d_s z(t), \end{aligned} \quad (4.122)$$

wobei

$$x(t) = \begin{pmatrix} x_s(t) \\ x_r(t) \end{pmatrix}, \quad A = \begin{bmatrix} A_s + d_r b_s c_s & b_s c_r \\ b_r c_s & A_r \end{bmatrix}, \quad \bar{b} = \begin{pmatrix} d_r d_s b_s + \bar{b}_s \\ d_s b_r \end{pmatrix}, \quad c = \begin{pmatrix} c_s & 0 \end{pmatrix}. \quad (4.123)$$

Für weitere Überlegungen erfolgt eine Fallunterscheidung, abhängig vom Relativgrad der Differentialgleichung (4.122).

Fall 1: $d_s \neq 0 \Rightarrow r = 0$

Auflösen der Auskoppelgleichung in (4.122) nach $z(t)$ und Einsetzen in die Zustandsgleichung ergibt:

$$\begin{aligned} \dot{x}(t) &= \left[A - \frac{1}{d_s} \bar{b}c \right] x(t) + \frac{1}{d_s} \bar{b}e(t), & x(0) &\in \mathbb{R}^{n_s+n_r} \\ e(t) &= cx(t) + d_s z(t) \end{aligned} \quad (4.124)$$

Eine exakte Unterdrückung der Störung schlägt sich in der Forderung $e(\cdot) \equiv 0$ nieder und verlangt

$$cx(\cdot) = -d_s z(\cdot). \quad (4.125)$$

Wegen $z(\cdot) \in \mathcal{W}$ gilt

$$z(t) = v(t) = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} e^{A_G t} \eta_o^* \quad \forall t \geq 0. \quad (4.126)$$

Die Störung wirkt sich deshalb nicht auf $y_s(\cdot)$ bzw. $e(\cdot)$ aus, wenn ein geeigneter Anfangswert $x(0) \in \mathbb{R}^{n_s+n_r}$ existiert und $[A - \bar{b}c/d_s]$ eine geeignete Eigendynamik festlegt, so dass für alle $t \geq 0$ die Gleichung

$$ce^{[A - \bar{b}c/d_s]t} x(0) = \begin{pmatrix} -d_s & 0 & \dots & 0 \end{pmatrix} e^{A_G t} \eta_o^* \quad (4.127)$$

erfüllt ist.

Fall 2: $d_s = 0 \Rightarrow r \geq 1$

Unter Ausnutzung der Ergebnisse aus Abschnitt A.8 erfolgt eine Transformation des Regelkreises (4.122), (4.123) in BINF. Deren Struktur lässt sich für geeignete Konstanten g , a_i , $h_j \in \mathbb{R}$, $n := n_s + n_r$, $i = 1, \dots, n$ und $j = 1, \dots, n - r$ durch die Matrizen

$$\begin{aligned}
 A_1 &= \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ a_1 & a_2 & a_3 & \dots & a_r \end{bmatrix} \in \mathbb{R}^{r \times r}, & A_2 &= \begin{bmatrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \\ a_{r+1} & \dots & a_n \end{bmatrix} \in \mathbb{R}^{r \times (n-r)}, \\
 A_3 &= \begin{bmatrix} h_1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ h_{n-r-1} & 0 & \dots & 0 \\ h_{n-r} & 0 & \dots & 0 \end{bmatrix} \in \mathbb{R}^{(n-r) \times r}, & A_4 &\in \mathbb{R}^{(n-r) \times (n-r)}
 \end{aligned} \tag{4.128}$$

beschreiben. Die gewählte Darstellung in BINF

$$\begin{aligned}
 \begin{pmatrix} \dot{\zeta}(t) \\ \dot{\xi}(t) \end{pmatrix} &= \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{pmatrix} \zeta(t) \\ \xi(t) \end{pmatrix} + \begin{pmatrix} O_{1 \times (r-1)} \\ g \\ O_{1 \times (n-r)} \end{pmatrix} \cdot z(t), & \begin{pmatrix} \zeta(0) \\ \xi(0) \end{pmatrix} \in \mathbb{R}^n \\
 e(t) &= \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} \zeta(t) \\ \xi(t) \end{pmatrix}
 \end{aligned} \tag{4.129}$$

teilt den Regelkreis in zwei Subsysteme auf, d.h. der Zustandsvektor $(\zeta(t)^T \ \xi(t)^T)^T$ zerfällt in die beiden Anteile

$$\zeta(t) = \begin{pmatrix} e(t) & \dot{e}(t) & \dots & e^{(r-1)}(t) \end{pmatrix} \in \mathbb{R}^r \quad \text{und} \quad \xi(t) \in \mathbb{R}^{n-r}. \tag{4.130}$$

Das Problem der Störgrößenunterdrückung verlangt

$$e(t) = \dot{e}(t) = \ddot{e}(t) = \dots = e^{(r)}(t) = 0 \quad \forall t \geq 0 \tag{4.131}$$

und ist daher gleichbedeutend mit der Nullung des Ausgangs (siehe Abschnitt A.9). Damit eine Störung $z(\cdot) \in \mathcal{W}$ keine Auswirkung auf $y_s(\cdot)$ bzw. $e(\cdot)$ erhält, muss die Bedingung

$$\begin{pmatrix} a_{r+1} & \dots & a_n \end{pmatrix} \xi(t) = -gz(t) \tag{4.132}$$

für alle $t \geq 0$ erfüllt werden. Die Trajektorien $\xi(\cdot)$ ergeben sich als Lösung der Differentialgleichung

$$\dot{\xi}(t) = A_4 \xi(t) + A_3 \zeta(t), \quad \xi(0) \in \mathbb{R}^{n-1}, \tag{4.133}$$

wobei deren Eingangsgröße $\zeta(t)$ – durch die Forderung (4.131) bedingt – entfällt. Deshalb muss zur Unterdrückung der Störgröße ein geeigneter Anfangswert $\xi(0)$ existieren und gleichzeitig A_4 die Gleichung

$$\begin{pmatrix} a_{r+1} & \dots & a_n \end{pmatrix} e^{A_4 t} \xi(0) = \begin{pmatrix} -g & 0 & \dots & 0 \end{pmatrix} e^{A_G t} \eta_o^* \quad (4.134)$$

für alle $t \geq 0$ erfüllen.

Die Vorgehensweise im Beweis von Theorem 4.8 ist auf das Problem der Störgrößenunterdrückung übertragbar. Daraus lassen sich die Bedingungen $\sigma(A_G) \subset \sigma(A - \bar{b}c/d_s)$ bzw. $\sigma(A_G) \subset \sigma(A_4)$ entnehmen, unter welchen die Gleichungen (4.127) bzw. (4.134) eingehalten werden können. Diese Tatsache zeigt, dass ein Modell der Störgröße im Regelkreis enthalten sein muss.

Mit Bezug auf Funnel-Control bleibt festzuhalten, dass ein zusätzliches internes Modell keine signifikante Verbesserung bringt, falls ein schmaler Trichter eingesetzt werden kann, weil durch diese Maßnahme bereits ein kleiner Regelfehler erzwungen wird. Ein internes Modell erzielt vorwiegend dann einen positiven Effekt, wenn der Trichter eine große Mindestbreite beibehalten muss. Ein breiter Trichter kann als Regelfall für die praktische Anwendung betrachtet werden.

4.7 Festwertregelung

Aus regelungstechnischer Sicht ist die Festwertregelung ein spezieller Fall der Folgeregelung mit konstantem oder zumindest stückweise konstantem Sollwert. Dennoch soll die Festwertregelung nicht als unbedeutender Sonderfall missverstanden werden – ein erheblicher Anteil industriell geregelter Anlagen verlangt nach solchen Reglern. An dieser Stelle sind exemplarisch kontinuierliche Fertigungsanlagen zur Papiererzeugung und Weiterverarbeitung zu nennen. Prozessgrößen wie die Transportgeschwindigkeit oder die Dehnung der Papierbahn sollen zumeist auf konstanten Werten gehalten werden. Auch Antriebe in Werkzeugmaschinen stellen einen umfassenden Anwendungsbereich für Festwertregelungen dar. Ebenso treten Festwertregelungen in biologisch/chemischen Prozessen auf, für die häufig konstante Sollgrößen für Druck, Mischungsverhältnis oder Temperatur vorgegeben sind. Die Liste ließe sich beliebig fortsetzen bis hin zu mikromechatronischen Systemen wie Computerfestplatten, deren Schreib-/Lesekopf durch eine Regelung in einem konstanten Abstand zur Magnetplatte gehalten wird. Aufgrund der Bedeutung von Festwertregelungen ist diesem „Spezialfall“ hier ein gesonderter Abschnitt gewidmet.

Damit ein Prozess stationär genau¹⁾ geregelt werden kann, muss dem Prinzip des internen Modells zufolge ein integral wirkender Anteil im Regelkreis vorhanden sein. Im ungestörten Regelkreis besteht kein Unterschied, ob das integrale Verhalten im Regler oder in der Strecke lokalisiert ist. Reibungsfreie Antriebssysteme sind ein Beispiel für

¹⁾ Der Begriff „stationär genau“ ist in diesem Zusammenhang gleichbedeutend mit asymptotisch verschwindendem Regelfehler bei Vorgabe eines konstanten Sollwertes.

solche I-Strecken. Ein proportionaler Regler ist hier ausreichend, damit der Fehler (bei konstantem Sollwert) asymptotisch verschwindet. Dagegen ist unter Störeinwirkung (z.B. Reibung oder Lastmoment) die Position des I-Anteils entscheidend. Ein proportionaler Regler ist nicht in der Lage, eine integrale Strecke unter dem Einfluss einer konstanten Störung stationär genau zu regeln. Um hier einen verschwindenden Regelfehler zu erhalten, muss auf einen integralen Regler zurückgegriffen werden.

Damit bleibt das Endergebnis, dass unabhängig vom Streckentyp ohne zusätzlichem Integralanteil im Regler keine stationäre Genauigkeit unter Störeinwirkung erreicht werden kann. Bei Verwendung von Funnel-Control (Proportionalregler) bleibt daher stets ein Fehler. Es muss sich also der Trichter sehr stark verengen, um einen akzeptablen Regelfehler zu erhalten. In der Praxis liegt das Problem dabei weniger in der Tatsache, dass ein (kleiner) Regelfehler bestehen bleibt. Durch die Verengung des Trichters wird der Fehler beliebig weit reduziert und darf als tolerierbar gering gelten. Das Problem liegt vielmehr darin, dass in der Praxis der Trichter nicht beliebig schmal gewählt werden darf, weil Messrauschen nicht vollständig unterdrückt werden kann. Der Trichter muss das Rauschsignal umschließen und ist somit auf eine Mindestbreite festgelegt. Das begrenzt die Verengung des Trichters und damit auch die Verkleinerung des Regelfehlers. Abhängig von der Amplitude des Rauschens muss also ein großer Fehler in Kauf genommen werden. Aus diesem Grund ist der proportionale Regler für die praktische Anwendung nicht ausreichend und muss um einen integralen Anteil erweitert werden, wobei die Stabilität der Regelung dadurch nicht beeinträchtigt werden darf. Für die Erweiterung des Regelkreises um einen integralen Anteil bestehen zwei unterschiedliche Lösungsmöglichkeiten. Zum einen kann ein Führungsintegrator vor den Funnel-Regelkreis gesetzt werden, so dass eine Kaskadenstruktur aus zwei Regelkreisen entsteht. Diese Vorgehensweise wird standardmäßig in der linearen Regelungstechnik angewandt, vorzugsweise dann, wenn eine zustandsgeregelte Strecke asymptotisch einen vorgegebenen Betriebspunkt erreichen soll. Zum anderen kann der Integralanteil direkt in den Funnel-Regelkreis eingebunden werden. Diese Lösung ermöglicht das Zusammenfassen von Internem Modell und Regelstrecke zu einem dynamischem Gesamtsystem. Der erforderliche Stabilitätsbeweis besteht in diesem Falle darin, die Zugehörigkeit des Gesamtsystems zur Systemklasse \mathcal{P} zu zeigen.

In den folgenden Unterabschnitten werden die beiden genannten Varianten zur Einbringung des integralen Anteils näher diskutiert und verglichen.

4.7.1 Führungsintegrator/Kaskadenstruktur

Eine aus der linearen Regelungstheorie bekannte Abhilfemaßnahme gegen einen stationären Regelfehler ist die Anpassung des Sollwertes durch einen überlagerten Regelkreis. Jene kann sowohl durch einen Führungsintegrator als auch durch einen PI-Regler realisiert werden. Damit ergibt sich eine Kaskadenstruktur aus Funnel- und PI-Regler, die in Abbildung 4.7 dargestellt ist.

Für ein Zweimassensystem als Regelstrecke ist diese Variante in [28] betrachtet. Im Gegensatz zur üblichen Anwendung der Kaskadenstruktur, bei der das dynamische

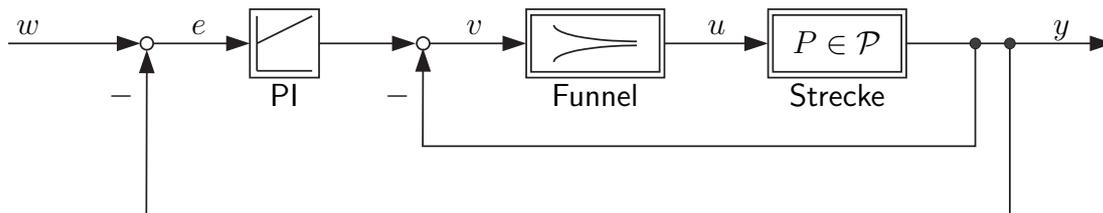


Abbildung 4.7: Zweischleifige Kaskadenstruktur: dem inneren Regelkreis – bestehend aus Strecke und Funnel-Regler – ist eine äußere Schleife mit PI-Baustein überlagert.

Verhalten des inneren Regelkreises als bekannt und meist auch als linear vorausgesetzt wird, ist im Zusammenhang mit einem zeitvarianten Regelgesetz und einer Strecke mit unsicheren Parametern diese Annahme unzutreffend. Daraus erwächst die Frage nach einer geeigneten Verstärkung für die äußere Regelschleife, mit der sich das Gesamtsystem stabil verhält.

In [28] ist der Versuch unternommen, diese Problemstellung nicht mittels zeitkontinuierlich arbeitendem Adaptionsgesetz, sondern durch eine stückweise konstante äußere Kreisverstärkung zu lösen. Ein solcher Ansatz birgt jedoch den Nachteil, dass eine sprungförmige Änderung der Verstärkung eine unzulässige sprungförmige Veränderung des Sollwertes für den inneren Regelkreis nach sich zieht. Daher ist grundsätzlich die Möglichkeit gegeben, dass der Regelfehler durch einen Sprung den Trichter verlässt, ohne dass der Regler eingreifen könnte. Dieser unerwünschte Fall muss durch zusätzliche, bisher noch nicht entwickelte Maßnahmen verhindert werden, damit ein tragfähiger Stabilitätsbeweis für die Kaskadenstruktur geführt werden kann. Desweiteren stützt sich die dort geführte Argumentation auf die Kombination aus Stellgrößenbeschränkung und stabiler Strecke im inneren Regelkreis, die selbst bei ungünstiger Einstellung des Führungsintegrators ein unbeschränktes Anwachsen der Zustandsgrößen ausschließt. Gerät jedoch die Stellgröße des inneren Regelkreises in Sättigung, besteht prinzipiell die Gefahr, dass der Fehler den Trichter verlässt.

Ohne diese Herangehensweise eingehend zu diskutieren und deren Schwachstellen zu beheben, wird diese Methode hier aufgegeben und stattdessen der Integralanteil in den Rahmen des internen Modells eingebettet.

4.7.2 Integraler Anteil im Funnel-Regelkreis

Die geeignetere von beiden Lösungsmöglichkeiten ist, um die Stabilität der Regelkreises nicht zu gefährden, den integralen Anteil als internes Modell zu betrachten, zur Regelstrecke hinzuzurechnen und damit direkt in den Funnel-Regelkreis zu implementieren. Diese Vorgehensweise ist in Abbildung 4.8 dargestellt. Dadurch bleibt weiterhin der einschleifige Regelkreis bestehen, es erfolgt eine Reihenschaltung aus internem Modell und Regler. Nachdem der PI-Baustein

$$\begin{aligned} \dot{x}_I(t) &= v(t), & x_I(0) &\in \mathbb{R} \\ u(t) &= k_I x_I(t) + k_P v(t) \end{aligned} \quad (4.135)$$

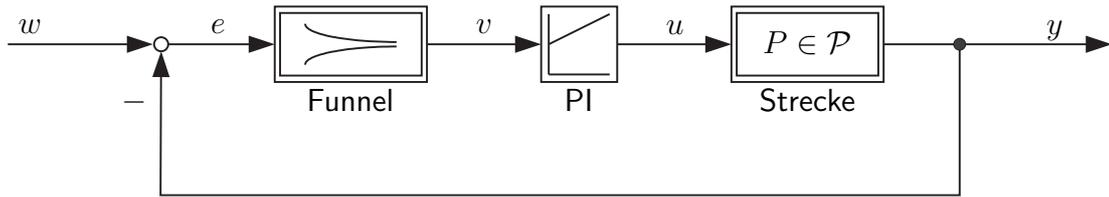


Abbildung 4.8: Einschleifige Regelung mit Reihenschaltung aus Proportionalregler und internem Modell.

einen Durchgriff mit dem Faktor k_P aufweist, besteht ein Relativer Grad von 0. Solange die Verstärkungen k_I und k_P identisches Vorzeichen besitzen, ist der Realteil der Nullstelle $q = -k_I/k_P$ negativ. Daraus folgt die Minimalphasigkeit des PI-Gliedes, die nach Satz A.30 mit der Stabilität der Nulldynamik äquivalent ist. Es ist im Einzelfall zu überprüfen, ob der PI-Baustein zur Strecke hinzugefügt werden darf, ohne deren Zugehörigkeit zu \mathcal{S} zu beeinträchtigen. Für lineare Strecken ist diese Aufgabe trivial. Es lässt sich zeigen, dass jede lineare Strecke aus \mathcal{S} durch (4.135) erweitert werden kann und dabei die notwendigen strukturellen Eigenschaften beibehält.

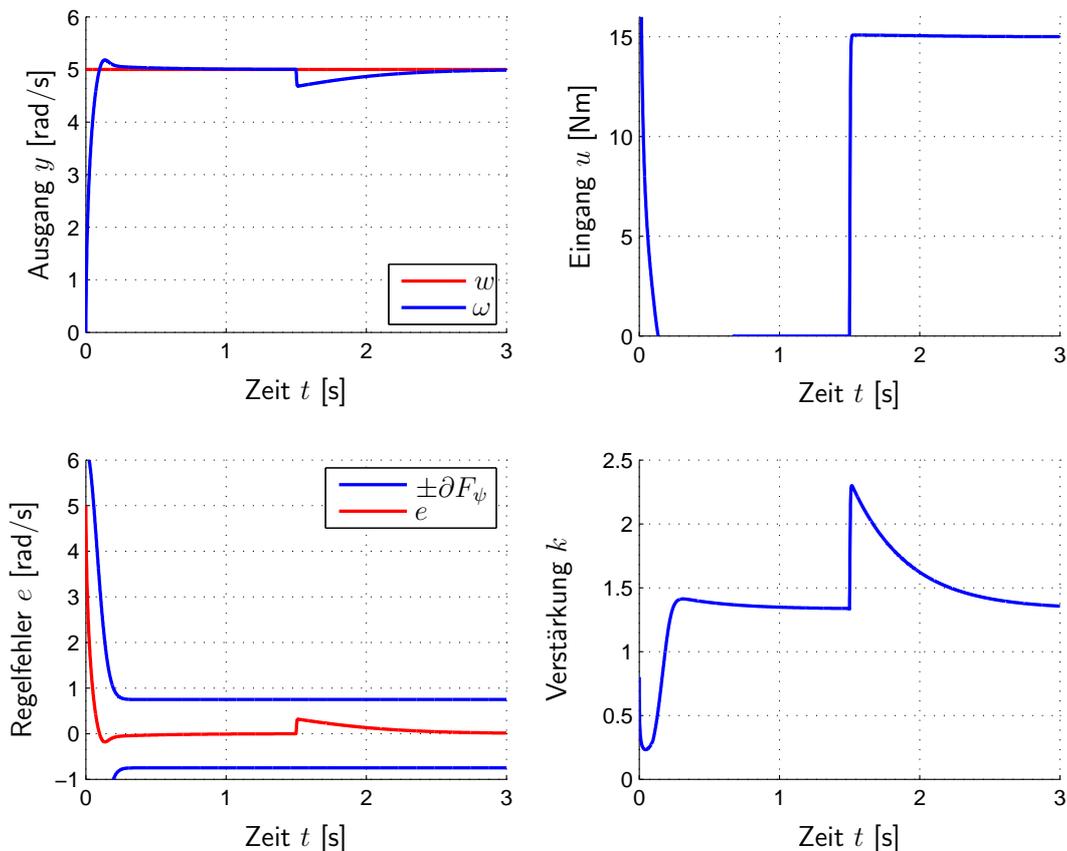


Abbildung 4.9: Hochverstärkungsbasierte Regelung mit nicht-monotoner Verstärkungsfunktion für das Einmassensystem (4.71) mit internem Modell (PI-Block, $k_P = 20$, $k_I = 50$).

Der Integrator im PI-Baustein fordert $v = 0$ als Zwangsbedingung, damit der Regelkreis in seiner Ruhelage verharrt. Das proportionale Verhalten des Reglers gestattet die Einhaltung dieser Restriktion lediglich für $e = 0$. Aus diesem Grund muss in der Gleichgewichtslage $e = w - y = 0$ gelten, es liegt daher stationäre Genauigkeit vor.

Bestätigt werden diese Überlegungen durch die Simulation in Abbildung 4.9. Hier ist die Simulation aus Abschnitt 4.5 wiederholt, wobei ein internes Modell gemäß Gleichung (4.135) mit den Parametern $k_P = 20$ und $k_I = 50$ eingesetzt ist. Die Trichterbreite ist für $t \rightarrow \infty$ durch $\Delta_F = 0.75$ rad/s gegeben. Deutlich zu erkennen ist das asymptotische Verschwinden des Regelfehlers, selbst unter dem Einfluss einer konstanten Laststörung von 15 Nm.

4.8 Zeitvariante Regelung mit internem Modell bei linearen Prozessen

Der in Abschnitt 4.1 erläuterte, zeitvariable Proportionalregler setzt wegen der Vorgabe $\psi(\cdot) \in \mathcal{L}^\infty$ einen Trichter voraus, dessen Ränder $\pm \partial F_\psi$ nicht gegen 0 konvergieren dürfen. Damit gestattet dieser Regler zwar, einem Sollwert $w(\cdot)$ mit beliebig geringer Abweichung zu folgen, asymptotisch exaktes Folgeverhalten wird dadurch aber nicht garantiert.

Die Veröffentlichung [38] bildet die Grundlage für asymptotische Folgeregelung mittels Funnel-Control. Durch die Erweiterung des proportionalen Reglers mit einem internen Modell wird das asymptotische Verschwinden des Regelfehlers erreicht, wobei sich die dort erzielten Ergebnisse auf die Kombination von linearen Modellen mit linearen Regelstrecken aus der Systemklasse \mathcal{S} beziehen. Es lässt sich zeigen, dass der Regelfehler verschwindet, weil mit $\psi(\cdot) \notin \mathcal{L}^\infty$ ein Trichter mit der Endbreite 0 verwendet werden darf [38].

Den nachfolgenden Betrachtungen liege aus der Systemklasse \mathcal{S} die lineare Regelstrecke

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), & x(0) &= x_o \in \mathbb{R}^n, & (A, b, c) &\in \mathcal{S} \\ y(t) &= cx(t), & (A, b, c^T) &\in \mathbb{R}^{n \times n} \times \mathbb{R}^n \times \mathbb{R}^n \end{aligned} \quad (4.136)$$

zugrunde, der vorgegebene Sollwertverlauf $w(\cdot)$ genüge der homogenen Differentialgleichung (4.80). Mit Hilfe des Polynoms $N : s \mapsto \sum_{i=0}^l \gamma_i s^i$ lässt sich die Gesamtheit aller Sollwertsignale, welchen die Strecke asymptotisch exakt folgen können soll, durch die Menge

$$\mathcal{W}(N) := \left\{ w \in \mathcal{C}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R}) \mid N \left(\frac{d}{dt} \right) w(\cdot) = 0 \right\}, \quad N \in \mathbb{R}[s], \quad N(s) \neq 0 \quad \forall s \in \mathbb{C}_+ \quad (4.137)$$

ausdrücken. Unbeschränkte Sollwertverläufe $w(\cdot) \notin \mathcal{L}^\infty$ werden ausgeschlossen, da für $N(\cdot)$ keine Wurzel mit positivem Realteil zugelassen ist. Für die Realisierung eines geeigneten Sollwertgenerators ist eine Matrix $\tilde{A} \in \mathbb{R}^{\nu \times \nu}$ zu wählen, deren charakteristisches Polynom mit $N(\cdot)$ übereinstimmt.

Um eine ideale Stellgröße zu erzeugen, die den Regelfehler $e(t) = y(t) - w(t)$ asymptotisch zum Verschwinden bringt, wird das interne Modell

$$\begin{aligned}\dot{\xi}(t) &= \tilde{A}\xi(t) + \tilde{b}v(t), & \xi(0) &= \xi_o \in \mathbb{R}^\nu \\ u(t) &= \tilde{c}\xi(t) + v(t), & (\tilde{A}, \tilde{b}, \tilde{c}^T) &\in \mathbb{R}^{\nu \times \nu} \times \mathbb{R}^\nu \times \mathbb{R}^\nu\end{aligned}\quad (4.138)$$

als minimalphasiges System mit Durchgriff (Relativgrad 0) angesetzt und erfüllt daher die Bedingung

$$\det \begin{bmatrix} sI_\nu - \tilde{A} & -\tilde{b} \\ \tilde{c} & 1 \end{bmatrix} \neq 0 \quad \forall s \in \mathbb{C} \text{ mit } \operatorname{Re}\{s\} \geq 0. \quad (4.139)$$

Mit einem proportionalen Regler $v(t) = -k(t)e(t)$ wird der Regelkreis geschlossen, der sich durch das Anfangswertproblem

$$\begin{aligned}\begin{pmatrix} \dot{x}(t) \\ \dot{\xi}(t) \end{pmatrix} &= \begin{bmatrix} A & b\tilde{c} \\ 0 & \tilde{A} \end{bmatrix} \begin{pmatrix} x(t) \\ \xi(t) \end{pmatrix} - k(t) \begin{pmatrix} b \\ \tilde{b} \end{pmatrix} e(t), & \begin{pmatrix} x(0) \\ \xi(0) \end{pmatrix} \in \mathbb{R}^{n+\nu} \\ y(t) &= \begin{pmatrix} c & 0 \end{pmatrix} \begin{pmatrix} x(t) \\ \xi(t) \end{pmatrix}\end{aligned}\quad (4.140)$$

darstellen lässt. Für eine abkürzende Notation werden die Größen

$$\bar{A} := \begin{bmatrix} A & b\tilde{c} \\ 0 & \tilde{A} \end{bmatrix}, \quad \bar{b} := \begin{pmatrix} b \\ \tilde{b} \end{pmatrix}, \quad \bar{c} := \begin{pmatrix} c & 0 \end{pmatrix} \quad \text{und} \quad \bar{x}(t) := \begin{pmatrix} x(t) \\ \xi(t) \end{pmatrix} \quad (4.141)$$

eingeführt.

Theorem 4.10

Wird das Stellsignal $u(\cdot)$ der linearen Regelstrecke (4.136) durch das interne Modell (4.138) generiert, dessen Eingangsgröße dem zeitvarianten, proportionalen Regelgesetz

$$v(t) = -k(t)e(t) \quad \text{mit} \quad k(t) = \frac{1}{1 - (\psi(t)e(t))^2} \quad \text{und} \quad e(t) = y(t) - w(t) \quad (4.142)$$

entspringt, ergibt sich das Anfangswertproblem (4.140). Dieses besitzt für jeden Anfangswert $(x(0)^T, \xi(0)^T)^T \in \mathbb{R}^n \times \mathbb{R}^\nu$ eine eindeutige maximale Lösung $\bar{x} : [0, T[\rightarrow \mathbb{R}^{n+\nu}$ mit $T \in]0, \infty]$, wenn folgende Voraussetzungen gegeben sind:

(i) Die Funktion $\psi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ ist absolut stetig, monoton wachsend, unbeschränkt und erfüllt $\psi(t) = 0 \Leftrightarrow t = 0$. Jeder anfängliche Fehler $e(0) \in \mathbb{R}$ liegt damit zum Zeitpunkt $t = 0$ innerhalb des Trichters.

(ii) Es gibt eine Konstante $\gamma > 1$, für welche die Funktion $\psi(\cdot)$ den Wachstumsbedingungen

$$\begin{aligned}\psi(t) &\leq \gamma \cdot \psi(t/2) & \forall t \in \mathbb{R}_+ \\ \dot{\psi}(t) &\leq \gamma \cdot [1 + \psi(t)] & \text{für fast alle } t \in \mathbb{R}_+\end{aligned}\quad (4.143)$$

genügt.

- (iii) Das Sollwertsignal $w : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$, $w \in \mathcal{W}$ ist eine Linearkombination aus Konstanten und sinusidalen Schwingungen und lässt sich als Ausgang einer homogenen linearen Differentialgleichung der Ordnung $\nu \in \mathbb{N}$ darstellen.
- (iv) Das interne Modell ist in geeigneter Weise parametrisiert und auf den Sollwertverlauf abgestimmt, d.h. N erfüllt $N(d/dt) \cdot w(\cdot) = 0$.

Für die Lösung gelten dabei folgende Eigenschaften:

- (i) Die Lösung ist für alle Zeiten $t \geq 0$ definiert, d.h. $T = \infty$.
- (ii) Der Fehler $e : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ bewegt sich für alle Zeiten $t \geq 0$ im Trichter, weil eine Zahl $\varepsilon \in]0, 1[$ existiert, die für alle $t > 0$ die Ungleichung $\psi(t)|e(t)| \leq 1 - \varepsilon$ erfüllt. Dadurch kann mit den Trichterrändern das transiente Verhalten, d.h. das gewünschte Abklingen des Fehlersignales festgelegt werden. Weil darüberhinaus $(e(t), v(t)) \rightarrow (0, 0)$ gilt, verschwindet der Fehler asymptotisch.
- (iii) Die Verstärkung $k(\cdot)$, die Stellgröße $u(\cdot)$ und die Zustände $\bar{x}(\cdot)$ sind beschränkte Funktionen.

Die Wachstumsbedingungen (4.143) schließen ein exponentielles Abklingen des Trichters aus und verbieten den Trichterrand $\psi : t \mapsto \exp\{at\} - 1$, $a > 0$. Es existiert keine reelle Zahl γ , welche auch für $t \rightarrow \infty$ der Forderung $\exp\{at\} - 1 \leq \gamma(\exp\{at/2\} - 1)$ genügt. Ein zulässiger Rand ist zum Beispiel die Funktion $t \mapsto \psi(t) = t^a$, $a \geq 1$, die für $\gamma = 2^a < \infty$ die Voraussetzungen erfüllt.

Beweis

Zunächst wird abhängig vom Sollwertverlauf $w(\cdot) \in \mathcal{W}$ eine zusammenhängende, offene Menge $\mathcal{D}_w \subset \mathbb{R}_+ \times \mathbb{R}^{n+\nu}$

$$\mathcal{D}_w := \{(t, \zeta) \in \mathbb{R}_+ \times \mathbb{R}^{n+\nu} \mid \psi(t) \cdot |\bar{c}\zeta - w(t)| < 1\} \quad (4.144)$$

definiert. Wird (4.140) als Anfangswertproblem in \mathcal{D}_w dargestellt, lässt sich in kompakter Form

$$\dot{\bar{x}}(t) = f(t, \bar{x}(t)), \quad \bar{x}(0) = \begin{pmatrix} x(0) \\ \xi(0) \end{pmatrix} \quad (4.145)$$

angeben, wobei $f : \mathcal{D}_w \rightarrow \mathbb{R}^{n+\nu}$ durch

$$f(t, \zeta) := \bar{A}\zeta - \frac{1}{1 - [\psi(t)|\bar{c}\zeta - w(t)]^2} \bar{b}[\bar{c}\zeta - w(t)] \quad (4.146)$$

gegeben ist. Als Lösung von (4.145), (4.146), (4.141) wird eine stetig differenzierbare Funktion $\bar{x} : [0, T[\rightarrow \mathbb{R}^{n+\nu}$ mit $0 < T \leq \infty$ und $(t, \bar{x}(t)) \in \mathcal{D}_w$ für alle $t \in [0, T[$ bezeichnet, die (4.145) erfüllt. Weiterhin heißt \bar{x} eindeutige maximale Lösung, wenn

$$\tilde{x} : [0, \tilde{T}[\rightarrow \mathbb{R}^{n+\nu} \quad (4.147)$$

eine Lösung von (4.145), (4.146), (4.141) ist und daraus

$$\tilde{T} \leq T \quad \text{und} \quad \bar{x}|_{[0, \tilde{T}[} = \tilde{x} \quad (4.148)$$

folgt. Da f als rechte Seite von (4.145) in \mathcal{D}_w lokal Lipschitz-stetig ist, folgt aus der Theorie gewöhnlicher Differentialgleichungen, dass für jeden Sollwertverlauf $w \in \mathcal{W}$ und für alle Anfangswerte $(x(0), \xi(0)) \in \mathbb{R}^n \times \mathbb{R}^\nu$ das Anfangswertproblem (4.145), (4.146), (4.141) eine eindeutige maximale Lösung $\bar{x} : [0, T[\rightarrow \mathbb{R}^{n+\nu}$ mit $0 < T \leq \infty$ besitzt. Falls $T < \infty$ ist, verlässt $(t, \bar{x}(t))$ jede kompakte Menge $C \subset \mathcal{D}_w$, weswegen ein Zeitpunkt $t^* \in [0, T[$ existiert, für den $(t^*, \bar{x}(t^*)) \notin C$ gilt [38, Prop. 3.1].

Das Interne Modell ist an den vorgegebenen Sollwertverlauf $w(\cdot)$ angepasst, was für alle $w \in \mathcal{W}$ die Existenz eines geeigneten Anfangswertes $\rho_o \in \mathbb{R}^{n+\nu}$ sichert, für den die homogene Differentialgleichung

$$\begin{aligned} \dot{\rho}(t) &= \bar{A}\rho(t), & \rho(0) &= \rho_o \in \mathbb{R}^{n+\nu} \\ w(t) &= \bar{c}\rho(t) \end{aligned} \quad (4.149)$$

das Referenzsignal $w(\cdot)$ generiert. Die Funktion $\rho : t \mapsto \exp\{\bar{A}t\}\rho_o$ markiert den idealen Verlauf der Zustandsgrößen in der Serienschaltung aus internem Modell und Strecke. Bewegt sich der Zustandsvektor $\bar{x}(t) = (x(t)^T, \xi(t)^T)^T$ entlang der durch $\rho(\cdot)$ vorgezeichneten Trajektorie, folgt der Ausgang der Strecke dem Referenzsignal exakt, was sich in $y(\cdot) = w(\cdot) \Leftrightarrow e(\cdot) = 0$ niederschlägt. Ziel ist daher, dass die Zustandsgrößen im geschlossenen Regelkreis asymptotisch gegen den Verlauf $\rho(\cdot)$ konvergieren.

Mit $e(t) = y(t) - w(t)$ und $x_e(t) = \bar{x}(t) - \rho(t)$ lässt sich $e(t) = \bar{c}x_e(t)$ und die Fehlerdifferentialgleichung

$$\begin{aligned} \dot{x}_e(t) &= \dot{\bar{x}}(t) - \dot{\rho}(t) = (\bar{A} - k(t)\bar{b}\bar{c})x_e(t) = \bar{A}x_e(t) - k(t)\bar{b}e(t), \\ x_e(0) &= \bar{x}(0) - \rho(0) \quad \forall t \in [0, T[\end{aligned} \quad (4.150)$$

angeben.

Durch Lemma 3.5 in [38] ist gewährleistet, dass die Kombination aus internem Modell und Strecke – beschrieben durch das Triple $(\bar{A}, \bar{b}, \bar{c})$ – Relativgrad 1, eine positive instantane Verstärkung und minimalphasiges Verhalten besitzt. Diese Eigenschaften erlauben die Anwendung von Lemma 3.4 in [38], womit eine invertierbare Matrix $S \in \mathbf{GL}_{n+\nu}(\mathbb{R})$ berechnet wird, mit deren Hilfe eine lineare Koordinatentransformation $(e(t), z(t)^T)^T = Sx_e(t)$ erfolgt. Es ergibt sich

$$\begin{pmatrix} \dot{e}(t) \\ \dot{z}(t) \end{pmatrix} = S\bar{A}S^{-1} \begin{pmatrix} e(t) \\ z(t) \end{pmatrix} - k(t)S\bar{b}e(t) \quad \forall t \in [0, T[\quad (4.151)$$

mit

$$S\bar{A}S^{-1} = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}, \quad \text{wobei } A_1 \in \mathbb{R}, \quad A_2^T, A_3 \in \mathbb{R}^{n+\nu-1}, \quad A_4 \in \mathbb{R}^{(n+\nu-1) \times (n+\nu-1)} \quad (4.152)$$

und

$$S\bar{b} = \begin{pmatrix} \bar{c}\bar{b} \\ 0 \end{pmatrix} = \begin{pmatrix} cb \\ 0 \end{pmatrix}. \quad (4.153)$$

Dadurch wird die Aufspaltung in eine zeitvariante Differentialgleichung erster Ordnung und in die lineare Nulldynamik erreicht. Daraus resultiert eine Darstellung des geschlossenen Regelkreises in Byrnes-Isidori-Normalform

$$\begin{aligned} \dot{e}(t) &= A_1 e(t) + A_2 z(t) - k(t)cb e(t) \\ \dot{z}(t) &= A_3 e(t) + A_4 z(t) \\ k(t) &= [1 - (\psi(t)e(t))^2]^{-1} \quad \forall t \in [0, T[, \end{aligned} \quad (4.154)$$

von der ausgehend der Verlauf der Nulldynamik im maximalen Existenzintervall mittels Variation der Konstanten formal angeschrieben wird:

$$z(t) = \exp\{A_4 t\}z(0) + \int_0^t \exp\{A_4(t - \tau)\}A_3 e(\tau)d\tau \quad \forall t \in [0, T[\quad (4.155)$$

Nachdem eine Lösungstrajektorie \bar{x} definitionsgemäß derart verläuft, dass $(t, \bar{x}(t)) \in \mathcal{D}_w$ für alle $t \in [0, T[$ gilt, trifft die Ungleichung

$$\psi(t)|e(t)| < 1 \quad \forall t \in [0, T[\quad (4.156)$$

zu, aus der mit $\psi(t) > 0$ für alle $t > 0$ die Aussage $|e(t)| < \infty$ für alle $t \in]0, T[$ zu folgern ist. Zusammen mit $|e(0)| < \infty$ ergibt sich daraus $e \in \mathcal{L}^\infty([0, T[; \mathbb{R})$. Wegen $e(t) = \bar{c}x_e(t)$ folgt daraus $x_e \in \mathcal{L}^\infty([0, T[; \mathbb{R}^{n+\nu})$. Bedingt durch die bestehende Minimalphasigkeit ist A_4 eine Hurwitz-Matrix, so dass $e(\cdot) \in \mathcal{L}^\infty$ die Aussage $z(\cdot) \in \mathcal{L}^\infty$ nach sich zieht.

Werden die Abkürzungen

$$q_0(t) := A_2 \exp\{A_4 t\}z(0) \quad \rightarrow \quad 0 \quad (4.157)$$

und

$$q_1(t) := A_2 \int_0^t \exp\{A_4(t - \tau)\}A_3 e(\tau)d\tau \quad \forall t \in [0, T[\quad (4.158)$$

genutzt und in (4.154) eingesetzt, entsteht die Differentialgleichung

$$\dot{e}(t) = [A_1 - k(t)cb]e(t) + q_0(t) + q_1(t) \quad \forall t \in [0, T[. \quad (4.159)$$

Nachdem der Einfluss der Größe

$$|q_0(t)| \leq \|A_2\| \cdot \|\exp\{A_4 t\}\| \cdot \|z(0)\| \leq c_1 \exp\{-\mu t\}, \quad c_1, \mu > 0 \quad (4.160)$$

wegen $\sigma(A_4) \subset \mathbb{C}_-$ exponentiell abklingt, geht die Differentialgleichung (4.159) asymptotisch in ein homogenes System über. Weil q_1 kein unabhängiger Eingang ist, sondern gemäß (4.158) als Ausgang eines dynamischen Systems mit Eingang e berechnet wird,

liegt an dieser Stelle eine Rückkopplung vor, von der gegebenenfalls eine destabilisierende Wirkung ausgehen kann. Der Schluss $k(t) > A_1/(cb) \Rightarrow e(\cdot) \in \mathcal{L}^\infty$ ist daher nicht zulässig. Im Folgenden ist zu zeigen, dass trotz Rückkopplung von e über das Signal q_1 die Stabilität der Fehlerdifferentialgleichung für große Verstärkungen gegeben ist. Wie Theorem A.42 aufzeigt, stellt die nach [38, Lem. 3.5] vorliegende Minimalphasigkeit für das Tuple (\bar{A}, \bar{c}) Detektierbarkeit sicher. Detektierbarkeit schließt aus, dass (\bar{A}, \bar{c}) ein dynamisches System mit instabilem, unbeobachtbarem Systemanteil beschreibt. Falls ein nicht beobachtbarer Unterraum des Zustandsraumes existiert, sind darin ausschließlich stabile Moden enthalten. Damit gestattet ein beschränktes Ausgangssignal den Schluss auf einen beschränkten Zustandsvektor. Unter der Voraussetzung $w \in \mathcal{L}^\infty$ lässt sich aus diesem Grund an (4.149) die Aussage $\rho \in \mathcal{L}^\infty$ ablesen. Wegen $x_e \in \mathcal{L}^\infty$ resultiert unmittelbar $\bar{x} = (x_e + \rho) \in \mathcal{L}^\infty$ und $e = y - w = (\bar{c}\bar{x} - w) \in \mathcal{L}^\infty$.

Im Zusammenhang mit der zeitvarianten Einstellung der Verstärkung ist zu klären, ob die Funktion $k(\cdot)$ beschränkt bleibt. Diese Fragestellung wird durch einen Widerspruchsbeweis beantwortet, dem die Annahme $k(\cdot) = [1 - (\psi(\cdot)e(\cdot))]^{-1} \notin \mathcal{L}^\infty$ zugrunde gelegt wird.

Eine streng monoton wachsende Folge (k_j) mit $(k_j) \rightarrow \infty$ für $j \rightarrow \infty$ und $k_0 > 1$ definiert analog zum Beweis von Theorem 4.6 für jede Zahl $j \in \mathbb{N}$ zwei Intervallgrenzen:

$$\tau_j := \inf \{t \in [0, T[\mid k(t) = k_{j+1}\} \quad (4.161)$$

$$\sigma_j := \sup \{t \in [0, \tau_j] \mid k(t) = k_j > 1\} \quad (4.162)$$

Aus dieser Festlegung geht die Aussage

$$k(\tau_j) = k_{j+1} > k_j = k(\sigma_j) \quad \forall j \in \mathbb{N} \quad (4.163)$$

hervor, zu der – basierend auf der Annahme $k(\cdot) \notin \mathcal{L}^\infty$ – in drei Teilschritten ein Widerspruch aufgestellt wird.

Schritt 1: zeitliche Ableitung von k

Die zeitliche Ableitung der Verstärkung lässt sich durch

$$\begin{aligned} \frac{d}{dt}k(t) &= \frac{1}{[1 - (\psi(t)e(t))^2]^2} 2\psi(t)e(t) \frac{d}{dt}(\psi(t)e(t)) = \\ &= k(t)^2 2\psi(t)e(t) \left(\dot{\psi}(t)e(t) + \psi(t)\dot{e}(t) \right) = \\ &= k(t)^2 \left[2\psi(t)\dot{\psi}(t)e(t)^2 + 2\psi(t)^2e(t)\dot{e}(t) \right] \quad \forall t \in [0, T[\end{aligned} \quad (4.164)$$

berechnen. Wird darüberhinaus Gleichung (4.159) und $q_2(t) := q_0(t) + q_1(t) + A_1e(t)$ eingesetzt, ergibt sich

$$\frac{d}{dt}k(t) = k(t)^2 \left[2\psi(t)\dot{\psi}(t)e(t)^2 + 2\psi(t)^2e(t)q_2(t) - 2\psi(t)^2k(t)cbe(t)^2 \right] \quad \forall t \in [0, T[. \quad (4.165)$$

Die Wachstumsbedingungen (4.143) setzen abhängig von $\psi(t)$ eine obere Grenze für $\dot{\psi}(t)$ und erlauben mit $\psi(t)|e(t)| < 1$ die Ungleichung

$$\psi(t)\dot{\psi}(t)e(t)^2 \leq \psi(t)\gamma[1 + \psi(t)]e(t)^2 < \gamma[1 + \psi(t)]|e(t)| < \gamma[|e(t)| + 1] \quad (4.166)$$

anzuschreiben. Wegen $e \in \mathcal{L}^\infty$ gibt es eine reelle Konstante $c_2 > 0$, die den Term $\gamma[|e(t)| + 1]$ beschränkt und dadurch

$$\psi(t)\dot{\psi}(t)e(t)^2 \leq c_2 \quad \forall t \in [0, T[\quad (4.167)$$

garantiert. Mit (4.160) und $|q_2(t)| \leq |q_0(t)| + |q_1(t)| + |A_1||e(t)|$ lässt sich für den Ausdruck $\psi(t)^2 e(t) q_2(t)$ durch

$$\begin{aligned} \psi(t)^2 e(t) q_2(t) &\leq \psi(t)^2 |e(t)| (|q_0(t)| + |q_1(t)| + |A_1||e(t)|) \leq \\ &\leq \psi(t) |q_0(t)| + \psi(t) |q_1(t)| + |A_1| \leq \\ &\leq c_1 \psi(t) \exp\{-\mu t\} + \psi(t) |q_1(t)| + |A_1| \quad \forall t \in [0, T[\end{aligned} \quad (4.168)$$

eine obere Grenze finden. Für $|q_1(t)|$ wird ausgehend von Gleichung (4.158) eine Schranke durch die Berechnung

$$\begin{aligned} |q_1(t)| &\leq \|A_2\| \int_0^t \|e^{A_4(t-\tau)}\| \cdot \|A_3\| \cdot |e(\tau)| d\tau \leq \\ &\leq \|A_2\| \int_0^{\frac{t}{2}} e^{-\mu(t-\tau)} d\tau \|A_3\| \max_{\tau \in [0, t/2]} |e(\tau)| + \\ &\quad + \|A_2\| \int_{\frac{t}{2}}^t e^{-\mu(t-\tau)} d\tau \|A_3\| \max_{\tau \in [t/2, t]} |e(\tau)| = \\ &= \|A_2\| \frac{1}{\mu} \left[e^{-\mu(t-\tau)} \right]_0^{\frac{t}{2}} \|A_3\| \max_{\tau \in [0, t/2]} |e(\tau)| + \\ &\quad + \|A_2\| \frac{1}{\mu} \left[e^{-\mu(t-\tau)} \right]_{\frac{t}{2}}^t \|A_3\| \max_{\tau \in [t/2, t]} |e(\tau)| = \\ &= \frac{\|A_2\| \cdot \|A_3\|}{\mu} \left[(e^{-\mu t/2} - e^{-\mu t}) \max_{\tau \in [0, t/2]} |e(\tau)| + (1 - e^{-\mu t/2}) \max_{\tau \in [t/2, t]} |e(\tau)| \right] \leq \\ &\leq \frac{\|A_2\| \cdot \|A_3\|}{\mu} \left[e^{-\mu t/2} \max_{\tau \in [0, t/2]} |e(\tau)| + \max_{\tau \in [t/2, t]} |e(\tau)| \right] \quad \forall t \in [0, T[\quad (4.169) \end{aligned}$$

ermittelt. Eingesetzt in (4.168) ergibt sich mit $c_3 := (\|A_2\| \cdot \|A_3\|)/\mu$ die Ungleichung

$$\begin{aligned} \psi(t)^2 e(t) q_2(t) &\leq c_1 \psi(t) e^{-\mu t} + \psi(t) c_3 \left[e^{-\mu t/2} \max_{\tau \in [0, t/2]} |e(\tau)| + \max_{\tau \in [t/2, t]} |e(\tau)| \right] + |A_1| \\ &\quad \forall t \in [0, T[\quad (4.170) \end{aligned}$$

in der $\psi(t)$ als Faktor auftritt. Da die Funktion $\psi(\cdot)$ über alle Grenzen wächst, ist zu prüfen, ob die jeweiligen Produkte für alle $t \in [0, T[$ beschränkt bleiben.

Aufgrund der Wachstumsbeschränkung (4.143) nimmt $\psi(\cdot)$ nicht mit exponentieller Rate zu. Daher sind die Summanden $m_1 : t \mapsto \psi(t)e^{-\mu t}$ und $m_2 : t \mapsto \psi(t)e^{-\mu t/2}$ beschränkt. Für $\psi(t) \max_{\tau \in [t/2, t]} |e(\tau)|$ dagegen ist eine weitere Umformung nötig. Für die monoton steigende Funktion $\psi(\cdot)$ gilt $\min_{\tau \in [t/2, t]} \psi(\tau) = \psi(t/2)$. Zusammen mit (4.156) folgt

$$|e(t)| < \frac{1}{\psi(t)} \quad \Rightarrow \quad \max_{\tau \in [t/2, t]} |e(\tau)| < \frac{1}{\psi(t/2)}. \quad (4.171)$$

Daraus erklärt sich die Zerlegung des Integrales in (4.169) auf zwei Teilintervalle. Eine sinnvolle Abschätzung wird nur erreicht, wenn das Maximum über einem streng positiven Intervall gesucht wird, da andernfalls wegen $\psi(0) = 0$ die triviale Aussage $\max_{\tau \in [0, t]} |e(\tau)| < 1/\psi(0) = \infty$ resultieren würde.

Aus (4.171) lässt sich $\psi(t) \max_{\tau \in [t/2, t]} |e(\tau)| < \psi(t)/\psi(t/2)$ schreiben, was mit (4.143) auf

$$\psi(t) \max_{\tau \in [t/2, t]} |e(\tau)| < \frac{\psi(t)}{\psi(t/2)} \leq \gamma < \infty \quad (4.172)$$

führt. Einsetzen der berechneten Abschätzungen in die Ableitung der Verstärkung liefert den Ausdruck

$$\frac{d}{dt} k(t) \leq 2k(t)^2 \left[c_2 + c_1 m_1(t) + c_3 m_2(t) \max_{\tau \in [0, t/2]} |e(\tau)| + c_3 \gamma + |A_1| - \psi(t)^2 k(t) c b e(t)^2 \right]. \quad (4.173)$$

Die beschränkten Größen werden durch eine Konstante $K_1 \in \mathbb{R}$ mit

$$K_1 \geq c_2 + c_1 m_1(t) + c_3 m_2(t) \max_{\tau \in [0, t/2]} |e(\tau)| + c_3 \gamma + |A_1| \quad \forall t \in [0, T[\quad (4.174)$$

zusammengefasst, woraus schließlich

$$\frac{d}{dt} k(t) < 2k(t)^2 \left[K_1 - c b (\psi(t) e(t))^2 k(t) \right] \quad \forall t \in [0, T[\quad (4.175)$$

als Abschätzung für die Ableitung von $k(t)$ resultiert.

Schritt 2: es gibt eine Konstante $\eta > 0$, so dass $\psi(t)|e(t)| \geq \eta > 0$ für alle $t \in [\sigma_j, \tau_j]$. Wegen $k_0 > 1$ definiert die Wurzel

$$\eta := \sqrt{1 - \frac{1}{k_0}} > 0 \quad (4.176)$$

eine positive Zahl $\eta > 0$. Durch Umstellen wird daraus der Zusammenhang $k_0 = 1/(1 - \eta^2)$ gewonnen. Aufgrund der Monotonie der Folge (k_j) gilt für alle $j \geq 0$ die Ungleichung $k_j \geq k_0$. Folglich lässt sich im Intervall $[\sigma_j, \tau_j]$

$$k(t) = \frac{1}{1 - (\psi(t)e(t))^2} \geq k_j \geq k_0 = \frac{1}{1 - \eta^2} \quad \forall t \in [\sigma_j, \tau_j] \quad (4.177)$$

angeben. Daraus ist

$$\psi(t)|e(t)| \geq \eta > 0 \quad \forall t \in [\sigma_j, \tau_j] \quad (4.178)$$

abzulesen. Diese Ungleichung zeigt, dass die Funktion $(\psi(\cdot)|e(\cdot)|)^2$ im Intervall $[\sigma_j, \tau_j]$ von 0 wegbeschränkt ist.

Schritt 3: $\dot{k}(t) \leq 0$, wenn $k(t)$ hinreichend groß

Um einen Widerspruch zur Annahme aufzustellen, ist die Aussage

$$\exists j^* \in \mathbb{N} : \dot{k}(t) \leq 0 \quad \forall t \in [\sigma_{j^*}, \tau_{j^*}] \quad (4.179)$$

zu beweisen. Aus Ungleichung (4.175) ist ersichtlich, dass hierzu eine entsprechend große Verstärkung

$$k(t) \geq \frac{K_1}{(\psi(t)e(t))^2 cb} \quad (4.180)$$

hinreichend ist, weil nach Schritt 3 das Produkt $(\psi(\cdot)e(\cdot))^2$ von 0 wegbeschränkt ist. Daher erreicht der Term $(\psi(t)e(t))^2 k(t)$ beliebig hohe Werte, wenn die Annahme $k(\cdot) \notin \mathcal{L}^\infty$ zutrifft.

Wird eine Zahl j^* hinreichend groß gewählt, garantiert die Monotonie der Folge (k_j) , dass $k_{j^*} = k(\sigma_{j^*})$ die Bedingung

$$k(t) \geq k(\sigma_{j^*}) > \frac{K_1}{\eta^2 cb} \geq \frac{K_1}{(\psi(t)e(t))^2 cb} \quad \forall t \in [\sigma_{j^*}, \tau_{j^*}] \quad (4.181)$$

erfüllt. Eingesetzt in (4.175) folgt schließlich

$$\frac{d}{dt} k(t) \leq 0 \quad \forall t \in [\sigma_{j^*}, \tau_{j^*}] \quad \Rightarrow \quad k_{j^*+1} = k(\tau_{j^*}) \leq k(\sigma_{j^*}) = k_{j^*} \quad (4.182)$$

und lässt damit ein Widerspruch zur Aussage (4.163) entstehen. Als Konsequenz resultiert daraus $k(\cdot) \in \mathcal{L}^\infty$.

Weil $k(\cdot)$ beschränkt bleibt, gibt es eine Zahl $\varepsilon > 0$, die für alle $t \in [0, T[$ die Ungleichung $\psi(t)|e(t)| \leq 1 - \varepsilon$ erfüllt.

Das Regelgesetz $v(t) = -k(t)e(t)$ liefert eine beschränkte Funktion $v \in \mathcal{L}^\infty([0, T[; \mathbb{R})$, da sowohl $k(\cdot)$ als auch $e(\cdot)$ auf dem Intervall $[0, T[$ beschränkt sind. Ebenfalls ist die Stellgröße $u \in \mathcal{L}^\infty([0, T[; \mathbb{R})$ beschränkt, was sich aus $v \in \mathcal{L}^\infty([0, T[; \mathbb{R})$ und der Tatsache erschließt, dass \tilde{A} in (4.138) als Hurwitz-Matrix vorausgesetzt ist.

Um das Existenzintervall der Lösung auf die positive reelle Achse auszudehnen, wird zum Aufbau eines Widerspruches $T < \infty$ angenommen. Wegen $x_e \in \mathcal{L}^\infty([0, T[; \mathbb{R}^{n+\nu})$ kann eine Konstante $K_2 > 0$ eingeführt werden, die $\|x_e(t)\| \leq K_2$ für alle $t \in [0, T[$ erfüllt. Wird

$$C := \{(t, \zeta) \in \mathcal{D}_w \mid \psi(t) \cdot |\bar{c}\zeta - w(t)| \leq 1 - \varepsilon, \|\zeta\| \leq K_2, t \in [0, T[\} \quad (4.183)$$

definiert, stellt C eine kompakte Untermenge von \mathcal{D}_w mit der Eigenschaft $(t, \bar{x}(t)) \in C$ für alle $t \in [0, T[$ dar und widerspricht damit [38, Prop. 3.1], die unter der Annahme

$T < \infty$ die Existenz eines Zeitpunktes $t^* \in [0, T[$ mit $(t^*, \bar{x}(t^*)) \notin C$ fordert. Folglich gilt $T = \infty$.

Unter der Voraussetzung einer unbeschränkten, monoton wachsenden Funktion $\psi(\cdot)$ ergibt sich

$$|e(t)| < 1/\psi(t) \rightarrow 0 \quad \text{wenn } t \rightarrow \infty \quad (4.184)$$

und zeigt, dass asymptotisch exaktes Folgeverhalten erreicht wird. Zusammen mit $k \in \mathcal{L}^\infty$ ist unmittelbar die Aussage $v(t) = -k(t)e(t) \rightarrow 0$ für $t \rightarrow \infty$ zu entnehmen. ■

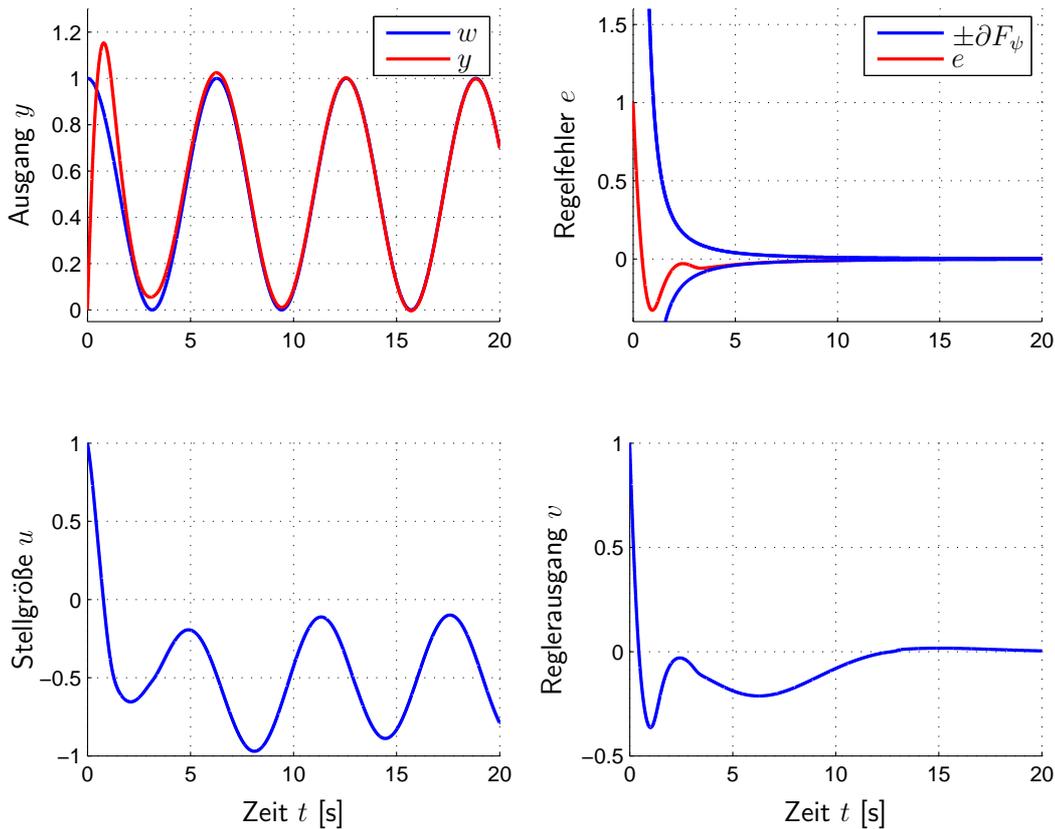


Abbildung 4.10: Hochverstärkungsbasierte Regelung mit internem Modell und asymptotisch verschwindendem Trichter.

Um die Wirksamkeit eines internen Modells zu demonstrieren, sei exemplarisch die Regelstrecke²⁾

$$F(s) = \frac{3s^2 + 2s + 1}{s^3 + s^2 - 1} \quad (4.185)$$

betrachtet, die dem Sollwertverlauf

$$w : t \mapsto \frac{1 + \cos(t)}{2} \quad \circlearrowleft \mathcal{L} \bullet \quad \frac{1}{2s} + \frac{s}{2(s^2 + 1)} = \frac{s^2 + 0.5}{s(s + j)(s - j)} \quad (4.186)$$

folgen soll. Im Sollwert enthalten ist ein konstanter Anteil mit der Frequenz 0 und eine Kosinusschwingung mit der Frequenz 1 rad/s. Der Nenner des Laplace-transformierten

²⁾ Die Darstellung der Regelstrecke als Übertragungsfunktion setzt implizit den Anfangszustand 0 voraus.

Sollsignales drückt dies durch die Polstellen bei $s = 0$ und $s = \pm j$ aus und legt dadurch das Polynom $N \in \mathbb{R}[s]$ zu $N : s \mapsto s^3 + s$ fest. Wegen

$$w^{(3)}(t) = -\dot{w}(t) = \frac{1}{2} \sin(t) \quad (4.187)$$

ist für den vorliegenden Sollwert $N \left(\frac{d}{dt} \right) w(\cdot) = 0$ erfüllt, so dass $w(\cdot) \in \mathcal{W}(N)$ zutrifft. Durch $N(\cdot)$ ist das Nennerpolynom des internen Modells bereits vorgegeben, da dessen Eigendynamik das Signal $w(\cdot)$ reproduzieren muss. Um Forderung (4.139) zu erfüllen, muss als Zähler ein Hurwitzpolynom dritter Ordnung gewählt werden. Zweckmäßig ist beispielsweise ein Modell mit der Übertragungsfunktion

$$F_{IM}(s) = \frac{(s+1)(s+0.75)(s+0.5)}{s(s+j)(s-j)}. \quad (4.188)$$

Der asymptotisch gegen 0 konvergierende Trichterrand ist Einschränkung (4.143) unterworfen, die von der absolut stetigen Funktion

$$\psi : t \mapsto t^2 \quad (4.189)$$

jedoch für die Konstante $\gamma = 4$ eingehalten wird. In Abbildung 4.10 ist ersichtlich, dass asymptotisch der Regelfehler abklingt. Folglich verschwindet auch der Ausgang v des Proportionalreglers, der gleichzeitig die Eingangsgröße für das interne Modell darstellt. Angeregt durch das Signal $v(\cdot)$ bewegt sich die Eigendynamik des Modells in einer Weise, dass trotz $v \rightarrow 0$ die Stellgröße $u(t)$ gegen ihren Idealverlauf strebt und diesem dauerhaft folgt. Indem ein Signal $u(\cdot)$ generiert wird, das $e(t) \rightarrow 0$ erreicht, übernimmt das interne Modell asymptotisch die Steuerung der Strecke und setzt für $t \rightarrow \infty$ den Regler außer Kraft.

Kapitel 5

Einfluss hoher Relativgrade

5.1 Die Problematik hoher Relativgrade

Die beiden Kapitel 3 und 4 befassen sich mit diversen Varianten der hochverstärkungs-basierten Regelung und beschränken sich dabei auf Regelstrecken mit Relativgrad 1. Um die bei höheren Relativgraden entstehende Problematik anschaulich darzustellen, sei zunächst eine lineare Regelstrecke betrachtet. Für nichtlineare Strecken gilt Ähnliches, auch wenn die verwendete Argumentationsweise mittels Wurzelortskurven in diesen Fällen nicht greift.

Gehört die Regelstrecke der Systemklasse \mathcal{S} an, geht von hohen Rückführverstärkungen im Regelkreis keine destabilisierende Wirkung aus und die in Abschnitt 3.1 aufgezählten Eigenschaften liegen vor. Im Allgemeinen kann jedoch nicht erwartet werden, dass große Verstärkungen grundsätzlich ohne nachteilige Folgen bleiben. Besonders bei Strecken mit höheren Relativgraden werden bei zu großen Kreisverstärkungen Stabilitätsprobleme verursacht, wie anhand von Wurzelortskurven gezeigt werden kann. Wegen der Verwendung linearer Analysemethoden erhalten die gewonnenen Aussagen über die Position der Polstellen ausschließlich für konstante Verstärkungen Bedeutung, da nur dann ein linearer, zeitinvarianter Regelkreis vorliegt. Die Betrachtung einer konstanten Verstärkung ist allerdings nicht realitätsfern und wird durch Theorem 3.7 legitimiert, weil im Falle einer monotonen Adaption asymptotisch ein konstanter, endlicher Grenzwert $\lim_{t \rightarrow \infty} k(t) = k_\infty < \infty$ erreicht wird. Somit nimmt der Regelkreis (nachdem die Strecke hier linear ist) asymptotisch lineares Verhalten an. Aussagen über das transiente Verhalten müssen dagegen auf andere Weise getroffen werden.

Um den Einfluss hoher Relativgrade auf die Stabilität des Regelkreises zu veranschaulichen, wird eine minimalphasige Strecke in Linearfaktor-Form

$$F_S(s) = \frac{y(s)}{u(s)} = \frac{Z(s)}{N(s)} = g \frac{(s - q_1)(s - q_2) \cdots (s - q_m)}{(s - p_1)(s - p_2) \cdots (s - p_n)}, \quad \begin{array}{l} q_1, \dots, q_m \in \mathbb{C}_-, \\ p_1, \dots, p_n \in \mathbb{C}, \\ g > 0 \end{array} \quad (5.1)$$

zugrunde gelegt, die sich aus den $m \in \mathbb{N}$ Nullstellen q_1, \dots, q_m und den $n \in \mathbb{N}$ Polen p_1, \dots, p_n zusammensetzt. Sowohl das Zählerpolynom $Z \in \mathbb{R}[s]$ als auch das Nennerpo-

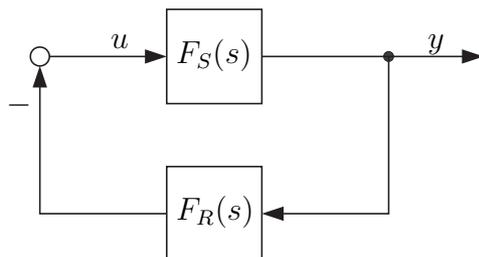


Abbildung 5.1: Regelkreis zur Gewinnung der Wurzelortskurve.

lynom $N \in \mathbb{R}[s]$ der Strecke sei auf reelle Koeffizienten beschränkt, weshalb komplexe Nullstellen und Pole ausschließlich konjugiert komplex auftreten können. Der Begriff „Relativgrad“ bezeichnet den Polüberschuss, so dass bei den nachfolgenden Betrachtungen Strecken mit $n > m$ im Fokus stehen. Durch negative Rückkopplung des Ausgangs y einer solchen Strecke gemäß Abbildung 5.1 mit einem proportionalen Regler

$$-F_R(s) = \frac{u(s)}{y(s)} = -k, \quad k \in \mathbb{R} \quad (5.2)$$

entsteht ein Regelkreis mit der charakteristischen Gleichung $1 + kF_S(s) = 0$. Wird die Verstärkung k von 0 bis ∞ variiert, zeichnen die n Wurzeln der charakteristischen Gleichung Trajektorien in der komplexen s -Ebene, die in ihrer Gesamtheit die Wurzelortskurve (WOK) bilden. Deren Bedeutung für die Regelungstechnik leitet sich daraus ab, dass die Wurzelortskurve den geometrischen Ort der Pole des geschlossenen Regelkreises angibt. Ohne im Detail auf die Eigenschaften von Wurzelortskurven einzugehen – ausführliche Herleitungen und Beweise finden sich z.B. in [59, S. 415ff], [79, S. 157ff] und [23, S. 200ff] – sei hier lediglich die Symmetrie zur reellen Achse genannt und der prinzipielle Verlauf skizziert.

Für die Verstärkung $k = 0$ ist der Regler außer Kraft gesetzt, der Rückführzweig entfällt. Der Regelkreis besteht damit aus der unregelmäßigten Strecke und zeigt das durch $F_S(s)$ vorgegebene Verhalten. Aus diesem Grund entspringen die n Äste der WOK in den n Polstellen p_1, \dots, p_n der Regelstrecke.

Wie in Abschnitt 3.2 ausgeführt, zieht für $k \rightarrow \infty$ jede Übertragungsnullstelle gemäß ihrer Vielfachheit eine entsprechende Anzahl von Polstellen des geschlossenen Regelkreises an. Deshalb münden genau m Äste der WOK in die vorhandenen m Nullstellen q_1, \dots, q_m der Regelstrecke. Die restlichen $r = n - m$ Äste streben mit steigender Kreisverstärkung ins Unendliche und konvergieren asymptotisch gegen Geraden, die sich im Wurzelschwerpunkt

$$W = \frac{\sum_{j=1}^n p_j - \sum_{i=1}^m q_i}{n - m} \quad (5.3)$$

auf der reellen Achse schneiden. Die Winkel der Asymptoten zur positiven reellen Achse sind durch

$$\phi_\nu = \frac{(2\nu - 1)\pi}{r}, \quad \nu = 1, 2, \dots, r, \quad r = n - m \quad (5.4)$$

gegeben.

Für Strecken mit Relativgrad 1 (d.h. $r = n - m = 1$) ergibt sich ein Asymptotenwinkel $\phi_1 = \pi$, die Bahnkurve der überzähligen Polstelle verläuft entlang der negativen reellen Achse. Dadurch besitzen sämtliche Pole einer minimalphasigen Strecke aus der Klasse \mathcal{S} negative Realteile, wenn die konstante Kreisverstärkung einen Mindestwert überschreitet. Eine weitere Vergrößerung der Verstärkung wird die negativen Realteile der Polstellen beibehalten, es kann aus diesem Grund keine Destabilisierung durch eine zu hohe Verstärkung auftreten. Abbildung 5.2 enthält die Wurzelortskurven für zwei Beispielsysteme mit Relativgrad 1 und stellt den beschriebenen Sachverhalt graphisch dar.

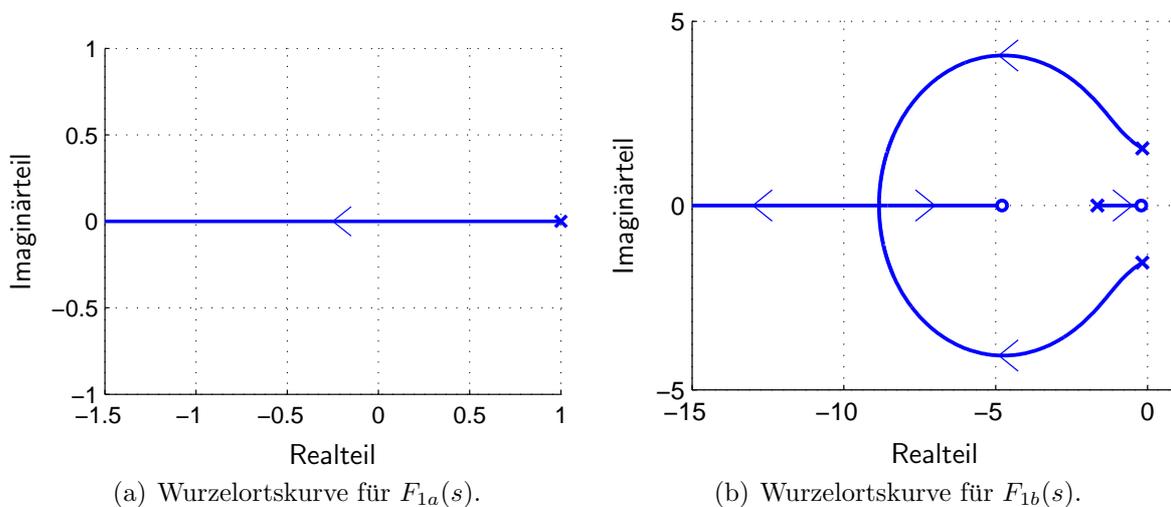


Abbildung 5.2: Wurzelortskurven für die beiden Beispielsysteme $F_{1a}(s) = 1/(s - 1)$ und $F_{1b}(s) = (s^2 + 5s + 1)/(s^3 + 2s^2 + 3s + 4)$.

Für Regelstrecken mit Relativgrad 2 kann es durchaus zu einem instabilen Kreis kommen, wenn die Verstärkung zu groß gewählt wird. Aus Gleichung (5.4) errechnen sich die beiden Asymptotenwinkel $\phi_1 = \pi/2$ und $\phi_2 = 3\pi/2$. Das entsprechende konjugiert komplexe Polpaar wird demnach parallel zur imaginären Achse gegen unendlich streben, eine Verschlechterung der Dämpfung ist die Folge. Der Realteil dieser beiden Pole nähert sich den senkrechten Asymptoten und konvergiert gegen den Wurzelschwerpunkt. Dessen Vorzeichen, das von den (unbekannten) Prozessparametern abhängt, bestimmt daher für $k \rightarrow \infty$ die Stabilität des Regelkreises. Zum Beispiel für das Zweimassensystem mit den Parametern aus Tabelle 2.2 liegt der Wurzelschwerpunkt in der rechten Halbebene. Obwohl die Asymptoten parallel zur imaginären Achse verlaufen und dadurch nicht in die instabile Halbebene weisen, darf dennoch die Kreisverstärkung nicht beliebig ansteigen. Allein aus der Richtung der Asymptoten kann nicht auf Stabilität geschlossen werden. Eine exemplarische Wurzelortskurve für ein solches Zweimassensystem mit positivem Wurzelschwerpunkt bei $s = 1$ ist in Abbildung 5.3(a) aufgetragen. Das konjugiert komplexe Polpaar wird in die instabile Halbebene verschoben, sobald die Verstärkung den kritischen Wert bei $k \approx 204$ übersteigt. Gleichzeitig zeigt Abbildung 5.3(b) den Fall für einen Wurzelschwerpunkt mit negativem Vorzeichen bei $s = -1$. Hier bleibt der Regelkreis auch für $k \rightarrow \infty$ stabil.

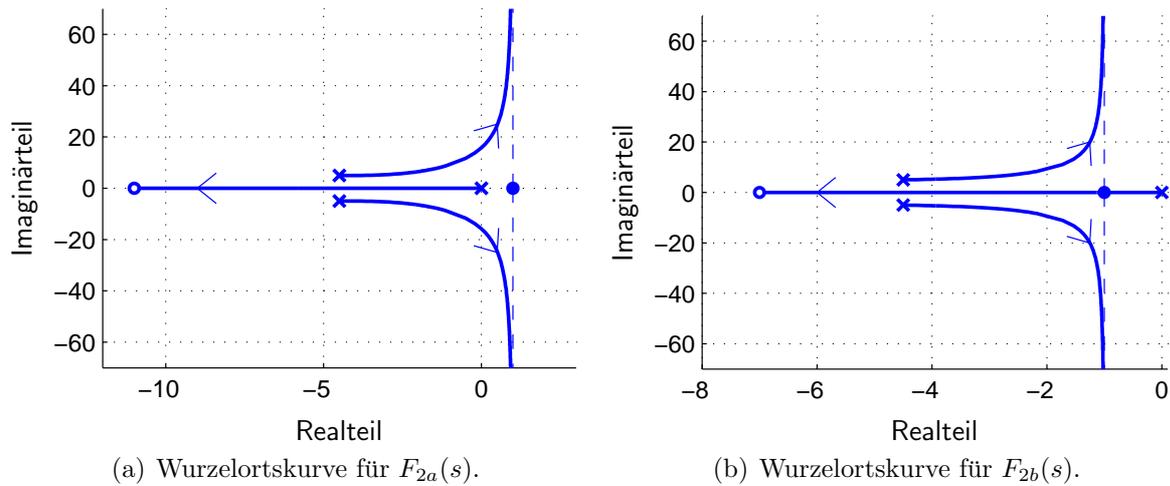


Abbildung 5.3: Wurzelortskurven für die beiden Beispielsysteme

$F_{2a}(s) = (s + 11)/(s(s + 4.5 + 5j)(s + 4.5 - 5j))$ mit Wurzelschwerpunkt bei $s = +1$ und $F_{2b}(s) = (s + 7)/(s(s + 4.5 + 5j)(s + 4.5 - 5j))$ mit Wurzelschwerpunkt bei $s = -1$.

Bei Systemen, deren Relativgrad den Wert 2 übersteigt, spielt die Lage des Wurzelschwerpunktes keine Rolle mehr, da der Winkel zwischen positiver reeller Achse und mindestens einer der Asymptoten kleiner als $\pi/2$ ist. Aus Symmetriegründen weisen deshalb mindestens zwei Asymptoten in die instabile rechte Halbebene. Es tritt demzufolge stets instabiles Verhalten für zu hohe Verstärkungen auf. Auch aus Abbildung 5.4 ist zu entnehmen, dass im Fall $r > 2$ die Kreisverstärkung nicht beliebig ansteigen darf. Wegen der gewählten Relativgrade $r = 3$ bzw. $r = 4$ besitzen die strichliert eingetragenen Asymptoten Winkel von $\pm\pi/3$ und π bzw. $\pm\pi/4$ und $\pm 3\pi/4$ zur reellen Achse.

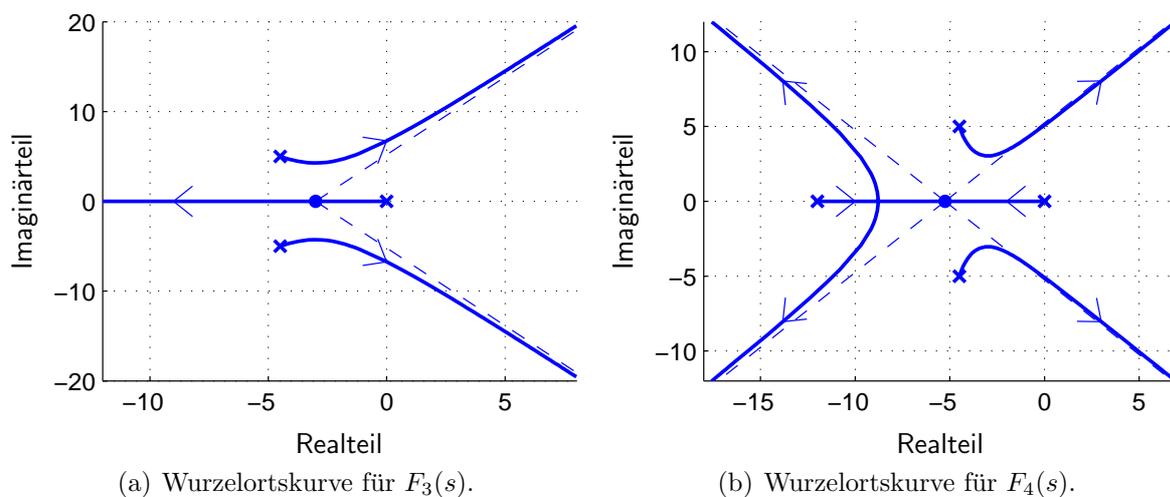


Abbildung 5.4: Wurzelortskurven für die beiden Beispielsysteme

$F_3(s) = 1/(s(s + 4.5 + 5j)(s + 4.5 - 5j))$ und $F_4(s) = 1/(s(s + 4.5 + 5j)(s + 4.5 - 5j)(s + 12))$.

Der Verlauf der Wurzelortskurven begründet anschaulich, warum Regelstrecken mit hohem Relativgrad nicht mit hochverstärkungs-basierten Proportionalreglern zu beherrschen sind. Bei hinreichend großer Rückführverstärkung befindet sich ein Teil der Polstellen in der instabilen rechten Halbebene. Daher werden in den folgenden Abschnitten geeignete Möglichkeiten zur Beeinflussung der Wurzelortskurven untersucht, so dass die Polstellen auch für große Rückführverstärkungen einen negativen Realteil besitzen.

5.2 Lösungsansätze im Falle hoher Relativgrade

Das Stabilitätsproblem bei hohen Relativgraden wird beseitigt, indem $r - 1$ zusätzliche Übertragungsnullstellen in der linken Halbebene plaziert werden. Die Ordnung des Zählerpolynomes wächst durch diese Maßnahme auf $n - 1$, weshalb sich der Relativgrad auf 1 reduziert. Allerdings bleibt zu klären, wie die Nullstellen in den Regelkreis einzufügen sind, da Realisierbarkeit kausale Systeme fordert: demzufolge kann ein realisierbares dynamisches System höchstens so viele Nullstellen wie Polstellen enthalten, ein Nullstellenüberschuss ist nicht erreichbar.

5.2.1 Dynamischer Kompensator

In Veröffentlichungen mit theoretischem Schwerpunkt wird zur Relativgradreduzierung meist ein Differenzierer

$$\tilde{F}_K(s) = \frac{\tilde{y}_K(s)}{y(s)} = (s - \tilde{q}_1)(s - \tilde{q}_2) \cdots (s - \tilde{q}_{r-1}) = s^{r-1} + \bar{q}_{r-2}s^{r-2} + \cdots + \bar{q}_1s + \bar{q}_0, \\ \tilde{q}_1, \dots, \tilde{q}_{r-1} \in \mathbb{C}_-, \quad \bar{q}_0, \dots, \bar{q}_{r-2} \in \mathbb{R}_+ \quad (5.5)$$

der Ordnung $r - 1$ eingesetzt und als „dynamisches Feedback“ oder „Kompensator“ bezeichnet. In Serienschaltung mit der Strecke (5.1) bestimmt der Kompensator (5.5) die ersten $r - 1$ Ableitungen des Streckenausganges y und errechnet daraus die Linearkombination

$$\tilde{y}_K(t) = y^{(r-1)}(t) + \bar{q}_{r-2}y^{(r-2)}(t) + \cdots + \bar{q}_1\dot{y}(t) + \bar{q}_0y(t). \quad (5.6)$$

In diesem Zusammenhang ist unter dem Begriff „Kompensator“ ein (nicht realisierbares) System mit überzähligen Nullstellen zu verstehen. Durch die $r - 1$ zusätzlich eingebrachten Nullstellen entsteht der gewünschte Relativgrad von $r = 1$ für die erweiterte Regelstrecke (bestehend aus Prozess und Kompensator). Nachdem wegen der erforderlichen Minimalphasigkeit sämtliche Nullstellen in der linken Halbebene liegen müssen, wird $\operatorname{Re}\{\tilde{q}_i\} < 0$ gefordert. Hierfür müssen die Koeffizienten \bar{q}_i zwangsweise positiv sein und sind so zu wählen, dass $\tilde{F}_K(\cdot)$ ein Hurwitzpolynom wird. Es ist sofort offensichtlich, dass die Strecke (5.1) mit Relativgrad $r = n - m > 1$ in Verbindung mit dem Kompensator (5.5) ein Gesamtsystem $F_S(s) \cdot \tilde{F}_K(s)$ ergibt, das der Systemklasse \mathcal{S} angehört.

Allerdings birgt ein solches Vorgehen grundsätzliche Schwierigkeiten. Zum einen erweist sich unter dem Einfluss von Messrauschen jegliche Art des Differenzierens als nachteilig, weil hochfrequente Rauschteile eine überaus hohe Verstärkung erfahren und dadurch den Nutzanteil des Messsignales vollständig überlagern können. Zum anderen verletzen differenzierende Systeme mit einem Zählerpolynom von höherem Grad als das Nennerpolynom die Bedingung der Kausalität.

Die Reduktion des Relativgrades mittels dynamischem Kompensator scheitert daher an der Realisierung der Übertragungsfunktion (5.5). Das Realisierbarkeitsproblem besteht bei allen differenzierenden Systemen und wird in der Regelungstechnik üblicherweise durch Tiefpassfilterung mit hoher Eckfrequenz umgangen. Unter Anwendung dieser Methode geht der nicht realisierbare Kompensator (5.5) in den realisierbaren Kompensator der Form

$$F_K(s) = \frac{y_K(s)}{y(s)} = \frac{(s - \tilde{q}_1)(s - \tilde{q}_2) \cdots (s - \tilde{q}_{r-1})}{(s - \tilde{p}_1)(s - \tilde{p}_2) \cdots (s - \tilde{p}_{r-1})}, \quad \tilde{p}_1, \dots, \tilde{p}_{r-1} \in \mathbb{R}_- \quad (5.7)$$

über. Allerdings bringt der kausale Kompensator (5.7) nicht nur $r - 1$ Nullstellen, sondern auch ebensoviele Polstellen ein, so dass der hohe Relativgrad unverändert bestehen bleibt. Isidori zeigt in [47, S. 189ff] unter Anwendung der Theorie singulärer Perturbationen, dass die Regelstrecke zusammen mit dem realisierbaren Kompensator dennoch hochverstärkungsbasiert regelbar ist. Eine näherungsweise (und damit realisierbare) Differentiation des Ausgangssignales ist ausreichend, um durch ein Regelgesetz

$$u(t) = -k y_K(t), \quad k \geq k^* > 0 \quad (5.8)$$

lokal asymptotische Stabilität der Ruhelage gewährleisten zu können. Unter der Bedingung

$$\tilde{p}_1 = \tilde{p}_2 = \cdots = \tilde{p}_{r-1} < 0 \quad \text{und} \quad |\tilde{p}_1| \text{ hinreichend groß} \quad (5.9)$$

erhalten die Pole \tilde{p}_i nur marginalen Einfluss und $F_K(s)$ approximiert den nicht-kausalen Kompensator $\tilde{F}_K(s)$, wobei sich die unumgängliche Tiefpassfilterung durch das Nennerpolynom in (5.7) nicht störend auswirkt. Die Vernachlässigung der parasitären Dynamik ist jedoch nicht für beliebig große Verstärkungen k zulässig. Solange die Verstärkung eine Grenze nicht überschreitet, gilt $y_K(\cdot) \approx \tilde{y}_K(\cdot)$ und die Dynamik des Regelkreises bleibt langsam im Vergleich zur Dynamik der zusätzlich eingebrachten Pole \tilde{p}_i . Die Stellgröße verläuft in guter Näherung derartig, als ob das Nennerpolynom mit den Polstellen \tilde{p}_i nicht existent wäre. Für zu große Verstärkungen liegt dagegen die Dynamik des geregelten Prozesses im Bereich der Pole \tilde{p}_i , so dass diese nun nicht mehr vernachlässigt werden dürfen und die Filterung instabiles Verhalten verursacht.

Proposition 4.7.3 in [47, S. 194] fasst diesen Sachverhalt zusammen und schließt in die Betrachtungen auch nichtlineare Regelstrecken vom Typ

$$\begin{aligned} \dot{x}(t) &= f(x(t)) + g(x(t)) \cdot u(t), & x(0) &= x_o \in \mathbb{R}^n \\ y(t) &= h(x(t)) \end{aligned} \quad (5.10)$$

$$\text{wobei } f, g \in \mathcal{C}^0(\mathbb{R}^n; \mathbb{R}^n), \quad h \in \mathcal{C}^0(\mathbb{R}^n; \mathbb{R})$$

ein, wobei ein wohldefinierter Relativgrad und eine stabile Nulldynamik vorausgesetzt werden.

Theorem 5.1 [47, S. 194]

Sofern die Regelstrecke (5.10) an der Stelle $x_o \in \mathbb{R}^n$ den Relativgrad $r \leq n$ aufweist und dort in erster Näherung eine asymptotisch stabile Nulldynamik besitzt, kann ein hochverstärkungsbasierter Regler gefunden werden. Werden die Koeffizienten des Polynoms

$$Z : s \mapsto s^{r-1} + \bar{q}_{r-2}s^{r-2} + \dots + \bar{q}_1s + \bar{q}_0, \quad \bar{q}_0, \dots, \bar{q}_{r-2} \in \mathbb{R}_+ \quad (5.11)$$

derart festgelegt, dass $Z(\cdot)$ ein Hurwitzpolynom ist und daher ausschließlich Wurzeln mit negativen Realteilen besitzt, stabilisiert der kausale Regler

$$F_R(s) = \frac{u(s)}{y(s)} = -k \frac{Z(s)}{(1 + sT)^{r-1}} \quad \text{mit} \quad k \in \mathbb{R}, T > 0, \quad \text{sign}(k) = \text{sign}(L_g L_f^{r-1} h(0)) \quad (5.12)$$

die Strecke (5.10), falls die konstante Verstärkung k hinreichend groß und Zeitkonstante T hinreichend klein gewählt werden.

Die Tatsache, dass beim Einsatz des Kompensators (5.7) weiterhin eine kritische Obergrenze für die Verstärkung besteht, ab der ein Teil der Pole des geschlossenen Regelkreises in der rechten Halbebene zu liegen kommt, wird anhand von Wurzelortskurven deutlich. Abbildung 5.5 enthält die Wurzelortskurve für das Beispielsystem $F_3(s)$ mit dem Kompensator

$$F_K(s) = \frac{(s + 20 + 10j)(s + 20 - 10j)}{(s + 500)^2}. \quad (5.13)$$

Zwei der drei Asymptoten weisen im Winkel $\pm\pi/3$ zur reellen Achse in die rechte Halbebene. Für die Verstärkung $k \approx 2 \cdot 10^8$ überschreitet das zugehörige Polpaar bei ± 460 die imaginäre Achse. Damit ist die grundlegende Idee der hochverstärkungsbasierten Regelung nicht mehr global gültig, sondern nur noch lokal. Durch den Kompensator

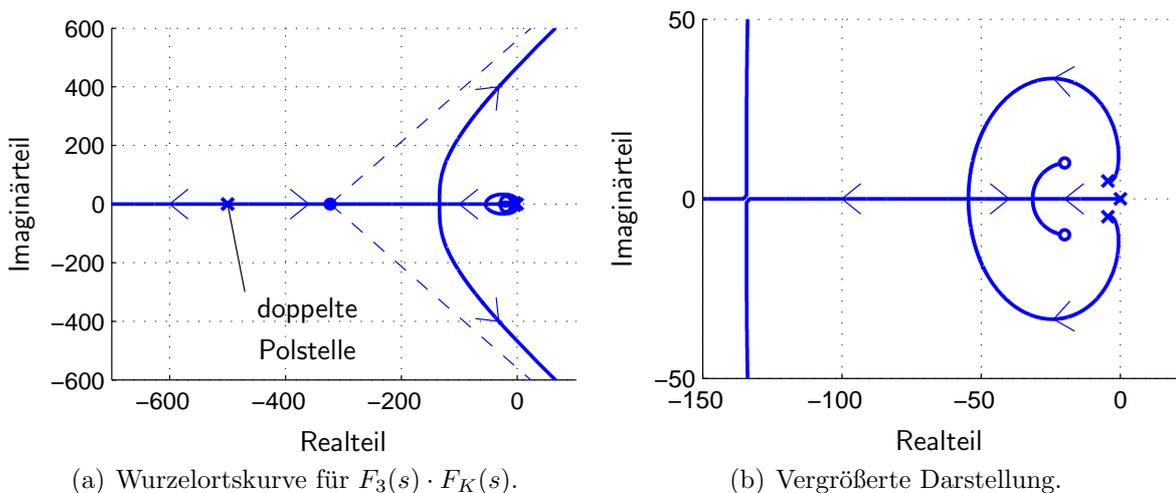


Abbildung 5.5: Wurzelortskurve für das Beispielsystem $F_3(s) = \frac{1}{s(s+4.5+5j)(s+4.5-5j)}$ mit Kompensator $F_K(s) = \frac{(s+20+10j)(s+20-10j)}{(s+500)^2}$.

wird das prinzipielle Problem daher nicht gelöst, sondern tritt lediglich erst bei größeren Verstärkungen in Erscheinung. Im Gegensatz hierzu verlässt die Wurzelortskurve für $F_3(s)$ ohne Kompensator, die in Abbildung 5.4(a) eingetragen ist, die stabile Halbebene bereits bei $k \approx 4 \cdot 10^2$. Der dynamische Kompensator erhöht die kritische Verstärkung um einige Zehnerpotenzen und ermöglicht für hinreichend schnelle Pole die Anwendung hochverstärkungsbasierter Regler. Ohne Systemkenntnis kann jedoch nicht garantiert werden, dass das Kriterium (5.9) von den gewählten Polstellen erfüllt wird.

Zusammenfassend ist festzustellen, dass die benötigten Nullstellen theoretisch durch einen (realisierbaren) Kompensator erzeugt werden können. Allerdings bestehen die beiden Nachteile, dass lediglich lokale Stabilität erreichbar ist, die Verstärkung daher nicht beliebig erhöht werden darf. Vor Inbetriebnahme der Regelung kann die Forderung (5.9) jedoch nicht überprüft werden. Desweiteren wird durch den Kompensator das Ausgangssignal (näherungsweise) differenziert, was bei realen Anlagen und Prozessen vermieden werden muss, weil in der Praxis stets Messrauschen zu erwarten ist.

5.2.2 Backstepping

Vom Ansatz abweichend, den Relativgrad der Strecke zu reduzieren, wird in [42] ein hochverstärkungsbasiertes Regelgesetz entwickelt, das sich im Falle beliebig hoher Relativgrade $r \in \mathbb{N}$ anwenden lässt. Der wesentliche Gedanke hierbei ist, die Stellgröße aus zwei Komponenten zusammensetzen: zum einen generiert der in Abschnitt 4.1 erläuterte Proportionalregler eine Stellgröße, die das gewünschte Verhalten einer Strecke mit Relativgrad 1 aufprägt. Diesem Signal wird ein geeigneter Korrekturterm hinzuaddiert, der den destabilisierenden Effekt hoher Relativgrade aufhebt und einen stabilen Regelkreis sicherstellt.

Obwohl die Systemklasse in [42] Mehrgrößensysteme (MIMO) mit nichtlinearer Störung umfasst und ein unbekanntes Vorzeichen der instantanen Verstärkung zulässt, beschränken sich die nachfolgenden Ausführungen auf lineare SISO-Systeme mit einer positiven instantanen Verstärkung. Den Betrachtungen liegt daher weiterhin die Regelstrecke (5.1) der Ordnung $n \in \mathbb{N}$ zugrunde, deren Zustandsbeschreibung

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), & x(0) &= x_o \in \mathbb{R}^n \\ y(t) &= cx(t), & cA^{r-1}b &= g > 0 \end{aligned} \quad (5.14)$$

minimal durch $A \in \mathbb{R}^{n \times n}$ und $b, c^T \in \mathbb{R}^n$ realisiert ist. An dieser Stelle sei zusätzlich vorausgesetzt, dass die Polynome Z und N in Gleichung (5.1) teilerfremd sind und $q_i \neq p_j$ für alle $i = 1, \dots, m$ und $j = 1, \dots, n$ gilt. Daher wird die in (5.1) durch $q_1, \dots, q_m \in \mathbb{C}_-$ geforderte Minimalphasigkeit im Zustandsraummodell (5.14) durch die Bedingung

$$\det \begin{bmatrix} sI_n - A & -b \\ c & 0 \end{bmatrix} \neq 0 \quad \forall s \in \mathbb{C} \text{ mit } \operatorname{Re}\{s\} \geq 0 \quad (5.15)$$

ausgedrückt. Sofern die Strecke den Relativgrad $r = 1$ besitzt, kann unter Verwendung der Funktion

$$\gamma_1 : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}, \quad (k, e) \mapsto \gamma_1(k, e) = ke \quad (5.16)$$

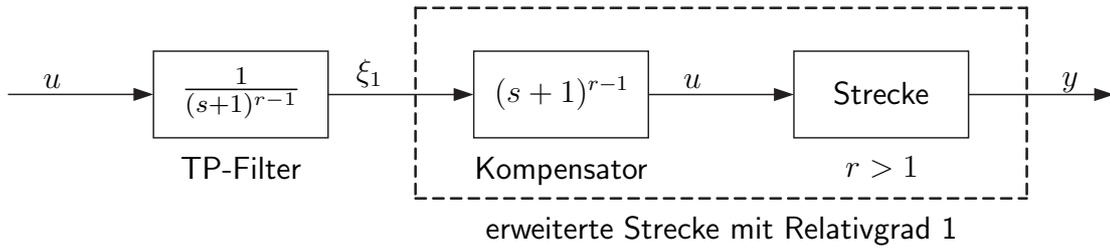


Abbildung 5.6: Prinzip der Relativgradreduzierung durch einen dynamischen Kompensator.

und des Sollwertes $w(\cdot) \in \mathcal{L}^\infty$ direkt das stabilisierende Regelgesetz

$$u(t) = \gamma_1(k(t), e(t)), \quad e(t) = w(t) - y(t) \quad (5.17)$$

angesetzt werden. Übersteigt der Relativgrad den Wert 1, ist dagegen die Stellgröße entsprechend anzupassen. Hierzu wird auf der Grundlage der Matrizen

$$F = \begin{bmatrix} -1 & 1 & 0 & \dots \\ 0 & \ddots & \ddots & 0 \\ \vdots & & \ddots & 1 \\ 0 & \dots & 0 & -1 \end{bmatrix} \in \mathbb{R}^{(r-1) \times (r-1)}, \quad G = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^{r-1}, \quad H^T = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{r-1} \quad (5.18)$$

und des Zustandsvektors

$$\xi(t) = \left(\xi_1(t) \quad \dots \quad \xi_{r-1}(t) \right)^T \in \mathbb{R}^{r-1} \quad (5.19)$$

zunächst eine Filterkette der Länge $r - 1$ festgelegt, die durch die Differentialgleichung

$$\begin{aligned} \dot{\xi}(t) &= F\xi(t) + Gu(t), & \xi(0) &= \xi_o \in \mathbb{R}^{r-1} \\ \xi_1(t) &= H\xi(t) \end{aligned} \quad (5.20)$$

beschrieben ist. Die Motivation für den Einsatz einer solchen Kette aus Tiefpässen erster Ordnung erklärt sich aus folgendem Gedankengang: der Vorgehensweise in Abschnitt 5.2.1 folgend, lässt sich der Relativgrad durch einen Kompensator der Form (5.5) auf 1 verringern, weshalb die Reihenschaltung aus Kompensator und Strecke in die Systemklasse \mathcal{S} fällt. Um die Realisierbarkeit des Kompensators zu wahren, sind, wie in Abbildung 5.6 gezeigt, $r - 1$ Polstellen in Form eines Tiefpassfilters (TP-Filter) hinzuzufügen. Gesetzt den Fall, $\xi_1(t)$ wäre die frei wählbare Stellgröße, könnte das Regelgesetz

$$\xi_1(t) = \gamma_1(k(t), e(t)) = k(t)e(t) \quad (5.21)$$

angewendet werden. In der Realität ergibt sich das Signal $\xi_1(\cdot)$ jedoch als Ausgang des Tiefpassfilters, der mit der freien Größe $u(\cdot)$ beaufschlagt wird. Deshalb besteht die Aufgabe darin, mittels Backstepping die Eingangsgröße $u(\cdot)$ der Filterkette so zu bestimmen, dass deren Ausgang $\xi_1(\cdot)$ dem in Gleichung (5.21) definierten Sollverlauf

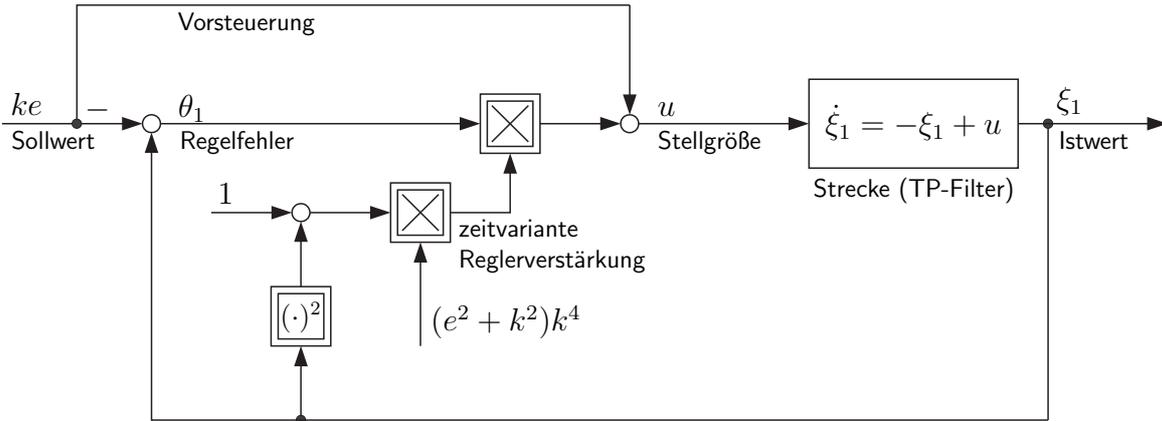


Abbildung 5.7: Regelkreis für den im Falle $r = 2$ benötigten Korrekturterm.

$\gamma_1(k(\cdot)e(\cdot))$ hinreichend nahe kommt. In anderen Worten, ein Mechanismus ist zu entwerfen, der durch Vorgabe eines Signales $u(\cdot)$ die Differenz $\theta_1(t) := \xi_1(t) - k(t)e(t)$ eliminiert. Ist ein solcher Verlauf für $u(\cdot)$ gefunden, kann dieser direkt, d.h. ohne den Filter oder den Kompensator zu passieren, der Regelstrecke aufgeschaltet werden.

Diese Problemstellung wird in [42] durch einen Regelkreis gelöst, der in Abbildung 5.7 exemplarisch für den Fall $r = 2$ dargestellt ist. Der Wert $k(t)e(t)$ ist als Sollvorgabe für die Regelgröße $\xi_1(t)$ zu betrachten. Der Tiefpassfilter als Regelstrecke ist der Stellgröße $u(t)$ unterworfen, die sich aus einer Vorsteuerung mit dem Sollwert $k(t)e(t)$ und dem Ausgang eines Proportionalreglers mit zeitvarianter Verstärkung zusammensetzt. Die Vorsteuerung ist maßgebend, damit im Falle $\theta_1(\cdot) \equiv 0$ (wenn der proportionale Regler die Stellgröße 0 generiert) der Ausgang der Filterkette bei $\xi_1(t) = k(t)e(t)$ gehalten werden kann. Der Regler nimmt lediglich dann Einfluss, sobald ein Regelfehler $\theta_1(t) \neq 0$ auftritt. Lemma 7 in [38] zeigt, dass die von der Ausgangsgröße abhängige Reglerverstärkung $(e(t)^2 + k(t)^2)k(t)^4(1 + \xi_1(t)^2)$ zu einem stabilen Regelkreis führt und damit $\theta_1 \in \mathcal{L}^\infty$ garantiert.

Der für $r = 2$ beschriebene Regelkreis wird für höhere Relativgrade $r > 2$ verallgemeinert, indem zusätzlich auch die Regelfehler $\theta_{i-1}(t) := \xi_{i-1}(t) - \gamma_{i-1}(k(t), e(t))$, $i = 3, \dots, r$ berücksichtigt werden. Die daraus resultierende Gesamtstruktur zeigt Abbildung 5.8.

Auf Basis der Einheitsmatrix I_i , $i = 1, \dots, r - 1$, und einer Nullmatrix $O_{i \times (r-1-i)}$ lässt sich die Matrix

$$V_i = \begin{bmatrix} I_i & O_i \end{bmatrix} \in \mathbb{R}^{i \times (r-1)}, \quad i = 1, \dots, r - 1 \quad (5.22)$$

zusammenstellen, die über das Matrix-Vektor-Produkt

$$v_i : \xi \mapsto V_i \xi \quad (5.23)$$

die Projektionen $v_i : \mathbb{R}^{r-1} \rightarrow \mathbb{R}^i$ definiert. Für Strecken mit Relativgrad $r = 2$ wird

auf die Regelstrecke (5.14) mit dem Relativgrad $r \geq 2$ und der Eigenschaft (5.15) angewandt, ergibt sich ein Anfangswertproblem, welches das Verhalten des geschlossenen Regelkreises durch

$$\begin{aligned} \dot{x}(t) &= Ax(t) + b\gamma_r(k(t), e(t), \xi(t)), & x(0) &= x_o \in \mathbb{R}^n \\ \dot{\xi}(t) &= F\xi(t) + G\gamma_r(k(t), e(t), \xi(t)), & \xi(0) &= \xi_o \in \mathbb{R}^{r-1} \\ k(t) &= \left(1 - (\psi(t)|e(t)|)^2\right)^{-1} \end{aligned} \quad (5.29)$$

beschreibt. Unter den Voraussetzungen

- (i) die Funktion $\psi(\cdot) \in \mathcal{F}$ besitzt die Eigenschaften $\psi(t) > 0$ für alle $t > 0$, $\psi(0) = 0$ und $\liminf_{t \rightarrow \infty} \psi(t) > 0$
- (ii) das Sollsignal $w(\cdot) \in \mathcal{F}$ stammt aus der selben Klasse wie $\psi(\cdot)$

besitzt das Anfangswertproblem (5.29) für alle Anfangswerte $(x_o, \xi_o) \in \mathbb{R}^n \times \mathbb{R}^{r-1}$ eine Lösung, wobei jede Lösung maximal erweiterbar ist. Mit $T \in [0, \infty]$ hat jede maximale Lösung $(x, \xi) : [0, T[\rightarrow \mathbb{R}^n \times \mathbb{R}^{r-1}$ die Eigenschaften:

- (i) Die Lösung ist für alle Zeiten $t \geq 0$ definiert, d.h. $T = \infty$.
- (ii) Es existiert ein $\varepsilon \in]0, 1[$, so dass $\psi(t)|e(t)| \leq 1 - \varepsilon$ für alle $t \geq 0$. Der Fehler bewegt sich nicht nur für alle Zeiten im Trichter, sondern hält einen Mindestabstand zum Trichterrand.
- (iii) Die Funktionen $u : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$, $k : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ und $(x, \xi) : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^n \times \mathbb{R}^{r-1}$ sind beschränkt.

Beweis

siehe [42, Thm. 2]. ■

Der ausschlaggebende Vorteil gegenüber der Reduktion des Relativgrades mit Hilfe eines realisierbaren Kompensators ist die Tatsache, dass hier keine Bedingung der Art (5.9) eingehalten werden muss. Weil im Korrekturkreis die Reglerverstärkung $\|\nabla\gamma_{r-1}(k, e, V_{r-2}\xi)\|^2 k^4 (1 + \|V_{r-1}\xi\|^2)$ explizit mit dem Faktor $k(t)$ gewichtet ist, beschleunigt sich die Dynamik dieses Regelkreises mit wachsendem $k(t)$. Im Gegensatz zum realisierbaren Kompensator, dessen Polstellen unabhängig vom Momentanwert $k(t)$ fixiert sind, passt sich der Korrekturmechanismus an hohe Verstärkungen an.

Problematisch hingegen sind die Exponenten von k im Regelgesetz. Gleichung (5.24) zeigt, dass für $r = 2$ bereits die Potenz k^6 auftritt, bei $r = 3$ ergibt sich der Faktor k^{18} . Dadurch wird die Regelung in höchstem Maße empfindlich gegenüber Veränderungen von $k(t)$. Selbst eine nur unwesentliche Zunahme kann durch die hohe Potenz eine bedeutende Vergrößerung der Reglerverstärkung und damit auch der Stellgröße nach sich ziehen. Insofern können beispielsweise Änderungen der Trichterbreite oder auch geringfügige Störungen zu Spitzenwerten in der Stellgröße führen, die in kaum einer realen Anwendung tragbar sind. Darüberhinaus ergeben sich gerade für höhere Relativgrade durch den Backstepping-Ansatz komplexe, unübersichtliche Regelgesetze [39].

5.2.3 Zustandsauskopplung

Die beiden vorangegangenen Abschnitte stellen zwei unterschiedliche Lösungsmöglichkeiten für das Relativgradproblem vor. Isidori löst die fehlende Kausalität des dynamischen Kompensators durch Tiefpassfilterung. Ilchmann et al. nutzen dagegen Backstepping, um Strecken mit hohen Relativgraden hochverstärkungsbasiert zu regeln. In beiden Fällen ist das Endergebnis aus Sicht des Anwenders jedoch nicht zufriedenstellend. Sowohl das Problem der Differentiation des gemessenen Ausgangssignales, als auch die hohen Potenzen der Verstärkung und die komplexen Regelgesetze wirken für die Anwendung hemmend. Um diese Schwierigkeiten zu vermeiden, wird deshalb zur Relativgradreduzierung eine Zustandsauskopplung eingeführt. Diese ist aus technischer Sicht die geeignetste Möglichkeit für die erforderliche Relativgradreduktion, da keine Differentiation notwendig ist und daher Messrauschen nicht übersteigert hoch verstärkt wird.

Beispiel

Als Beispiel soll mittels hochverstärkungsbasierter Methoden die instabile Strecke

$$F_S(s) = \frac{y(s)}{u(s)} = \frac{1}{s^2 - s} = \frac{Z(s)}{N(s)} \quad (5.30)$$

mit Relativgrad 2 stabilisiert werden. Mit einem Regelgesetz $u(t) = -ky(t)$, das auf statischer Ausgangsrückführung beruht, kann – unabhängig von der gewählten Verstärkung $k \in \mathbb{R}$ – die Strecke nicht stabilisiert werden. Die charakteristische Gleichung

$$N(s) + k \cdot Z(s) = s^2 - s + k \quad (5.31)$$

des geschlossenen Regelkreises besitzt für alle $k > 0$ zwei und für alle $k \leq 0$ eine Wurzel mit positivem Realteil. Die zugehörige Wurzelortskurve ist in Abbildung 5.9(a) aufgetragen und zeigt graphisch, dass keine Stabilisierung erreichbar ist.

Als Abhilfemaßnahme für dieses Problem wird die Regelstrecke (5.30) um einen dynamischen Kompensator erster Ordnung ergänzt, der durch Einfügen einer minimalphasigen Nullstelle den Relativgrad von 2 auf 1 reduziert. Obwohl der Kompensator

$$\tilde{F}_K(s) = \frac{\tilde{y}_K(s)}{y(s)} = s + v, \quad v > 0 \quad (5.32)$$

als eigenständiges System nicht realisierbar ist, bildet die Kombination aus Strecke und Kompensator ein kausales Gesamtsystem mit einem relativen Grad von 1. Wird dessen Ausgang über das Regelgesetz $u(t) = -k\tilde{y}_K(t)$ zurückgekoppelt, entsteht eine Kreisstruktur mit der Übertragungsfunktion

$$F(s) = \frac{\tilde{F}_K(s)F_S(s)}{1 + k\tilde{F}_K(s)F_S(s)} = \frac{s + v}{s^2 + (k - 1)s + kv}. \quad (5.33)$$

Für den Fall $v = 1$ ist die Wurzelortskurve des erweiterten Prozesses in Abbildung 5.9(b) dargestellt. In deren Verlauf vereinigen sich zunächst die beiden reellen

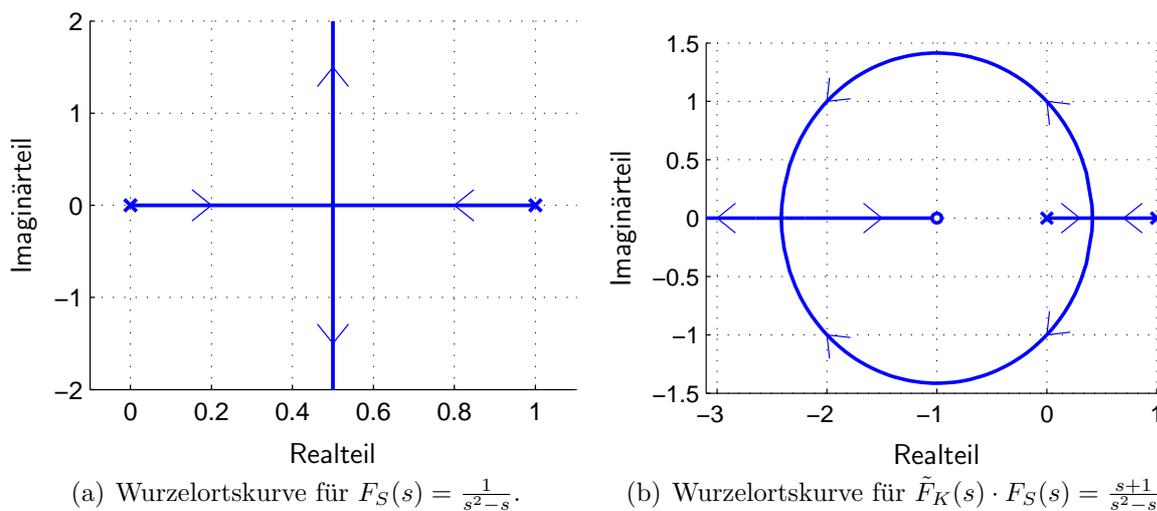


Abbildung 5.9: Wurzelortskurve für das Beispielsystem (5.30) mit und ohne Kompensator.

Polstellen in der instabilen Halbebene zu einem reellen Doppelpol, bevor sich ein konjugiert komplexes Polpaar mit positivem Realteil bildet. Im Gegensatz zur Wurzelortskurve 5.9(a) nimmt ab dem Verzweigungspunkt der Realteil beständig ab, das Polpaar überquert bei $k = 1$ die imaginäre Achse und tritt in die stabile Halbebene ein. Im Vereinigungspunkt in der linken Halbebene treffen sich beide Pole wieder und verbleiben auch für höhere Verstärkungen auf der reellen Achse. Einer der beiden Pole konvergiert gegen die Nullstelle $s = -1$, wohingegen der zweite ins Unendliche strebt. Für keine Verstärkung $k > 1$ wird die linke Halbebene verlassen. Daran ist ersichtlich, dass es eine kritische Verstärkung gibt, die für einen stabilen Regelkreis überschritten werden muss und hochverstärkungs-basierte Regelung durch den Kompensator möglich wird.

Wegen fehlender Kausalität ist die gewünschte Veränderung der Wurzelortskurve auf diese Weise jedoch nicht erreichbar. Allerdings lässt sich durch Zustandsauskopplung ebenfalls eine Nullstelle einfügen, so dass der Kompensator durch eine kausale Struktur ersetzt werden kann. An dieser Stelle wird der Übergang auf eine zustandsbasierte Beschreibung der Strecke (5.30) notwendig, für die sich beispielsweise die Regelungsnormform

$$\begin{aligned} \begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} &= \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot u(t), & \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} = 0 \in \mathbb{R}^2 \\ y(t) &= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \end{aligned} \tag{5.34}$$

eignet. Der Polverschiebungssatz [95, S. 11, 186] beschreibt die Fähigkeit eines Zustandsreglers, jede beliebige Polkonfiguration zu erzwingen. Vor diesem Hintergrund ist ein Zustandsregler prinzipiell in der Lage, die Pole auf jedem beliebigen Punkt der Wurzelortskurve zu plazieren, die von der Kombination aus Strecke und Kompensator definiert wird und damit dem Prozess identisches Verhalten aufzuprägen wie der

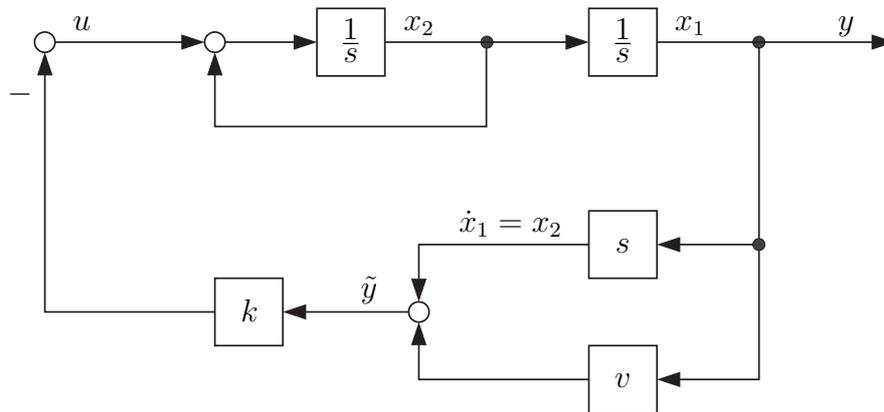


Abbildung 5.10: Regelstrecke (5.34) in Regulationsnormalform mit Kompensator (5.32).

Kompensator. Der Kompensator (5.32) beeinflusst die Strecke in gewünschter Weise und legt dadurch gewissermaßen ein Sollverhalten fest, das sich in den Eigenschaften von (5.33) niederschlägt. Die Zustandsauskopplung

$$\tilde{y}(t) = \begin{pmatrix} v & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \quad (5.35)$$

ersetzt den Kompensator (5.32) und fügt dem Übertragungsverhalten von $u(\cdot)$ nach $\tilde{y}(\cdot)$ eine Nullstelle bei $s = -v$ hinzu. Im Pfad zwischen $u(\cdot)$ und $y(\cdot)$ bleiben hingegen die von der Regelstrecke vorgegebenen Nullstellen unberührt. Eine Rückführung mittels Zustandsregler

$$u(t) = -k \cdot \tilde{y}(t) = -k \begin{pmatrix} v & 1 \end{pmatrix} \begin{pmatrix} x_1(t) & x_2(t) \end{pmatrix}^T \quad (5.36)$$

prägt der Strecke die erforderliche Dynamik auf, damit die Eigenwerte der zustandsregelten Strecke für alle $k \in \mathbb{R}$ mit den Wurzeln des Nennerpolynomes in (5.33) übereinstimmen. Weil die Wurzelortskurve für (5.33) durch die Veränderung eines einzigen Parameters k entsteht und das zustandsregelte System

$$\begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ -kv & 1 - k \end{bmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}, \quad \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} \in \mathbb{R}^2 \quad (5.37)$$

$$\begin{pmatrix} \tilde{y}(t) \\ y(t) \end{pmatrix} = \begin{bmatrix} v & 1 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$$

das selbe E/A-Verhalten von $u(\cdot)$ nach $\tilde{y}(\cdot)$ wie (5.33) besitzt, genügt auch beim Zustandsregler die Variation eines Parameters, um die Wurzelortskurve zu generieren.

Bei Verwendung des dynamischen Kompensators liegt die Struktur aus Abbildung 5.10 vor. Die dort auftretende Differentiation im Rückführzweig muss nicht realisiert werden, wenn die Zustandsgröße x_2 direkt aus der Integrator-kette abgegriffen werden kann. Durch Rückkopplung von x_2 wird das nicht kausale Übertragungsglied eliminiert und

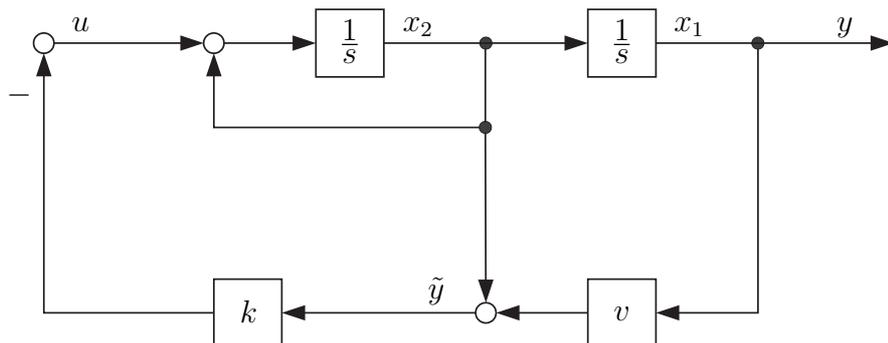


Abbildung 5.11: Regelstrecke (5.34) in Regelungsnormform mit Zustandsauskopplung (5.36).

es entsteht, wie in Abbildung 5.11 dargestellt, ein Zustandsregler. Demzufolge ist die nicht realisierbare Kompensatorstruktur äquivalent zu einem realisierbaren Zustandsregelgesetz. \square

Dieser, am konkreten Beispiel aufgezeigte Sachverhalt, wird durch nachfolgende Ausführungen verallgemeinert.

Für eine Zustandsauskopplung ist der Verlauf der inneren Systemzustände relevant, allein die Kenntnis des E/A-Verhaltens ist nicht ausreichend. Deshalb wird neben der Systembeschreibung durch Übertragungsfunktionen eine Zustandsdarstellung eingeführt. Diese basiert auf den Koordinaten $z_q \in \mathbb{R}$, $q = 1, 2, \dots, n$ und wird durch die Größen $A \in \mathbb{R}^{n \times n}$ und $b, c^T \in \mathbb{R}^n$ charakterisiert. Die reellen Koeffizienten $g > 0$ und $a_i, \rho_j, \eta_{jp} \in \mathbb{R}$, $i = 1, 2, \dots, n$, $j, p = 1, 2, \dots, m$ erfüllen die Forderung, dass die Übertragungsfunktion der Differentialgleichung

$$\begin{pmatrix} \dot{z}_1(t) \\ \vdots \\ \dot{z}_{r-1}(t) \\ \dot{z}_r(t) \\ \dot{z}_{r+1}(t) \\ \vdots \\ \dot{z}_n(t) \end{pmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 & \vdots & & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ a_1 & a_2 & a_3 & \dots & a_r & a_{r+1} & \dots & a_n \\ \rho_1 & 0 & \dots & \dots & 0 & \eta_{11} & \dots & \eta_{1m} \\ \vdots & \vdots & & & \vdots & \vdots & & \vdots \\ \rho_m & 0 & \dots & \dots & 0 & \eta_{m1} & \dots & \eta_{mm} \end{bmatrix}}_A \cdot \begin{pmatrix} z_1(t) \\ \vdots \\ z_{r-1}(t) \\ z_r(t) \\ z_{r+1}(t) \\ \vdots \\ z_n(t) \end{pmatrix} + \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ g \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_b u(t)$$

$$y(t) = \underbrace{\begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}}_c \begin{pmatrix} z_1(t) \\ \vdots \\ z_n(t) \end{pmatrix} = z_1(t), \quad \begin{pmatrix} z_1(0) \\ \vdots \\ z_n(0) \end{pmatrix} \in \mathbb{R}^n$$

(5.38)

mit derjenigen der Strecke (5.1) übereinstimmt und daher

$$c(sI_n - A)^{-1}b = F_S(s) = g \frac{(s - q_1)(s - q_2) \cdots (s - q_m)}{(s - p_1)(s - p_2) \cdots (s - p_n)}, \quad g > 0, \quad m < n \quad (5.39)$$

gilt. Die Zustandsbeschreibung beinhaltet in der Form (5.38) im Vorwärtszweig ein Subsystem in Regelungsnormalform (Integratorreihe der Ordnung r) und erlaubt, die Beziehung

$$\begin{pmatrix} z_1(t) \\ \vdots \\ z_r(t) \end{pmatrix} = \begin{pmatrix} y(t) \\ \vdots \\ y^{(r-1)}(t) \end{pmatrix} \in \mathbb{R}^r \quad (5.40)$$

abzulesen. Diese zeigt, dass die Kenntnis der Systemzustände im Vorwärtszweig der zugehörigen Byrnes-Isidori-Normalform gleichbedeutend ist mit dem Informationsgehalt, der vom Kompensator durch $r - 1$ -fache Differentiation des Ausgangssignales $y(\cdot)$ generiert wird. Deswegen kann das in Gleichung (5.6) definierte Signal $\tilde{y}_K(\cdot)$ auch ohne explizite Verwendung der Ableitungen ermittelt werden, wenn anstelle des Kompensators die Zustandsauskopplung

$$\tilde{y}(t) = \begin{pmatrix} \bar{q}_0 & \cdots & \bar{q}_{r-2} & 1 \end{pmatrix} \cdot \begin{pmatrix} z_1(t) & \cdots & z_{r-1}(t) & z_r(t) \end{pmatrix}^T \quad (5.41)$$

eingesetzt wird. In diesem Sinne stellt die Zustandsauskopplung aus regelungstheoretischer Sicht eine mögliche Realisierung des dynamischen Kompensators $\tilde{F}_K(s)$ dar.

Im Grundsatz ist keine spezielle Struktur des Zustandsraummodelles

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), & x(0) &= x_o \in \mathbb{R}^n, & n &\in \mathbb{N} \\ y(t) &= cx(t), & A &\in \mathbb{R}^{n \times n}, & b, c^T &\in \mathbb{R}^n \end{aligned} \quad (5.42)$$

erforderlich, um durch Zustandsauskopplung den Relativgrad auf 1 zu verringern. Sofern $b \neq 0$ lässt sich unter der Voraussetzung der Zustandsmessbarkeit stets ein Auskoppelvektor $\tilde{c}^T \in \mathbb{R}^n$ definieren, der $\tilde{c}b \neq 0$ erfüllt und damit der Strecke

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), & x(0) &= x_o \in \mathbb{R}^n \\ \tilde{y}(t) &= \tilde{c}x(t), & \tilde{c}^T &\in \mathbb{R}^n \end{aligned} \quad (5.43)$$

den Relativgrad 1 zuordnet. Allerdings ist neben der Relativgradbedingung durch die Wahl der Auskoppelvektoren gleichzeitig auch Minimalphasigkeit zu garantieren. Diese Forderung ist problemlos zu erfüllen, wenn die Strecke entweder in Byrnes-Isidori-Normalform oder in Regelungsnormalform vorliegt.

Ist der Zustandsvektor des Vorwärtspfades einer Byrnes-Isidori-Normalform zugänglich, lässt sich die Linearkombination (5.41) berechnen und es entsteht dasselbe Verhalten wie beim Einsatz des dynamischen Kompensators (5.5). Dessen Nullstellen liegen in der linken Halbebene, wenn $\tilde{F}_K : s \mapsto s^{r-1} + \bar{q}_{r-2}s^{r-2} + \cdots + \bar{q}_1s + \bar{q}_0$ durch die Koeffizienten $\bar{q}_i > 0$ als Hurwitzpolynom festgelegt wird. Daher ist unabhängig von den Parametern in (5.38) bei geeigneter Wahl der Konstanten \bar{q}_i eine minimalphasige Auskopplung sichergestellt.

Ähnliches trifft für Prozesse zu, deren Zustände $x(t) = (x_1(t), \dots, x_n(t))^T \in \mathbb{R}^n$ in Regelungsnormalform verfügbar sind. Unter dieser Voraussetzung erhält die Strecke durch die Auskopplung

$$\tilde{y}_R(t) = \begin{pmatrix} \tilde{c}_1 & \tilde{c}_2 & \dots & \tilde{c}_n \end{pmatrix} \cdot \begin{pmatrix} x_1(t) & x_2(t) & \dots & x_n(t) \end{pmatrix}^T \quad (5.44)$$

mit den Faktoren $\tilde{c}_1, \dots, \tilde{c}_n \in \mathbb{R}$ zwischen Eingang $u(\cdot)$ und Ausgang $\tilde{y}_R(\cdot)$ die Übertragungsfunktion

$$\tilde{F}(s) = \frac{\tilde{y}_R(s)}{u(s)} = c(sI_n - A)^{-1}b = \frac{\tilde{c}_n s^{n-1} + \dots + \tilde{c}_2 s + \tilde{c}_1}{s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0}. \quad (5.45)$$

Weil es die Zustände einer Regelungsnormalform sind, die mit den Faktoren $\tilde{c}_1, \dots, \tilde{c}_n$ gewichtet werden, hängt das Zählerpolynom einzig vom frei wählbaren Auskoppelvektor $\tilde{c} = (\tilde{c}_1, \dots, \tilde{c}_n) \in \mathbb{R}^n$ ab, der damit die Nullstellen festlegt. Die unbekannt Parameter $a_1, \dots, a_n \in \mathbb{R}$ im Nennerpolynom der Strecke beeinflussen den Zähler nicht, sondern wirken sich allein auf die Polstellen aus. Aus diesem Grund kann bei Auskopplung aus einer Regelungsnormalform stets die Einhaltung der Minimalphasigkeit garantiert werden, indem $\tilde{c}_n s^{n-1} + \dots + \tilde{c}_2 s + \tilde{c}_1$ als Hurwitzpolynom gewählt wird. Der Hauptvorteil der Regelungsnormalform liegt in der Tatsache, dass keine Kenntnis der Systemparameter notwendig ist, um die Nullstellen beliebig zu positionieren. Die Allgemeingültigkeit dieser Aussage ist in Abschnitt A.12.2 gezeigt.

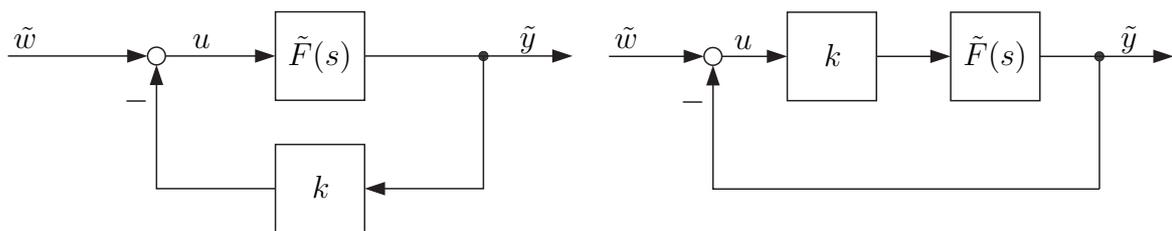
In der überwiegenden Mehrzahl aller Fälle liegt die Strecke jedoch weder in Regelungsnormalform noch in Byrnes-Isidori-Normalform vor. Entgegen der gewohnten Vorgehensweise ist für die Strecke die Berechnung einer Ähnlichkeitstransformation in eine der beiden Normalformen unmöglich, da hierzu Eigenwerte und Eigenvektoren bekannt sein müssen, was der Annahme unbekannter Parameterwerte entgegensteht. So stimmt zum Beispiel die physikalisch motivierte Zustandsbeschreibung des betrachteten Zweimassensystems mit keiner der beiden Normalformen überein. Dennoch kann, wie Theorem 6.3 zeigt, auch ohne präzise Kenntnis der Systemparameter eine minimalphasige Auskopplung gewählt werden. Es handelt sich hier zwar um eine Besonderheit der vorliegenden Regelstrecke und ist daher nicht allgemeingültig. Weil ein Zweimassensystem jedoch in der Mechatronik als grundlegendes Modell für viele mechanische Anordnungen betrachtet wird, ist eine solche Einschränkung an dieser Stelle durchaus akzeptabel.

Für andere Anwendungsfälle muss in einer Einzelfallprüfung untersucht werden, ob sich trotz unbekannter Parameter eine Zustandsauskopplung bestimmen lässt, die eine minimalphasige Regelstrecke ergibt.

Laut Polverschiebungssatz [95, S. 186] gestattet ein Zustandsregler vollständige Freiheit bei der Festlegung der Pole des geschlossenen Regelkreises, wenn n Freiheitsgrade bei der Wahl der Gewichtungsfaktoren in der Linearkombination der n Systemzustände gegeben sind. Wird eine Adaption mit n Freiheitsgraden zugelassen, kann der Regelstrecke zwar asymptotisch eine beliebige Dynamik aufgezwungen werden, die Wahl eines geeigneten Adaptionsgesetzes und die Stabilitätsanalyse erschweren sich dadurch erheblich. Wie die Verwendung eines Kompensators mit konstanten Polynomkoeffizienten zeigt, reicht jedoch ein einziger Freiheitsgrad in der Adaption aus, um das

gewünschte Verhalten zu erzielen. Deshalb werden konstante Verhältnisse zwischen den Rückführungen festgelegt und die gesamte Linearkombination mit dem zeitvarianten Faktor $k(t)$ gewichtet. Die signifikante Vereinfachung besteht darin, dass eine Strecke mit einer solchen Auskopplung strukturell die selbe Übertragungsfunktion wie die Kombination aus Strecke (5.1) und Kompensator (5.5) annimmt, dadurch der Systemklasse \mathcal{S} angehört und mit hochverstärkungsbasierten Reglern beherrscht werden kann. Es muss demzufolge kein neuer Stabilitätsbeweis geführt werden.

Die Zustandsauskopplungen (5.41) bzw. (5.44) reduzieren den hohen Relativgrad einer Regelstrecke unter der Nebenbedingung einer stabilen Nulldynamik. Deshalb sind die hier angestellten Überlegungen zielführend, um hochverstärkungsbasierte Regler anzuwenden. Die Linearkombination der Zustände ergibt ein Signal, das eine geeignete Regelgröße darstellt. Jene kann als Systemausgang betrachtet werden und ist auf einen Sollwert zu regeln. Nachdem die Strecke mit Zustandsauskopplung zur Systemklasse \mathcal{S} gehört, führt ein hochverstärkungsbasierter Regler nach Theorem 3.7 auf einen stabilen Regelkreis und stellt damit die essentielle Eigenschaft sicher. Dennoch ist nicht zwangsläufig gesichert, dass der Zustandsvektor gegen eine Ruhelage konvergiert. Im Zusammenhang mit der Bestimmung des stationären Regelfehlers muss daher die Existenz des Grenzwertes $\lim_{t \rightarrow \infty} y(t) = y_\infty \in \mathbb{R}$ vorausgesetzt werden dürfen.



(a) Regelkreis mit P-Regler im Rückführzweig. (b) Regelkreis mit P-Regler im Vorwärtszweig.

Abbildung 5.12: Regelkreise mit unterschiedlicher stationärer Verstärkung.

Wie das Kapitel 3 für adaptive Regler und das Kapitel 4 für Funnel-Control zeigt, erreichen Proportionalregler nicht notwendigerweise stationäre Genauigkeit. Soll die Linearkombination $\tilde{y}(t)$ asymptotisch einen konstanten Sollwert \tilde{w} erreichen, ist laut Prinzip des internen Modells (siehe Abschnitt 4.6) ein Übertragungsglied mit integralem Verhalten im Regelkreis erforderlich. Falls dieses fehlt, ergibt sich in der Gleichgewichtslage eine Abweichung zwischen Soll- und Istwert. Durch die stationäre Verstärkung V_∞ wird das Verhältnis $\tilde{y}_\infty/\tilde{w}_\infty$ beschrieben, das idealerweise den Wert 1 annehmen soll. Für den Regelkreis in Abbildung 5.12(a) mit der Regelstrecke $\tilde{F}(s) = Z(s)/N(s)$ bestimmt sich die stationäre Verstärkung zu und verkleinert sich mit anwachsender Verstärkung k . Je höher die Verstärkung im Regelkreis, desto geringer fällt der Istwert im Verhältnis zum Sollwert aus.

$$V_\infty = F_{RK}(0) = \frac{\tilde{F}(0)}{1 + k\tilde{F}(0)} = \frac{Z(0)}{N(0) + kZ(0)} \quad (5.46)$$

Besteht die Aufgabe im Stabilisieren der Strecke, wird durch eine verschwindend kleine stationäre Verstärkung kein Problem verursacht. Weil ein Sollwert $\tilde{w} = 0$ vorgegeben

wird, ist stationäre Genauigkeit für alle Werte der Verstärkungen gegeben, selbst im Grenzfall $V_\infty = 0$.

Für eine Festwertregelung mit $\tilde{w} \neq 0$ ist $V_\infty = 0$ jedoch nicht tragbar. Folglich ist der proportionale Regler in den Vorwärtszweig zu verschieben (siehe Abbildung 5.12(b)), um durch diese Modifikation die stationäre Verstärkung auf

$$V_\infty = \frac{kZ(0)}{N(0) + kZ(0)} \quad (5.47)$$

zu ändern. Steigt die Verstärkung k an, ergibt sich wunschgemäß

$$V_\infty = \lim_{k \rightarrow \infty} \frac{kZ(0)}{N(0) + kZ(0)} = \frac{kZ(0)}{kZ(0)} = 1. \quad (5.48)$$

Für hohe Verstärkungen, bzw. im Grenzfall $k \rightarrow \infty$ liegt damit stationäre Genauigkeit vor, was mit der Aussage von Abschnitt 3.5.2 übereinstimmt.

Für endliche Werte der Verstärkung $k \in \mathbb{R}$ hängt das stationäre Verhältnis zwischen Ist- und Sollwert von den Eigenschaften der Regelstrecke ab. Sind die Polynome $Z, N \in \mathbb{R}[s]$ mit den Koeffizienten $a_0, \dots, a_{n-1} \in \mathbb{R}$ und $c_0, \dots, c_m \in \mathbb{R}$ durch

$$Z(s) = c_m s^m + \dots + c_2 s^2 + c_1 s + c_0 \quad \text{und} \quad N(s) = s^n + \dots + a_2 s^2 + a_1 s + a_0 \quad (5.49)$$

festgelegt, resultiert aus Gleichung (5.47) eine stationäre Verstärkung des Regelkreises von

$$V_\infty = \frac{kc_0}{a_0 + kc_0}. \quad (5.50)$$

Offensichtlich arbeitet der Regler stationär genau, wenn eine der drei Bedingungen zutrifft:

- (i) $a_0 = 0$. Das ist die charakteristische Eigenschaft einer Regelstrecke mit global integralem Verhalten. Durch den integralen Anteil verschwindet der Regelfehler, so dass die Adaption bereits bei endlichen Werten konvergiert. Es gilt $V_\infty = 1$ für alle $k \in \mathbb{R}$. Dieser Fall trifft z.B. beim Zweimassensystem zu, oder auch dann, wenn ein internes Modell (Integrator) eingesetzt wird.
- (ii) Sollwertvorverstärkung mit dem Faktor $V_\infty^{-1} = (a_0 + k_\infty c_0)/(k_\infty c_0)$. Weil der Endwert k_∞ vorab nicht bekannt ist, muss dieser durch die momentane Verstärkung $k(t)$ substituiert werden. Das Produkt $k(t)c_0$ im Nenner fordert, dass ein Anfangswert $k(0) > 0$ festgesetzt wird. Nachteilig wirkt sich aus, dass zur Berechnung der Vorverstärkung der Koeffizient a_0 bekannt sein muss, was im Allgemeinen nicht zutrifft.
- (iii) $k = \infty$. Dieser Grenzfall bedeutet zwar das Verschwinden des Regelfehlers, ist jedoch für praktische Anwendungen belanglos und wird überdies durch Zusatzmaßnahmen unterbunden. Beispielsweise durch die Einführung eines λ -Bandes verhindert das Adaptionsgesetz

$$\dot{k}(t) = \begin{cases} (\tilde{y}(t) - \tilde{w}(t))^2 & \text{für } \tilde{y}(t) - \tilde{w}(t) > \lambda \\ 0 & \text{für } \tilde{y}(t) - \tilde{w}(t) \leq \lambda \end{cases} \quad (5.51)$$

mit $\lambda > 0$ das unbeschränkte Anwachsen der Verstärkung. Nur wenn der Regelfehler die Konstante λ überschreitet, erhöht sich die Verstärkung. Solange der Fehler im λ -Band verläuft, bleibt die Reglerverstärkung dagegen unverändert. Dadurch wird zwar die Verstärkung beschränkt gehalten, allerdings mit der Folge, dass ein (beliebig kleiner) Regelfehler in Kauf genommen werden muss.

Wie bei Zustandsreglern allgemein üblich, ist neben dem betrachteten Zusammenhang zwischen Sollwert \tilde{w} und Istwert \tilde{y} zusätzlich auch die Beziehung zwischen Regelgröße \tilde{y} und dem technologisch relevanten Prozessausgang y von Interesse. Dieses Verhältnis ist wesentlich, um aus dem gegebenen Sollwert w für den ursprünglichen Ausgang y eine passende Vorgabe \tilde{w} für die Regelgröße \tilde{y} zu ermitteln. Ziel ist, dass der Prozessfehler $e(t) = w(t) - y(t)$ gegen 0 konvergiert und damit der physikalische Prozessausgang y mit dem gewünschten Sollwert w übereinstimmt, wenn der Regelfehler $\tilde{e}(t) = \tilde{w}(t) - \tilde{y}(t)$ ausgeregelt ist. Dies führt auf die Frage, wie sich die Fehlersignale $e(\cdot)$ und $\tilde{e}(\cdot)$ zueinander verhalten und wie aus $w(\cdot)$ das Signal $\tilde{w}(\cdot)$ zu berechnen ist. In Anlehnung an die bewährte Vorgehensweise [23, S. 467] soll ein Vorfilter eingesetzt werden, der die Umrechnung durch einen geeigneten Faktor vornimmt. Ein frei wählbarer Vektor $\tilde{c}^T \in \mathbb{R}^n$, $n \in \mathbb{N}$, gewichtet die Zustände der Regelstrecke

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), & x(0) &= x_o \in \mathbb{R}^n \\ \tilde{y}(t) &= \tilde{c}x(t), & A &\in \mathbb{R}^{n \times n}, b, \tilde{c}^T \in \mathbb{R}^n \end{aligned} \quad (5.52)$$

und definiert auf diese Weise eine Linearkombination $\tilde{y}(\cdot)$, die als Regelgröße betrachtet wird und einem vorgegebenen Sollwert $\tilde{w}(t)$ folgen soll. Das Zustandsregelgesetz $u(t) = k \cdot (\tilde{w}(t) - \tilde{y}(t))$, $k \in \mathbb{R}$ führt die Regelgröße $\tilde{y}(t)$ zurück und legt dadurch den Regelkreis

$$\begin{aligned} \dot{x}(t) &= (A - kb\tilde{c})x(t) + bk\tilde{w}(t), & x(0) &= x_o \in \mathbb{R}^n \\ y(t) &= cx(t) \end{aligned} \quad (5.53)$$

fest. Auf der Grundlage, dass der gewählte Auskoppelvektor $\tilde{c}^T \in \mathbb{R}^n$ zu einem stabilen Regelkreis führt, besitzt die quadratische Matrix $(A - kb\tilde{c}) \in \mathbb{R}^{n \times n}$ ausschließlich Eigenwerte mit negativen Realteilen und ist daher invertierbar. Wird ein Sollwertverlauf vorgegeben, der asymptotisch gegen die Konstante \tilde{w}_∞ konvergiert, stellt sich im Regelkreis für $t \rightarrow \infty$ die Ruhelage $x_\infty = -(A - kb\tilde{c})^{-1}bk\tilde{w}_\infty$ ein. Damit die zugehörige Ausgangsgröße $\lim_{t \rightarrow \infty} y(t) = y_\infty = cx_\infty$ den gewünschten Wert $w_\infty \in \mathbb{R}$ annimmt, muss

$$\tilde{w}_\infty = V \cdot w_\infty \quad \text{mit} \quad V = -\frac{1}{c(A - kb\tilde{c})^{-1}bk} \quad (5.54)$$

gewählt werden. Die Sollwertverstärkung V , die den Sollwert $w(t)$ für den Prozessausgang $y(t)$ in den Sollwert $\tilde{w}(t)$ umrechnet, hängt von den (bekannten) Reglerparametern k und \tilde{c} ab, sowie von den (unbekannten) Streckenparametern A, b, c . Daher ist die Größe V für allgemeine Fälle nicht berechenbar.

Als vorteilhaft erweist sich an dieser Stelle wiederum die Regelungsnormalform. In einer Gleichgewichtslage $y_\infty \neq 0$ wird der Quotient aus Regel- und Prozessgröße als stationäres Verstärkungsverhältnis V_w

$$V_w := \frac{\tilde{y}_\infty}{y_\infty} = \frac{\tilde{c}x_\infty}{cx_\infty}, \quad y_\infty \neq 0 \quad (5.55)$$

definiert. Die Realisierung der Strecke in Regelungsnormalform (=Integrator-kette) erfordert in der Gleichgewichtslage die Bedingung $x_i = 0$ für alle $1 < i \leq n$. Dadurch kürzt sich der Zustand $\lim_{t \rightarrow \infty} x_1(t) = x_{1,\infty}$ heraus und es entsteht ein von der Ruhelage unabhängiges Verhältnis

$$V_w = \frac{\tilde{c}_0 \cdot x_{1,\infty} + \tilde{c}_1 \cdot 0 + \cdots + \tilde{c}_{n-1} \cdot 0}{c_0 \cdot x_{1,\infty} + c_1 \cdot 0 + \cdots + c_{n-1} \cdot 0} = \frac{\tilde{c}_0}{c_0} \quad \Rightarrow \quad \tilde{y}_\infty = \frac{\tilde{c}_0}{c_0} y_\infty. \quad (5.56)$$

Als Endergebnis resultiert daraus die gesuchte Umrechnung

$$\tilde{w}_\infty = \frac{\tilde{c}_0}{c_0} w_\infty, \quad (5.57)$$

wobei w_∞ den Sollwert für den physikalischen Ausgang $y(t \rightarrow \infty)$ angibt. Die technologische Aufgabenstellung wird daher erreicht, wenn die Größe \tilde{y} auf den Wert \tilde{w} geregelt wird.

Die Division durch c_0 ist als unproblematisch zu erachten, da Prozesse mit $c_0 = 0$ keine konstante Eingangsgröße u besitzen, die eine Ruhelage mit der Eigenschaft $y \neq 0$ hervorruft. Für solche Prozesse ist eine Festwertregelung aus physikalischen Gründen unmöglich. Eine Stabilisierung dagegen bereitet auch im Falle $c_0 = 0$ keine Probleme, weil eine Umrechnung des Sollwertes entfallen kann.

Nachdem der Skalierungsfaktor zwischen den beiden Ausgangssignalen im stationären Zustand vom unbekanntem Streckenparameter c_0 beeinflusst wird, bleibt das ursprüngliche Problem zwar bestehen, vereinfacht sich jedoch dahingehend, dass lediglich ein Systemparameter als unbekannte Größe auftritt. Ein möglicher Lösungsansatz hierfür ist die Verwendung eines hochverstärkungs-basierten Beobachters [9], [17], [51]. Ein solcher besteht aus einer Integrator-kette, in die der Beobachterfehler $\hat{e}(t) = y(t) - \hat{y}(t)$ über die Faktoren $h_0, \dots, h_{r-1} \in \mathbb{R}$ eingekoppelt wird. Die Differentialgleichung ist durch

$$\begin{pmatrix} \dot{\hat{x}}_1(t) \\ \vdots \\ \dot{\hat{x}}_r(t) \end{pmatrix} = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix} \cdot \begin{pmatrix} \hat{x}_1(t) \\ \vdots \\ \hat{x}_r(t) \end{pmatrix} + \begin{pmatrix} h_0 \\ \vdots \\ h_{r-1} \end{pmatrix} \cdot \hat{e}(t), \quad \begin{pmatrix} \hat{x}_1(0) \\ \vdots \\ \hat{x}_r(0) \end{pmatrix} \in \mathbb{R}^r$$

$$\hat{y}(t) = \hat{x}_1(t) \quad (5.58)$$

gegeben, die Kombination aus Strecke und Beobachter ist in Abbildung 5.13 dargestellt.

Durch den Einsatz eines Beobachters ergeben sich zwei Vorteile. Einerseits müssen die Zustände der Regelstrecke nicht gemessen werden. Dies ist wesentlich, da die Strecke im Allgemeinen nicht in RNF vorliegt und daher die benötigten Zustände physikalisch nicht existieren. Andererseits leistet der Beobachter implizit die notwendige Skalierung der Zustände mit der Unbekannten c_0 . Wenn anstelle der Rückführung $\tilde{y}(t) = \tilde{c}x(t)$ die Auskopplung $\hat{y}(t) = \tilde{c}\hat{x}(t)$ angesetzt wird, ergibt sich in der Ruhelage:

$$\begin{aligned} y_\infty &= c_0 \cdot x_{1,\infty} \\ \hat{y}_\infty &= \hat{x}_{1,\infty} \end{aligned} \quad \Rightarrow \quad V_w = \frac{\tilde{c}_0 \cdot \hat{x}_{1,\infty} + \tilde{c}_1 \cdot \hat{x}_{2,\infty} + \cdots + \tilde{c}_{r-1} \cdot \hat{x}_{r,\infty}}{c_0 \cdot x_{1,\infty} + c_1 \cdot 0 + \cdots + c_{r-1} \cdot 0} \quad (5.59)$$

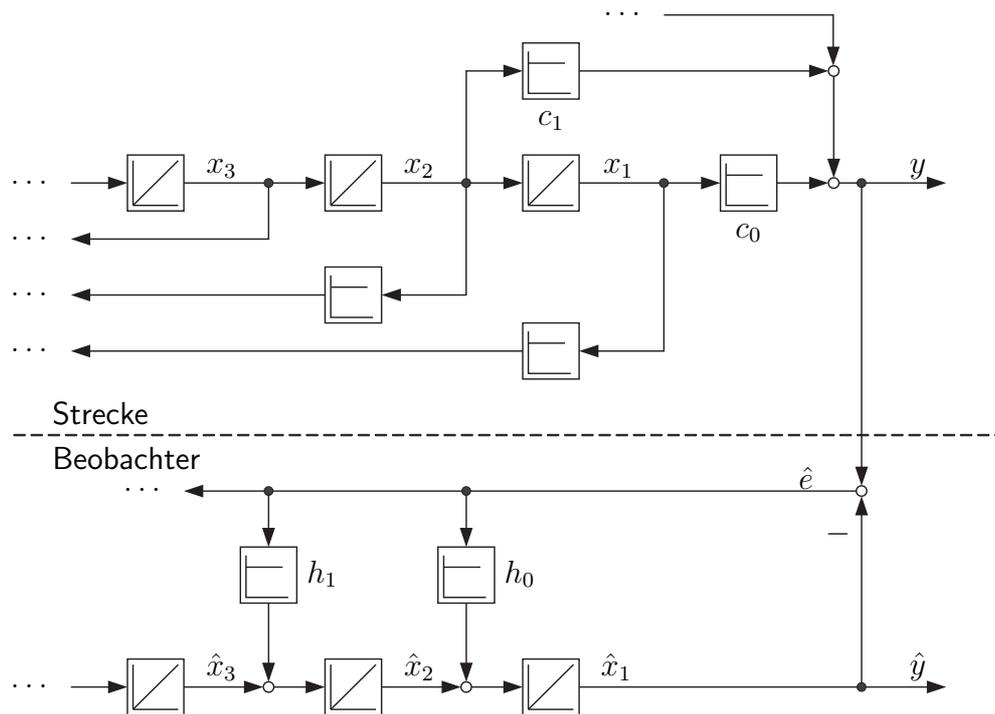


Abbildung 5.13: Strecke in Regulationsnormalform mit hochverstärkungsbasiertem Beobachter.

Unter der Voraussetzung eines korrekt eingestellten Beobachters wird ein stabiler Regelkreis angenommen, d.h. das Erreichen einer Gleichgewichtslage ist unter diesen Umständen garantiert. Weil aus (5.58) die Gleichung

$$\frac{d}{dt} \hat{x}_r(t) = h_{r-1} \cdot \hat{e}(t) \quad (5.60)$$

entnommen werden kann, wird dadurch für $h_{r-1} \neq 0$ in der Ruhelage $\hat{e}(t) = 0$ erzwungen. Aus diesem Grund wirkt keine Einkopplung auf die Integratorkette, so dass die Bedingungen $\hat{x}_i = 0$ für alle $1 < i \leq r$ erfüllt sind und sich das stationäre Verstärkungsverhältnis zu

$$V_w = \frac{\tilde{c}_0 \cdot \hat{x}_{1,\infty}}{c_0 \cdot x_{1,\infty}} \quad (5.61)$$

vereinfacht. Mit der Beziehung

$$\hat{e}_\infty = 0 \quad \Rightarrow \quad y_\infty = c_0 \cdot x_{1,\infty} = \hat{x}_{1,\infty} = \hat{y}_\infty \quad (5.62)$$

ergibt sich

$$V_w = \frac{\tilde{c}_0 \cdot \hat{x}_{1,\infty}}{\hat{x}_{1,\infty}} = \tilde{c}_0. \quad (5.63)$$

Die gesuchte Verstärkung für den Sollwert hängt ausschließlich von einem Koeffizienten der frei wählbaren Zustandsrückführung ab und ist somit bekannt.

In den Beiträgen [87] und [88] wird ein hochverstärkungsbasierter Beobachter genutzt, um den Relativgrad des vorliegenden Zweimassensystems unter Beibehaltung der Minimalphasigkeit zu reduzieren.

Allerdings wirft ein solcher Beobachter eine zusätzliche Problemstellung auf. Die Verstärkungsfaktoren h_0, \dots, h_{r-1} für die Fehlerrückführung beeinflussen die Qualität der Zustandsschätzung und dadurch auch die Stabilität der Regelkreises. Bei deren Einstellung ist als besondere Schwierigkeit ein Zielkonflikt zu lösen. Einerseits muss der Beobachter hinreichend schnell einschwingen (hierzu sind hohe Verstärkungen nötig), andererseits darf Messrauschen nicht in störendem Maße verstärkt werden (hierzu müssen niedrige Verstärkungen gewählt werden). Falls der Zielkonflikt eine Lösung zulässt, bleibt die Fragestellung, was quantitativ unter „hinreichend schnell“ zu verstehen ist. Die Antwort darauf hängt maßgeblich von der Dynamik der Regelstrecke ab. Sofern mittels worst-case-Abschätzung eine Eingrenzung möglich ist, lassen sich Mindestwerte für die Fehlerrückführungen ermitteln. Andernfalls müssen adaptive Einstellregeln in Erwägung gezogen werden, wobei die Wahl eines Adaptionsgesetzes, sowie die zugehörige Stabilitätsanalyse nicht-triviale Problemstellungen beinhalten.

Die beschriebene Möglichkeit, mit dem Auskoppelvektor \tilde{c} eine beliebige Linearkombination der Zustände zu generieren und als neue Ausgangsgröße zu definieren, erlaubt eine weitreichende Modifikation der Regelstrecke. Unter der wenig restriktiven Annahme eines Einkoppelvektors $b \in \mathbb{R}^n$ mit der Eigenschaft $b \neq 0$ existiert ein nicht-orthogonaler Auskoppelvektor $\tilde{c}^T \in \mathbb{R}^n$, der zu einem Skalarprodukt $\tilde{c}b \neq 0$ führt. Dadurch lässt sich jeder beliebige Relativgrad $r \in \mathbb{N}$ auf 1 reduzieren. Gleichzeitig lässt sich durch \tilde{c} auch Minimalphasigkeit erzwingen.

Die Eigenschaften des physikalischen Systems mit ursprünglichem Auskoppelvektor c und Ausgang $y(t)$ verlieren ihre Bedeutung und werden durch jene Attribute des modifizierten Systems mit Ausgang $\tilde{y}(t) = \tilde{c}x(t)$ ersetzt. Weil die erforderliche Minimalphasigkeit von \tilde{c} bestimmt wird, kann durch Zustandsauskopplung grundsätzlich auch ein System geregelt werden, das bezüglich des technologischen Ausgangs y nicht-minimalphasiges Verhalten zeigt.

Kapitel 6

Hochverstärkungsbasierte Regelung des Zweimassensystems

Konventionelle Regler für das Zweimassensystem, wie sie in Abschnitt 2.3 beschrieben sind, verlangen eine vollständig bekannte Regelstrecke und sind daher nicht länger einsetzbar, wenn ein Zweimassensystem mit unbekanntem Parametern zu regeln ist. Soll die Voraussetzung der exakten Systemkenntnis fallen gelassen werden, bieten sich hochverstärkungsbasierte Regler an. Wie eine Streckenanalyse für das Zweimassensystem aufzeigt, ist eine Regelung der Motordrehzahl, sowie eine Zustandsregelung mit Funnel-Control möglich. Die Lastdrehzahl lässt sich allerdings mit dieser Methode nicht regeln, da hierfür ein zu hoher Relativgrad vorliegt.

Es ist hervorzuheben, dass weder die physikalischen Parameter der Regelstrecke, noch der Verlauf der Reibcharakteristik als bekannt vorausgesetzt werden. Die hier betrachteten hochverstärkungsbasierten Regler verlangen lediglich strukturelle Eigenschaften der Strecke, wie bekanntes und konstantes Vorzeichen der instantanen Verstärkung, stabile Nulldynamik und Relativgrad 1.

Wird als Regelgröße eine der beiden Drehzahlen ω_1 oder ω_2 vorgegeben, bestehen die in Kapitel 2 dargestellten Nachteile. Diese lassen sich durch den Einsatz einer Zustandsregelung vermeiden, für die mit $\hat{y}(t) := y(t) = cx(t)$, $c^T \in \mathbb{R}^3$, eine Linearkombination aller Zustände als Regelgröße angesetzt wird. Dadurch tritt der Regelfehler $e(t) = \hat{y}(t) - y_{\text{ref}}(t)$ an die Stelle der aus technologischer Sicht relevanten Abweichung $\omega_{\text{ref}}(t) - \omega_2(t)$ zwischen Prozessgröße $\omega_2(t)$ und deren Vorgabe $\omega_{\text{ref}}(t)$.

Das Ziel ist der Entwurf eines Reglers, der für jeden beliebigen Sollwert

$$y_{\text{ref}} \in W^{1,\infty}(\mathbb{R}_{\geq 0}; \mathbb{R}) \quad (6.1)$$

(beschränkte Funktion mit essentiell beschränkter Ableitung) und für jedes beschränkte Störsignal $\hat{d} \in \mathcal{L}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R})$ die folgenden Forderungen garantiert, die in Abbildung 6.1 illustriert sind:

- (i) Der Regelfehler $e(t) = \hat{y}(t) - y_{\text{ref}}(t)$ verläuft in einem vorgegebenen Trichter.

- (ii) Ab einem festgelegten Zeitpunkt $\tau > 0$ bleibt der Regelfehler $|e(t)|$ kleiner als eine vorgegebene Schranke $\lambda_\tau > 0$.
- (iii) Sämtliche Zustandsgrößen des Regelkreises bleiben beschränkt.
- (iv) In der Nähe der Gleichgewichtslage (für „große“ t) ist der Regelfehler „klein“.
- (v) Stationär, d.h. für $\dot{x}(t) = 0$, ist die Abweichung $\omega_2(t) - \omega_{\text{ref}}(t)$ proportional zum Regelfehler $e(t)$.
- (vi) Torsionsschwingungen der elastischen Welle werden aktiv bedämpft.
- (vii) Sämtliche Störungen $\hat{d} \in \mathcal{L}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R})$ werden unterdrückt.

Die Regelziele (i)–(iii) und (vii) werden durch einen hochverstärkungsbasierten, zeitvarianten Proportionalregler

$$\begin{aligned}
 e(t) &= \hat{y}(t) - y_{\text{ref}}(t) \\
 v(t) &= -k(t) e(t) \\
 k(t) &= \frac{1}{1 - \psi(t)|e(t)|}
 \end{aligned} \tag{6.2}$$

erreicht, der in Kapitel 4 beschrieben ist. Aufgrund von Theorem 4.6 ist für einen solchen Regler die Einhaltung von Punkt (i) und (iii) garantiert. Wird beispielsweise, wie in Abbildung 6.1 erfolgt, als Trichterrand die Funktion $\psi : t \mapsto \min\{t/\tau, 1\}/\lambda_\tau$ angesetzt, folgt aus (i), dass mit dem Freiheitsgrad $\lambda_\tau > 0$ ab dem Zeitpunkt τ ein gewünschtes Maß an Genauigkeit eingestellt werden kann und (ii) somit zutrifft. Durch die Verwendung eines PI-Reglers in Reihe mit dem Proportionalregler (6.2) wird Ziel (iv) erreicht. Die im Punkt (vi) geforderte Dämpfung entsteht durch Gewichtung des Torsionswinkels in der Regelgröße. Um dadurch die stationäre Genauigkeit in der Prozessgröße nicht zu beeinträchtigen und (v) zu garantieren, wird der Torsionswinkel einer Hochpassfilterung unterzogen. Durch die Erweiterung des Reglers mittels PI-Regler und Hochpassfilter wird Ziel (iii), die Stabilität des Regelkreises, nicht gefährdet.

Der Funktionsweise des Regelungskonzeptes liegt nachfolgend skizziertes Prinzip zu Grunde, das in den Theoremen 6.5 bis 6.9 für nichtlineare Zweimassensysteme formuliert und bewiesen ist.

Für die dort betrachteten Prozesse lässt sich der Regler (6.2) mit $y_{\text{ref}} \in W^{1,\infty}(\mathbb{R}_{\geq 0}; \mathbb{R})$ und $u(t) = v(t)$ ansetzen, wobei sich der resultierende Regelkreis in der Form

$$\begin{aligned}
 \dot{e}(t) &= -\alpha k(t)e(t) + \zeta(t), \quad \alpha > 0 \\
 k(t) &= \frac{1}{1 - \psi(t)|e(t)|} \\
 \zeta(t) &= (\mathbf{T}e + y_{\text{ref}})(t) + p(t) - \dot{y}_{\text{ref}}(t)
 \end{aligned} \tag{6.3}$$

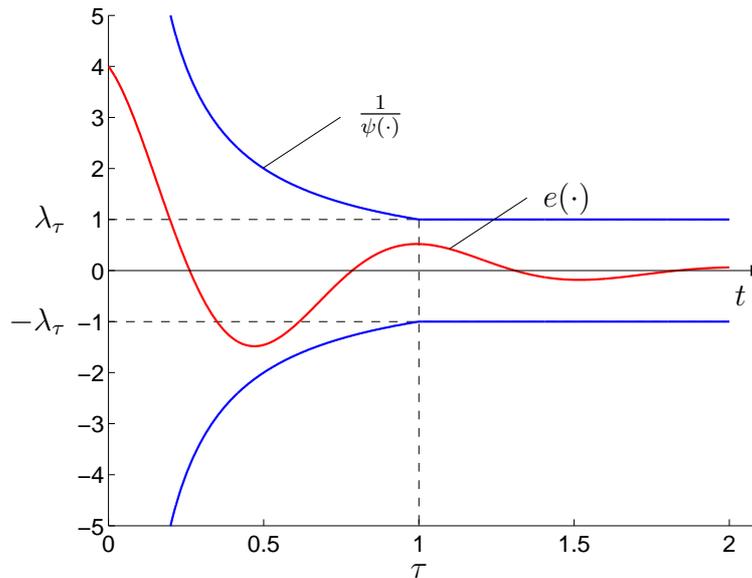


Abbildung 6.1: Beispiel eines Trichters \mathcal{F}_φ , der die Regelziele (i)–(iii) garantiert.

darstellen lässt. Darüberhinaus lässt sich zeigen, dass der Operator \mathbf{T} der in Definition 4.2 angegebenen Operatorklasse \mathcal{Z} angehört und darüberhinaus auch $p \in \mathcal{L}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R})$ gilt, so dass zusammen mit $y_{\text{ref}} \in W^{1,\infty}(\mathbb{R}_{\geq 0}; \mathbb{R})$ das essentielle Ergebnis $\zeta \in \mathcal{L}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R})$ festzuhalten ist, wenn an dieser Stelle die (technische) Problematik einer möglicherweise endlichen Entweichzeit ignoriert wird.

Aus der Darstellung (6.3) ist folgendes Verhalten des Regelkreises abzulesen. Sobald sich der Regelfehler $|e(t)|$ hinreichend nahe auf $1/\psi(t)$ zubewegt, d.h. der Punkt $(t, e(t))$ befindet sich in der Nähe des Trichterrandes, nimmt $\psi(t)|e(t)|$ näherungsweise den Wert 1 an. Durch die Struktur der Funktion $s \mapsto 1/(1-s)$ entsteht in dieser Situation eine ausreichend hohe Verstärkung $k(t)$, die

$$\dot{e}(t) < 0 \quad \text{für } e(t) > 0 \quad \text{und} \quad \dot{e}(t) > 0 \quad \text{für } e(t) < 0 \quad (6.4)$$

erzwingt und dadurch $(t, e(t))$ am Verlassen des Trichters hindert. Die Struktur der Differentialgleichung (6.3) ist ausschlaggebend, damit durch eine hohe Verstärkung der Regelfehler klein gehalten wird. Darin ist jedoch keine Aussage über asymptotische Konvergenz enthalten. Ein beliebig kleiner Regelfehler impliziert nicht, dass der Regelkreis eine Gleichgewichtslage erreicht.

6.1 Regelung des linearen Modells

An dieser Stelle wird auf das Ergebnis der Modellbildung in Kapitel 2 zurückgegriffen und die dort erarbeitete Zustandsbeschreibung

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t) + b_L M_2(t), & x(0) &= x_o \in \mathbb{R}^3 \\ y(t) &= cx(t), & c^T &\in \mathbb{R}^3 \end{aligned} \quad (6.5)$$

mit der Stellgröße $u(t) \in \mathbb{R}$ (Drehmoment der Antriebsmaschine), der stückweise stetigen Störgröße $M_2 \in \mathcal{L}_{lok}^1(\mathbb{R}_{\geq 0}; \mathbb{R})$ (Lastmoment) und den Größen

$$A = \begin{bmatrix} -d_w/J_2 & k_w/J_2 & d_w/J_2 \\ -1 & 0 & 1 \\ d_w/J_1 & -k_w/J_1 & -d_w/J_1 \end{bmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ 1/J_1 \end{pmatrix} \quad \text{und} \quad b_L = \begin{pmatrix} -1/J_2 \\ 0 \\ 0 \end{pmatrix} \quad (6.6)$$

angesetzt. Der Auskoppelvektor $c^T \in \mathbb{R}^3$ ist von der Wahl der Regelgröße abhängig und beeinflusst maßgeblich den Relativgrad der Regelstrecke sowie die Stabilität der Nulldynamik.

Obwohl in den Ergebnissen für das nichtlineare Zweimassensystem diejenigen für das lineare Zweimassensystem enthalten sind, erfolgt dennoch eine gesonderte Darstellung des linearen Falles. Vor dem Hintergrund, dass sich am reibungsfreien Zweimassensystem die Zustandsverläufe und damit vor allem die Torsionsschwingungen der Welle explizit berechnen lassen, wird die Wirkungsweise der in Abschnitt 6.2 vorgestellten Methode zur aktiven Bedämpfung der Torsionsschwingung anhand der linearen Darstellung transparent.

6.1.1 Regelung der Motordrehzahl

Für den Fall, dass die Motordrehzahl $\omega_1(t)$ als Regelgröße definiert wird, ist der Auskoppelvektor

$$c = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \quad (6.7)$$

festzulegen.

Theorem 6.1

Das lineare Zweimassensystem (6.5), (6.6) mit dem durch (6.7) gegebenen Auskoppelvektor gehört der Systemklasse \mathcal{S} an. Demzufolge garantiert das Regelgesetz (6.2), in dem $u(t) = v(t)$ und $\hat{y}(t) = y(t)$ zu setzen ist, die Einhaltung der Regelziele (i), (iii) und (vii).

Beweis

Aus dem Zustandsmodell (6.5), (6.6), (6.7) ergibt sich durch

$$F(s) = c(sI_3 - A)^{-1}b \quad (6.8)$$

die zugehörige Übertragungsfunktion zu

$$F(s) = \frac{\omega_1(s)}{u(s)} = \frac{\frac{1}{J_1}s^2 + \frac{d_w}{J_1 J_2}s + \frac{k_w}{J_1 J_2}}{s^3 + d_w \frac{J_1 + J_2}{J_1 J_2} s^2 + k_w \frac{J_1 + J_2}{J_1 J_2} s}. \quad (6.9)$$

Diese bietet sich für die Untersuchung an, ob das lineare Modell die strukturellen Eigenschaften erfüllt, um der Systemklasse \mathcal{S} anzugehören und damit für hochverstärkungsbasierte Regler zugänglich zu sein.

(i) Instantane Verstärkung:

Aus der Ordnung von Zähler- und Nennerpolynom errechnet sich gemäß Gleichung (A.17) die instantane Verstärkung zu

$$V_0 = \lim_{s \rightarrow \infty} s^{n-m} F(s) = \frac{1}{J_1} \quad (6.10)$$

und ist wegen $J_1 > 0$ positiv.

(ii) Relativgrad:

Abschnitt A.6 stellt den Relativgrad linearer Systeme als Differenz der Ordnungen von Nenner- und Zählerpolynom dar. Damit ergibt sich der Relativgrad zu $r = n - m = 1$.

(iii) Minimalphasigkeit:

Minimalphasigkeit liegt nach Definition A.27 genau dann vor, wenn das Zählerpolynom der Übertragungsfunktion ein Hurwitzpolynom ist. Nachdem sämtliche Systemparameter positive Werte annehmen, besteht das Zählerpolynom in der Übertragungsfunktion (6.9) ausschließlich aus positiven Koeffizienten. Dadurch ist die notwendige und hinreichende Bedingung [29, S. 23] für ein Hurwitzpolynom erfüllt, so dass minimalphasiges Verhalten gegeben ist.

Wegen $(6.5) \in \mathcal{S}$ und $\mathcal{S} \subset \mathcal{P}$ wird durch Anwendung von Theorem 4.6 ersichtlich, dass die Motordrehzahl des Zweimassensystems (6.5) mit dem Regler (6.2) zu beherrschen ist. ■

Anhand einer Simulation soll das Verhalten des geregelten Zweimassensystems untersucht werden, dessen Parameter in Tabelle 2.2 festgelegt sind. Der Trichterrand ist mit der Funktion

$$\partial F_\psi : t \mapsto 11 \cdot e^{-20t^2} + \Delta_F, \quad \Delta_F \in \{0.1, 2\} \quad (6.11)$$

beschrieben. Die Konstante Δ_F definiert die Breite des Trichters, die für $t \rightarrow \infty$ asymptotisch erreicht wird. Ziel sei eine Festwertregelung auf die Drehzahl $\omega_{\text{ref}} = 10$ rad/s, weshalb für den Regelkreis als Sollwert eine Sprungfunktion

$$t \mapsto y_{\text{ref}}(t) = \begin{cases} 0 & \forall t < 0 \\ \omega_{\text{ref}} & \forall t \geq 0 \end{cases} \quad (6.12)$$

vorgegeben wird. Das in Abbildung 6.2 aufgetragene Simulationsergebnis lässt vermuten, dass beide Drehzahlen (und damit auch der Regelfehler) gegen konstante Werte konvergieren. Die Beobachtung, dass der Regelkreis eine Ruhelage erreicht, stützt sich allein auf die Simulation und ist nicht durch theoretische Resultate untermauert. Zwar garantiert Theorem 4.6, dass der hier verwendete Regler den Fehler $e(t) = \omega_1(t) - y_{\text{ref}}(t)$ durch den Trichter beschränkt, d.h. es gilt $|e(t)| < 1/\psi(t)$ für alle $t > 0$. Allerdings ist in Theorem 4.6 keine Aussage über die Konvergenz der Zustandsgrößen bzw. des Regelfehlers enthalten.

Wie Abbildung 6.2 zeigt, treten trotz Sprunganregung während des Einschwingvorganges nur geringfügige Schwingungen in den Drehzahlen $\omega_1(\cdot)$ und $\omega_2(\cdot)$ auf, das

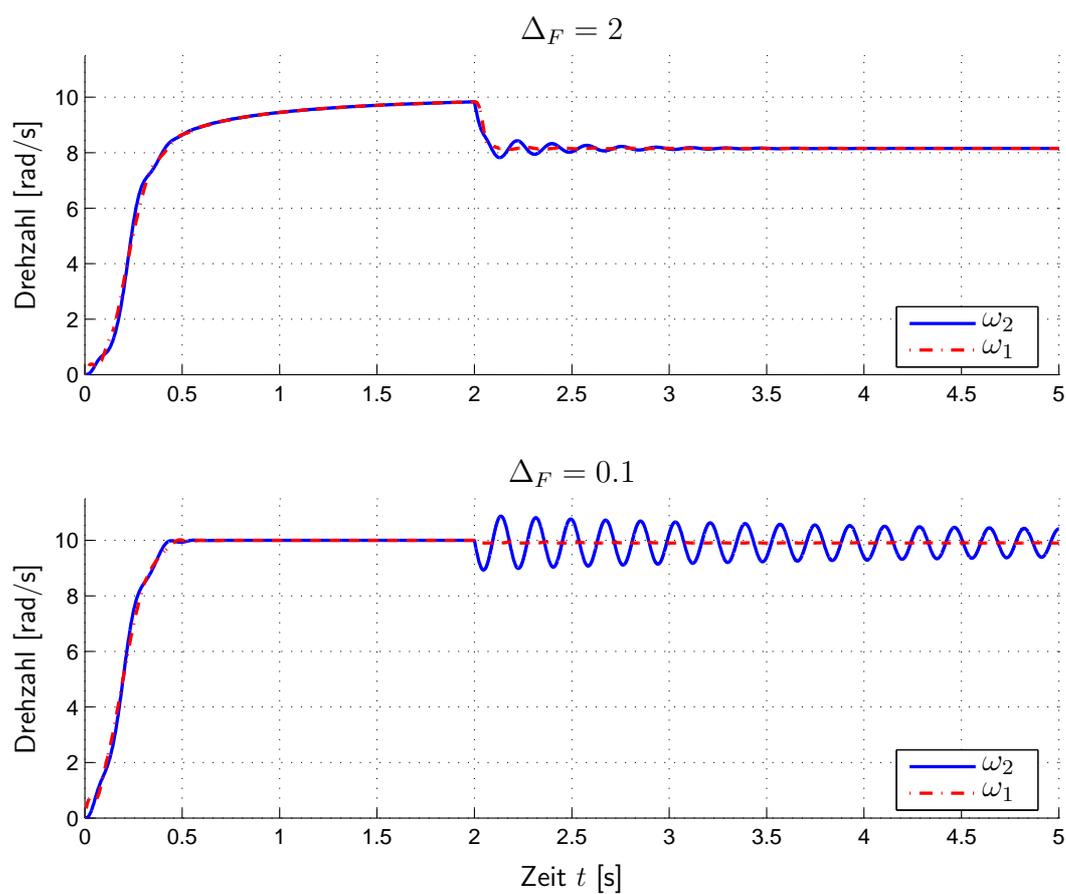


Abbildung 6.2: Funnel-Control für das Zweimassensystem bei Regelung der Motordrehzahl.

Überschwingen ist vernachlässigbar klein. Bedingt durch das integrale Verhalten der Strecke (Zweimassensystem ohne Reibung) erreicht der proportionale Regler stationäre Genauigkeit, solange kein Lastmoment angreift. Der Regelfehler $e(t)$ verschwindet asymptotisch, unabhängig von der Breite des Trichters.

Nachdem der Regelkreis eingeschwungen ist und näherungsweise die Ruhelage erreicht hat, wird ab dem Zeitpunkt $t = 2$ s ein konstantes Lastmoment von 12 Nm zugeschaltet, das ungefähr der Hälfte des maximalen Maschinenmomentes entspricht. Das obere Diagramm in Abbildung 6.2 zeigt, dass sich dadurch für $\Delta_F = 2$ die Drehzahl im Simulationsbeispiel um fast 20 % reduziert und auch für $t \rightarrow \infty$ nicht auf den Sollwert zurückkehrt. Solches Verhalten ist typisch für proportionale Regler, die ohne Integralteil den Regelfehler unter Störeinfluss nicht eliminieren können.

Im unteren Diagramm wird der Trichter weiter verengt ($\Delta_F = 0.1$) und dadurch der Regelfehler bei Belastung verringert. Die Trichterbreite entspricht für $t \rightarrow \infty$ genau 1 % vom Sollwert. Dadurch schlägt sich das Lastmoment nur in einer geringfügigen Verlangsamung der Regelgröße $\omega_1(t)$ nieder, eine Veränderung ist innerhalb der Zeichengenauigkeit nicht wahrnehmbar. Obwohl sich durch einen engeren Trichter der Regelfehler beliebig verbessern lässt, verschlechtert der enge Trichter das transiente Verhalten der Prozessgröße $\omega_2(t)$ nach dem Lastschlag deutlich. Weil der Abstand zum Trichterrand abnimmt und dadurch die Verstärkung steigt, wenn der Trichter schmaler gewählt wird, resultiert eine höhere Stellgröße. Auf diesem Wege regt die Regelung die Eigenschwingung der Welle und damit Oszillationen in $\omega_2(\cdot)$ umso ausgeprägter an, je enger sich der Trichter zusammenzieht. In der Ruhelage gilt $\omega_1(t) = \omega_2(t)$, so dass stationär auch das Verhalten der Prozessgröße durch einen kleinen Wert von Δ_F verbessert wird.

Bei der Beurteilung der Regelgüte ist das Verhalten der Regelgröße $\omega_1(t)$ bzw. des Regelfehlers $e(t) = \omega_1(t) - y_{\text{ref}}(t)$ zu bewerten, nicht der Verlauf der Prozessgröße $\omega_2(t)$. Auf dieser Grundlage ist festzustellen, dass alle Regelziele außer (vi) durch die Regelung erfüllt werden.

Dennoch ist das Ergebnis, dass die Prozessgröße entweder schlecht gedämpfte Schwingungen ausführt (d.h. (vi) ist nicht erfüllt) oder eine zu große stationäre Abweichung unter Last zeigt, zur Regelung eines Antriebes unbrauchbar.

6.1.2 Regelung der Lastdrehzahl

Wird die Lastdrehzahl $\omega_2(t)$ als Regelgröße festgelegt, ist der Auskoppelvektor

$$c = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \quad (6.13)$$

anzusetzen.

Theorem 6.2

Das lineare Zweimassensystem (6.5), (6.6) mit dem durch (6.13) gegebenen Auskoppelvektor gehört nicht der Systemklasse \mathcal{S} an.

Beweis

Aus dem Zustandsmodell (6.5), (6.6), (6.13) ergibt sich durch

$$F(s) = c(sI_3 - A)^{-1}b \quad (6.14)$$

die zugehörige Übertragungsfunktion zu

$$F(s) = \frac{\omega_2(s)}{u(s)} = \frac{\frac{d_w}{J_1 J_2} s + \frac{k_w}{J_1 J_2}}{s^3 + d_w \frac{J_1 + J_2}{J_1 J_2} s^2 + k_w \frac{J_1 + J_2}{J_1 J_2} s}. \quad (6.15)$$

Eine Analyse aller notwendigen Eigenschaften für die Zugehörigkeit zu \mathcal{S} ergibt folgendes Bild:

(i) Instantane Verstärkung:

Die instantane Verstärkung errechnet sich zu $V_0 = d_w / (J_1 J_2)$ und ist daher positiv.

(ii) Relativgrad:

Nachdem das Zählerpolynom erster Ordnung ist, das Nennerpolynom aber dritter Ordnung, ergibt sich ein Relativgrad von $r = 3 - 1 = 2$.

(iii) Minimalphasigkeit:

Das Zählerpolynom besitzt genau eine Nullstelle bei $s = -k_w / d_w < 0$. Deren Realteil ist negativ, die Regelstrecke ist somit minimalphasig.

■

Bedingt durch den Relativgrad erfüllt diese Regelstrecke nicht die notwendigen Voraussetzungen, um sich mit hochverstärkungsbasierten Reglern beherrschen zu lassen, ohne dass dabei auf die zeitliche Ableitung von $\omega_2(t)$ zurückgegriffen wird.

6.1.3 Zustandsregelung

Nicht alle der genannten Regelziele werden eingehalten, wenn lediglich ein Zustand als Regelgröße festgesetzt wird. Vor allem im Hinblick auf die Prozessgröße ist von der Regelung eines einzigen Zustandes Abstand zu nehmen und stattdessen eine Linearkombination der Systemzustände als Regelgröße zu definieren. Hierfür ist der Auskopplvektor

$$c = \begin{pmatrix} c_1 & c_2 & c_3 \end{pmatrix} \in \mathbb{R}^{1 \times 3} \quad (6.16)$$

zu wählen, der eine Linearkombination aus beiden Drehzahlen und dem Torsionswinkel festlegt. Gleichung (2.28) gibt die entsprechende Übertragungsfunktion durch

$$F(s) = \frac{\frac{c_3}{J_1} s^2 + \frac{(c_2 J_2 + (c_1 + c_3) d_w)}{J_1 J_2} s + \frac{(c_1 + c_3) k_w}{J_1 J_2}}{s^3 + d_w \frac{J_1 + J_2}{J_1 J_2} s^2 + k_w \frac{J_1 + J_2}{J_1 J_2} s} \quad (6.17)$$

vor, deren Zählerpolynom sich mit den drei Freiheitsgraden $c_1, c_2, c_3 \in \mathbb{R}$ beliebig einstellen lässt. Insofern ist eine Relativgradreduktion, sowie eine beliebige Platzierung

der Nullstellen durch die Zustandsrückführung möglich, und es ist daher zu erwarten, dass bei entsprechender Wahl der Auskopplung die Strecke in die Systemklasse \mathcal{S} fällt. Allerdings wird die Lage der Nullstellen (und damit die erforderliche Minimalphasigkeit der Regelstrecke) nicht nur von den Koeffizienten der Auskopplung $c^T \in \mathbb{R}^3$ beeinflusst, sondern hängt ebenso auch von den unbekanntem Parametern der Strecke ab.

Im Sonderfall einer Realisierung der Strecke in Regelungsnormalform (RNF) hängt das Zählerpolynom der Übertragungsfunktion ausschließlich vom frei wählbaren Rückführvektor $c^T \in \mathbb{R}^3$ ab. Weil die unbekanntem Systemparameter keinen Einfluss auf das Zählerpolynom erhalten, können dessen Wurzeln beliebig plaziert werden, die Gewährleistung der Minimalphasigkeit bereitet somit keine Schwierigkeit.

Die Realisierung des Zweimassensystems, welche auf physikalischen Zustandsgrößen basiert, stimmt jedoch nicht mit der Regelungsnormalform überein. In Gleichung (6.17) schlägt sich dies darin nieder, dass die unbekanntem Systemparameter J_1 , J_2 , k_w und d_w in den Koeffizienten des Zählerpolynoms vertreten sind. Dadurch geht die Kenntnis über die Positionen der Nullstellen zwar verloren, die erforderliche Minimalphasigkeit lässt sich dennoch garantieren. Allein die in (2.10) festgestellte Tatsache, dass sämtliche Parameter positive Vorzeichen besitzen, ist als Voraussetzung notwendig, um im Zähler ein Hurwitzpolynom zu erhalten und dadurch die Nullstellen unabhängig von den unbekanntem Beträgen der Parameter in der stabilen linken Halbebene zu plazieren.

Theorem 6.3

Die Bedingungen

$$c_3 > 0, \quad c_1 > -c_3, \quad c_2 \geq 0 \quad (6.18)$$

stellen ein hinreichendes Kriterium dar, damit das lineare Zweimassensystem (6.5), (6.6) mit dem durch (6.16) gegebenen Auskoppelvektor der Systemklasse \mathcal{S} angehört und demzufolge das Regelgesetz (6.2), in dem $v(t) = u(t)$, $\hat{y}(t) = y(t)$ und $y_{\text{ref}}(t) = (c_1 + c_3)\omega_{\text{ref}}(t)$ zu setzen ist, die Einhaltung der Regelziele (i), (iii) und (vii) garantiert.

Beweis

Aus der Übertragungsfunktion (6.17) lassen sich folgende Eigenschaften ablesen:

- (i) Instantane Verstärkung:

Unter der Bedingung $c_3 \neq 0$ errechnet sich die instantane Verstärkung zu $V_0 = c_3/J_1$. Weil physikalisch bedingt das Massenträgheitsmoment J_1 positiv ist, besitzt die instantane Verstärkung das positive Vorzeichen von c_3 .

- (ii) Relativgrad:

Unter der Bedingung $c_3 \neq 0$ ergibt sich ein Relativgrad von $r = 3 - 2 = 1$.

- (iii) Minimalphasigkeit:

Minimalphasigkeit erfordert ein Zählerpolynom mit Wurzeln, die ausschließlich negative Realteile aufweisen. Als Kriterium hierfür eignet sich beispielsweise der Hurwitz-Test [29, S. 23]. Daraus lassen sich zusammen mit Ungleichung (2.10) die notwendigen und hinreichenden Bedingungen

$$c_3 > 0, \quad c_1 > -c_3, \quad c_2 > -\frac{d_w}{J_2}(c_1 + c_3) \quad (6.19)$$

für die Auskoppelkoeffizienten ableiten. Die Koeffizienten des Zählerpolynoms sind von den unbekanntem Systemparametern abhängig, weshalb diese auch in Ungleichung (6.19) auftreten. Im Falle von c_2 ist die untere Schranke daher nicht als Zahlenwert explizit bestimmbar. Allerdings folgt aus (2.10) und $c_1 + c_3 > 0$, dass c_2 einen unbekanntem, aber *negativen* Wert übersteigen muss. Damit sind die Bedingungen (6.19) auch erfüllt, wenn stattdessen die restriktiveren, aber parameterunabhängigen Ungleichungen (6.18) angesetzt werden. Deren Einhaltung garantiert ebenfalls eine minimalphasige Übertragungsfunktion. ■

Theorem 6.3 zufolge ist es möglich, durch Zustandsauskopplung den Relativgrad zu reduzieren und dabei die Nebenbedingung der Minimalphasigkeit einzuhalten, obwohl weder die Systemparameter bekannt sind, noch das Zweimassensystem in Regelungsnormalform vorliegt. Durch die vorgeschlagene Zustandsauskopplung ist das Zweimassensystem der Systemklasse \mathcal{S} zuzuordnen.

Bemerkenswert ist die Tatsache, dass Bedingung (6.18) keine vollständige Zustandsrückführung fordert, die Parametrierung $c_1 = 0$ bzw. $c_2 = 0$ ist eine zulässige Wahl. Sowohl die Lastdrehzahl, als auch der Verdrehwinkel können damit unberücksichtigt bleiben, wenn allein die Stabilität des Regelkreises maßgebend ist. Daran zeigt sich, dass sich die Regelung der Motordrehzahl als Sonderfall der Zustandsregelung auffassen lässt.

Insbesondere muss wegen $c_2 \geq 0$ der Verdrehwinkel nicht notwendigerweise gemessen werden, um ein stabiles Gesamtsystem zu erhalten. Auf einen entsprechenden Sensor kann aus Sicht der Stabilität verzichtet werden. Dennoch kann die Gewichtung des Verdrehwinkels mit einem Faktor $c_2 > 0$ sinnvoll sein, speziell wenn, wie im Fall des vorliegenden Zweimassensystems, kein zusätzlicher Messaufwand entsteht. Durch das Einbeziehen des Verdrehwinkels in die Linearkombination erhält der Regler Information über Schwingungen der elastischen Welle und kann deshalb dämpfend einwirken.

Die Freiheitsgrade, die durch (6.18) gegeben sind, werfen die Frage nach einer geeigneten Wahl des Auskoppelvektors $c^T \in \mathbb{R}^3$ auf. Im relevanten Falle, dass Minimalphasigkeit vorliegt und die Auskoppelfaktoren den Ungleichungen (6.19) genügen, tritt eine Pol-Nullstellen-Kürzung und damit der Verlust der vollständigen Beobachtbarkeit ausschließlich für die Konstellation

$$c_2 = 0 \quad \text{und} \quad c_1 = c_3 \frac{J_2}{J_1} \quad (6.20)$$

ein. Daher ist $c_2 \neq 0$, also die Gewichtung des Torsionswinkels in der Rückführung, hinreichend für den Erhalt der vollständigen Zustandsbeobachtbarkeit. Dennoch besteht auch ohne Verdrehwinkel Beobachtbarkeit, solange $c_1 \neq J_2/J_1 \cdot c_3$ gesetzt wird. Wenn beide Auskoppelkoeffizienten unterschiedliches Vorzeichen aufweisen ($\text{sign}(c_1) < 1$, $\text{sign}(c_3) = 1$) liegt Beobachtbarkeit vor, unabhängig von c_2 . In diesem Falle kann ohne Bedenken der Torsionswinkel unberücksichtigt bleiben.

Falls ein Wertebereich für das Verhältnis J_2/J_1 abgeschätzt werden kann, darf die Verstärkung der Lastdrehzahl ein positives Vorzeichen besitzen ($\text{sign}(c_1) = 1$). Ohne

eine solche Abschätzung kann Beobachtbarkeit nicht garantiert werden, oder es muss der Verdrehwinkel gewichtet werden ($c_2 \neq 0$).

Damit ist gezeigt, dass lediglich für den Fall $c_2 = 0$ Beobachtbarkeit verloren gehen kann, für $c_2 \neq 0$ ist dagegen stets vollständige Beobachtbarkeit gegeben.

Allerdings ist durch vorhandene Beobachtbarkeit nicht notwendigerweise die Bedämpfung von Schwingungen in der Welle zufriedenstellend gut. Beobachtbarkeit ist lediglich eine notwendige Voraussetzung dafür, dass dem Regler Information über vorhandene Schwingungen zur Verfügung steht und dieser die Möglichkeit zu deren Bedämpfung erhält. Inwiefern eine Bedämpfung tatsächlich geschieht, hängt nicht allein von der Beobachtbarkeit ab.

Obwohl unbeobachtbare Vorgänge im System definitionsgemäß nicht am Ausgang sichtbar sind und daher im Übertragungsverhalten auch keinen störenden Einfluss besitzen, ist Beobachtbarkeit dennoch aus zwei Gründen zu befürworten:

- (i) Bei der Auslegung eines Reglers darf in der Praxis das Interesse nicht nur dem Systemausgang gelten, sondern muss sich auf das Verhalten des Gesamtsystems beziehen. Unnötige Schwingungen sind stets zu vermeiden, auch wenn diese in der Regelgröße selbst nicht auftreten, um das System nicht einer überflüssigen Belastung auszusetzen.
- (ii) Im untersuchten Fall tritt der Verlust der Beobachtbarkeit lediglich aus der Sicht der Rückführung $y(t) = cx(t)$ ein. Nur in diesem Signal sind die Eigenschwingungen der Welle unter den genannten Umständen nicht zu erkennen, weswegen der Regler diese nicht unterdrücken kann. Am physikalischen Ausgang $\omega_2(t)$ sind Schwingungen jedoch beobachtbar. Die zugehörige Beobachtbarkeitsmatrix lautet

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{d_w}{J_2} & \frac{k_w}{J_2} & \frac{d_w}{J_2} \\ -\frac{d_w^2 J_1 + k_w J_2 J_1 - d_w^2 J_2}{J_2^2 J_1} & -\frac{d_w k_w (J_1 + J_2)}{J_2^2 J_1} & \frac{-d_w^2 J_1 + k_w J_2 J_1 - d_w^2 J_2}{J_2^2 J_1} \end{bmatrix} \quad (6.21)$$

woraus sich die Determinante

$$\det(Q) = \frac{k_w^2}{J_2^2} \neq 0 \quad (6.22)$$

ergibt. Es kann aus technischen Gründen stets $k_w > 0$ und $J_2 < \infty$ vorausgesetzt werden: $k_w = 0$ beschreibt ein System ohne Kopplung der beiden Massen, also das Fehlen einer Welle; $J_2 = \infty$ ist der Grenzfall einer unendlich großen Trägheit, so dass keine Drehzahländerungen möglich sind. In beiden Fällen ist eine Drehzahlregelung für die Lastmasse hinfällig. Damit darf die Regelstrecke aus Sicht der Prozessgröße $\omega_2(t)$ stets als vollständig beobachtbar betrachtet werden, also dann, wenn die physikalisch gegebene Auskopplung $\omega_2(t) = (1, 0, 0)x(t)$ betrachtet wird.

Dies zeigt, dass die Eigenschwingungen beim Verlust der Beobachtbarkeit die Regelgüte verschlechtern, was vom Regler nicht verhindert werden kann.

Desweiteren ist für eine Anwendung des Regelgesetzes (6.2) zu bedenken, dass proportionale Regler prinzipbedingt kein internes Modell für konstante Anregungssignale enthalten und daher nicht für die Beseitigung stationärer Regelfehler ausgelegt sind. Aus diesem Grund ist (6.2) um einen integralen Anteil zu erweitern, wenn asymptotisch ein konstanter Sollwert erreicht werden soll. Diese Vorgehensweise ist in Abschnitt 4.6 erläutert und wird hier angewendet.

Um den Relativgrad des Zweimassensystems nicht zu erhöhen, ist ein internes Modell mit Durchgriff einzufügen, das als zusätzlicher PI-Regler agiert und in Reihe mit dem zeitvarianten Proportionalregler (6.2) anzuordnen ist. Ein solcher PI-Regler weist die Übertragungsfunktion

$$F_{PI}(s) = \frac{u(s)}{v(s)} = \frac{k_P s + k_I}{s}, \quad k_P, k_I > 0 \quad (6.23)$$

auf. Die Einschränkung von k_P und k_I auf positive Werte legt die Nullstelle der Übertragungsfunktion auf $s = -k_I/k_P < 0$, bedingt dadurch minimalphasiges Verhalten und führt zusätzlich auf eine positive instantane Verstärkung.

Es lässt sich auf direktem Wege zeigen, dass die erforderlichen Eigenschaften der Regelstrecke auch bei Verwendung eines zusätzlichen integralen Anteils erhalten bleiben. Die strukturellen Eigenschaften für den Einsatz hochverstärkungsbasierter Regler sind daher erfüllt.

Der Auskoppelvektor in Gleichung (6.16) definiert das Ausgangssignal

$$y(\cdot) = c_1 \omega_2(\cdot) + c_2 \varphi_w(\cdot) + c_3 \omega_1(\cdot) \quad (6.24)$$

als Linearkombination aller Zustände, das die Regelgröße bildet. Um im stationären Zustand eine Verstärkung von 1 für das zustandsgeregelte Zweimassensystem zu erhalten, muss (wie bei Zustandsreglern üblich und für den allgemeinen Fall in Abschnitt 5.2.3 beschrieben) eine Skalierung des Sollwertes vorgenommen werden. Falls der Torsionswinkel in der Rückführung gewichtet wird, hängt der benötigte Vorfaktor von dieser Zustandsgröße ab, die ihrerseits eine Funktion des unbekanntes Lastmomentes und der unbekanntes Steifigkeit der Welle ist. Daher kann unter Einwirkung eines Lastmomentes die passende Sollwertverstärkung nicht berechnet werden. Falls dagegen der Verdrehwinkel nicht gewichtet wird, verliert der Vorfaktor seine Abhängigkeit von den unbekanntes Größen. Unbeeinflusst von möglicherweise wirkenden Lastmomenten kann der Faktor

$$V = c_1 + c_3 > 0 \quad (6.25)$$

zur notwendigen Skalierung des Sollwertes aus bekannten Größen einfach errechnet werden. Wird für $\omega_2(t)$ die Drehzahl $\omega_{\text{ref}}(t)$ gewünscht, ist dem Regelkreis der Sollwert $y_{\text{ref}}(t) = V \omega_{\text{ref}}(t)$ vorzugeben.

Im Folgenden wird die beschriebene Regelstrategie auf ein reibungsfreies Zweimassensystem angewendet. Den Simulationen liegen die Parameter aus Tabelle 2.2 zugrunde, der Vorfilter ist gemäß Gleichung (6.25) gegeben.

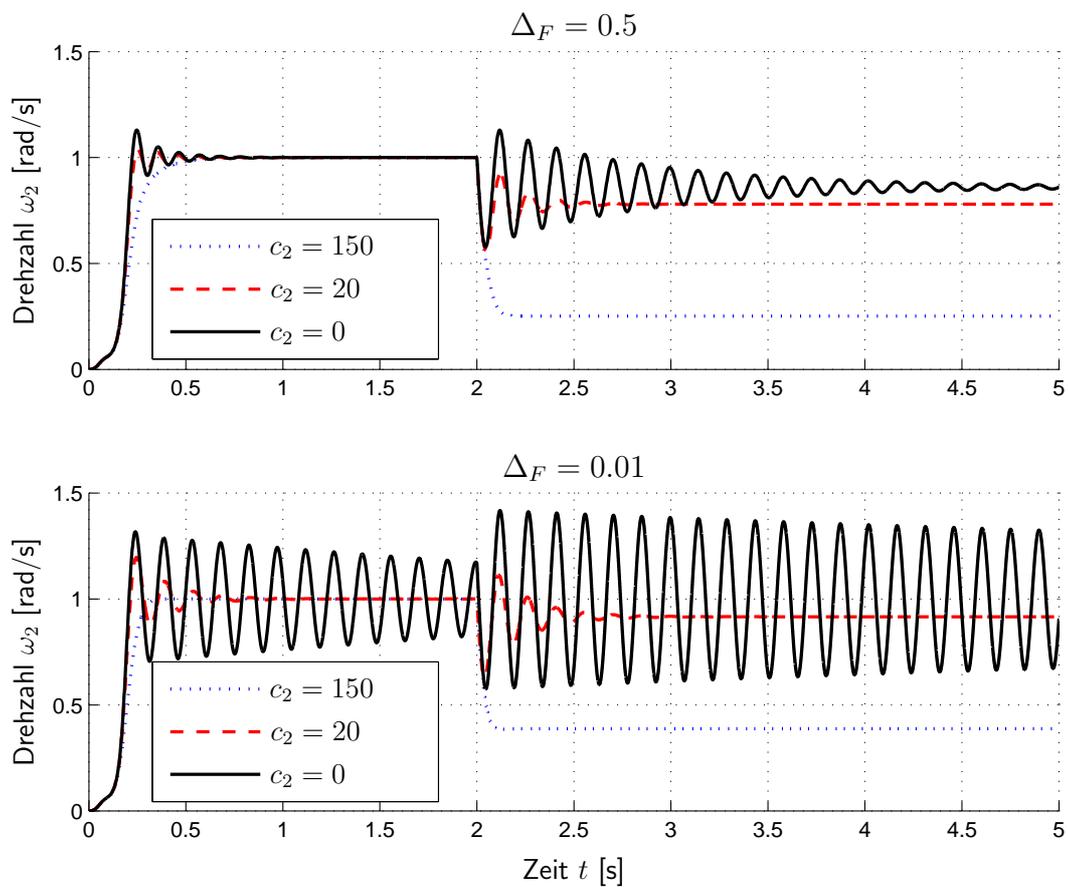


Abbildung 6.3: Funnel-Control für das Zweimassensystem mit Zustandsrückführung $y(t) = (1, c_2, 2) x(t)$.

Abbildung 6.3 zeigt das Simulationsergebnis für die Rückführung $c = (1, c_2, 2)$, wobei der Koeffizient c_2 zwischen den drei Werten $c_2 \in \{0, 20, 150\}$ variiert. Der Trichterrand ist durch die Funktion

$$\partial F_\psi : t \mapsto 15 \cdot e^{-80t^2} + \Delta_F, \quad \Delta_F \in \{0.01, 0.5\} \quad (6.26)$$

gegeben, deren Parameter Δ_F im oberen Diagramm auf den Wert 0.5 gesetzt ist, im unteren Diagramm auf 0.01. Als Vorgabe für die Prozessgröße sei $\omega_{\text{ref}} = 1$ rad/s gewählt, woraus sich mit $V = c_1 + c_3 = 3$ für den Sollwert

$$t \mapsto y_{\text{ref}}(t) = \begin{cases} 0 & \forall t < 0 \\ V\omega_{\text{ref}} & \forall t \geq 0 \end{cases} \quad (6.27)$$

ergibt. Ist der Torsionswinkel in der Linearkombination (6.24) z. B. durch $c_2 = 150$ ausreichend hoch gewichtet, dämpft der Regler die Eigenschwingung im erwünschten Maß. Durch den Sprung werden daher weder störende Oszillationen angeregt, noch entsteht Überschwingen. Je schwächer $\varphi_w(t)$ jedoch verstärkt wird, desto weniger greift der Regler dämpfend ein, so dass vermehrt Schwingungen entstehen.

Die Zustandsrückführung ändert nichts am integralen Verhalten der Strecke, so dass der proportionale Regler im eingeschwungenen Zustand den Fehler 0 erreicht. Ab dem Zeitpunkt $t = 2$ s greift ein konstantes Lastmoment von 5 Nm an. Auch hier zeigt sich die Dämpfung abhängig von der Verstärkung des Verdrehwinkels. Eine zu geringe Gewichtung bedingt einen Informationsverlust, der in der Folge die notwendige Reaktion auf die Eigenschwingung hemmt. Für $c_2 = 150$ kann dagegen die Eigenschwingung der Welle stark bedämpft werden, so dass der Lastschlag keine Schwingung anregt.

Allerdings entsteht eine erhebliche Abweichung zwischen Prozessgröße $\omega_2(t)$ und ω_{ref} , wenn auf der Lastseite des Antriebsstranges ein externes Widerstandsmoment bremst. Der proportionale Regler allein ist im Grundsatz nicht in der Lage, die Abweichung zwischen Prozessgröße $\omega_2(t)$ und deren Sollwert ω_{ref} zu beseitigen.

Wie ein Vergleich des oberen Diagrammes mit dem unteren in Abbildung 6.3 darlegt, nimmt trotz Verengung des Trichters im Falle $c_2 = 150$ die Differenz $\omega_2(t) - \omega_{\text{ref}}$ für $t \rightarrow \infty$ nicht in zufriedenstellendem Maße ab. Mit $\Delta_F = 0.01$ erreicht der Trichter asymptotisch ein fünfzigstel der Breite, die sich für $\Delta_F = 0.5$ einstellt. Dennoch reduziert sich die stationäre Abweichung nicht im selben Maßstab, sondern sinkt von 75 % lediglich auf 61 % ab.

Der Grund hierfür liegt darin, dass ein Zustandsregler auf das Verhalten einer Linearkombination aller Zustände reagiert, nicht auf einen einzelnen Zustandsverlauf $\omega_2(\cdot)$, der in Abbildung 6.3 eingetragen ist. Im konkreten Falle des Zweimassensystems wird anstelle der relevanten Prozessgröße $\omega_2(t)$ die gesamte Linearkombination (6.24) zurückgeführt. Daher wird an der Vergleichsstelle ein Regelfehler

$$\begin{aligned} e(t) &= y(t) - y_{\text{ref}}(t) = y(t) - \omega_{\text{ref}}(c_1 + c_3) = \\ &= c_1(\omega_2(t) - \omega_{\text{ref}}(t)) + c_3(\omega_1(t) - \omega_{\text{ref}}(t)) + c_2\varphi_w(t) \end{aligned} \quad (6.28)$$

gebildet, aus dem der Regler die Stellgröße ermittelt. Implizit ist die relevante Differenz $\omega_2(t) - \omega_{\text{ref}}(t)$ zwar im Regelfehler $e(t)$ enthalten, eine explizite Minimierung dieser

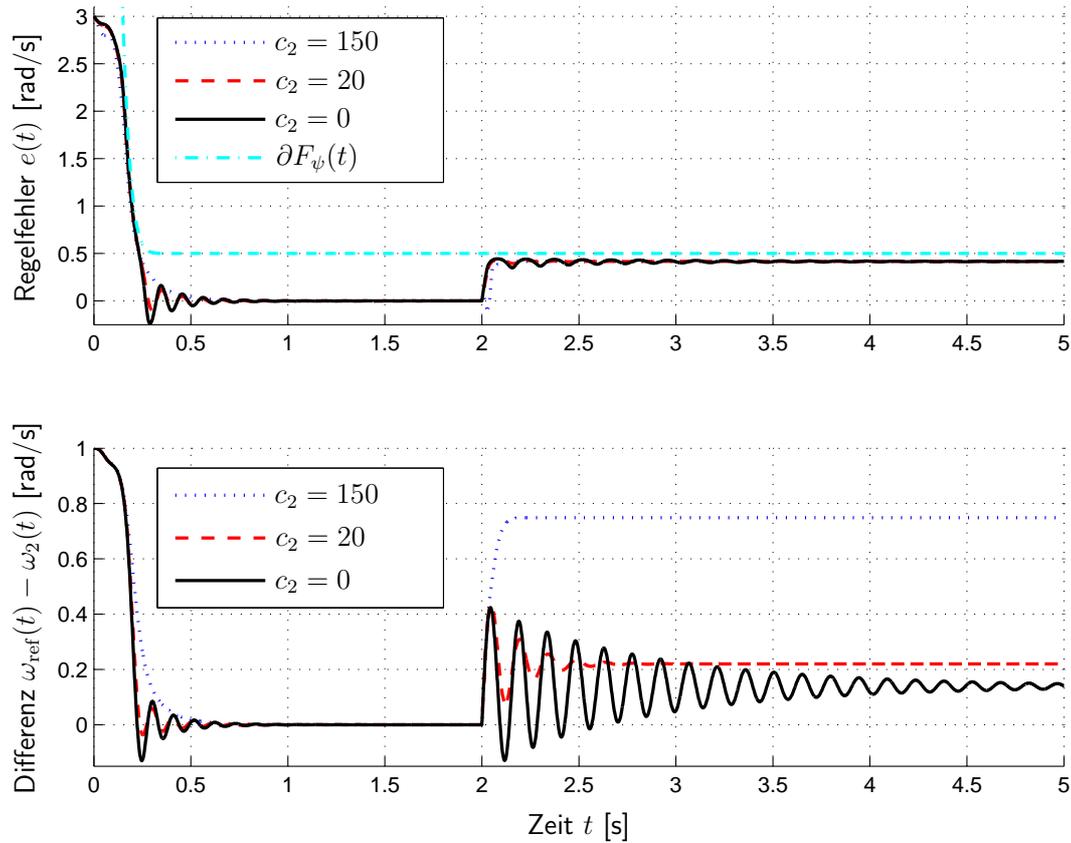


Abbildung 6.4: Funnel-Control für das Zweimassensystem mit Zustandsrückführung $y(t) = (1, c_2, 2) x(t)$ für $\Delta_F = 0.5$.

Differenz erfolgt durch die Zustandsregelung jedoch nicht. Der Unterschied zwischen den Signalen $\omega_2(\cdot) - \omega_{\text{ref}}(\cdot)$ und $e(\cdot)$ wird aus Abbildung 6.4 ersichtlich.

Damit erklärt sich das in der Simulation zu beobachtende Verhalten. Im stationären Zustand sind die Drehzahlen $\omega_1(t)$ und $\omega_2(t)$ gleich, so dass durch Umstellen von Gleichung (6.28) die Beziehung

$$\omega_2(t) - \omega_{\text{ref}}(t) = -\frac{e(t) + c_2 \varphi_w(t)}{c_1 + c_3} = -\frac{e(t)}{V} - \frac{c_2}{V} \varphi_w(t) \quad (6.29)$$

resultiert. Nur für den Sonderfall $c_2 = 0$ sind die Größen $e(t)$ und $\omega_2(t) - \omega_{\text{ref}}(t)$ über einen proportionalen Zusammenhang verknüpft, so dass aus $e(t) \rightarrow 0$ auch $\omega_2(t) - \omega_{\text{ref}}(t) \rightarrow 0$ folgt. Falls $c_2 \neq 0$ gesetzt ist, führt eine Verengung des Trichters unter Umständen zwar zu einer Verminderung des Regelfehlers $e(t)$. Doch selbst bei idealer Regelung ($e(\cdot) \equiv 0$) bleibt eine Abweichung in der Prozessgröße bestehen, die durch den Term $c_2 \varphi_w(t)/V$ in Gleichung (6.29) verursacht wird. Unter Einwirkung eines Lastmomentes M_2 entsteht stationär eine Wellenverdrehung $\varphi_w = M_2/k_w$, so dass die Gewichtung des Torsionswinkels damit die stationäre Genauigkeit beeinträchtigt.

Im Hinblick auf stationäre Genauigkeit stellt die Wahl $c_2 = 0$ eine Lösungsmöglichkeit dar, die, konform zur Nebenbedingung (6.18), darüberhinaus auf einen zusätzlichen Sensor für den Verdrehwinkel verzichtet. Als Resultat sind Regelfehler und Prozess-

größenabweichung miteinander über einen proportionalen Zusammenhang mit dem Faktor $V = c_1 + c_3 > 0$ verknüpft. Die Vorverstärkung V kann somit unabhängig vom Lastmoment und von unbekanntem Systemparametern ausgelegt werden. Verkleinert der Regler den Fehler $e(t)$, verringert sich gleichzeitig im selben Verhältnis auch die relevante Abweichung $\omega_2(t) - \omega_{\text{ref}}(t)$. Über die Reduktion von Δ_F und die damit verbundene Verengung des Trichters ist im Falle $c_2 = 0$ daher eine beliebige Verbesserung der stationären Genauigkeit für die Prozessgröße erreichbar. Dieser Effekt ist in Abbildung 6.3 für den Fall $\Delta_F = 0.5$ erkennbar. Je geringer der Verdrehwinkel gewichtet ist, desto kleiner bleibt die stationäre Abweichung in der Prozessgröße.

Allerdings sind in Abbildung 6.3 schlecht gedämpfte Oszillationen zu beobachten, je weniger der Torsionswinkel berücksichtigt wird. Ein schmalerer Trichter führt tendenziell zu einer höheren Verstärkung und damit zu mehr Stellgröße, die das System verstärkt anregt. Insofern erhöhen kleine Werte für Δ_F die Amplitude der Eigenschwingung, das Verhalten im transienten Vorgang verschlechtert sich. Nur asymptotisch verbessert ein schmaler Trichter das Regelergebnis.

Ursächlich für die unzureichende Bedämpfung der Eigenschwingung ist der potentielle Verlust der Beobachtbarkeit, wenn der Verdrehwinkel nicht gewichtet ist. Die Ausführungen in Abschnitt 2.2.3.1 analysieren, unter welchen Umständen die Dynamik der Zustände x nur unvollständig im Systemausgang y sichtbar ist. Anschaulich erklärt sich dieser Effekt durch eine Betrachtung der homogenen Eigendynamik des Zweimassensystems.

Wird die Torsionsschwingung der Welle angeregt, treten in den Geschwindigkeitsverläufen beider Massen gegenphasige Schwingungen auf. Dieser Sachverhalt zeigt sich in der Lösung der homogenen Differentialgleichung für den Anfangswert

$$x_o = \begin{pmatrix} \omega_o & \varphi_{w,o} & \omega_o \end{pmatrix}^T \in \mathbb{R}^3. \quad (6.30)$$

Mittels Auswertung des Matrixexponential $x(t) = e^{At}x_o$ errechnen sich die gesuchten Drehzahlverläufe

$$\omega_1 : t \mapsto \omega_o - J_2 \frac{2k_w}{\sqrt{-R}} e^{-\lambda t} \sin(\omega_e t) \varphi_{w,o} \quad (6.31)$$

und

$$\omega_2 : t \mapsto \omega_o + J_1 \frac{2k_w}{\sqrt{-R}} e^{-\lambda t} \sin(\omega_e t) \varphi_{w,o}. \quad (6.32)$$

Dabei setzt sich die Konstante $R \in \mathbb{R}$ aus

$$R := (J_1 + J_2)[(J_1 + J_2)d_w^2 - 4k_w J_1 J_2] < 0 \quad (6.33)$$

zusammen und ist für schwingungsfähige Systeme negativ, wie aus Ungleichung (2.14) zu entnehmen ist. Die Abklingzeitkonstante λ und die Eigenfrequenz ω_e ergeben sich zu

$$\lambda = \frac{1}{2} \frac{J_1 + J_2}{J_1 J_2} d_w > 0 \quad \text{und} \quad \omega_e = \frac{1}{2} \frac{1}{J_1 J_2} \sqrt{-R} > 0 \quad (6.34)$$

und stimmen folgerichtig mit Real- und Imaginärteil der Polstellen in Gleichung (2.13) überein.

Ein Vergleich beider Drehzahlverläufe zeigt, dass $\omega_1(\cdot)$ und $\omega_2(\cdot)$ mit identischer Frequenz ω_e oszillieren. Die unterschiedlichen Vorzeichen vor den Trägheitsmomenten verursachen jedoch eine Phasenverschiebung von 180° und führen so zu gegenphasigen Schwingungen. Daraus ergeben sich folgende Konsequenzen für unterschiedliche Auskopplungen:

(a) $c_1 > 0$:

Besitzen beide Auskoppelkoeffizienten c_1 und c_3 positives Vorzeichen, werden bei der Bildung der Linearkombination zwei gegenphasige Schwingungen addiert. In der Folge reduziert sich die Amplitude der resultierenden Schwingung durch Interferenz, bis hin zur vollständigen Auslöschung, wenn

$$c_1 = c_3 \frac{J_2}{J_1} \quad (6.35)$$

gewählt wird. In diesem Falle nimmt der Ausgang

$$y(t) = c_1 \omega_2(t) + c_3 \omega_1(t) = c_3 \left(\frac{J_2}{J_1} \omega_2(t) + \omega_1(t) \right) = c_3 \frac{J_1 + J_2}{J_1} \omega_o \quad (6.36)$$

einen konstanten Wert an, die Oszillationen der Drehzahlverläufe sind nicht enthalten. Gleichung (2.33) stützt diese Aussage und zeigt, dass die Konstellation (6.35) zum Verlust der Beobachtbarkeit führt. Dieser Effekt entspricht in der Übertragungsfunktion (2.34) der Kürzung des konjugiert komplexen Polpaares.

Die Eigenschwingung tritt umso schwächer in Erscheinung, je weiter sich die Auskoppelkoeffizienten an den Grenzfall (6.35) annähern. Durch die nicht sichtbare Eigenschwingung in der Rückkopplung bleibt die Dynamik des Zweimasensystems verborgen – dem Regler wird lediglich das Verhalten des Einmasensystems vorgespiegelt. Gleichung (6.36) ist zu entnehmen, dass ausschließlich die Geschwindigkeit ω_o am Arbeitspunkt bewertet wird. Folglich entspricht die Festsetzung (6.35) dem Wunsch nach einem kleinen stationären Regelfehler, die Torsionsschwingungen der Welle bleiben dagegen vom Regler unberücksichtigt.

(b) $c_1 < 0$:

Eine Gewichtung beider Geschwindigkeiten mit unterschiedlichem Vorzeichen ($c_1 < 0$ und $c_3 > 0$) belässt selbst ohne Auskopplung des Verdrehwinkels im Ausgangssignal die erforderlichen Informationen über Schwingungen in der Welle. Das negative Vorzeichen von c_1 spiegelt die Funktion $|c_1| \omega_2(\cdot)$ an der Zeitachse und bewirkt damit eine Phasenverschiebung um 180° . Dadurch sind die zu summierenden Drehzahlen phasengleich, so dass die Überlagerung beider Signale eine Erhöhung der Amplitude im Rückführsignal bewirkt. Demzufolge bleibt die Eigenschwingung am Ausgang stets beobachtbar.

Andererseits verliert die Geschwindigkeit am Arbeitspunkt ω_o in der Linearkombination an Bedeutung. Im Grenzfall

$$c_1 = -c_3 \quad (6.37)$$

bleibt der Ausgang

$$\begin{aligned} y(t) &= c_1\omega_2(t) + c_3\omega_1(t) = c_3(-\omega_2(t) + \omega_1(t)) = \\ &= -c_3(J_1 + J_2) \frac{2k_w}{\sqrt{-R}} e^{-\lambda t} \sin(\omega_\epsilon t) \varphi_{w,o} \quad (6.38) \end{aligned}$$

gänzlich unabhängig von der Drehzahl ω_o . Wie auch das Fehlen des integralen Anteils im Nenner der Übertragungsfunktion (2.34) andeutet, ist ω_o nicht beobachtbar. Der Regler ist damit ausgelegt auf die Bedämpfung der Eigenschwingung und lässt die Drehzahl im Arbeitspunkt völlig unbeachtet – die Geschwindigkeit im stationären Zustand spielt für den Regler keine Rolle. Dadurch sind große Regelfehler zu erwarten, besonders wenn Last auftritt. Zusätzlich wird die Anregelzeit erhöht, diese Gewichtung bedingt einen langsamen Regelkreis.

Liegen die Auskoppelkoeffizienten nahe am Grenzfall (6.37), erreicht die Regelung zwar eine Dämpfung der Torsionsschwingung, die Drehzahl am Arbeitspunkt wird aber nicht berücksichtigt. Werden die Koeffizienten an den Fall (6.35) verschoben, tritt die Reduzierung des stationären Regelfehlers umso mehr in den Vordergrund, allerdings auf Kosten der Bedämpfung. Damit kann durch die Wahl von c_1 und c_3 ein gleitender Übergang zwischen diesen beiden konkurrierenden Zielen erfolgen. Ohne Gewichtung des Verdrehwinkels kann der Zustandsregler aus seinem potenziellen Vorteil gegenüber der Kaskadenregelung keinen Nutzen ziehen.

Wie Abbildung 6.3 zeigt, lässt sich durch die Gewichtung des Torsionswinkels mit $c_2 \neq 0$ zwar die Dämpfung erhöhen, gleichzeitig wächst aber auch die stationäre Differenz $\omega_2(t) - \omega_{\text{ref}}(t)$ an. Der Parameter c_2 dient lediglich als zusätzlicher Freiheitsgrad, um den Kompromiss zwischen Dämpfung und stationärer Genauigkeit auszubalancieren.

Allein durch geeignete Wahl des Auskoppelvektors lassen sich beide Zielsetzungen nicht simultan in befriedigender Weise erreichen. Es besteht ein Zielkonflikt zwischen transientem und stationärem Verhalten der Regelung. Um diesen zu beheben, sind zusätzliche Maßnahmen erforderlich.

6.2 Behebung des Zielkonfliktes durch dynamische Zustandsrückführung

Die statische Zustandsrückführung $y(t) = (c_1, c_2, c_3) x(t)$ mit einem konstanten Faktor $c_2 \in \mathbb{R}$ verlangt eine Abwägung zwischen stationärer Genauigkeit und aktiver Dämpfung von mechanischen Schwingungen in der Antriebswelle. Dieser Zielkonflikt legt die Verwendung einer geeigneten Funktion $c_2 : t \mapsto c_2(t)$ nahe, die den Torsionswinkel $\varphi_w(t)$ abhängig vom Betriebszustand gewichtet und dadurch anstelle von (6.24) die Rückkopplung

$$y(\cdot) = c_1\omega_2(\cdot) + c_2(\cdot)\varphi_w(\cdot) + c_3\omega_1(\cdot) \quad (6.39)$$

verwirklicht. Treten Schwingungen auf, muss die Funktion $c_2(\cdot)$ große Werte annehmen, hat sich das System beruhigt, soll $c_2(\cdot)$ auf 0 abklingen, um einen bleibenden Regelfehler zu vermeiden. Im Folgenden wird eine Realisierungsmöglichkeit für die situationsabhängige Gewichtung des Verdrehwinkels diskutiert.

Wird die Funktion $c_2 : t \mapsto c_2(t)$ bereits vor Inbetriebnahme der Regelung festgelegt, fehlt die notwendige Flexibilität, um auch auf unvorhersehbare Störanregungen durch $M_2(t)$ mit einer Erhöhung von $c_2(t)$ zu reagieren. Gängige Anwendungsfälle verlangen daher, den Faktor $c_2(t)$ abhängig vom momentanen Systemverhalten einzustellen.

Diese Aufgabe kann von einem geeigneten Adaptionsgesetz übernommen werden, wobei weiterhin stabiles Betriebsverhalten gewährleistet werden muss. Der Entwurf einer passenden Adaption, sowie die zugehörige Beweisführung für Stabilität sind allerdings aufwändig.

Aus diesem Grund ist eine abweichende Herangehensweise vorzuziehen. Die Aufgabenstellung verlangt, den Torsionswinkel ausschließlich dann zu berücksichtigen, wenn Oszillationen auftreten. Daher besteht die Schlüsselidee zur Lösung des Zielkonfliktes darin, weiterhin eine konstante Rückführverstärkung $c_2 > 0$ zu verwenden, dafür aber die Zustandsgröße $\varphi_w(t)$ mittels DT₁-Baustein (Hochpassfilter, HPF)

$$\begin{aligned} \dot{x}_F(t) &= -T_F^{-1}x_F(t) + \varphi_w(t), & x_F(0) &= 0 \\ q(t) &= -T_F^{-1}x_F(t) + \varphi_w(t), & T_F &> 0 \end{aligned} \quad (6.40)$$

zu filtern und für das Zweimassensystem ein Ausgangssignal

$$\begin{aligned} y_F(\cdot) &= c_1\omega_2(\cdot) + c_2q(\cdot) + c_3\omega_1(\cdot) = c_1\omega_2(\cdot) + c_2\varphi_w(\cdot) + c_3\omega_1(\cdot) - c_2\frac{1}{T_F}x_F(\cdot) = \\ &= y(\cdot) - c_2\frac{1}{T_F}x_F(\cdot) \end{aligned} \quad (6.41)$$

zu definieren. In der Gleichgewichtslage

$$q(t) = \dot{x}_F(t) = 0 \quad \Rightarrow \quad x_F(t) = T_F\varphi_w(t) \quad (6.42)$$

führt der Filter zum gewünschten Effekt, indem die Regelgröße

$$y_F(\cdot) = c_1\omega_2(\cdot) + c_3\omega_1(\cdot) \quad \text{für } x_F(t) = T_F\varphi_w(t) \quad (6.43)$$

im stationären Zustand wunschgemäß nicht vom Verdrehwinkel abhängt. Dadurch wird der selbe Effekt erzielt, wie durch eine Reduktion von $c_2(t)$ auf 0 in Gleichung (6.39). Folglich kann die stationäre Abweichung, die von einem Lastmoment verursacht wird, durch einen schmalen Trichter beliebig klein gehalten werden.

Wird für $t \geq 0$ der Eingang des Filters (6.40) durch

$$\varphi_w(t) = A \sin(\omega t), \quad A, \omega > 0 \quad (6.44)$$

festgelegt, errechnet sich der Zustandsverlauf zu

$$x_F(t) = \frac{AT_F}{1 + T_F^2\omega^2} \left[\sin(\omega t) - T_F\omega \cos(\omega t) + T_F\omega e^{-t/T_F} \right]. \quad (6.45)$$

Für hohe Anregungsfrequenzen $\omega \gg 1/T_F$ dominiert der Summand $T_F^2 \omega^2 \gg 1$, so dass für den Zustand näherungsweise

$$x_F(t) \approx \frac{A}{T_F \omega^2} \sin(\omega t) - \frac{A}{\omega} \cos(\omega t) + \frac{A}{\omega} e^{-t/T_F} \approx 0 \quad (6.46)$$

gilt. Die Ausgangsgleichung in (6.40) führt folglich für hohe Frequenzen auf $q(t) \approx \varphi_w(t)$, worin sich die Hochpasscharakteristik der Filterung äußert. Deshalb werden hochfrequente Oszillationen im Verdrehwinkel vom Filter übertragen und treten damit im Signal $q(\cdot)$ und folglich auch in der Linearkombination (6.41) auf. Vorhandene Schwingungen werden vom Filter dynamisch mit c_2 gewichtet, was analog zur Verstärkung des Verdrehwinkels mit dem Faktor c_2 zur gewünschten Dämpfung führt.

Daraus ist ersichtlich, dass die Hochpassfilterung die Aufgabenstellung erfüllt, Schwingungen des Verdrehwinkels $\varphi_w(\cdot)$ zu detektieren und daher einer geeigneten Adaption des Verstärkungsfaktors c_2 entspricht. Die Verwendung des Filters besitzt den entscheidenden Vorteil, dass kein Adaptionsgesetz entwickelt und aufwändig auf stabiles Verhalten hin untersucht werden muss und stellt daher den deutlich einfacheren Weg dar. Der sich ergebende Signalfussplan ist in Abbildung 6.5 gezeigt.

Um nachweisen zu können, dass die dynamische Zustandsrückführung in Form des DT_1 -Filters in den Regelkreis eingebracht werden darf, ohne dadurch die Stabilität zu gefährden, muss die Kombination aus Zweimassensystem (6.5), (6.6), (6.16) und Hochpassfilter (6.40), die durch

$$\begin{aligned} \begin{pmatrix} \dot{\omega}_2(t) \\ \dot{\varphi}_w(t) \\ \dot{\omega}_1(t) \\ \dot{x}_F(t) \end{pmatrix} &= \begin{bmatrix} -d_w/J_2 & k_w/J_2 & d_w/J_2 & 0 \\ -1 & 0 & 1 & 0 \\ d_w/J_1 & -k_w/J_1 & -d_w/J_1 & 0 \\ 0 & 1 & 0 & -1/T_F \end{bmatrix} \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ \omega_1(t) \\ x_F(t) \end{pmatrix} + \\ &+ \begin{pmatrix} 0 \\ 0 \\ 1/J_1 \\ 0 \end{pmatrix} u(t) + \begin{pmatrix} -1/J_2 \\ 0 \\ 0 \\ 0 \end{pmatrix} M_2(t) \\ y_F(t) &= \begin{pmatrix} c_1 & c_2 & c_3 & -c_2/T_F \end{pmatrix} \begin{pmatrix} \omega_2(t) & \varphi_w(t) & \omega_1(t) & x_F(t) \end{pmatrix}^T \end{aligned} \quad (6.47)$$

beschrieben wird, auf Zugehörigkeit zur Systemklasse \mathcal{S} geprüft werden.

Theorem 6.4

Die Bedingungen

$$c_3 > 0, \quad c_1 > -c_3, \quad c_2 \geq 0 \quad (6.48)$$

und

$$T_F > 0 \quad (6.49)$$

stellen ein hinreichendes Kriterium dar, damit (6.47) der Systemklasse \mathcal{S} angehört und demzufolge das Regelgesetz (6.2), in dem $v(t) = u(t)$, $\hat{y}(t) = y_F(t)$ und $y_{\text{ref}}(t) = (c_1 + c_3)\omega_{\text{ref}}(t)$ zu setzen ist, die Einhaltung der Regelziele (i), (iii), (v) und (vii) garantiert.

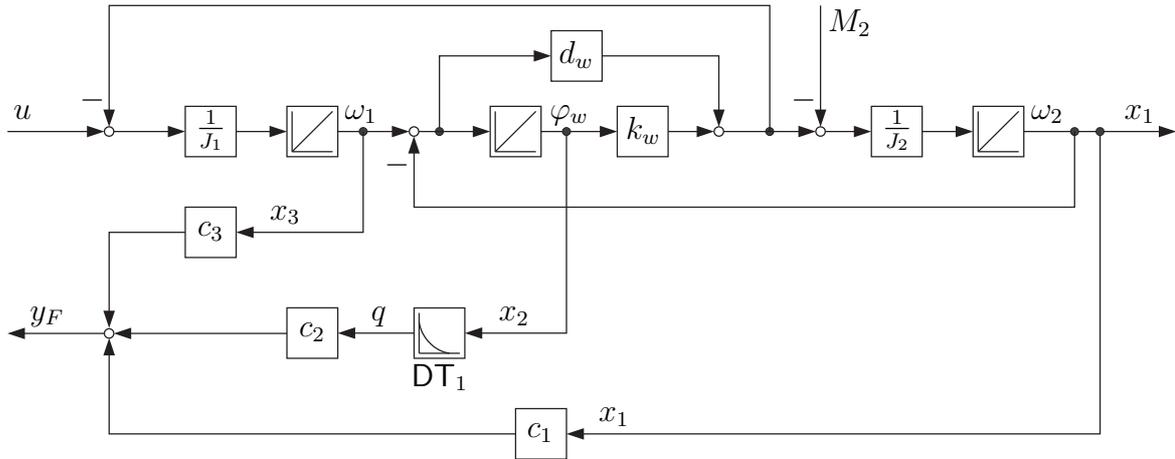


Abbildung 6.5: Signalflussplan für das Gesamtsystem, bestehend aus Zweimassensystem und DT_1 -Filter.

Beweis

Im Laplace-Bereich lässt sich das Übertragungsverhalten der Regelstrecke (6.47) durch

$$F(s) = \frac{y_F(s)}{u(s)} = \frac{c_3 J_2 T_F s^3 + [(c_1 + c_3) d_w T_F + c_2 T_F J_2 + c_3 J_2] s^2 + (c_1 + c_3) (k_w T_F + d_w) s + (c_1 + c_3) k_w}{s [J_1 J_2 s^2 + d_w (J_1 + J_2) s + k_w (J_1 + J_2)] (T_F s + 1)} \quad (6.50)$$

beschreiben.

(i) Instantane Verstärkung:

Die instantane Verstärkung bleibt mit $V_0 = c_3/J_1$ unverändert bestehen, der Hochpassfilter nimmt hier keinen Einfluss. Wegen $J_1 > 0$ wird das Vorzeichen von V_0 einzig von c_3 bestimmt.

(ii) Relativgrad:

Unter der Bedingung $c_3 \neq 0$ ist das Zählerpolynom dritter Ordnung, weshalb sich ein Relativgrad $r = 1$ ergibt.

(iii) Minimalphasigkeit:

Der Aufwand für die Überprüfung der Minimalphasigkeit steigt durch die Verwendung des Filters. Weil das Zählerpolynom die Ordnung 2 überschreitet, müssen nach Hurwitz nicht nur positive Polynomkoeffizienten

$$\begin{aligned} c_3 J_2 T_F > 0 &\Rightarrow c_3 > 0 \\ (c_1 + c_3) d_w T_F + c_2 T_F J_2 + c_3 J_2 > 0 &\Rightarrow c_2 > -[(c_1 + c_3) d_w T_F + c_3 J_2] / (T_F J_2) \\ (c_1 + c_3) (k_w T_F + d_w) > 0 &\Rightarrow c_1 > -c_3 \\ (c_1 + c_3) k_w > 0 &\Rightarrow c_1 > -c_3 \end{aligned} \quad (6.51)$$

vorliegen, sondern zusätzlich auch die Bedingung

$$(c_1 + c_3)(k_w T_F + d_w) \cdot [(c_1 + c_3)d_w T_F + c_2 T_F J_2 + c_3 J_2] - (c_1 + c_3)k_w \cdot c_3 J_2 T_F > 0 \quad (6.52)$$

zutreffen. Ausmultiplizieren führt auf

$$(c_1 + c_3)[k_w T_F [(c_1 + c_3)d_w T_F + c_2 T_F J_2] + d_w [(c_1 + c_3)d_w T_F + c_2 T_F J_2 + c_3 J_2]] \\ + (c_1 + c_3)k_w c_3 J_2 T_F - (c_1 + c_3)k_w c_3 J_2 T_F > 0 \quad (6.53)$$

und zeigt, dass sich der Summand mit negativem Vorzeichen heraushebt. Unter der Maßgabe

$$c_3 > 0, \quad c_1 > -c_3, \quad c_2 \geq 0 \quad (6.54)$$

die mit (6.18) übereinstimmt, verbleiben ausschließlich positive Terme, die Hurwitz-Bedingungen sind daher erfüllt. ■

Durch den Hochpassfilter wird die Zugehörigkeit zur Systemklasse \mathcal{S} nicht beeinträchtigt.

Um unnötige Wiederholungen zu vermeiden, wird die Verwendung des integralen Anteils und das Zusammenspiel von PI-Regler und HPF beim linearen Zweimassensystem nicht weiter vertieft. Die Tatsache, dass der PI-Regler die Stabilität des Regelkreises nicht beeinträchtigt, sowie dessen Effekt auf die stationäre Genauigkeit lässt sich aus den Theoremen, Simulationen und Messergebnissen für das reibungsbehaftete Zweimassensystem entnehmen.

6.3 Regelung des nichtlinearen Modells

Die beschriebenen Regler für das lineare Zweimassensystem sind prinzipiell auch am nichtlinearen Zweimassensystem einsetzbar.

Wie in Kapitel 2 ausgeführt, wird die nichtlineare Reibkennlinie in einen linearen und einen nichtlinearen Anteil getrennt, wobei letzterer durch den in Gleichung (2.95) definierten Operator beschrieben wird. Für den linearen Anteil wird die Steigung $\mu_g \geq 0$ angesetzt. Die Ausgliederung des linear steigenden Terms aus dem Operator \mathbf{N} ist notwendig, damit die Forderung (2.96) erfüllt wird und \mathbf{N} demzufolge global beschränkt ist. Im Ergebnis lässt sich das nichtlineare Zustandsraummodell

$$\begin{aligned} \dot{x}(t) &= Ax(t) + b_L [(\mathbf{N}\omega_2)(t) + M_2(t)] + b[u(t) + \hat{d}(t)], & x(0) &\in \mathbb{R}^3 \\ y(t) &= cx(t), & c &\in \mathbb{R}^{1 \times 3} \end{aligned} \quad (6.55)$$

mit $M_2 \in \mathcal{L}_{lok}^1(\mathbb{R}_{\geq 0}; \mathbb{R})$ und den Matrizen

$$A = \begin{bmatrix} -(d_w + \mu_g)/J_2 & k_w/J_2 & d_w/J_2 \\ -1 & 0 & 1 \\ d_w/J_1 & -k_w/J_1 & -d_w/J_1 \end{bmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ 1/J_1 \end{pmatrix} \quad \text{und} \quad b_L = \begin{pmatrix} -1/J_2 \\ 0 \\ 0 \end{pmatrix} \quad (6.56)$$

anschreiben, das den nachfolgenden Betrachtungen zu Grunde liegt.

Als Lösung des Anfangswertproblems (6.55) auf dem Intervall $[0, T[$ mit $0 < T \leq \infty$, wird eine absolut stetige Funktion $x \in \mathcal{AC}([0, T[; \mathbb{R}^3)$ bezeichnet, welche die Differentialgleichung (6.55) für fast alle Zeitpunkte $t \in [0, T[$ erfüllt. Theorem 4.2 in [43] zeigt, dass das Anfangswertproblem (6.55) für alle Anfangswerte $x(0) \in \mathbb{R}^3$ eine eindeutige Lösung besitzt, die im Intervall $[0, T[$ maximal erweiterbar ist. Falls $x \in \mathcal{L}^\infty$ gilt $T = \infty$.

Analog zur Darstellungsweise des linearen Systems wird auch für die Untersuchung des nichtlinearen Systems eine Fallunterscheidung anhand der ausgekoppelten Regelgröße vorgenommen.

6.3.1 Regelung der Motordrehzahl

Die Regelung der Motordrehzahl mit hochverstärkungsbasierten Verfahren erfordert eine Untersuchung der Differentialgleichung (6.55) mit dem Ausgang

$$y(t) = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} x(t). \quad (6.57)$$

Theorem 6.5

Das nichtlineare Zweimassensystem (6.55), (6.56) mit

$$y(t) = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} x(t) \quad (6.58)$$

ist in einer einzigen Zustandsvariablen durch die Differentialgleichung

$$\dot{y}(t) = (\mathbf{T}y)(t) + p(t) + \frac{1}{J_1} u(t), \quad y(0) = \omega_1(0) \quad (6.59)$$

darstellbar, in der $\mathbf{T} \in \mathcal{Z}$ und $p \in \mathcal{L}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R})$ gilt. Als solches gehört das nichtlineare Zweimassensystem der Systemklasse \mathcal{P} an. Folglich garantiert das Regelgesetz (6.2), in dem $u(t) = v(t)$, $\hat{y}(t) = y(t)$ und $y_{\text{ref}}(t) = \omega_{\text{ref}}(t)$ zu setzen ist, die Einhaltung der Regelziele (i), (iii), (v) und (vii).

Beweis

Theorem 6.5 ergibt sich als Spezialfall aus Theorem 6.6 für $c_1 = c_2 = 0$ und $c_3 = 1$. ■

6.3.2 Zustandsregelung

Durch die Zustandsauskopplung

$$y(t) = \begin{pmatrix} c_1 & c_2 & c_3 \end{pmatrix} x(t), \quad c_1, c_2, c_3 \in \mathbb{R} \quad (6.60)$$

wird die notwendige Relativgradreduzierung ermöglicht, wobei der Wahl der Rückkopplungskoeffizienten besondere Beachtung zukommt, da diese der Regelstrecke eine stabile Nulldynamik zuweisen müssen.

Theorem 6.6

Das nichtlineare Zweimassensystem (6.55), (6.56) mit einem Auskoppelvektor

$$c = \begin{pmatrix} c_1 & c_2 & c_3 \end{pmatrix} \in \mathbb{R}^{1 \times 3}, \quad \text{wobei} \quad c_3 > 0, \quad c_2 \geq 0, \quad c_1 > -c_3, \quad (6.61)$$

ist in einer einzigen Zustandsvariablen durch die Differentialgleichung

$$\dot{y}(t) = (\mathbf{T}y)(t) + p(t) + \frac{c_3}{J_1} u(t), \quad y(0) = c \begin{pmatrix} \omega_2(0) & \varphi_w(0) & \omega_1(0) \end{pmatrix}^T \quad (6.62)$$

darstellbar, in der $\mathbf{T} \in \mathcal{Z}$ und $p \in \mathcal{L}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R})$ gilt. Als solches gehört das nichtlineare Zweimassensystem der Systemklasse \mathcal{P} an. Folglich garantiert das Regelgesetz (6.2), in dem $v(t) = u(t)$, $\hat{y}(t) = y(t)$ und $y_{\text{ref}}(t) = (c_1 + c_3)\omega_{\text{ref}}(t)$ zu setzen ist, die Einhaltung der Regelziele (i), (iii), (v) und (vii).

Beweis

Differenzieren des Ausgangs führt auf die Gleichung

$$\dot{y}(t) = c A x(t) - \frac{c_1}{J_2} [(\mathbf{N}\omega_2)(t) + M_2(t)] + \frac{c_3}{J_1} \hat{d}(t) + \frac{c_3}{J_1} u(t), \quad (6.63)$$

in der die Stellgröße $u(t)$ mit c_3/J_1 gewichtet ist. Definitionsgemäß besitzt ein dynamisches System den Relativgrad 1, wenn die Stellgröße – unabhängig vom Zustandsvektor $x(t) \in \mathbb{R}^3$ – direkten Eingriff auf die erste Ableitung von $y(t)$ erhält. Gleichung (6.63) ist zu entnehmen, dass hierfür eine instantane Verstärkung $c_3/J_1 \neq 0$ erforderlich ist. Folglich garantiert $c_3 \neq 0$ einen global definierten Relativgrad $r = 1$. Die in (6.61) geforderte Bedingung $c_3 > 0$ führt darüberhinaus auf ein positives Vorzeichen für die instantane Verstärkung. Daher ermöglicht ein globaler Diffeomorphismus eine global gültige Transformation in Byrnes-Isidori-Normalform (BINF).

Bedingt durch $c_3 \neq 0$ ist die Matrix

$$S = \begin{bmatrix} c_1 & c_2 & c_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \in \mathbf{GL}_3(\mathbb{R}) \quad (6.64)$$

regulär mit der Inversen

$$S^{-1} = \frac{1}{c_3} \cdot \begin{bmatrix} 0 & c_3 & 0 \\ 0 & 0 & c_3 \\ 1 & -c_1 & -c_2 \end{bmatrix} \quad (6.65)$$

und eignet sich daher für eine global gültige, lineare Koordinatentransformation. Aus der Abbildungsvorschrift $z(t) := Sx(t)$ ergibt sich der Zustandsvektor der BINF zu

$$z(t) := \begin{bmatrix} c_1 & c_2 & c_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ \omega_1(t) \end{pmatrix} = \begin{pmatrix} y(t) \\ \omega_2(t) \\ \varphi_w(t) \end{pmatrix}. \quad (6.66)$$

Durch Einsetzen der inversen Abbildungsvorschrift $x(t) = S^{-1}z(t)$ in das Zustandsraummodell (6.55), (6.61) ergibt sich eine äquivalente Beschreibung in z -Koordinaten, die in zwei Differentialgleichungen partitioniert ist:

$$\dot{y}(t) = \rho y(t) + \delta \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \end{pmatrix} - \frac{c_1}{J_2} [(\mathbf{N}\omega_2)(t) + M_2(t)] + \frac{c_3}{J_1} [\hat{d}(t) + u(t)] \quad (6.67a)$$

$$\begin{pmatrix} \dot{\omega}_2(t) \\ \dot{\varphi}_w(t) \end{pmatrix} = \Lambda \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \end{pmatrix} + \begin{pmatrix} \frac{d_w}{c_3 J_2} \\ \frac{1}{c_3} \end{pmatrix} y(t) - \begin{pmatrix} \frac{1}{J_2} \\ 0 \end{pmatrix} [(\mathbf{N}\omega_2)(t) + M_2(t)] \quad (6.67b)$$

wobei

$$\rho = \frac{(c_1 - c_3 J_2 / J_1) d_w + c_2 J_2}{c_3 J_2} \quad (6.68)$$

$$\delta = \begin{pmatrix} -\frac{d_w(c_1 c_3 + c_1^2)}{c_3 J_2} + \frac{d_w(c_1 + c_3)}{J_1} - c_2 - \frac{c_1 c_2}{c_3} - \frac{c_1 \mu_g}{J_2} \\ \frac{c_1 c_3 k_w - c_1 c_2 d_w - 2c_2^2 J_2}{c_3 J_2} + \frac{c_2 d_w - c_3 k_w}{J_1} + \frac{c_2^2}{c_3} \end{pmatrix}^T \quad (6.69)$$

$$\Lambda = \begin{bmatrix} -\frac{d_w}{J_2} \left(1 + \frac{c_1}{c_3}\right) - \frac{\mu_g}{J_2} & \frac{k_w}{J_2} - \frac{d_w}{J_2} \frac{c_2}{c_3} \\ -\left(1 + \frac{c_1}{c_3}\right) & -\frac{c_2}{c_3} \end{bmatrix} \quad (6.70)$$

Durch (6.67b) wird die Nulldynamik des Zweimassensystems beschrieben. Weil in Gleichung (2.96) ein global beschränkter Operator gefordert wird, bleibt die Funktion $(\mathbf{N}\omega_2)(\cdot)$ unabhängig vom Verlauf $\omega_2(\cdot)$ stets beschränkt. Weil gleichzeitig auch $M_2 \in \mathcal{L}^\infty$ gilt, darf $[(\mathbf{N}\omega_2)(\cdot) + M_2(\cdot)]$ als beschränktes Eingangssignal aufgefasst werden, das auf eine lineare Differentialgleichung mit der Systemmatrix Λ einwirkt.

Falls Λ negative Eigenwertrealteile aufweist und $\mathbf{N} \equiv 0$ gilt, ist die lineare Differentialgleichung (6.67b) und damit die Nulldynamik des linearen Zweimassensystems asymptotisch stabil. Durch die Einwirkung eines nichtlinearen Störterms $\mathbf{N} \neq 0$, $M_2 \neq 0$ geht möglicherweise die asymptotische Stabilität verloren. Da $[(\mathbf{N}\omega_2)(\cdot) + M_2(\cdot)] \in \mathcal{L}^\infty$ bleibt aber die notwendige Stabilität erhalten, die Zustandstrajektorien $\omega_2(\cdot)$ und $\varphi_w(\cdot)$ werden unter der Annahme $y(\cdot) \in \mathcal{L}^\infty$ durch beschränkte Funktionen beschrieben. Deshalb ist die Nulldynamik des nichtlinearen Zweimassensystem unter der Voraussetzung (2.96) stabil.

Aus dem Ansatz $\det(sI_2 - \Lambda) = 0$ errechnet sich die charakteristische Gleichung der Systemmatrix Λ zu

$$\lambda^2 + \frac{J_2 c_2 + d_w(c_1 + c_3) + \mu_g c_3}{c_3 J_2} \lambda + \frac{\mu_g c_2 + k_w(c_1 + c_3)}{c_3 J_2} = 0. \quad (6.71)$$

Nach Hurwitz besitzen die Wurzeln des charakteristischen Polynoms und damit die Eigenwerte der Systemmatrix Λ negative Realteile, wenn die Ungleichungen

$$\frac{J_2 c_2 + d_w(c_1 + c_3) + \mu_g c_3}{c_3 J_2} > 0 \quad \text{und} \quad \frac{\mu_g c_2 + k_w(c_1 + c_3)}{c_3 J_2} > 0 \quad (6.72)$$

erfüllt sind. Wegen (2.10) und $\mu_g \geq 0$ sind hierfür die in (6.61) aufgestellten, parameterunabhängigen Bedingungen

$$c_3 > 0, \quad c_2 \geq 0, \quad c_1 > -c_3 \quad \Rightarrow \quad \sigma(\Lambda) \subset \mathbb{C}_- \quad (6.73)$$

hinreichend. Allein durch die physikalisch bedingte Tatsache, dass sämtliche Parameter positiv sind, kann auch für nichtlineare Zweimassensysteme allgemein eine Auskoppung bestimmt werden, welche die erforderliche Stabilität der Nulldynamik garantiert.

Aus $r = 1$ resultiert als wesentliche Konsequenz die Möglichkeit, das System in einer Variablen (Zustandsgröße y) darzustellen.

Mittels Variation der Konstanten ergibt sich als Lösung für die lineare Differentialgleichung (6.67b) mit den beiden Eingangssignalen $y(\cdot)$ und $[(\mathbf{N}\omega_2)(\cdot) + M_2(\cdot)]$:

$$\begin{aligned} \delta \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \end{pmatrix} &= \delta e^{\Lambda t} \begin{pmatrix} \omega_2(0) \\ \varphi_w(0) \end{pmatrix} + \delta \int_0^t e^{\Lambda(t-\tau)} \begin{pmatrix} \frac{d_w}{c_3 J_2} \\ \frac{1}{c_3} \end{pmatrix} y(\tau) d\tau + \\ &+ \delta \int_0^t e^{\Lambda(t-\tau)} \begin{pmatrix} \frac{-1}{J_2} \\ 0 \end{pmatrix} [(\mathbf{N}\omega_2)(\tau) + M_2(\tau)] d\tau \end{aligned} \quad (6.74)$$

Einsetzen in (6.67a) führt zusammen mit den Definitionen

$$\begin{aligned} \mathbf{T}_1 &: \mathcal{L}_{\text{lok}}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R}^2) \rightarrow \mathcal{L}_{\text{lok}}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R}^2), \quad \zeta(\cdot) \mapsto \int_0^\cdot e^{\Lambda(\cdot-\tau)} \zeta(\tau) d\tau, \\ p(\cdot) &:= \delta e^{\Lambda \cdot} \begin{pmatrix} \omega_2(0) \\ \varphi_w(0) \end{pmatrix} + \delta \left(\mathbf{T}_1 \begin{pmatrix} \frac{-1}{J_2} \\ 0 \end{pmatrix} [(\mathbf{N}\omega_2) + M_2] \right) (\cdot) + \\ &+ \frac{c_3}{J_1} \hat{d}(\cdot) - \frac{c_1}{J_2} [(\mathbf{N}\omega_2)(\cdot) + M_2(\cdot)], \end{aligned} \quad (6.75)$$

$$(\mathbf{T}y)(\cdot) := \rho y(\cdot) + \delta \left(\mathbf{T}_1 \begin{pmatrix} \frac{d_w}{c_3 J_2} \\ \frac{1}{c_3} \end{pmatrix} y \right) (\cdot)$$

auf Gleichung (6.62). Bedingt durch (2.96),

$$M_2 \in \mathcal{L}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R}) \quad \text{und} \quad \hat{d} \in \mathcal{L}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R}) \quad \text{und} \quad \sigma(\Lambda) \subset \mathbb{C}_- \quad (6.76)$$

folgt $\mathbf{T}_1 \in \mathcal{Z}$, $p \in \mathcal{L}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R})$ und $\mathbf{T} \in \mathcal{Z}$ (siehe [41, S. 476f]). Der lineare Operator \mathbf{T} beschreibt in der Darstellung (6.62) den linearen Anteil der Nulldynamik des Zweimassensystems, der wegen $\mathbf{T} \in \mathcal{Z}$ ein beschränktes Eingangssignal in ein beschränktes Ausgangssignal abbildet. Die nichtlinearen Effekte sind im Störterm $p(\cdot)$ enthalten und besitzen wegen $p \in \mathcal{L}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R})$ nur beschränkten Einfluss. Daher erfüllt das reibungsbehaftete Zweimassensystem die erforderlichen Voraussetzungen, um den in Kapitel 4 beschriebenen Regler anwenden zu können. ■

In Abbildung 6.6 ist das Simulationsergebnis für das Zweimassensystem (6.55) mit nichtlinearer Reibkennlinie

$$\mathbf{N} : \omega_2(\cdot) \mapsto \frac{2}{\pi} \cdot \arctan(10 \omega_2(\cdot)) \quad (6.77)$$

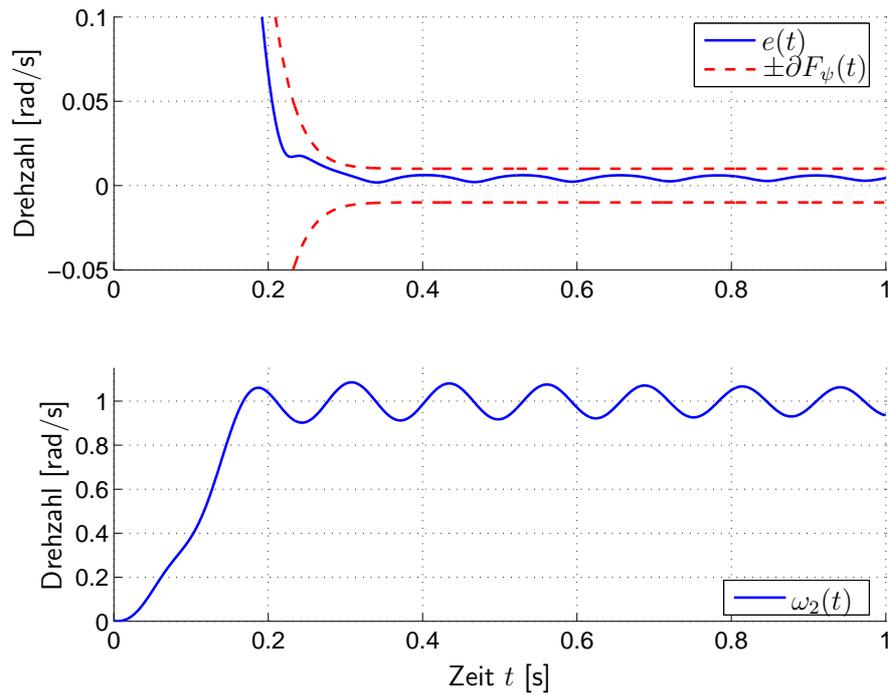


Abbildung 6.6: Regelung des Zweimassensystems ohne Berücksichtigung des Verdrehwinkels mittels Zustandsauskopplung $c = (1, 0, 1)$.

dargestellt. Damit durch Verengung des Trichters die Differenz $|\omega_2(t) - \omega_{\text{ref}}(t)|$ aus Sollwert $\omega_{\text{ref}} = 1$ rad/s und Lastdrehzahl $\omega_2(t)$ beliebig weit reduziert werden kann, wird der Verdrehwinkel in der Zustandsauskopplung

$$y(t) = \begin{pmatrix} 1 & 0 & 1 \end{pmatrix} x(t) \quad (6.78)$$

nicht gewichtet. Als Trichterrand wird dem Regler (6.2) mit $u(t) = v(t)$, $\hat{y}(t) = y(t)$ und $y_{\text{ref}}(t) = (c_1 + c_3)\omega_{\text{ref}}(t)$ die Funktion

$$\psi : t \mapsto (3 \cdot e^{-80t^2} + \Delta_F)^{-1}, \quad \Delta_F = 0.01 \quad (6.79)$$

vorgegeben, die für $t \rightarrow \infty$ den Betrag des Regelfehlers durch die Schranke $\Delta_F = 0.01$ begrenzt.

Gleichung (6.29) ergibt daraus in der Gleichgewichtslage eine Abweichung

$$|\omega_2(t) - \omega_{\text{ref}}(t)| = \frac{|e(t)|}{c_1 + c_3} < \frac{\Delta_F}{c_1 + c_3} = \frac{0.01}{2} = 0.005, \quad (6.80)$$

die maximal 0.5% vom Sollwert beträgt und sich wegen $c_2 = 0$ proportional zum Regelfehler $e(t)$ verhält. Dies eröffnet die Möglichkeit, durch Verringerung der Trichterbreite Δ_F den Wert $|\omega_2(t) - \omega_{\text{ref}}(t)|$ beliebig zu verkleinern.

Allerdings resultiert aus $c_2 = 0$ ein Informationsverlust, die Torsionsschwingung ist im Regelfehler $e(\cdot)$ nicht enthalten. Deshalb kann der Regler nicht aktiv eingreifen, so dass die Schwingung nur aufgrund der Eigendämpfung $d_w > 0$ der Welle abklingt.

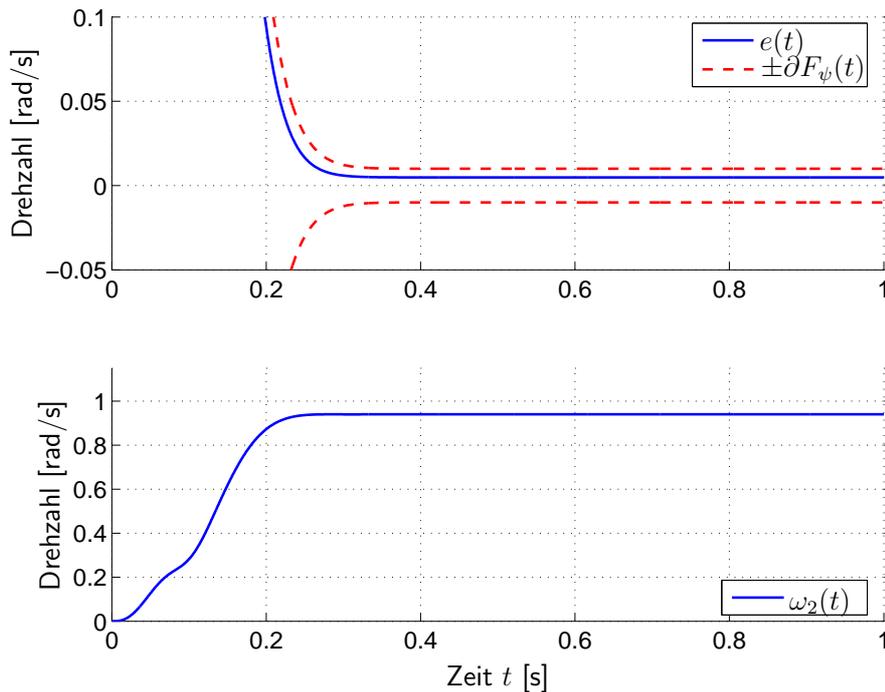


Abbildung 6.7: Regelung des Zweimassensystems mit Berücksichtigung des Verdrehwinkels mittels Zustandsauskopplung $c = (1, 50, 1)$.

Um die unerwünschten mechanischen Schwingungen der Antriebswelle zu eliminieren, wird der Verdrehwinkel in der Linearkombination

$$y(t) = \begin{pmatrix} 1 & 50 & 1 \end{pmatrix} x(t) \quad (6.81)$$

mit $c_2 = 50$ gewichtet, so dass die erforderliche Information in das Fehlersignal $e(\cdot)$ eingebracht wird. Abbildung 6.7 bestätigt, dass diese Maßnahme eine aktive Bedämpfung durch den Regler ermöglicht, der Regelfehler $e(\cdot)$ zeigt keine Schwingungen mehr.

Aufgrund des Zusammenhangs zwischen $e(t)$ und $\omega_2(t) - \omega_{\text{ref}}(t)$ aus Gleichung (6.29) löst $c_2 \neq 0$ die proportionale Beziehung zwischen diesen beiden Größen auf. Es verbleibt daher – abhängig vom Lastmoment – eine deutliche Differenz. Diese Abweichung besteht selbst im Falle $e(t) = 0$. Eine Verschmälerung des Trichters durch den Parameter Δ_F kann daher den Wert $|\omega_2(t) - \omega_{\text{ref}}(t)|$ nicht beliebig verkleinern.

6.3.3 Zustandsregelung mit Hochpassfilter und I-Anteil

Der am linearen Zweimassensystem diskutierte Zielkonflikt besteht prinzipiell auch im Falle von Reibung. Deshalb wird der Hochpassfilter (6.40) in das Zustandsraummodell des nichtlinearen Zweimassensystems eingefügt, indem die Beschreibung des Zweimassensystems um die Zustandsgröße $x_F(t)$ des Filters ergänzt wird. Daraus resultiert die

Darstellung:

$$\begin{aligned}
 \begin{pmatrix} \dot{\omega}_2(t) \\ \dot{\varphi}_w(t) \\ \dot{\omega}_1(t) \\ \dot{x}_F(t) \end{pmatrix} &= \begin{bmatrix} -(d_w + \mu_g)/J_2 & k_w/J_2 & d_w/J_2 & 0 \\ -1 & 0 & 1 & 0 \\ d_w/J_1 & -k_w/J_1 & -d_w/J_1 & 0 \\ 0 & 1 & 0 & -1/T_F \end{bmatrix} \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ \omega_1(t) \\ x_F(t) \end{pmatrix} + \\
 &+ \begin{pmatrix} -1/J_2 \\ 0 \\ 0 \\ 0 \end{pmatrix} [(\mathbf{N}\omega_2)(t) + M_2(t)] + \begin{pmatrix} 0 \\ 0 \\ 1/J_1 \\ 0 \end{pmatrix} [u(t) + \hat{d}(t)] \\
 y_F(t) &= \begin{pmatrix} c_1 & c_2 & c_3 & -c_2/T_F \end{pmatrix} \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ \omega_1(t) \\ x_F(t) \end{pmatrix} = y(t) - \frac{c_2}{T_F} x_F(t)
 \end{aligned} \tag{6.82}$$

Theorem 6.7

Unter den Voraussetzungen

$$c_3 > 0, \quad c_2 \geq 0, \quad c_1 > -c_3 \tag{6.83}$$

und

$$T_F > 0 \tag{6.84}$$

ist die in Gleichung (6.82) angegebene Kombination aus Hochpassfilter und nichtlinearem Zweimassensystem in einer einzigen Zustandsvariablen durch die Differentialgleichung

$$\dot{y}_F(t) = (\mathbf{T}_F y_F)(t) + p_F(t) + \frac{c_3}{J_1} u(t), \quad y_F(0) = c \begin{pmatrix} \omega_2(0) & \varphi_w(0) & \omega_1(0) \end{pmatrix}^T \tag{6.85}$$

darstellbar, in der $\mathbf{T}_F \in \mathcal{Z}$ und $p_F \in \mathcal{L}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R})$ gilt. Als solches gehört (6.82) der Systemklasse \mathcal{P} an. Folglich garantiert das Regelgesetz (6.2), in dem $u(t) = v(t)$, $\hat{y}(t) = y_F(t)$ und $y_{\text{ref}}(t) = (c_1 + c_3)\omega_{\text{ref}}(t)$ zu setzen ist, die Einhaltung der Regelziele (i), (iii), (v) und (vii).

Beweis

Im Filter (6.40) besteht zwischen Eingang $\varphi_w(\cdot)$ und Ausgangssignal $q(\cdot)$ ein Durchgriff, weshalb der Relativgrad 0 vorliegt. Daher kann der Filter den Relativgrad des Zweimassensystems nicht erhöhen. Es gilt für das Gesamtsystem (6.82)

$$\begin{aligned}
 \dot{y}_F(t) = \dot{y}(t) - \frac{c_2}{T_F} \dot{x}_F(t) &= c A x(t) - \frac{c_1}{J_2} [(\mathbf{N}\omega_2)(t) + M_2(t)] + \frac{c_3}{J_1} \hat{d}(t) + \\
 &+ \frac{c_3}{J_1} u(t) + \frac{c_2}{T_F} \left(\frac{1}{T_F} x_F(t) - \varphi_w(t) \right), \tag{6.86}
 \end{aligned}$$

wobei die Bedingung $c_3 > 0$ auf $r = 1$ führt. Darüberhinaus garantiert $c_3 > 0$ eine reguläre Matrix

$$S = \begin{bmatrix} c_1 & c_2 & c_3 & -c_2/T_F \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad S \in \mathbf{GL}_4(\mathbb{R}), \quad (6.87)$$

mit deren Hilfe über die Abbildungsvorschrift $z_F(t) := Sx_F(t)$ ein Wechsel von x_F - in z_F -Koordinaten vollzogen wird. Die lineare Abbildung ist so angesetzt, dass lediglich eine Umbenennung der einzelnen Koordinaten erfolgt, eine physikalische Interpretation der z_F -Koordinaten

$$z_F(t) := \begin{bmatrix} c_1 & c_2 & c_3 & -c_2/T_F \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ \omega_1(t) \\ x_F(t) \end{pmatrix} = \begin{pmatrix} y_F(t) \\ \omega_2(t) \\ \varphi_w(t) \\ x_F(t) \end{pmatrix} \quad (6.88)$$

bleibt daher möglich. Gleichung (6.88) erlaubt, das Argument der Reibung – den Drehzahlverlauf $\omega_2(\cdot)$ – durch die zweite Komponente des Zustandsvektors $z_F(\cdot)$ zu substituieren. Der nichtlineare Operator \mathbf{N} bleibt dagegen von der Transformation völlig unberührt. Die Zustandsbeschreibung in BINF ist somit durch

$$\begin{aligned} \dot{y}_F(t) = \rho_F y_F(t) + \delta_F \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ x_F(t) \end{pmatrix} - \frac{c_1}{J_2} [(\mathbf{N}\omega_2)(t) + M_2(t)] + \\ + \frac{c_3}{J_1} [\hat{d}(t) + u(t)] \end{aligned} \quad (6.89a)$$

$$\begin{pmatrix} \dot{\omega}_2(t) \\ \dot{\varphi}_w(t) \\ \dot{x}_F(t) \end{pmatrix} = \Lambda_F \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ x_F(t) \end{pmatrix} + \begin{pmatrix} \frac{d_w}{c_3 J_2} \\ \frac{1}{c_3} \\ 0 \end{pmatrix} y_F(t) - \begin{pmatrix} \frac{1}{J_2} \\ 0 \\ 0 \end{pmatrix} [(\mathbf{N}\omega_2)(t) + M_2(t)] \quad (6.89b)$$

gegeben, wobei

$$\rho_F = \frac{(c_1 J_1 - c_3 J_2) d_w}{c_3 J_1 J_2} + \frac{c_2}{c_3}, \quad (6.90)$$

$$\delta_F = \begin{pmatrix} -\frac{(c_1 c_3 + c_1^2) d_w}{c_3 J_2} - \frac{c_2 c_3 + c_1 c_2}{J_1} + \frac{(c_1 + c_3) d_w}{J_1} - \frac{c_1 \mu_g}{J_2} \\ \frac{(c_1 J_1 - c_3 J_2) k_w}{J_1 J_2} + \frac{c_2 (d_w T_F - J_1)}{J_1 T_F} - \frac{c_1 c_2 d_w}{c_3 J_2} - \frac{c_2^2}{c_3} \\ \frac{c_1 c_2 d_w}{c_3 J_2 T_F} + \frac{c_2^2}{c_3 T_F} - \frac{c_2 d_w}{J_1 T_F} + \frac{c_2}{T_F^2} \end{pmatrix}^T \quad (6.91)$$

und

$$\Lambda_F = \begin{bmatrix} -\frac{d_w}{J_2} \left(1 + \frac{c_1}{c_3}\right) - \frac{\mu_g}{J_2} & \frac{k_w}{J_2} - \frac{d_w c_2}{J_2 c_3} & \frac{d_w}{J_2} \frac{1}{T_F} \frac{c_2}{c_3} \\ -\left(1 + \frac{c_1}{c_3}\right) & -\frac{c_2}{c_3} & \frac{1}{T_F} \frac{c_2}{c_3} \\ 0 & 1 & -\frac{1}{T_F} \end{bmatrix} \quad (6.92)$$

gilt. Gleichung (6.89b) beschreibt hierbei die Nulldynamik, die aus den drei Zuständen $\omega_2(t)$, $\varphi_w(t)$ und $x_F(t)$ besteht.

Eine direkte Berechnung und Untersuchung der Eigenwertrealteile von Λ_F erfordert hohen Aufwand, weswegen für die Matrix $c_3 T_F J_2 \Lambda_F$ die charakteristische Gleichung

$$\begin{aligned} \det(\lambda I_3 - c_3 T_F J_2 \Lambda_F) &= \lambda^3 + [c_3 J_2 + (c_1 + c_3) d_w T_F + c_3 \mu_g T_F + c_2 T_F J_2] \lambda^2 + \\ &+ [c_3 J_2 T_F (c_1 + c_3) (d_w + k_w T_F) + c_3 \mu_g J_2 T_F (c_2 T_F + c_3)] \lambda + c_3^2 k_w J_2^2 T_F^2 (c_1 + c_3) \end{aligned} \quad (6.93)$$

bestimmt und einem Hurwitz-Test unterzogen wird. Die Skalierung einer Matrix mit einem positiven Faktor beeinflusst zwar die Lage ihrer Eigenwerte, ändert aber das Vorzeichen von Realteil und Imaginärteil nicht. Daher ist Λ_F genau dann eine Hurwitz-Matrix, wenn auch $c_3 T_F J_2 \Lambda_F$ eine Hurwitz-Matrix ist. Weil sämtliche Koeffizienten des charakteristischen Polynoms wegen (2.10) positiv sind und deswegen gleichzeitig auch die Bedingung

$$\begin{aligned} &[c_3 J_2 + (c_1 + c_3) d_w T_F + c_3 \mu_g T_F + c_2 T_F J_2] \cdot \\ &\cdot [c_3 J_2 T_F (c_1 + c_3) (d_w + k_w T_F) + c_3 \mu_g J_2 T_F (c_2 T_F + c_3)] - c_3^2 k_w J_2^2 T_F^2 (c_1 + c_3) = \\ &= [c_3 J_2 + (c_1 + c_3) d_w T_F + c_3 \mu_g T_F + c_2 T_F J_2] \cdot [c_3 J_2 T_F (c_1 + c_3) k_w T_F] + \\ &+ [c_3 J_2 + (c_1 + c_3) d_w T_F + c_3 \mu_g T_F + c_2 T_F J_2] \cdot \\ &\cdot [c_3 J_2 T_F (c_1 + c_3) d_w + c_3 \mu_g J_2 T_F (c_2 T_F + c_3)] - c_3^2 k_w J_2^2 T_F^2 (c_1 + c_3) = \\ &= [(c_1 + c_3) d_w T_F + c_3 \mu_g T_F + c_2 T_F J_2] \cdot [c_3 J_2 T_F (c_1 + c_3) k_w T_F] + \\ &+ [c_3 J_2 + (c_1 + c_3) d_w T_F + c_3 \mu_g T_F + c_2 T_F J_2] \cdot \\ &\cdot [c_3 J_2 T_F (c_1 + c_3) d_w + c_3 \mu_g J_2 T_F (c_2 T_F + c_3)] + \\ &+ c_3^2 k_w J_2^2 T_F^2 (c_1 + c_3) - c_3^2 k_w J_2^2 T_F^2 (c_1 + c_3) > 0 \end{aligned} \quad (6.94)$$

erfüllt ist, erhalten nach Hurwitz sämtliche Wurzeln negative Realteile. Der lineare Teil der Nulldynamik wird von einer Systemmatrix mit der Eigenschaft

$$\sigma(\Lambda_F) \subset \mathbb{C}_- \quad (6.95)$$

festgelegt und ist deshalb exponentiell stabil. Der restliche Verlauf der Beweisführung folgt exakt dem Fortgang des Beweises von Theorem 6.6, um das System in einer einzigen Zustandsvariablen $y_F(t)$ darzustellen. ■

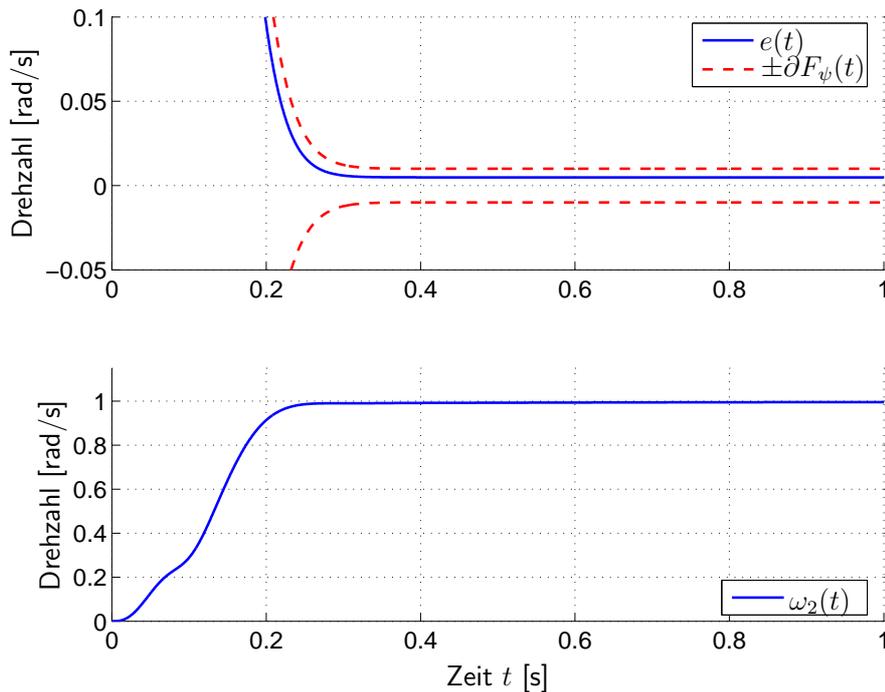


Abbildung 6.8: Auflösung des Zielkonfliktes: Ausgeprägte Bedämpfung des schwingungsfähigen Systems und beliebig kleine Abweichung $\omega_2(t) - \omega_{\text{ref}}(t)$ durch den Einsatz des DT_1 -Filters.

Mittels Simulation wird bestätigt, dass der Zielkonflikt zwischen ausgeprägter Dämpfung und geringer Abweichung $\omega_2(t) - \omega_{\text{ref}}(t)$ im stationären Zustand durch den beschriebenen DT_1 -Filter (6.40) mit $T_F = 0.5$ aufgelöst wird. Es ist durch den Filter gewährleistet, dass im stationären Zustand der proportionale Zusammenhang aus Gleichung (6.29) Gültigkeit besitzt. Daher kann durch einen entsprechend schmal ausgelegten Trichter die Prozessgröße $\omega_2(t)$ beliebig nahe am Sollwert $\omega_{\text{ref}}(t)$ gehalten werden. Dennoch werden während der transienten Phase Torsionsschwingungen unterdrückt, in Abbildung 6.8 sind keine Schwingungen erkennbar.

Ein Proportionalregler besitzt kein internes Modell für konstante Anregungen und kann daher die Regeldifferenz stationär nicht beseitigen, sobald ein Lastmoment angreift. Zwar wird durch Verschmälern des Trichters der Regelfehler $e(t)$ und durch den gewählten Ansatz mit dem Hochpassfilter damit in der Folge auch $\omega_2(t) - \omega_{\text{ref}}(t)$ beliebig reduziert. Allerdings basiert dies auf der unerfüllbaren Voraussetzung, dass eine rauschfreie Messung der Regelgröße realisierbar ist. Unter wirklichkeitsnahen Bedingungen muss dagegen sichergestellt sein, dass der Trichter eine ausreichende Mindestbreite beibehält, um das rauschbehaftete Fehlersignal $e(\cdot)$ für alle Zeitpunkte zu umschließen. Als Folge eines weiten Trichters ergibt sich ein möglicherweise großer Regelfehler und, dadurch bedingt, eine ebenfalls große Abweichung $\omega_2(t) - \omega_{\text{ref}}(t)$ im stationären Zustand.

Weil der Simulation in Abbildung 6.9 der Wert $\Delta_F = 0.75$ zugrundegelegt ist, verjüngt sich der Trichter nicht in ausreichendem Maße und toleriert vor allem bei Lasteinwirkung einen beträchtlichen Regelfehler.

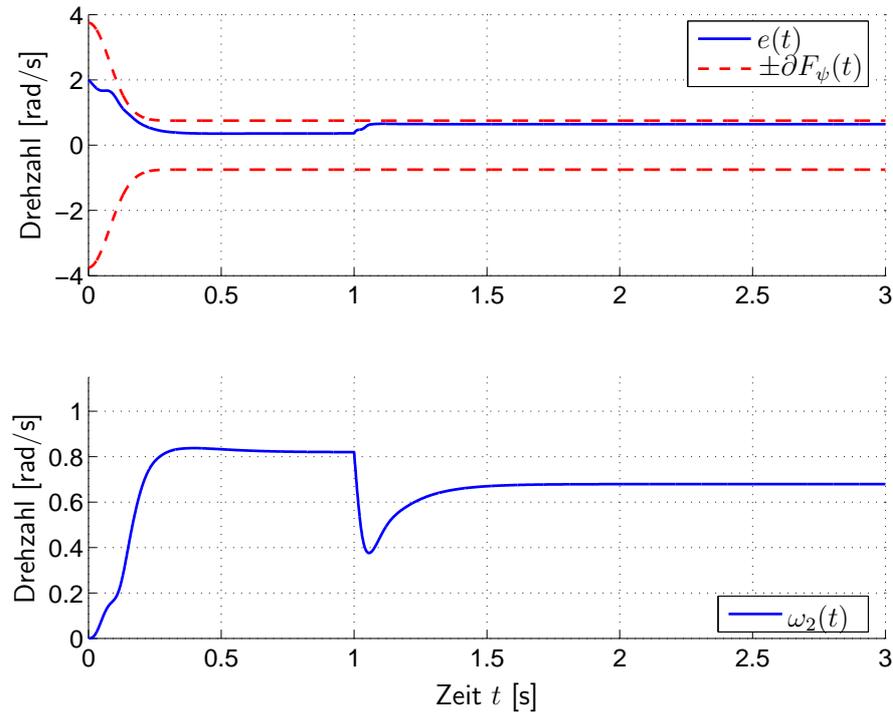


Abbildung 6.9: Die Verwendung eines breiten Trichters führt unter Last auf einen großen Regelfehler und damit auf eine beträchtliche Abweichung $\omega_2(t) - \omega_{\text{ref}}(t)$ in der Ruhelage.

Aufgrund dieser Tatsache ist der proportionale Regler nur von theoretischem Interesse. Für die Anwendung von Funnel-Control in der Praxis muss ein integraler Anteil im Sinne eines internen Modells in den Regelkreis eingebracht werden, ohne dabei die Stabilität des Gesamtsystems zu gefährden. Hierzu wird dem Proportionalregler (6.2) ein PI-Regler¹⁾

$$\begin{aligned} \dot{x}_I(t) &= v(t), & x_I(0) &= 0 \\ u(t) &= k_I x_I(t) + k_P v(t), & k_I, k_P &> 0 \end{aligned} \quad (6.96)$$

nachgeschaltet und zum nichtlinearen Zweimassensystem mit dynamischer Zustandsauskopplung hinzugerechnet. Im stationären Zustand, der durch $\dot{x}_I(t) = 0$ gekennzeichnet ist, erzwingt der Integralanteil in (6.96) die Gleichung $v(t) = 0$. Daher führt der Proportionalregler $v(t) = -k(t)e(t)$ unter der Voraussetzung $\limsup_{t \rightarrow \infty} k(t) > 0$ in der Ruhelage auf das gewünschte Ergebnis $e(t) = 0$.

¹⁾ Das Zustandsraummodell (6.96) stellt die Realisierung der Übertragungsfunktion (6.23) dar und gestattet, den integralen Anteil mit dem nichtlinearen Zweimassensystem zusammenzufassen. Wie auch im Falle des linearen Zweimassensystems stellen die Einschränkungen $k_I > 0$ und $k_P > 0$ in Gleichung (6.96) sicher, dass trotz zusätzlichem PI-Regler die BIBO-Stabilität der nichtlinearen Nulldynamik bzw. die asymptotische Stabilität des linearen Anteils der Nulldynamik erhalten bleibt. Die frei wählbaren Reglerparameter $k_I, k_P > 0$ bieten damit Freiheitsgrade zur Verbesserung der Regelgüte, ohne dabei die Stabilität des Regelkreises zu beeinträchtigen.

Die daraus resultierende Struktur der Regelung kann durch

$$\begin{aligned}
\begin{pmatrix} \dot{\omega}_2(t) \\ \dot{\varphi}_w(t) \\ \dot{\omega}_1(t) \\ \dot{x}_F(t) \\ \dot{x}_I(t) \end{pmatrix} &= \begin{bmatrix} -(d_w + \mu_g)/J_2 & k_w/J_2 & d_w/J_2 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ d_w/J_1 & -k_w/J_1 & -d_w/J_1 & 0 & k_I/J_1 \\ 0 & 1 & 0 & -1/T_F & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ \omega_1(t) \\ x_F(t) \\ x_I(t) \end{pmatrix} + \\
&+ \begin{pmatrix} 0 \\ 0 \\ k_P/J_1 \\ 0 \\ 1 \end{pmatrix} v(t) + \begin{pmatrix} 0 \\ 0 \\ 1/J_1 \\ 0 \\ 0 \end{pmatrix} \hat{d}(t) - \begin{pmatrix} 1/J_2 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} [(\mathbf{N}\omega_2)(t) + M_2(t)] \\
y_{FI}(t) &= \begin{pmatrix} c_1 & c_2 & c_3 & -c_2/T_F & 0 \end{pmatrix} \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ \omega_1(t) \\ x_F(t) \\ x_I(t) \end{pmatrix} = y_F(t)
\end{aligned} \tag{6.97}$$

beschrieben werden und ist in Abbildung 6.10 als Signalfussplan dargestellt.

Nachfolgendes Theorem zeigt, dass die Kombination von nichtlinearem Zweimassensystem mit DT_1 -Filter und PI-Block durch Funnel-Control stabil beherrscht werden kann.

Theorem 6.8

Unter den Voraussetzungen

$$c_3 > 0, \quad c_2 \geq 0, \quad c_1 > -c_3 \tag{6.98}$$

und

$$T_F > 0, \quad k_I > 0, \quad k_P > 0 \tag{6.99}$$

ist die in Gleichung (6.97) angegebene Kombination aus Hochpassfilter, PI-Regler und nichtlinearem Zweimassensystem in einer einzigen Zustandsvariablen durch die Differentialgleichung

$$\dot{y}_{FI}(t) = (\mathbf{T}_{\mathbf{FI}} y_{FI})(t) + p_{FI}(t) + \frac{c_3 k_P}{J_1} v(t), \quad y_{FI}(0) = c \begin{pmatrix} \omega_2(0) & \varphi_w(0) & \omega_1(0) \end{pmatrix}^T \tag{6.100}$$

darstellbar, in der $\mathbf{T}_{\mathbf{FI}} \in \mathcal{Z}$ und $p_{FI} \in \mathcal{L}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R})$ gilt. Als solches gehört (6.97) der Systemklasse \mathcal{P} an. Folglich garantiert das Regelgesetz (6.2) mit (6.96), in welchem $\hat{y}(t) = y_{FI}(t)$ und $y_{\text{ref}}(t) = (c_1 + c_3)\omega_{\text{ref}}(t)$ zu setzen ist, die Einhaltung der Regelziele (i), (iii)-(v) und (vii).

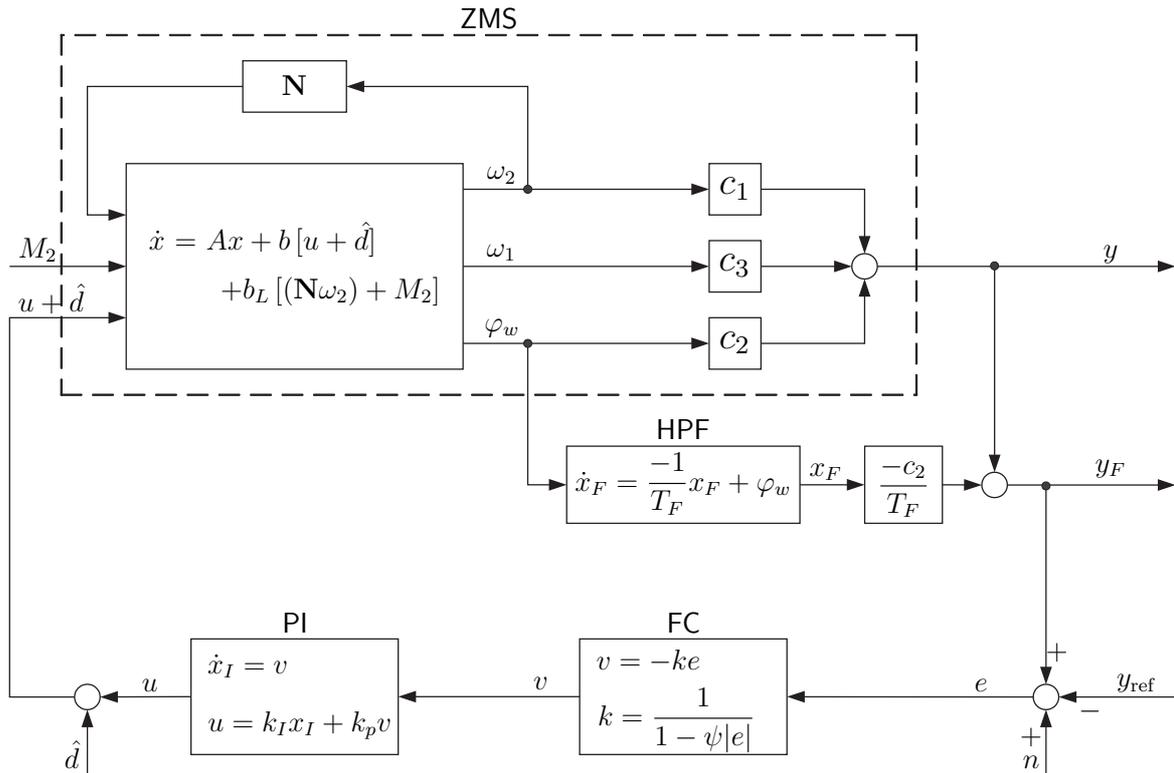


Abbildung 6.10: Signalflussplan der Gesamtstruktur der Regelung. Obwohl die mit HPF gekennzeichnete Differentialgleichung einen Tiefpass beschreibt, wird die Bezeichnung „Hochpass“ gewählt, weil das Übertragungsverhalten von $\varphi_w(\cdot)$ nach $y_F(\cdot)$ den Durchgriff über c_2 beinhaltet und somit im Gesamten Hochpassverhalten (DT_1 -Verhalten) besteht.

Beweis

Differenzieren von (6.97) ergibt

$$\begin{aligned} \dot{y}_{FI}(t) = \dot{y}_F(t) = cAx(t) - \frac{c_1}{J_2} [(N\omega_2)(t) + M_2(t)] + \frac{c_3}{J_1} \hat{d}(t) + \frac{c_3}{J_1} u(t) + \\ + \frac{c_2}{T_F} \left(\frac{1}{T_F} x_F(t) - \varphi_w(t) \right), \quad (6.101) \end{aligned}$$

wobei die Bedingung $c_3 > 0$ auf $r = 1$ führt. Darüberhinaus garantiert $c_3 > 0$ eine reguläre Matrix

$$S = \begin{bmatrix} c_1 & c_2 & c_3 & -c_2/T_F & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ c_1 & c_2 & c_3 & -c_2/T_F & -c_3 k_P / J_1 \end{bmatrix}, \quad S \in \mathbf{GL}_5(\mathbb{R}), \quad (6.102)$$

um das Anfangswertproblem (6.97) durch eine globale Transformation

$$\begin{pmatrix} y_{FI}(t) & \omega_2(t) & \varphi_w(t) & x_F(t) & \xi(t) \end{pmatrix}^T := S \begin{pmatrix} \omega_2(t) & \varphi_w(t) & \omega_1(t) & x_F(t) & x_I(t) \end{pmatrix}^T \quad (6.103)$$

in BINF zu überführen:

$$\begin{aligned} \dot{y}_{FI}(t) = & \rho_{FI} y_{FI}(t) + \delta_{FI} \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ x_F(t) \\ \xi(t) \end{pmatrix} - \frac{c_1}{J_2} [(\mathbf{N}\omega_2)(t) + M_2(t)] + \\ & + \frac{c_3}{J_1} [k_P v(t) + \hat{d}(t)] \quad (6.104a) \end{aligned}$$

$$\begin{aligned} \begin{pmatrix} \dot{\omega}_2(t) \\ \dot{\varphi}_w(t) \\ \dot{x}_F(t) \\ \dot{\xi}(t) \end{pmatrix} = & \Lambda_{FI} \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ x_F(t) \\ \xi(t) \end{pmatrix} + \begin{pmatrix} \frac{d_w}{J_2 c_3} \\ \frac{1}{c_3} \\ 0 \\ \rho_{FI} \end{pmatrix} y_{FI}(t) - \begin{pmatrix} \frac{1}{J_2} \\ 0 \\ 0 \\ \frac{c_1}{J_2} \end{pmatrix} [(\mathbf{N}\omega_2)(t) + M_2(t)] + \\ & + \begin{pmatrix} 0 \\ 0 \\ 0 \\ \frac{c_3}{J_1} \end{pmatrix} \hat{d}(t) \quad (6.104b) \end{aligned}$$

Dabei gelten die Abkürzungen:

$$\rho_{FI} = \frac{c_1 d_w}{c_3 J_2} + \frac{c_2}{c_3} - \frac{d_w}{J_1} + \frac{k_I}{k_P} \quad (6.105)$$

$$\delta_{FI} = \begin{pmatrix} -\frac{c_1 d_w}{J_2} - c_2 - \frac{c_1 c_2}{c_3} + \frac{(c_1 + c_3) d_w}{J_1} - \frac{c_1^2 d_w}{c_3 J_2} - \frac{c_1 \mu_g}{J_2} \\ \frac{c_1 k_w}{J_2} - \frac{c_3 k_w}{J_1} - \frac{c_2}{T_F} - \frac{c_1 c_2 d_w}{c_3 J_2} - \frac{c_2^2}{c_3} + \frac{c_2 d_w}{J_1} \\ \frac{c_1 c_2 d_w}{c_3 J_2 T_F} + \frac{c_2^2}{c_3 T_F} - \frac{c_2 d_w}{J_1 T_F} + \frac{c_2}{T_F^2} \\ -\frac{k_I}{k_P} \end{pmatrix}^T \quad (6.106)$$

$$\Lambda_{FI} = \begin{bmatrix} -\frac{d_w}{J_2} \left(1 + \frac{c_1}{c_3}\right) - \frac{\mu_g}{J_2} & \frac{k_w}{J_2} - \frac{d_w}{J_2} \frac{c_2}{c_3} & \frac{d_w}{J_2} \frac{1}{T_F} \frac{c_2}{c_3} & 0 \\ -\left(1 + \frac{c_1}{c_3}\right) & -\frac{c_2}{c_3} & \frac{1}{T_F} \frac{c_2}{c_3} & 0 \\ 0 & 1 & \frac{-1}{T_F} & 0 \\ & \delta_{FI} & & \end{bmatrix} \quad (6.107)$$

Wegen $\sigma(\Lambda_{FI}) = \sigma(\Lambda_F) \cup \{-k_I/k_P\} \subset \mathbb{C}_-$ ist durch den Beweis von Theorem 6.6 der weitere Verlauf vorgezeichnet, um das System in einer einzigen Zustandsvariablen $y_{FI}(t)$ darzustellen. ■

Auch ohne Hochpassfilter erfüllt die Kombination aus Zweimassensystem und PI-Regler die notwendigen Voraussetzungen. Wird in Gleichung (6.97) der Zustand x_F gestrichen, entsteht die Beschreibung:

$$\begin{pmatrix} \dot{\omega}_2(t) \\ \dot{\varphi}_w(t) \\ \dot{\omega}_1(t) \\ \dot{x}_I(t) \end{pmatrix} = \begin{bmatrix} -(d_w + \mu_g)/J_2 & k_w/J_2 & d_w/J_2 & 0 \\ -1 & 0 & 1 & 0 \\ d_w/J_1 & -k_w/J_1 & -d_w/J_1 & k_I/J_1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ \omega_1(t) \\ x_I(t) \end{pmatrix} + \\ + \begin{pmatrix} 0 \\ 0 \\ k_P/J_1 \\ 1 \end{pmatrix} v(t) + \begin{pmatrix} 0 \\ 0 \\ 1/J_1 \\ 0 \end{pmatrix} \hat{d}(t) - \begin{pmatrix} 1/J_2 \\ 0 \\ 0 \\ 0 \end{pmatrix} [(\mathbf{N}\omega_2)(t) + M_2(t)]$$

$$y_I(t) = \begin{pmatrix} c_1 & c_2 & c_3 & 0 \end{pmatrix} \begin{pmatrix} \omega_2(t) & \varphi_w(t) & \omega_1(t) & x_I(t) \end{pmatrix}^T = y(t) \quad (6.108)$$

Obwohl eine Regelungsstrategie ohne Hochpassfilter am beschriebenen Zielkonflikt leidet, sei aus Gründen der Vollständigkeit diese Möglichkeit erfasst.

Theorem 6.9

Unter den Voraussetzungen

$$c_3 > 0, \quad c_2 \geq 0, \quad c_1 > -c_3 \quad (6.109)$$

und

$$k_I > 0, \quad k_P > 0 \quad (6.110)$$

ist die in Gleichung (6.108) angegebene Kombination aus PI-Regler und nichtlinearem Zweimassensystem in einer einzigen Zustandsvariablen durch die Differentialgleichung

$$\dot{y}_I(t) = (\mathbf{T}_I y_I)(t) + p_I(t) + \frac{c_3 k_P}{J_1} v(t), \quad y_I(0) = c \begin{pmatrix} \omega_2(0) & \varphi_w(0) & \omega_1(0) \end{pmatrix}^T \quad (6.111)$$

darstellbar, in der $\mathbf{T}_I \in \mathcal{Z}$ und $p_I \in \mathcal{L}^\infty(\mathbb{R}_{\geq 0}; \mathbb{R})$ gilt. Als solches gehört (6.108) der Systemklasse \mathcal{P} an. Folglich garantiert das Regelgesetz (6.2) mit (6.96), in dem $\hat{y}(t) = y_I(t)$ und $y_{\text{ref}}(t) = (c_1 + c_3)\omega_{\text{ref}}(t)$ zu setzen ist, die Einhaltung der Regelziele (i), (iii), (iv) und (vii).

Beweis

Weil (6.63) die zeitliche Ableitung von $y(t)$ angibt, und $c_3 > 0$ gilt, liegt im System (6.108) der Relativgrad $r = 1$ vor. Wird auf (6.108) die Transformation

$$\begin{pmatrix} y_I(t) \\ \omega_2(t) \\ \varphi_w(t) \\ \xi(t) \end{pmatrix} := \begin{bmatrix} c_1 & c_2 & c_3 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ c_1 & c_2 & c_3 & -c_3 k_P/J_1 \end{bmatrix} \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ \omega_1(t) \\ x_I(t) \end{pmatrix} \quad (6.112)$$

angewendet, entsteht analog zu (6.104a), (6.104b) die BINF, wobei wiederum der Zustand x_F entfällt. Mit den Abkürzungen ρ_{FI} aus Gleichung (6.105) und

$$\Lambda_I = \begin{bmatrix} -\frac{d_w}{J_2} \left(1 + \frac{c_1}{c_3}\right) - \frac{\mu_g}{J_2} & \frac{k_w}{J_2} - \frac{d_w}{J_2} \frac{c_2}{c_3} & 0 \\ -\left(1 + \frac{c_1}{c_3}\right) & -\frac{c_2}{c_3} & 0 \\ * & * & -\frac{k_I}{k_P} \end{bmatrix} \quad (6.113)$$

lautet die Differentialgleichung der Nulldynamik:

$$\begin{pmatrix} \dot{\omega}_2(t) \\ \dot{\varphi}_w(t) \\ \dot{\xi}(t) \end{pmatrix} = \Lambda_I \begin{pmatrix} \omega_2(t) \\ \varphi_w(t) \\ \xi(t) \end{pmatrix} + \begin{pmatrix} \frac{d_w}{J_2 c_3} \\ \frac{1}{c_3} \\ \rho_{FI} \end{pmatrix} y_I(t) - \begin{pmatrix} \frac{1}{J_2} \\ 0 \\ \frac{c_1}{J_2} \end{pmatrix} [(\mathbf{N}\omega_2)(t) + M_2(t)] + \begin{pmatrix} 0 \\ 0 \\ \frac{c_3}{J_1} \end{pmatrix} \hat{d}(t) \quad (6.114)$$

Deren Zustandsmatrix Λ_I besitzt ein Spektrum, das sich wegen

$$\sigma(\Lambda_I) = \sigma \left(\begin{bmatrix} -\frac{d_w}{J_2} \left(1 + \frac{c_1}{c_3}\right) - \frac{\mu_g}{J_2} & \frac{k_w}{J_2} - \frac{d_w}{J_2} \frac{c_2}{c_3} \\ -\left(1 + \frac{c_1}{c_3}\right) & -\frac{c_2}{c_3} \end{bmatrix} \right) \cup \{-k_I/k_P\} = \sigma(\Lambda) \cup \{-k_I/k_P\} \quad (6.115)$$

aus dem Spektrum der in Gleichung (6.70) definierten Hurwitzmatrix Λ und aus $-k_I/k_P < 0$ zusammensetzt. Weil folglich Λ_I ebenfalls ein stabiles Spektrum besitzt, ist der Rest des Beweises in identischer Weise wie in Theorem 6.6 zu führen. ■

Die in den Theoremen 6.6 bis 6.9 betrachteten, unterschiedlichen Kombinationsmöglichkeiten von Regler (6.2), Hochpassfilter (6.40) und integralem Anteil (6.96) werden im nachfolgenden Theorem auf ihre Auswirkung bezüglich der eingangs des Kapitels festgelegten Regelziele untersucht.

Theorem 6.10

Aus dem in Definition 4.1 festgelegten Funktionsraum \mathcal{F} sei mit $\psi \in \mathcal{F}$ eine beschränkte und stetige Funktion $\psi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ mit der Eigenschaft

$$\forall t > 0 : \psi(t) > 0 \quad (6.116)$$

vorgegeben, die durch ihren Kehrwert

$$\partial F_\psi(t) := \frac{1}{\psi(t)} > 0 \quad \forall t > 0 \quad (6.117)$$

einen Trichterrand beschreibt. Sofern das Zweimassensystem (6.55), (6.56) die Bedingung

$$c_3 > 0, \quad c_2 \geq 0, \quad c_1 > -c_3, \quad (6.118)$$

erfüllt, führt jeder der folgenden Regler

(FC): Proportionalregler (6.2) mit $\hat{y}(t) = y(t)$, $y_{\text{ref}}(t) = (c_1 + c_3)\omega_{\text{ref}}(t)$,
 $v(t) = u(t)$

(HPF-FC): Proportionalregler (6.2) und HPF (6.40) mit $\hat{y}(t) = y_F(t)$, $y_{\text{ref}}(t) =$
 $(c_1 + c_3)\omega_{\text{ref}}(t)$, $v(t) = u(t)$

(PI-FC): Proportionalregler (6.2) und PI-Regler (6.96) mit $\hat{y}(t) = y_I(t)$,
 $y_{\text{ref}}(t) = (c_1 + c_3)\omega_{\text{ref}}(t)$

(HPF-PI-FC): Proportionalregler (6.2), HPF (6.40) und PI-Regler (6.96) mit
 $\hat{y}(t) = y_{FI}(t)$, $y_{\text{ref}}(t) = (c_1 + c_3)\omega_{\text{ref}}(t)$

auf einen geschlossenen Regelkreis, der für alle Sollwertverläufe $y_{\text{ref}} \in W^{1,\infty}(\mathbb{R}_{\geq 0}; \mathbb{R})$
und für alle beliebigen Anfangswerte $x(0) = (\omega_2(0), \varphi(0), \omega_1(0))^T \in \mathbb{R}^3$ eine Lösung
besitzt, die auf dem Intervall $[0, \infty[$ maximal erweiterbar ist, wobei die Funktionen
 $\omega_2(\cdot)$, $\varphi_w(\cdot)$ und $\omega_1(\cdot)$ auf dem Intervall $[0, \infty[$ beschränkt sind. Desweiteren gilt:

(i) es existiert eine Zahl $\varepsilon \in]0, 1[$ mit der Eigenschaft $t \geq 0$: $|e(t)| < 1/\psi(t) - \varepsilon$;

(ii) falls der PI-Regler involviert ist, d.h. im Falle (PI-FC) oder (HPF-PI-FC), gilt
für alle $t \geq 0$ die Aussage $|e(t)| < |\dot{x}_I(t)|$.

Durch (i) ist gewährleistet, dass der Regelfehler innerhalb des Trichters verläuft und
dabei vom Rand wegbeschränkt bleibt. Theorem 4.6 beinhaltet die Aussage, dass der
Fehler einen Mindestabstand zum Trichterrand einhält. Aussage (ii) beschränkt den
Betrag des Regelfehlers durch die zeitliche Ableitung der Zustandsvariablen $x_I(t)$ des
PI-Reglers. Daran wird offensichtlich, dass der Regelfehler verschwindet, falls der Re-
gelkreis gegen eine Ruhelage konvergiert.

Beweis

(FC): Mit Anwendung von Theorem 6.6 lässt sich die Strecke (6.55), (6.56) in der
Form (6.62) darstellen. Letztere genügt den in Theorem 4.6 geforderten Voraussetzungen,
so dass der Regler (FC) anwendbar ist und, mit Ausnahme von (ii), die Behauptungen
von Theorem 6.10 erfüllt.

(HPF-FC): Es gilt mit Verwendung von Theorem 6.7 anstelle von 6.6 dieselbe Argu-
mentation wie im Falle (FC).

(PI-FC): Es gilt mit Verwendung von Theorem 6.9 anstelle von 6.6 dieselbe Argumen-
tation wie im Falle (FC).

(HPF-PI-FC): Es gilt mit Verwendung von Theorem 6.8 anstelle von 6.6 dieselbe Argu-
mentation wie im Falle (FC).

Aussage (ii), die lediglich in den Fällen (PI-FC) und (HPF-PI-FC) Gültigkeit besitzt,
ergibt sich direkt aus $\dot{x}_I(t) = v(t) = -k(t)e(t)$ und $k(t) > 1$. ■

Fundamental und einschränkend für die Anwendung von Funnel-Control ist die Bedin-
gung an den Relativgrad der Strecke. Weil sich für $y(t) = \omega_2(t)$ ein Relativgrad von 2
ergibt, kann die relevante Lastdrehzahl $\omega_2(t)$ nicht als Regelgröße angesetzt werden.
Wird eine Zustandsauskopplung zur Relativgradreduktion verwendet, verliert die Last-

drehzahl regelungstechnisch ihre Bedeutung, weil sich der Regler auf die Linearkombination aller Zustände fokussiert. Weiterhin bleibt aber aus Sicht der Antriebstechnik das Interesse am Verlauf von $\omega_2(\cdot)$ bestehen. Insofern ist der Zusammenhang zwischen Regelgröße $y(t)$ bzw. Regelfehler $e(t) = \hat{y}(t) - y_{\text{ref}}(t)$, den Zustandsgrößen und der Prozessgröße $\omega_2(t)$ bzw. $\omega_2(t) - \omega_{\text{ref}}(t)$ zu klären.

Theorem 6.11 Sei $\omega_{\text{ref}} \in W^{1,\infty}(\mathbb{R}_{\geq 0}; \mathbb{R})$ der gewünschte Verlauf für die Lastdrehzahl $\omega_2(\cdot)$. Wird daraus über

$$y_{\text{ref}}(\cdot) := (c_1 + c_3)\omega_{\text{ref}}(\cdot) \quad (6.119)$$

eine Sollwertvorgabe für den Regelkreis ermittelt, so ergibt unter den in Theorem 6.10 getroffenen Voraussetzungen die Anwendung der Regler (FC), (HPF-FC), (PI-FC) bzw. (HPF-PI-FC) auf ein System (6.55), (6.56) für jeden beliebigen Anfangszustand $x(0) = (\omega_2(0), \varphi_w(0), \omega_1(0))^T \in \mathbb{R}^3$ einen geschlossenen Kreis, der die in Theorem 6.10 genannten Eigenschaften besitzt. Darüberhinaus gelten für alle Zeitpunkte $t \geq 0$ die Zusammenhänge:

$$(FC): \quad \omega_2(t) - \omega_{\text{ref}}(t) = \frac{1}{c_1 + c_3} [e(t) - c_2\varphi_w(t) - c_3\dot{\varphi}_w(t)]$$

$$(HPF-FC): \quad \omega_2(t) - \omega_{\text{ref}}(t) = \frac{1}{c_1 + c_3} [e(t) - c_2\varphi_w(t) - c_3\dot{\varphi}_w(t) + \frac{c_2}{T_F}x_F(t)]$$

$$(PI-FC): \quad \omega_2(t) - \omega_{\text{ref}}(t) = \frac{1}{c_1 + c_3} [-k(t)^{-1}\dot{x}_I(t) - c_2\varphi_w(t) - c_3\dot{\varphi}_w(t)]$$

$$(HPF-PI-FC): \quad \omega_2(t) - \omega_{\text{ref}}(t) = \frac{1}{c_1 + c_3} [-k(t)^{-1}\dot{x}_I(t) - c_2\varphi_w(t) - c_3\dot{\varphi}_w(t) + \frac{c_2}{T_F}x_F(t)]$$

Im Bezug auf stationäre Genauigkeit ist herauszuheben, dass in der Gleichgewichtslage $\dot{\varphi}_w(t) = 0$, $\dot{x}_I(t) = 0$ und $x_F(t) = T_F\varphi_w(t)$ gilt.

Beweis

(FC): Aus $y(t) = c_1\omega_2(t) + c_2\varphi_w(t) + c_3\omega_1(t)$ und (6.119) ergibt sich

$$y(t) - y_{\text{ref}}(t) = (c_1 + c_3)(\omega_2(t) - \omega_{\text{ref}}(t)) + c_2\varphi_w(t) + c_3(\omega_1(t) - \omega_2(t)), \quad (6.120)$$

was zusammen mit $\dot{\varphi}_w(t) = \omega_1(t) - \omega_2(t)$ und $e(t) = y(t) - y_{\text{ref}}(t)$ auf die Behauptung führt.

(HPF-FC): Aus $y_F(t) = y(t) - c_2/T_F x_F(t)$ und (6.119) ergibt sich

$$y_F(t) - y_{\text{ref}}(t) = (c_1 + c_3)(\omega_2(t) - \omega_{\text{ref}}(t)) + c_2\varphi_w(t) + c_3(\omega_1(t) - \omega_2(t)) - \frac{c_2}{T_F}x_F(t), \quad (6.121)$$

was zusammen mit $\dot{\varphi}_w(t) = \omega_1(t) - \omega_2(t)$ und $e(t) = y_F(t) - y_{\text{ref}}(t)$ auf die Behauptung führt.

(PI-FC): Aus Gleichung (6.96) und (6.2) ergibt sich $e(t) = -\dot{x}_I(t)/k(t)$. Einsetzen in die Aussage für (FC) führt auf die Behauptung.

(HPF-PI-FC): Aus Gleichung (6.96) und (6.2) ergibt sich $e(t) = -\dot{x}_I(t)/k(t)$. Einsetzen in die Aussage für (HPF-FC) führt auf die Behauptung. ■

Theorem 6.11 stellt eine Entscheidungshilfe zur Verfügung, um zu einem vorgegebenen Regelziel den geeigneten Regler zu wählen.

Zwar lässt sich der Regelfehler $e(t)$ theoretisch durch einen schmalen Trichter beliebig klein halten. Aus unterschiedlichen Gründen ist jedoch bei praktischen Anwendungen ein breiter Trichterrand vorzuziehen, der als unerwünschte Nebenwirkung einen erhöhten Regelfehler gestattet. Hier ist der Integralanteil gewinnbringend, der $e(t)$ durch $-\dot{x}_I(t)/k(t)$ ersetzt und somit in der Ruhelage $e(t) = 0$ garantiert, unabhängig von der Breite des Trichters.

Theorem 6.11 ist zu entnehmen, dass der Regler (FC) im Hinblick auf die prozesstechnische Aufgabenstellung kein zufriedenstellendes Ergebnis bringt. Selbst wenn $e(t)$ als vernachlässigbar klein betrachtet werden kann, muss der Verdrehwinkel der Differentialgleichung

$$\dot{\varphi}_w(t) = -\frac{c_2}{c_3} \varphi_w(t) + \frac{1}{c_3} e(t) \quad (6.122)$$

genügen, um den Idealfall $\omega_2(\cdot) - \omega_{\text{ref}}(\cdot) \equiv 0$ zu erreichen. Diese Forderung an das Verhalten des Torsionswinkels ist jedoch nicht notwendig stimmig mit der Eigendynamik des mechanischen Systems, die $\varphi_w(t) = \int_0^t [\omega_1(\tau) - \omega_2(\tau)] d\tau + \varphi_w(0)$ vorschreibt.

Weil im Allgemeinen $\varphi_w(\cdot)$ die Differentialgleichung (6.122) nicht erfüllt, treten hochfrequente Vibrationen der Welle mit der inhärenten Eigenschaft $\|\dot{\varphi}_w(\cdot)\|_\infty \gg 0$ unmittelbar in der Differenz $\omega_2(t) - \omega_{\text{ref}}(t)$ störend in Erscheinung.

Selbst in der Ruhelage ($\dot{\varphi}_w(t) = 0$) bleibt eine Abweichung bestehen, die durch den Verdrehwinkel $\varphi_w(t) \neq 0$ verursacht ist. Die lineare Federcharakteristik der Welle führt auf einen Torsionswinkel, der proportional mit dem übertragenen Antriebsmoment anwächst. Deshalb ist nicht zu erwarten, dass $\omega_2(t)$ gegen den gewünschten Sollwert $\omega_{\text{ref}}(t) = \text{const.}$ konvergiert. Um diesen Effekt zu unterbinden, wird der Hochpassfilter eingesetzt, dessen Aufgabe die Auslöschung des Terms $c_2\varphi_w(t)$ im stationären Zustand umfasst. Gleichung (6.41) zeigt, dass der Filter dem Ausgang $y(t)$ den Term $-c_2x_F(t)/T_F$ hinzufügt, der, ausgewertet in der Ruhelage $\dot{x}_F(t) = 0$, auf

$$c_2\varphi_w(t) - \frac{c_2}{T_F}x_F(t) = 0 \quad (6.123)$$

führt. Nachdem in der Ruhelage gleichzeitig auch $\dot{\varphi}_w(t) = 0$ gilt, ist die Abweichung $\omega_2(t) - \omega_{\text{ref}}(t)$ proportional zu $e(t)$ und daher beliebig klein einstellbar.

6.4 Messergebnisse

Der in Kapitel 2 vorgestellte Prüfstand des Lehrstuhls bietet die Möglichkeit, die in Theorem 6.11 beschriebenen Regelungsmethoden zu validieren.

Für alle Messungen ist der Trichterrand durch die Funktion $t \mapsto \psi(t) = (150 \cdot e^{-2t^2} + 80)^{-1}$ festgelegt, das Sollwertsignal $t \mapsto \omega_{\text{ref}}(t) = 30 \cdot (1 - e^{-t/0.3})$ rad/s konvergiert asymptotisch gegen einen konstanten Arbeitspunkt. Um das Störverhalten zu ermitteln, wird ab dem Zeitpunkt $t = 10$ s ein konstantes Lastmoment $M_2 = 10$ Nm zugeschaltet. Ein Störsignal $\hat{d}(\cdot)$ wird am Eingang nicht explizit aufgebracht, ist jedoch

implizit wirksam aufgrund unvermeidbarer Unterschiede zwischen idealem und realem Betriebsverhalten des Umrichters.

Die durchgeführten Simulationen sowie Experimente legen nahe, dass der Verdrehwinkel eine wesentliche Rolle spielt, um die vorgegebenen Regelziele zu erreichen. Deshalb findet der Verdrehwinkel in der Linearkombination

$$y(t) = \begin{pmatrix} 0.1 & 80 & 0.5 \end{pmatrix} \begin{pmatrix} \omega_2(t) & \varphi_w(t) & \omega_1(t) \end{pmatrix}^T \quad (6.124)$$

Berücksichtigung, wobei dessen nachteiliger Einfluss auf das stationäre Verhalten der Regelung durch einen Hochpassfilter eliminiert wird.

Die Auswirkungen des DT_1 -Filters und des PI-Reglers sind in den Abbildungen 6.11 bis 6.16 erkennbar. Das Diagramm in Abbildung 6.11 zeigt die Prozessgröße $\omega_2(\cdot)$, sowie den zugehörigen Referenzwert $\omega_{\text{ref}}(\cdot)$. Diagramm 6.12 enthält die Linearkombination $\hat{y}(\cdot)$ mit Referenzwert $y_{\text{ref}}(\cdot)$. Abbildung 6.13 stellt die prozesstechnisch relevante Abweichung $\omega_2(\cdot) - \omega_{\text{ref}}(\cdot)$ dar, Abbildung 6.14 den Regelfehler $e(\cdot)$.

Wegen der hohen Gewichtung des Verdrehwinkels mit $c_2 = 80$ werden sämtliche Schwingungen im Regelfehler sowie in der Prozessgröße von allen Reglern unterdrückt. Ohne Integralteil bleibt jedoch, verursacht durch Reib- und Lastmoment, stets ein stationärer Regelfehler bestehen, der sich in einer unerwünschten Abweichung der Prozessgröße niederschlägt. Die Variante (FC) führt unter dem Einfluss der Reibung auf eine Drehzahl $\omega_2 \approx 26.7$ rad/s, eine zusätzliche Last von $M_2 = 10$ Nm bremsen die Drehzahl auf $\omega_2 \approx 13.2$ rad/s ab, die einem inakzeptablen Regelfehler von ca. 56% entspricht.

Wie sich aus Abbildung 6.14 ablesen lässt, konvergiert $e(t)$ auch unter Lasteinfluss gegen 0, wenn ein I-Anteil in den Regelkreis eingebracht wird. Es ist allerdings aus den Abbildungen 6.13 und 6.14 ersichtlich, dass sich selbst für $e(t) = 0$ unter Last eine Differenz $\omega_2(t) - \omega_{\text{ref}}(t) \neq 0$ einstellt, sofern kein Hochpassfilter eingesetzt wird. Diese Beobachtung deckt sich mit Theorem 6.11, das für $\varphi_w \neq 0$ die Abweichung der Prozessgröße quantifiziert. Hinsichtlich einer Drehzahlregelung, die sich auf $\omega_2(t)$ fokussiert, ist daher die alleinige Erweiterung von (FC) mit einem PI-Baustein nicht zielführend.

Daher ist die Aufgabe des Hochpassfilters, den Term $c_2\varphi_w$, der im Falle von (PI-FC) auftritt, zu streichen, dabei jedoch keine Information über etwaige Schwingungen aus dem Signal $\hat{y}(\cdot)$ zu entfernen. Durch diese Erweiterung bleibt die gewünschte Dämpfung erhalten, in der Ruhelage wird dagegen der negative Einfluss des Verdrehwinkels unterdrückt. Dadurch wird von $\varphi_w \neq 0$ stationär keine Abweichung in der Prozessgröße verursacht. Es ist festzuhalten, dass (FC-HPF) mit $c_2 \neq 0$ die selbe stationäre Abweichung ergibt, wie (FC) mit $c_2 = 0$, jedoch mit dem Vorteil einer verbesserten Dämpfung.

Schließlich werden die Regelziele durch den Einsatz von (HPF-PI-FC) erreicht. Der integrale Anteil erzwingt $e(t) = 0$ in der Ruhelage, so dass, selbst bei Lasteinwirkung, die Prozessgröße asymptotisch gegen den Referenzwert konvergiert.

Die Diagramme 6.15 und 6.16 zeigen den Verlauf der Reglerverstärkung und der Stellgröße. Deutlich zu erkennen ist das nicht-monotone Verhalten der Verstärkung sowie deren Konvergenz; selbst für $e(t) \neq 0$ tritt kein unbeschränktes Driften auf.

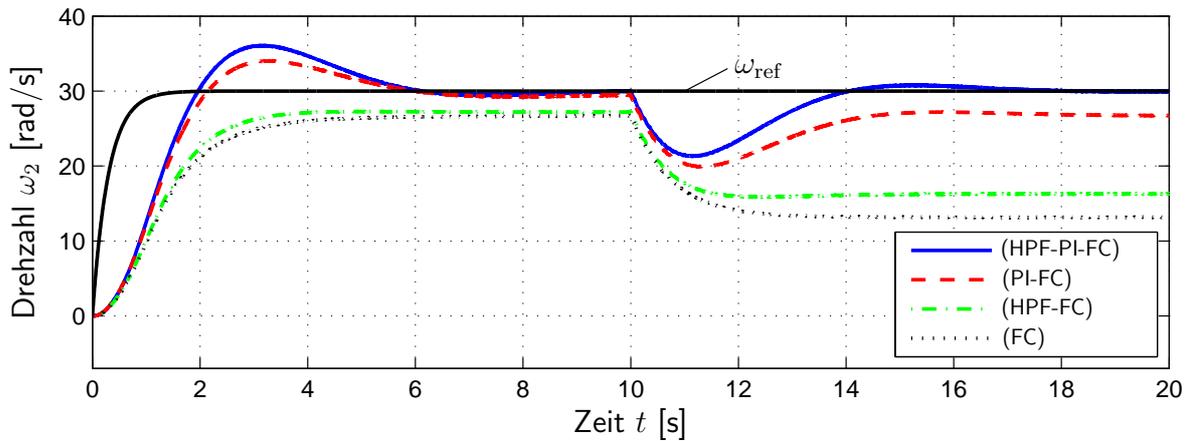


Abbildung 6.11: Messergebnis für (FC), (PI-FC), (HPF-FC) und (HPF-PI-FC) am Prüfstand Zweimassensystem.

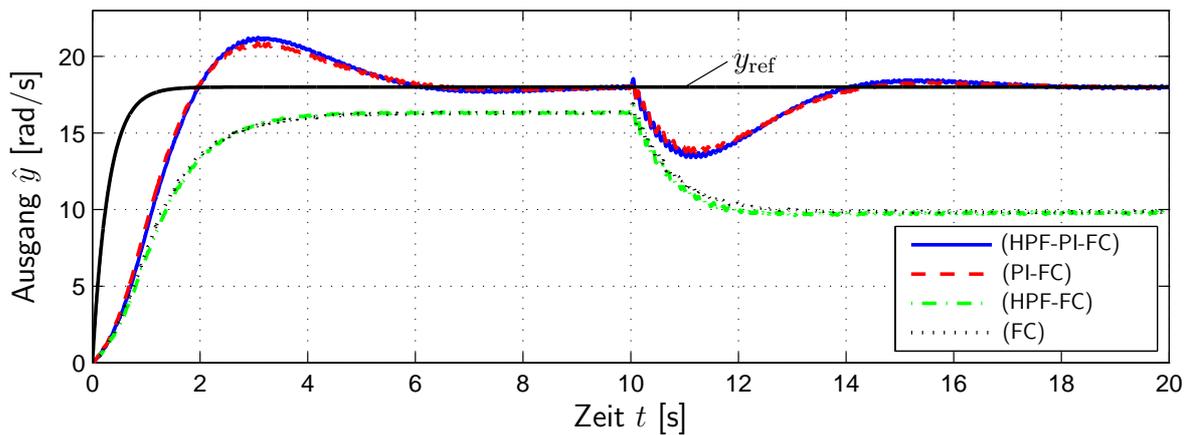


Abbildung 6.12: Messergebnis für (FC), (PI-FC), (HPF-FC) und (HPF-PI-FC) am Prüfstand Zweimassensystem.

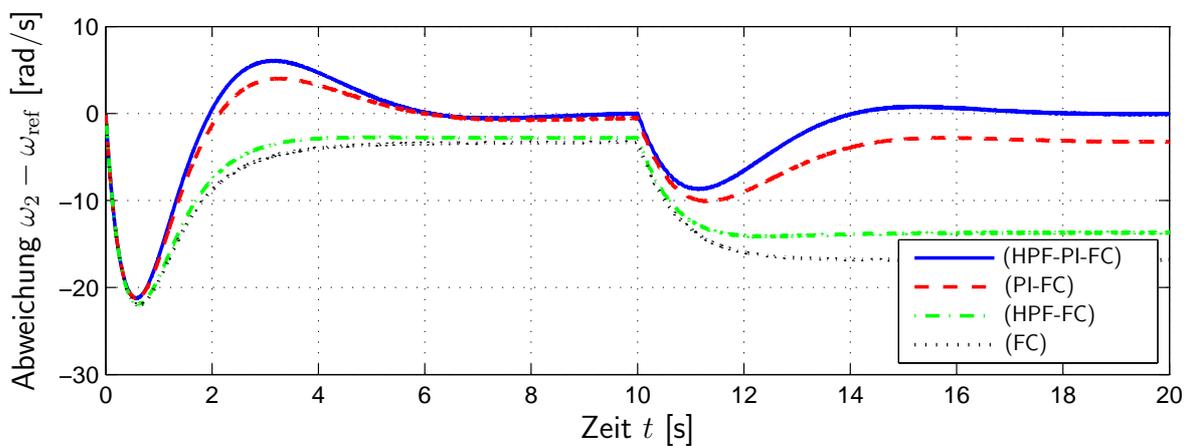


Abbildung 6.13: Messergebnis für (FC), (PI-FC), (HPF-FC) und (HPF-PI-FC) am Prüfstand Zweimassensystem.

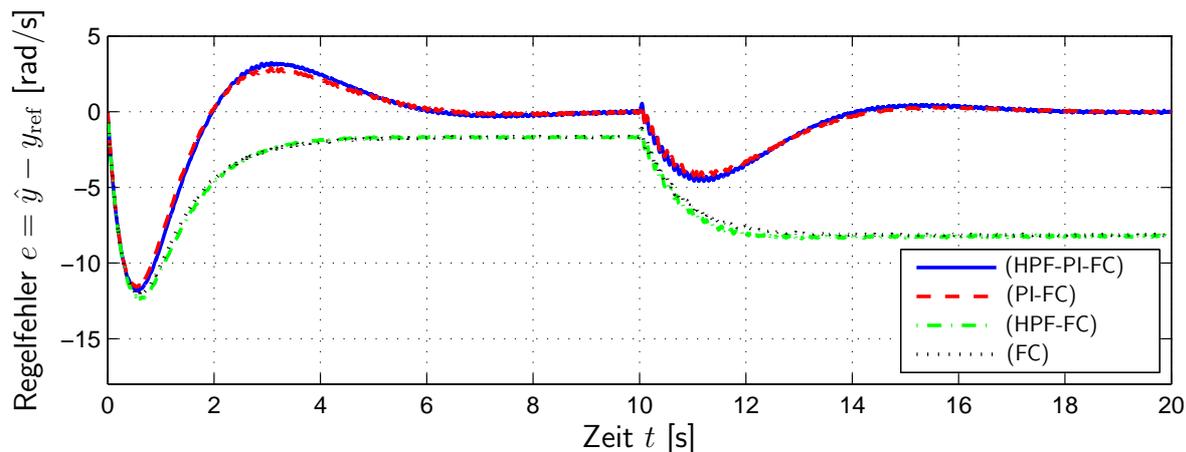


Abbildung 6.14: Messergebnis für (FC), (PI-FC), (HPF-FC) und (HPF-PI-FC) am Prüfstand Zweimassensystem.

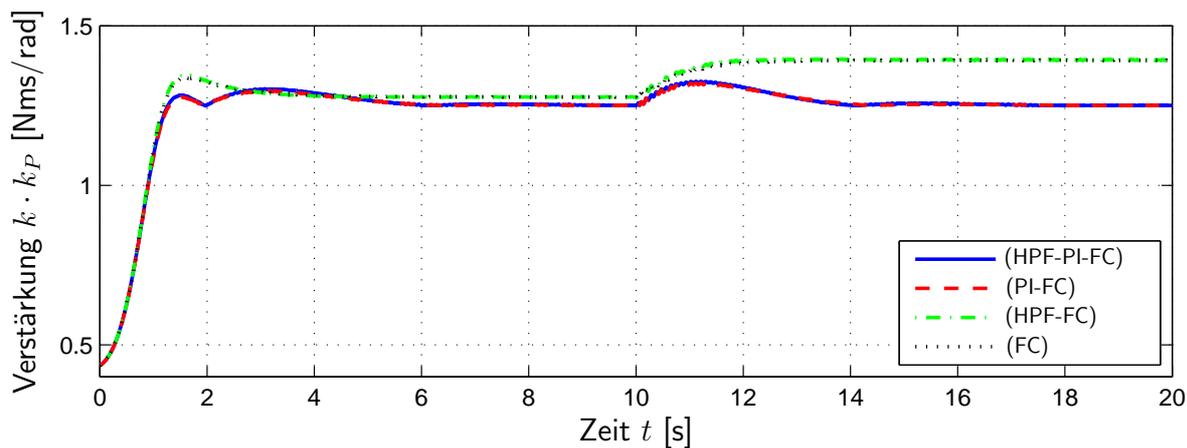


Abbildung 6.15: Messergebnis für (FC), (PI-FC), (HPF-FC) und (HPF-PI-FC) am Prüfstand Zweimassensystem.

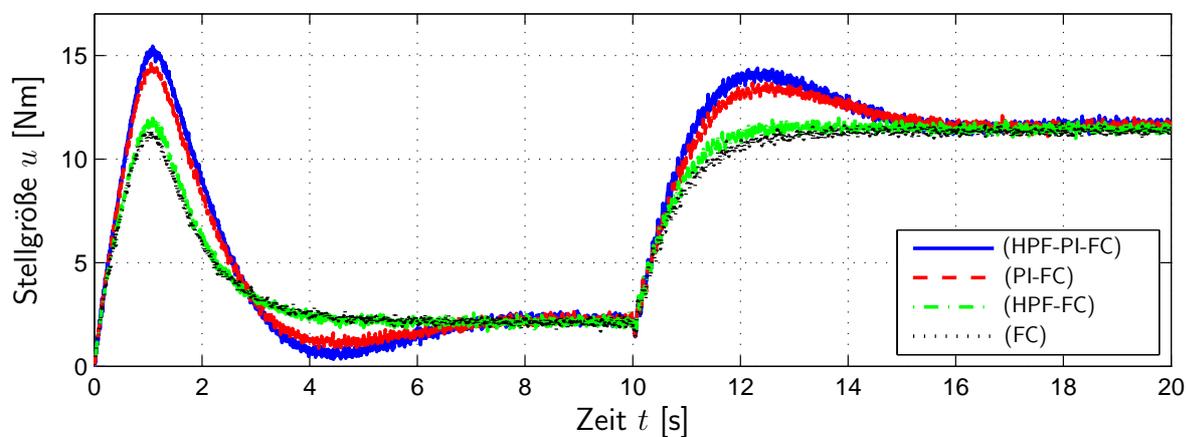


Abbildung 6.16: Messergebnis für (FC), (PI-FC), (HPF-FC) und (HPF-PI-FC) am Prüfstand Zweimassensystem.

Kapitel 7

Zusammenfassung und Ausblick

Die vorliegende Arbeit setzt den Schwerpunkt auf nicht-parameterbasierte Strategien zur Regelung nichtlinearer Antriebssysteme, die sich mit geringem Aufwand implementieren lassen, weil aus ingenieurtechnischer Perspektive letztendlich die Anwendbarkeit an realen Problemstellungen in den Vordergrund rückt und das entscheidende Kriterium für die Akzeptanz eines technologischen Systems darstellt. Der in Kapitel 3.4 beschriebene Ansatz mit monotonem Adaptionsgesetz weist nicht das geforderte Maß an Robustheit gegenüber Störeinflüssen und Sollwertänderungen auf, so dass beim Einsatz an realen Anlagen ein unbeschränktes Anwachsen der Verstärkung zu erwarten ist. Durch Einführung eines λ -Bandes [35, Kap. 5.2] kann das Driften der Verstärkung verhindert und die nötige Robustheit erreicht werden. Es ist zwar sichergestellt, dass der Regelfehler für $t \rightarrow \infty$ gegen das vorgegebene λ -Band, d.h. gegen das Intervall $[-\lambda, \lambda]$, $\lambda > 0$ konvergiert. Eine Aussage über das transiente Verhalten ist jedoch nicht impliziert. An dieser Stelle hat die in Kapitel 4 dargestellte Regelungsstrategie ihre Vorteile. Zum einen lässt sich das transiente Verhalten vor Inbetriebnahme der Regelung beliebig vorgeben, zum anderen ist die Monotonie der Verstärkung aufgehoben.

Im Sinne einer realitätsnahen Reglersynthese werden in der Struktur des Streckenmodells folgende Eigenschaften berücksichtigt:

- elastische Ankopplung der Arbeitsmaschine (Zweimassensystem)
- Parameter mit unbekanntem Wert
- nichtlineare Effekte (drehzahlabhängige Reibung an der Lastmasse mit unbekannter Charakteristik)
- Störgrößen (Lastmomente, Messrauschen, Verfälschung der Stellgröße durch den Aktor)

Die gewählte Methode zur Regelung solcher Prozesse nutzt das Modell einzig zur Analyse struktureller Eigenschaften, ohne sich dabei auf die expliziten Werte der Parameter zu stützen. Deshalb kann auf eine störempfindliche und aufwändige Schätzung der Parameterwerte verzichtet werden.

Für die Anwendung an Zweimassensystemen wird ein hochverstärkungsbasierter Drehzahlregler entworfen. Nachdem solche Strecken einen relativen Grad von 2 besitzen, wird mittels Zustandsauskopplung eine Relativgradreduzierung realisiert, wobei trotz unbekannter Streckenparameter die erforderliche Stabilität der Nulldynamik erhalten bleibt.

Zur Verbesserung des Regelergebnisses wird der hochverstärkungsbasierte Proportionalregler zu einem PI-Regler erweitert. Hierdurch zeigt sich vor allem unter Lasteinwirkung ein positiver Effekt, da in der Gleichgewichtslage stationäre Genauigkeit erreicht wird.

Zusätzlich wird eine Hochpassfilterung des Torsionswinkels eingeführt, um den Zielkonflikt zwischen aktiver Bedämpfung von mechanischen Schwingungen und kleinem Regelfehler in der Ruhelage aufzulösen.

Darüberhinaus ist der Nachweis auf mathematischem Wege erbracht, dass sowohl PI-Regler, als auch Hochpassfilter in den Regelkreis eingebracht werden dürfen, ohne dessen Stabilitätseigenschaften zu beeinträchtigen. Die Kombination aus hochverstärkungsbasiertem Proportionalregler mit zeitvarianter Verstärkung, PI-Regler und Hochpassfilter garantiert die Einhaltung einer vorgegebenen Fehlergrenze, gewährleistet aktive Dämpfung, sowie stationäre Genauigkeit und toleriert Störungen und Messrauschen. Damit ist der Weg für eine Umsetzung des entworfenen Konzeptes an der realen Anlage geebnet.

Die experimentelle Erprobung der entwickelten Regelungsstrategie bestätigt am Laborprüfstand die theoretischen Resultate und Simulationen. Trotz der bestehend einfachen Reglerstruktur sind die Messergebnisse vielversprechend hinsichtlich industrieller Einsatzfähigkeit.

Ein interessantes Feld für zukünftige Forschungsaktivitäten ist in der Erweiterung und Anwendung des Regelungskonzeptes auf Mehrmassensysteme und flexible Roboterarme gegeben. Solche Prozesse sind in wesentlichem Umfang in der Antriebstechnik vertreten und weisen ein verschärftes Relativgradproblem auf. Neben diesen praktisch orientieren Themen ist die Fragestellung nach dem tatsächlichen Erreichen einer Ruhelage offen. Die theoretische Analyse des Regelkreisverhaltens führt lediglich zu der Feststellung, dass der Regelfehler innerhalb eines wählbaren Toleranzbandes verbleibt. Damit ist jedoch keine Aussage über asymptotische Konvergenz verbunden. Die Ergebnisse der vorgenommenen Simulationen und Messungen zeigen zwar, dass in allen Fällen eine Gleichgewichtslage asymptotisch erreicht wird, ein theoretisches Fundament hierfür ist allerdings noch nicht geschaffen.

Anhang A

Grundlagen und Definitionen

Ausgewählte Definitionen sowie Grundlagen, die im Hinblick auf nichtlineare dynamische Systeme von Bedeutung sind, werden im Folgenden näher beschrieben. Weiterführende Literatur zur Systemtheorie findet sich beispielsweise in den Lehrbüchern von Desoer et al. [18], Föllinger [23], Hinrichsen et al. [31], Ioannou et al. [46], Isidori [47], Kailath [48], Khalil [50], Reinschke [79], Sastry [82], Schwarz [93], Slotine et al. [94] oder Sontag [95]. Die ausschließlich mathematischen Themengebiete sind unter anderem in den Büchern von Amann et al. [2], [3], [4], Königsberger [53], Lancaster et al. [55] und Meyberg et al. [64], [65] behandelt.

Die Intention dieses Anhangs ist einerseits Schreibweisen festzulegen und Definitionen einzuführen, die in der vorliegenden Arbeit verwendet werden. Andererseits soll die gewählte Darstellungsweise systemtheoretische Grundlagen von einem technisch-physikalischen Standpunkt aus beleuchten.

A.1 Abkürzungen und Schreibweisen

| | |
|------|---|
| ARMA | Auto-Regressive-Moving-Average Modell |
| HWP | Hurwitzpolynom, d.h. ein Polynom mit Wurzeln, die ausschließlich negative Realteile besitzen. |
| BNF | Beobachtungsnormalform |
| RNF | Regelungsnormalform |
| BINF | Byrnes-Isidori-Normalform |
| LTI | Zeitinvariantes lineares System (Linear Time-Invariant) |
| LTV | Zeitvariantes lineares System (Linear Time-Variant) |
| SISO | Eingrößensystem (Single-Input Single-Output) |
| MIMO | Mehrgrößensystem (Multi-Input Multi-Output) |
| BIBO | Eingangs-/Ausgangsstabil (Bounded-Input Bounded-Output) |
| WOK | Wurzelortskurve |

| | |
|-----------|---|
| SO | Symmetrisches Optimum |
| BO | Betragsoptimum |
| DO | Dämpfungsoptimum |
| EMS | Einmassensystem |
| ZMS | Zweimassensystem |
| PI | PI-Block (Parallelschaltung aus proportionalem und integralem Baustein) |
| HPF | Hochpassfilter |
| E/A-Verh. | Eingangs-/Ausgangsverhalten |

| | |
|---|---|
| $\mathbb{R}_{\geq 0}$ | $:= [0, \infty[$ (Intervall der nicht-negativen reellen Zahlen) |
| $\mathcal{L}^p(D; \mathbb{R})$ | Raum der (Lebesgue-)messbaren Funktionen $x : D \rightarrow \mathbb{R}$ |
| $\mathcal{L}_{lok}^\infty(D; \mathbb{R})$ | Raum der (Lebesgue-)messbaren, lokal essentiell beschränkten Funktionen $x : D \rightarrow \mathbb{R}$ |
| $\mathcal{L}_{lok}^1(D; \mathbb{R})$ | Raum der (Lebesgue-)messbaren, lokal integrierbaren Funktionen $x : D \rightarrow \mathbb{R}$ |
| $\mathcal{C}^0(D; \mathbb{R})$ | Raum der stetigen Funktionen $x : D \rightarrow \mathbb{R}$, synonym auch $\mathcal{C}(D; \mathbb{R})$ |
| $\mathcal{C}^l(D; \mathbb{R})$ | Raum der l -fach differenzierbaren Funktionen $x : D \rightarrow \mathbb{R}$, $l \in \mathbb{N} \cup \{\infty\}$ |
| $\mathcal{AC}(D; \mathbb{R})$ | Raum der absolut stetigen Funktionen $x : D \rightarrow \mathbb{R}$ |
| $\mathcal{AC}_{lok}(D; \mathbb{R})$ | Raum der lokal absolut stetigen Funktionen $x : D \rightarrow \mathbb{R}$ |
| $W^{1,\infty}(D; \mathbb{R})$ | Raum der absolut stetigen und beschränkten Funktionen $x : D \rightarrow \mathbb{R}$ mit essentiell beschränkter Ableitung |
| $B_R^n(x)$ | die offene Umgebung eines Punktes $x \in \mathbb{R}^n$ mit dem Radius $R > 0$, d.h. $B_R^n(x) = \{y \in \mathbb{R}^n \mid \ x - y\ < R\}$, $R > 0$ |
| $\bar{B}_R^n(x)$ | die abgeschlossene Umgebung eines Punktes $x \in \mathbb{R}^n$ mit dem Radius $R > 0$, d.h. $\bar{B}_R^n(x) = \{y \in \mathbb{R}^n \mid \ x - y\ \leq R\}$, $R > 0$ |
| $\sigma(A)$ | Spektrum der Matrix $A \in \mathbb{C}^{n \times n}$, d.h. $\sigma(A) := \{\lambda \in \mathbb{C} \mid \det(\lambda I_n - A) = 0\}$ |
| $\mathbf{GL}_n(\mathbb{C})$ | Menge der invertierbaren $n \times n$ Matrizen über dem Körper \mathbb{C} |
| $\mathbf{GL}_n(\mathbb{R})$ | Menge der invertierbaren $n \times n$ Matrizen über dem Körper \mathbb{R} |
| I_n | $:= \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix} \in \mathbb{R}^{n \times n}$, Einheitsmatrix der Dimension $n \times n$ |
| $O_{n \times m}$ | $:= \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & 0 \end{bmatrix} \in \mathbb{R}^{n \times m}$, Nullmatrix der Dimension $n \times m$ |
| e_i^n | Einheitsvektor im \mathbb{R}^n , d.h. die i -te Spalte der Einheitsmatrix I_n |

A.2 Existenz und Eindeutigkeit von Lösungen

Wenn bei Differentialgleichungen nach der Existenz und Eindeutigkeit von Lösungen gefragt wird, spielt Stetigkeit eine zentrale Rolle. So sichert nach dem Existenzsatz von Peano [65, S. 50] eine stetige rechte Seite in einem Anfangswertproblem der Form

$$\dot{x}(t) = f(t, x(t)), \quad x(0) = x_o \in \mathbb{R}^n \quad (\text{A.1})$$

die Existenz wenigstens einer Lösungskurve. Stetigkeit ist jedoch nicht ausreichend, um Eindeutigkeit zu garantieren. In der Theorie gewöhnlicher Differentialgleichungen erlangt die Lipschitz-Stetigkeit Bedeutung, da diese Bedingung für die rechte Seite nach dem Satz von Picard-Lindelöf [65, S. 53] zusätzlich zur Existenz auch die Eindeutigkeit der Lösung sicherstellt.

Der Begriff der Stetigkeit ist in der hier angegebenen Form auf Bernard Bolzano zurückzuführen und wurde von Karl Weierstraß als Epsilon-Delta-Kriterium formuliert [53, S. 92].

Definition A.1 (stetige Funktion)

Eine auf $D \subseteq \mathbb{R}^n$, $n \in \mathbb{N}$, definierte reellwertige Funktion $f : D \rightarrow \mathbb{R}^m$, $m \in \mathbb{N}$, ist nach dem Epsilon-Delta-Kriterium genau dann stetig im Punkt $x_o \in D$, wenn zu jeder beliebigen Zahl $\varepsilon > 0$ eine Zahl $\delta > 0$ gefunden werden kann, so dass für alle $x \in D$ mit $\|x - x_o\| < \delta$ die Ungleichung

$$\|f(x) - f(x_o)\| < \varepsilon \quad (\text{A.2})$$

gilt. Die Funktion f heißt stetig in D , wenn sie an jeder Stelle ihres Definitionsbereiches stetig ist.

Eine äquivalente Definition ist durch das Folgenkriterium gegeben. Demnach ist eine reellwertige Funktion $f : D \rightarrow \mathbb{R}$ im Punkt $x_o \in D$ genau dann stetig, wenn für den Grenzübergang $\lim_{x \rightarrow x_o} f(x) = f(x_o)$ gilt.

Die Auswirkung der Stetigkeit besteht darin, dass die Änderung im Bildbereich (Funktionswert) durch ε beschränkt ist, wenn die Änderung im Definitionsbereich (Argument) kleiner als der Wert δ bleibt. Folglich dürfen keine Sprünge im Funktionsgraphen enthalten sein.

Stetige Funktionen $f : D \rightarrow \mathbb{R}$ sind auf jedem Kompaktum beschränkt, das zum Definitionsbereich der Funktion f gehört. Diese Aussage schlägt sich im Schrankensatz nieder:

Theorem A.2 (Schrankensatz) [64, S. 108]

Für jede auf einem abgeschlossenen Intervall $a \leq x \leq b \in \mathbb{R}$ stetige Funktion $f : [a, b] \rightarrow \mathbb{R}$ gibt es eine Schranke $K \in \mathbb{R}$ mit $|f(x)| < K$ für alle $x \in [a, b]$.

Die Klasse der stetigen Funktionen $f : D \rightarrow Y$ wird mit $\mathcal{C}(D; Y)$, teils auch mit $\mathcal{C}^0(D; Y)$ gekennzeichnet.

Definition A.3 (gleichmäßig stetige Funktion)

Eine auf $D \subseteq \mathbb{R}^n$ definierte reellwertige Funktion $f : D \rightarrow \mathbb{R}^m$ heißt genau dann gleichmäßig stetig in D , wenn zu jeder beliebigen Zahl $\varepsilon > 0$ eine Zahl $\delta > 0$ gefunden werden kann, so dass für alle $x, x_o \in D$ mit $\|x - x_o\| < \delta$ die Ungleichung

$$\|f(x) - f(x_o)\| < \varepsilon \quad (\text{A.3})$$

gilt. Im Unterschied zur Stetigkeit fordert die gleichmäßige Stetigkeit, dass δ ausschließlich von ε abhängt, nicht aber von der Stelle x_o .

Ein Beispiel für eine zwar stetige, nicht aber gleichmäßig stetige Funktion ist die Normalparabel $f : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto x^2$. An jeder Stelle $x_o \in \mathbb{R}$ kann nach dem Epsilon-Delta-Kriterium ein $\varepsilon > 0$ vorgegeben und ein dazu passendes $\delta > 0$ bestimmt werden. Deshalb ist f stetig. Da für gleichmäßige Stetigkeit die Zahl δ nicht vom Ort x_o abhängen darf, muss das Kriterium für alle $x_o \in \mathbb{R}$ mit dem ermittelten δ erfüllt werden können. Unabhängig davon, wie klein δ gewählt ist, wird Ungleichung (A.3) für hinreichend große x_o stets verletzt. Allerdings ist jede Einschränkung von f auf ein kompaktes Intervall gleichmäßig stetig. Diese Aussage lässt sich auf alle stetigen Funktionen verallgemeinern: nach dem Satz von Heine [53, S. 92] ist auf einem Kompaktum $K \subset \mathbb{R}$ jede stetige Funktion $f : K \rightarrow \mathbb{R}$ gleichmäßig stetig.

Aus gleichmäßig stetig folgt allerdings nicht, dass sich die Funktion nur mit endlicher Steigung verändert. Offensichtlich ist für alle positiven Zahlen $\delta < \varepsilon^2$ das Epsilon-Delta-Kriterium für $f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$, $x \mapsto \sqrt{x}$ unabhängig vom Ort x_o erfüllt. Die Ableitung $f'(x) = (2\sqrt{x})^{-1}$ der gleichmäßig stetigen Funktion f wächst jedoch für $x \rightarrow 0$ unbeschränkt an.

Die Eigenschaft der (gleichmäßigen) Stetigkeit bedingt nicht notwendigerweise Differenzierbarkeit. Die Betragsfunktion $f : x \mapsto |x|$ als Beispiel ist an jedem Punkt $x \in \mathbb{R}$ und insbesondere bei $x = 0$ (gleichmäßig) stetig, dort aber nicht differenzierbar.

Definition A.4 (absolut stetige Funktion)

Eine reellwertige Funktion $f : D \rightarrow \mathbb{R}$, die auf einem Intervall $D \subseteq \mathbb{R}$ erklärt ist, ist genau dann absolut stetig über D , wenn für jedes beliebige $\varepsilon > 0$ eine Konstante $\delta > 0$ gefunden werden kann, so dass für jedes $n \in \mathbb{N}$ jede Folge paarweise disjunkter Teilintervalle

$$[x_1, y_1], [x_2, y_2], \dots, [x_n, y_n] \subseteq D, \quad (\text{A.4})$$

die der Bedingung

$$\sum_{k=1}^n |y_k - x_k| < \delta \quad (\text{A.5})$$

genügt, gleichzeitig auch die Forderung

$$\sum_{k=1}^n |f(y_k) - f(x_k)| < \varepsilon \quad (\text{A.6})$$

erfüllt.

Definition A.5 (lokal absolut stetige Funktion)

Eine reellwertige Funktion $f : D \rightarrow \mathbb{R}$, die auf einem Intervall $D \subseteq \mathbb{R}$ erklärt ist, heißt lokal absolut stetig auf K , wenn sie auf jedem Kompaktum $K \subseteq D$ absolut stetig ist.

Die Menge aller absolut stetigen Funktionen $f : D \rightarrow Y$ wird mit $\mathcal{AC}(D; Y)$ bezeichnet, die der lokal absolut stetigen Funktionen mit $\mathcal{AC}_{lok}(D; Y)$.

Eine Funktion $f : K \rightarrow \mathbb{R}$, die auf einem Kompaktum $K = [a, b]$ definiert ist, ist genau dann absolut stetig, wenn es eine integrierbare Funktion g gibt, die

$$f(t) = f(a) + \int_a^t g(\tau) d\tau \quad (\text{A.7})$$

für alle $t \in K$ erfüllt [95, S. 471]. Absolut stetige Funktionen sind stetig und gleichzeitig fast überall differenzierbar, weil

$$\dot{f}(t) = g(t) \quad (\text{A.8})$$

für fast alle t gilt [95, S. 471].

Definition A.6 (Lipschitz-Bedingung)

Sei $I \subseteq \mathbb{R}$ und $D \subseteq \mathbb{R}^n$, $n \in \mathbb{N}$. Eine auf $I \times D$ definierte Funktion $f : I \times D \rightarrow \mathbb{R}^n$ erfüllt die Lipschitz-Bedingung auf D , wenn eine Zahl $L \geq 0$ gefunden werden kann, so dass für alle $(t, x_1), (t, x_2) \in I \times D$ die Ungleichung

$$\|f(t, x_1) - f(t, x_2)\| \leq L \cdot \|x_1 - x_2\| \quad (\text{A.9})$$

gilt. Dabei wird L als Lipschitz-Konstante bezeichnet.

Definition A.7 (lokal Lipschitz-stetige Funktion)

Sei $I \subseteq \mathbb{R}$ und $D \subseteq \mathbb{R}^n$, $n \in \mathbb{N}$. Eine auf $I \times D$ definierte Funktion $f : I \times D \rightarrow \mathbb{R}^n$ heißt lokal Lipschitz-stetig in x , wenn es für alle $x_o \in D$ eine Zahl $\varepsilon > 0$ und eine lokal integrierbare Funktion $L : I \rightarrow \mathbb{R}_{>0}$ gibt, so dass der Ball $B_\varepsilon^n(x_o) \subset D$ in D enthalten ist und f der Lipschitz-Bedingung

$$\|f(t, x_1) - f(t, x_2)\| \leq L(t) \cdot \|x_1 - x_2\|, \quad \forall x_1, x_2 \in B_\varepsilon^n(x_o), \forall t \in I \quad (\text{A.10})$$

genügt.

Definition A.8 (global Lipschitz-stetige Funktion)

Eine auf $D \subseteq \mathbb{R}^n$, $n \in \mathbb{N}$, definierte Funktion $f : D \rightarrow \mathbb{R}^n$ heißt global Lipschitz-stetig auf D , wenn für alle $x_1, x_2 \in D$ die Lipschitz-Bedingung

$$\|f(x_1) - f(x_2)\| \leq L \cdot \|x_1 - x_2\|, \quad \forall x_1, x_2 \in D \quad (\text{A.11})$$

mit einer einzigen Konstanten $L > 0$ erfüllt werden kann. Die Lipschitz-Konstante L darf hierbei nicht vom Ort abhängen.

Wird in Definition A.8 $n = 1$ gesetzt, lässt sich die Bedeutung der Lipschitz-Stetigkeit prägnant illustrieren. Nachdem Forderung (A.11) für alle $x_1, x_2 \in D$ zutreffen muss, kann das Steigungsdreieck

$$\frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} = \frac{|\Delta y|}{|\Delta x|} \leq L \quad (\text{A.12})$$

auch infinitesimal klein angesetzt und der Grenzübergang $x_2 \rightarrow x_1$ betrachtet werden. Daran zeigt sich, dass an keinem Punkt die Steigung der Funktion f den Wert L überschreiten kann. Die Steigung einer Lipschitz-stetigen Funktion ist deshalb im Intervall $[x_1, x_2] \subset D$ durch die Lipschitz-Konstante $L > 0$ begrenzt. Aus diesem Grund ist eine differenzierbare Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ genau dann Lipschitz-stetig, wenn ihre erste Ableitung beschränkt ist.

Zwischen der Lipschitz-Stetigkeit und der gleichmäßigen Stetigkeit besteht eine Verbindung, weil das Epsilon-Delta-Kriterium in Definition A.3 erfüllt ist, wenn $\delta = \varepsilon/L$ (für $L \neq 0$ bzw. $\delta = 1$ für $L = 0$) gesetzt wird. Daraus wird offenkundig, dass jede Lipschitz-stetige Funktion gleichzeitig auch gleichmäßig stetig ist. Darüberhinaus folgt aus dem Satz von Rademacher [12], [77], dass Lipschitz-stetige Funktionen fast überall differenzierbar sind. Die gleiche Aussage ist auch in [95, S. 471] enthalten, da eine Funktion genau dann Lipschitz-stetig ist, wenn sie absolut stetig ist und eine essentiell beschränkte Ableitung aufweist.

Zusammenhänge zwischen den Typen der Stetigkeit im Fall reeller Funktionen:

- f ist Lipschitz-stetig $\Rightarrow f$ ist lokal Lipschitz-stetig $\Rightarrow f$ ist stetig
- f ist Lipschitz-stetig $\Rightarrow f$ ist absolut stetig $\Rightarrow f$ ist gleichmäßig stetig $\Rightarrow f$ ist stetig

Einfache Gegenbeispiele demonstrieren, dass die Folgerungen nicht generell umkehrbar sind:

- $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto x^2$ ist stetig und lokal Lipschitz-stetig, aber weder Lipschitz-stetig noch gleichmäßig stetig
- $f : [-1, 1] \rightarrow \mathbb{R}, x \mapsto \sqrt[3]{x}$ ist stetig, aber nicht lokal Lipschitz-stetig im Nullpunkt.

Theorem A.9 (Existenz und Eindeutigkeit der Lösung) [95, S. 476]

Durch $I \subseteq \mathbb{R}$ werde ein Zeitintervall und durch $D \subseteq \mathbb{R}^n$ eine offene Teilmenge des \mathbb{R}^n dargestellt. Es sei angenommen, dass die rechte Seite $f : I \times D \rightarrow \mathbb{R}^n$ im Anfangswertproblem (A.1) die folgenden Eigenschaften erfüllt:

$$f(\cdot, x) : I \rightarrow \mathbb{R}^n \text{ ist für jedes feste } x \in D \text{ stückweise stetig} \quad (\text{A.13})$$

$$f(t, \cdot) : D \rightarrow \mathbb{R}^n \text{ ist für jedes feste } t \in I \text{ stetig} \quad (\text{A.14})$$

Sofern f den Voraussetzungen (A.13) und (A.14) genügt und darüberhinaus im Argument x lokal Lipschitz-stetig ist, sowie im Argument t für jedes konstante x lokal integrierbar ist, zeigt Theorem 54 in [95, S. 476], dass für das Anfangswertproblem (A.1) in $(t_o, x_o) \in I \times D$ eine maximale und eindeutige Lösung existiert.

A.3 Instantane Verstärkung

Bevor eine Generalisierung für nichtlineare Systeme erfolgt, sollen zunächst lineare Systeme betrachtet werden.

A.3.1 Instantane Verstärkung linearer Systeme

Den folgenden Ausführungen sei ein Zustandsraummodell

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), & x(0) &\in \mathbb{R}^n, \quad n \in \mathbb{N} \\ y(t) &= cx(t), & A &\in \mathbb{R}^{n \times n}, \quad b, c^T \in \mathbb{R}^n \end{aligned} \quad (\text{A.15})$$

zugrunde gelegt, dessen E/A-Verhalten für den Anfangswert $x(0) = 0$ durch die Übertragungsfunktion

$$F(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_0}, \quad \begin{array}{l} b_0, \dots, b_m \in \mathbb{R}, \quad b_m \neq 0, \\ a_0, \dots, a_n \in \mathbb{R}, \quad a_n \neq 0, \\ m \in \mathbb{N} \end{array} \quad (\text{A.16})$$

beschrieben werden kann.

Definition A.10 (Instantane Verstärkung linearer Systeme) [69, S. 405]

Das LTI-System (A.16) besitzt die instantane Verstärkung V_0 , die durch

$$V_0 = \lim_{s \rightarrow \infty} s^{n-m} \cdot F(s) = \lim_{s \rightarrow \infty} \frac{b_m + b_{m-1} s^{-1} + \dots + b_0 s^{-m}}{a_n + a_{n-1} s^{-1} + \dots + a_0 s^{-n}} = \frac{b_m}{a_n} \quad (\text{A.17})$$

gegeben ist.

Die instantane Verstärkung hängt somit ausschließlich von denjenigen Koeffizienten der Übertragungsfunktion ab, die als Faktoren die jeweils höchste Potenz s^m bzw. s^n gewichten. Falls die Übertragungsfunktion

$$F(s) = g \cdot \frac{Z(s)}{N(s)}, \quad g \neq 0 \quad (\text{A.18})$$

durch monische Polynome $Z, N \in \mathbb{R}[s]$ dargestellt ist, muss zur Erhaltung der Allgemeingültigkeit ein konstanter Vorfaktor $g \in \mathbb{R}$ zugelassen werden. Dieser stimmt im Falle monischer Polynome mit der instantanen Verstärkung V_0 überein [69, S. 183].

Um die physikalische Bedeutung der instantanen Verstärkung zu verdeutlichen, wird bei einem System (A.16) mit Relativgrad $r = n - m$ das Verhalten der r -ten Ableitung des Ausganges betrachtet. Durch Erweiterung von (A.16) mit einer Kette aus r Differentiatoren entsteht ein System mit der Übertragungsfunktion $F_r(s) := s^r F(s)$. Wird für dieses System der Anfangswertsatz der Laplace-Transformation

$$y(t = 0^+) = \lim_{s \rightarrow \infty} s \cdot F(s) \cdot U(s) \quad (\text{A.19})$$

für Sprunganregung ($U(s) = 1/s$) ausgewertet, ergibt sich

$$y_r(t = 0^+) = \lim_{s \rightarrow \infty} s F_r(s) \frac{1}{s} = \lim_{s \rightarrow \infty} s^r F(s) = V_0. \quad (\text{A.20})$$

Wegen $y_r(t) = y^{(r)}(t)$ gilt $V_0 = y^{(r)}(t = 0^+)$, die instantane Verstärkung entspricht daher allgemein dem Wert der r -ten Ableitung des Ausganges unmittelbar nach einer Sprunganregung. Weil sich das Eingangssignal durch $u(t = 0^+) = 1$ auszeichnet, gilt instantan $y^{(r)}(t = 0^+) = V_0 \cdot u(t = 0^+)$, woran die Namensgebung erklärbar wird.

Definition A.10 bezieht die instantane Verstärkung eines linearen Systems auf dessen Übertragungsfunktion. Alternativ zur Berechnung aus der Übertragungsfunktion kann die instantane Verstärkung ebenso aus der Zustandsdarstellung (A.15) direkt berechnet werden.

Theorem A.11

Die instantane Verstärkung des LTI-Systems (A.15) errechnet sich zu

$$V_0 = cA^{r-1}b = \frac{b_m}{a_n}, \quad (\text{A.21})$$

wobei $1 \leq r \leq n$ den Relativgrad beschreibt.

Beweis

Wird die Zustandsbeschreibung (A.15) einer Laplace-Transformation unterzogen, ergibt sich die Übertragungsfunktion

$$F(s) = c(sI_n - A)^{-1}b = \frac{cb}{s} + \frac{cAb}{s^2} + \frac{cA^2b}{s^3} + \dots \quad (\text{A.22})$$

in Reihendarstellung. Weil für Systeme mit Relativgrad $r \in \mathbb{N}$ die Aussage

$$cA^j b = 0 \quad \text{für } j = 0, 1, \dots, r-2 \quad (\text{A.23})$$

zutrifft, gilt

$$V_0 = \lim_{s \rightarrow \infty} s^r \sum_{j=r}^{\infty} \frac{cA^{j-1}b}{s^j} = \lim_{s \rightarrow \infty} \sum_{j=r}^{\infty} \frac{cA^{j-1}b}{s^{j-r}} = cA^{r-1}b. \quad (\text{A.24})$$

Ohne Beschränkung der Allgemeinheit kann eine Realisierung in RNF angenommen werden, so dass die Vektor-Matrix-Multiplikation cA^{r-1} eine Verschiebung der Einträge von c um $r-1$ Elemente nach rechts verursacht. Es gilt daher $cA^{r-1} = (*, \dots, *, b_m)$.

Desweiteren impliziert die RNF einen Einkoppelvektor $b^T = (0, \dots, 0, 1)$ und einen Koeffizienten $a_n = 1$ für das charakteristische Polynom der Systemmatrix. Aus diesem Grund gilt $V_0 = cA^{r-1}b = b_m/a_n$, daher ist Definition A.10 mit Theorem A.11 kompatibel. ■

Anmerkung

Die Realisierung (A.15), die in jeder beliebigen Form vorliegen kann, lässt sich mit einer geeigneten, regulären Matrix $S \in \mathbf{GL}_n(\mathbb{C})$ durch die Ähnlichkeitstransformation

$$\xi(t) = Sx(t) \quad (\text{A.25})$$

in RNF überführen. Dadurch wird das Modell (A.15) durch den Zustandsvektor $\xi(t) \in \mathbb{R}^n$ und die Größen

$$\tilde{A} = SAS^{-1}, \quad \tilde{b} = Sb, \quad \tilde{c} = cS^{-1} \quad (\text{A.26})$$

beschrieben. Weil sich das Produkt $S^{-1}S = I_n$ jeweils heraushebt, gilt

$$\tilde{c}\tilde{A}^{r-1}\tilde{b} = cS^{-1}(SAS^{-1})^{r-1}Sb = cA^{r-1}b = V_0. \quad (\text{A.27})$$

Deshalb ist die Annahme der RNF keine Einschränkung, sondern behält die Allgemeingültigkeit bei. □

A.3.2 Instantane Verstärkung nichtlinearer Systeme

Obwohl die instantane Verstärkung bei nichtlinearen Systemen nicht durch die Bestimmung des Grenzüberganges einer Übertragungsfunktion ermittelt werden kann, lässt sich dennoch eine geeignete Verallgemeinerung dieses Begriffes finden. Aus der Betrachtung des linearen Falles resultiert, dass die instantane Verstärkung als Faktor zwischen Eingang und r -ter Ableitung des Ausganges auftritt, wenn sich das System zum Zeitpunkt $t = 0$ in der Ruhelage $x = 0$ befindet. Anhand einer Darstellung in BINF (siehe Abschnitt A.8) lässt sich die instantane Verstärkung direkt ablesen. Diese Normalform zeichnet sich dadurch aus, dass im Vorwärtszweig eine Integratorkette mit den Zuständen $z_1(t) = y(t)$ bis $z_r(t) = y^{(r-1)}(t)$ enthalten ist, deren Eingang für geeignete Funktionen $\alpha, \beta : \mathbb{R}^n \rightarrow \mathbb{R}$ durch die Gleichung

$$y^{(r)}(t) = \dot{z}_r(t) = \beta(z(t)) + \alpha(z(t)) \cdot u(t), \quad z(t) \in \mathbb{R}^n \quad (\text{A.28})$$

gegeben ist und exakt die r -te Ableitung des Ausganges bildet.

Definition A.12 (Instantane Verstärkung nichtlinearer Systeme)

Die instantane Verstärkung eines Systems, das in BINF vorliegt, wird durch den Faktor $\alpha(\cdot)$ angegeben, der das Eingangssignal in Gleichung (A.28) gewichtet.

Wird ein LTI-System mittels Polynomdivision in die in Abschnitt A.8.2 beschriebene Form gebracht, so kann die Übereinstimmung zwischen $\alpha(z)$ und V_0 leicht verifiziert werden.

A.4 Stationäre/Statische Verstärkung

Im Gegensatz zur instantanen Verstärkung, die als Verstärkungsfaktor zwischen Eingang und r -ter Ableitung vom Ausgang unmittelbar nach einer Sprunganregung beschreibt, handelt es sich bei der stationären Verstärkung um einen Faktor zwischen Eingang und Ausgang, wenn die Sprunganregung bereits unendlich weit zurückliegt, also im eingeschwungenen Zustand. Diese Aussage bezieht sich auf stabile Systeme. Allgemein stellt die stationäre Verstärkung das Verhältnis von Eingangs- und Ausgangsgröße in der (möglicherweise instabilen) Gleichgewichtslage dar.

Mit den Funktionen $f, g \in \mathcal{C}^0(\mathbb{R}^n; \mathbb{R}^n)$ und $h, d : \mathbb{R}^n \rightarrow \mathbb{R}$ sei die Zustandsbeschreibung

$$\dot{x}(t) = f(x(t)) + g(x(t)) \cdot u(t), \quad x(0) \in \mathbb{R}^n \quad (\text{A.29})$$

$$y(t) = h(x(t)) + d(x(t)) \cdot u(t) \quad (\text{A.30})$$

eines dynamischen Systems festgelegt. Die Definition der stationären Verstärkung basiert auf der Existenz einer Ruhelage, weshalb die Annahme

$$\exists (x_\infty, u_\infty) \in \mathbb{R}^n \times \mathbb{R} \quad \text{mit} \quad f(x_\infty) + g(x_\infty) \cdot u_\infty = 0 \quad (\text{A.31})$$

zu treffen ist.

Definition A.13 (Gleichgewichtslage, Ruhelage)

Als Gleichgewichtslage oder Ruhelage von (A.29) wird ein Punkt $x_\infty \in \mathbb{R}^n$ bezeichnet, wenn für $u_\infty \in \mathbb{R}$

$$0 = f(x_\infty) + g(x_\infty) \cdot u_\infty \quad (\text{A.32})$$

gilt.

Wird der Zustand $x_\infty \in \mathbb{R}^n$ in (A.29) eingesetzt, ergibt sich in der Gleichgewichtslage die Ausgangsgröße $y_\infty = h(x_\infty) + d(x_\infty) \cdot u_\infty$.

Für die Bestimmung des Gleichgewichtspunktes x_∞ ist in (A.29) $\dot{x} \equiv 0$ zu setzen und diese Gleichung nach dem Zustandsvektor aufzulösen. Nicht in allen Fällen lässt sich dafür eine explizite Lösung x_∞ finden.

Unter der Voraussetzung $u_\infty \neq 0$ und

$$\begin{aligned} g(x_\infty) = 0 &\Rightarrow d(x_\infty) \neq 0 \\ d(x_\infty) = 0 &\Rightarrow g(x_\infty) \neq 0 \end{aligned} \quad (\text{A.33})$$

lässt sich die stationäre Verstärkung definieren.

Definition A.14 (Stationäre/Statische Verstärkung)

Das dynamische System (A.29) besitzt in der Ruhelage $x_\infty \in \mathbb{R}^n$ die stationäre Verstärkung V_∞ , die durch

$$V_\infty := \frac{y_\infty}{u_\infty} = \frac{h(x_\infty) + d(x_\infty) \cdot u_\infty}{u_\infty} \quad (\text{A.34})$$

gegeben ist.

Gleichung (A.34) offenbart, dass im Allgemeinen die stationäre Verstärkung sowohl vom Zustandsvektor in der Ruhelage als auch vom Eingangssignal abhängt. Der Grund hierfür liegt in der Tatsache, dass im Falle nichtlinearer Systeme unterschiedliche Ruhelagen möglich sind, die jeweils verschiedene stationäre Verstärkungen aufweisen können.

Lineare zeitinvariante Systeme (LTI) zeigen dagegen die Besonderheit, dass die stationäre Verstärkung weder vom Zustandsvektor in der Gleichgewichtslage, noch vom Eingangssignal abhängt, sondern sich ausschließlich durch die Systemparameter ausdrücken lässt:

Definition A.15 (Stationäre Verstärkung (LTI))

Ein LTI-System mit der Übertragungsfunktion

$$F(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_0} + d, \quad \begin{array}{l} b_0, \dots, b_m \in \mathbb{R}, \\ a_0, \dots, a_n \in \mathbb{R}, \\ d \in \mathbb{R}, a_0 \neq 0 \end{array} \quad (\text{A.35})$$

besitzt die stationäre Verstärkung

$$V_\infty = \lim_{s \rightarrow 0} F(s) = \frac{b_0}{a_0} + d. \quad (\text{A.36})$$

Theorem A.16

Falls (A.35) die Übertragungsfunktion der Zustandsdarstellung

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), & x(0) &\in \mathbb{R}^n \\ y(t) &= cx(t) + du(t), & n \in \mathbb{N}, \quad A &\in \mathbf{G}\mathbf{l}_n(\mathbb{R}), \quad b, c^T \in \mathbb{R}^n, \quad d \in \mathbb{R} \end{aligned} \quad (\text{A.37})$$

angibt, ist die stationäre Verstärkung in Gleichung (A.36) mit dem Ausdruck

$$V_\infty = -cA^{-1}b + d \quad (\text{A.38})$$

identisch.

Beweis

Anhand einer Darstellung von (A.37) in RNF lässt sich leicht verifizieren, dass der Ausdruck (A.38) mit Gleichung (A.36) übereinstimmt. ■

Wird das LTI-System (A.37) mit einem konstanten Eingangssignal $u(\cdot) = u_\infty$ beaufschlagt und in der Ruhelage $\dot{x}(t) = 0$ ausgewertet, ergibt sich $-Ax_\infty = bu_\infty$. Auflösen nach $x(t)$ und Einsetzen in (A.37) führt auf $y_\infty = (-cA^{-1}b + d)u_\infty = V_\infty \cdot u_\infty$. Darin spiegelt sich die Bedeutung der stationären Verstärkung wider: in der Gleichgewichtslage stimmt V_∞ mit dem Verstärkungsfaktor zwischen Eingang und Ausgang überein.

Für Systeme mit $a_0 = 0$ bzw. $A \notin \mathbf{G}\mathbf{l}_n(\mathbb{R})$ lässt sich keine stationäre Verstärkung definieren. Die Systemmatrix A eines solchen Systems ist nicht invertierbar und besitzt daher den Eigenwert $\lambda = 0$. Unter der Annahme einer minimalen Realisierung ist das

Spektrum der Systemmatrix identisch zur Menge der Polstellen der Übertragungsfunktion. Folglich ist $s = 0$ eine Polstelle von (A.35), woraus direkt $a_0 = 0$ resultiert. Dies zeigt, dass $a_0 \neq 0$ und $A \in \mathbf{Gl}_n(\mathbb{R})$ gleichbedeutende Forderungen sind.

Ein dynamisches System mit einer Polstelle bei $s = 0$ enthält notwendigerweise im Nennerpolynom den Koeffizienten $a_0 = 0$. Wie die Berechnung der Sprungantwort für ein solches System zeigt, wird bei Erregung mit einem konstanten Eingangssignal keine Ruhelage eingenommen, der Ausgang strebt gegen unendlich. In der Folge lässt sich keine stationäre Verstärkung definieren.

A.5 Lie-Ableitung

Die Lie-Ableitung differenziert ein Skalarfeld entlang einer beliebigen Richtung. Im Gegensatz zur partiellen Ableitung, deren Differentiationsrichtung entlang der jeweiligen Koordinatenachsen des Definitionsbereiches verläuft, wird bei der Lie-Ableitung die Richtung der Differentiation durch ein Vektorfeld beliebig vorgegeben. Daher lässt sich die Lie-Ableitung auch als Richtungsableitung eines Skalarfeldes bezeichnen. Im Zusammenhang mit der Berechnung von Lie-Ableitungen werden sowohl das Skalarfeld als auch das Vektorfeld zweckmäßigerweise als hinreichend oft differenzierbar, vorausgesetzt.

Definition A.17 (Skalarfeld)

Eine Funktion $h : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, $n \in \mathbb{N}$, die ein vektorielles Argument in einen skalaren Funktionswert abbildet, heißt Skalarfeld.

Definition A.18 (Vektorfeld)

Eine Funktion $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$, $n \in \mathbb{N}$, die vektoriellen Argumenten als Funktionswert einen Vektor zuordnet, heißt Vektorfeld.

Analog zur partiellen Differentiation, die sich aus einer Linearkombination des Gradienten¹⁾ und eines Einheitsvektors²⁾ errechnet, ergibt sich die Lie-Ableitung ebenfalls als Linearkombination von Gradient und Differentiationsrichtung.

Definition A.19 (Lie-Ableitung)

Als Lie-Ableitung wird die Richtungsableitung eines Skalarfeldes $h \in \mathcal{C}^l(\mathbb{R}^n; \mathbb{R})$ entlang eines Vektorfeldes $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ bezeichnet und berechnet sich durch

$$L_f h(x) := \frac{\partial}{\partial x} h(x) \cdot f(x) = \nabla h(x) \cdot f(x), \quad \forall x \in D \quad (\text{A.39})$$

aus dem Skalarprodukt von Gradient und Differentiationsrichtung.

¹⁾ Der Gradient ist der Vektor der partiellen Ableitungen.

²⁾ Die Differentiationsrichtung verläuft in diesem Fall entlang einer der Koordinatenachsen, also parallel zum entsprechenden Einheitsvektor. Damit kann die partielle Differentiation als Sonderfall der Lie-Ableitung aufgefasst werden.

Wird durch hinreichend große Zahlen $l_f, l_h \in \mathbb{N}$ und $f \in \mathcal{C}^{l_f}(\mathbb{R}^n; \mathbb{R}^n)$, $h \in \mathcal{C}^{l_h}(\mathbb{R}^n; \mathbb{R})$ Differenzierbarkeit garantiert, so kann die Lie-Ableitung wiederholt ausgeführt werden. In gleicher Weise wie f bezeichne auch $g : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ ein Vektorfeld gemäß Definition A.18. Es ergeben sich folgende Rechenregeln für mehrfache Richtungsableitungen:

$$L_f^0 h(x) = h(x) \quad (\text{A.40})$$

$$L_f^2 h(x) = L_f L_f h(x) = \frac{\partial}{\partial x} L_f h(x) \cdot f(x) \quad (\text{A.41})$$

$$L_f^k h(x) = L_f L_f^{k-1} h(x) = \frac{\partial}{\partial x} L_f^{k-1} h(x) \cdot f(x) \quad (\text{A.42})$$

$$L_g L_f h(x) = \frac{\partial}{\partial x} L_f h(x) \cdot g(x) \quad (\text{A.43})$$

Besonders in der Systemtheorie erhält die Lie-Ableitung Bedeutung, wenn von einem dynamischen System

$$\begin{aligned} \dot{x}(t) &= f(x(t)) + g(x(t)) \cdot u(t), & x(0) &= x_o \in \mathbb{R}^n \\ y(t) &= h(x(t)) \end{aligned} \quad (\text{A.44})$$

mit $f, g \in \mathcal{C}^l(\mathbb{R}^n; \mathbb{R}^n)$, $h \in \mathcal{C}^l(\mathbb{R}^n; \mathbb{R})$ und $l \in \mathbb{N}$ die zeitliche Ableitung des Ausgangssignales $y(\cdot)$ berechnet werden soll. Weil die Funktion h (ein Skalarfeld) als Argument den zeitabhängigen Zustandsvektor $x(t)$ besitzt, lässt sich die zeitliche Ableitung des Ausgangs mit Hilfe der Kettenregel der Differentiation durch

$$\dot{y}(t) = \frac{d}{dt} h(x(t)) = \nabla h(x(t)) \cdot \dot{x}(t) \quad (\text{A.45})$$

ermitteln. Der zeitlichen Differentiation des Ausgangssignales liegt eine Differentiation einer verketteten Funktion nach dem Argument der inneren Funktion zu Grunde. Ein Vergleich mit der Definition der Lie-Ableitung (A.39) offenbart, dass diese Operation formal mit einer Richtungsableitung übereinstimmt. Folglich kann $\dot{y}(t)$ durch Lie-Ableitungen

$$\dot{y}(t) = \nabla h(x(t)) \cdot \dot{x}(t) = L_{\dot{x}} h(x(t)) = L_f h(x(t)) + L_g h(x(t)) \cdot u(t) \quad (\text{A.46})$$

ausgedrückt werden.

Geometrisch betrachtet ist der Differentialquotient $\partial h / \partial x$ der Gradient der Funktion $h(\cdot)$ und zeigt somit die Richtung des steilsten Anstieges an. Jener ist in Abbildung A.1 als grüner Vektorpfeil eingetragen und mit ∇h beschriftet. Der zeitlich veränderliche Zustandsvektor $x(t)$ zeichnet gemäß der vorliegenden Systemdynamik in der Phasebene eine Trajektorie, die in Abbildung A.1 exemplarisch als Spirale in der Ebene $x_3 = 0$ eingezeichnet ist. Die zeitliche Ableitung $\dot{x}(t)$ ergibt den zugehörigen Geschwindigkeitsvektor und liegt stets tangential zur Bahnkurve orientiert, wodurch gleichzeitig auch die momentane Bewegungsrichtung beschrieben wird. Die Länge des Geschwindigkeitsvektors stellt ein Maß für die aktuelle Geschwindigkeit der Bewegung des Zustandsvektors entlang der Bahnkurve dar.

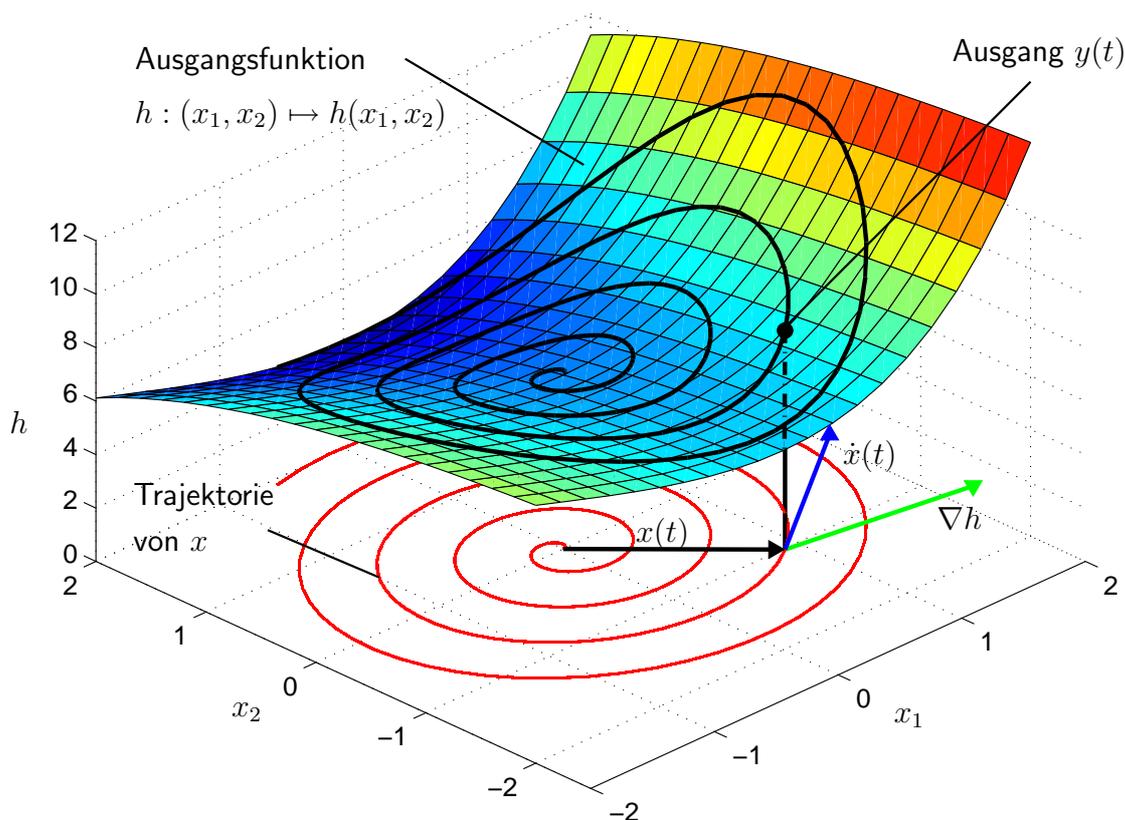


Abbildung A.1: Beispielhafte Trajektorie des Zustandsvektors $x(t)$ in der Phaseebene mit Projektion auf die Ausgangsfunktion h .

Durch das Skalarfeld $h(\cdot)$ wird die Trajektorie des Zustandsvektors $x(t)$ von der Phaseebene in ihren Wertebereich abgebildet, also auf die von $h(\cdot)$ festgelegte Fläche projiziert. Dadurch entsteht der Systemausgang $y(t) = h(x(t))$, der in Abbildung A.1 als schwarze Kurve erkennbar ist. Für die zeitliche Veränderung des Ausgangssignales ist sowohl die Steigung des Skalarfeldes $h(\cdot)$ verantwortlich, als auch die relative Bewegungsrichtung und deren Geschwindigkeit zwischen der Trajektorie und dem Gradienten ∇h . Daher ergibt sich die zeitliche Änderung von $y(t) = h(x(t))$ aus der geometrischen Anschauung heraus als das Skalarprodukt von Gradient und Bahngeschwindigkeit und führt somit zum identischen Ergebnis wie die Berechnung nach der Kettenregel.

A.6 Relativgrad

In der Literatur finden sich unterschiedliche Definitionen des Relativgrades $r \in \mathbb{N}$, der vor allem in deutschsprachigen Veröffentlichungen auch „Differenzgrad“ [93, S. 297] oder „Differenzordnung“ [23, S. 489], [58, S. 102, 299] genannt wird. Eine präzise Definition des Relativgrades wird zunächst zurückgestellt und einleitend der Einfluss des Relativgrades auf das Ein-/Ausgangs-Verhalten dynamischer Systeme diskutiert.

Der Relativgrad $r \in \mathbb{N}$ eines Zustandsraummodells

$$\begin{aligned}\dot{x}(t) &= f(x(t)) + g(x(t)) \cdot u(t), & x(0) &= x_o \in \mathbb{R}^n \\ y(t) &= h(x(t))\end{aligned}\tag{A.47}$$

wobei $f, g \in \mathcal{C}^l(\mathbb{R}^n; \mathbb{R}^n)$, $h \in \mathcal{C}^l(\mathbb{R}^n; \mathbb{R})$, $l \in \mathbb{N}$, gibt, grob formuliert³⁾, die Anzahl von Integratoren an, die sich auf dem kürzest möglichen Pfad zwischen Systemeingang und Systemausgang befinden. Demzufolge müssen auf dem Weg durch das System mindestens r Integratoren überwunden werden. In gewisser Weise beschreibt der Relativgrad die „Verzögerung“, die ein Signal beim Durchlaufen eines dynamischen Systems erfährt und stellt damit ein Maß für dessen Trägheit dar.

Wenn ein System den relativen Grad r besitzt, trennen r Integratoren den Ausgang vom Eingang (siehe Normalform in Abbildung A.4). Jede Differentiation überwindet einen Integrator, folglich muss in der r -ten Ableitung $y^{(r)}$ die Eingangsgröße u explizit enthalten sein, nicht jedoch in niedrigeren Ableitungen. Um den Relativgrad zu berechnen, wird das Ausgangssignal $y(\cdot)$ daher so oft nach der Zeit differenziert, bis darin der Systemeingang u explizit als Faktor in Erscheinung tritt.

Im Folgenden werden für ein nichtlineares System der Form (A.47) die zeitlichen Ableitungen schrittweise bis zur Ordnung $j \in \mathbb{N}$ bestimmt. In jedem Schritt wird der Einfluss der Stellgröße u auf die entsprechende Ableitung überprüft. Falls der zugehörige Einkoppelfaktor von 0 verschieden ist, stimmt der Relativgrad mit der Ordnung der Ableitung überein und die Prozedur wird abgebrochen. Andernfalls wird die nächsthöhere Ableitung berechnet, solange bis die Stellgröße erscheint:

$$\begin{aligned}y &= h(x) = L_f^0 h(x) \quad \Rightarrow \quad r > 0 \\ \dot{y} &= \frac{dh}{dt} = \frac{d}{dt} L_f^0 h(x) = \frac{\partial}{\partial x} L_f^0 h(x) \dot{x} = L_f h(x) + L_g L_f^0 h(x) \cdot u \\ \text{falls } &\begin{cases} L_g L_f^0 h(x) \neq 0 \quad \forall x \in \mathbb{R}^n \quad \Rightarrow \quad r = 1 \\ L_g L_f^0 h(x) = 0 \quad \forall x \in \mathbb{R}^n \quad \Rightarrow \quad r > 1 \end{cases} \\ \ddot{y} &= \frac{d}{dt} L_f h(x) = \frac{\partial}{\partial x} L_f h(x) \dot{x} = \frac{\partial(L_f h(x))}{\partial x} f(x) + \frac{\partial(L_f h(x))}{\partial x} g(x) u = \\ &= L_f^2 h(x) + L_g L_f h(x) \cdot u \\ \text{falls } &\begin{cases} L_g L_f h(x) \neq 0 \quad \forall x \in \mathbb{R}^n \quad \Rightarrow \quad r = 2 \\ L_g L_f h(x) = 0 \quad \forall x \in \mathbb{R}^n \quad \Rightarrow \quad r > 2 \end{cases}\end{aligned}$$

³⁾ Falls im System parallele Zweige enthalten sind, wie beispielsweise bei der Modalform oder Jordanform, ist eine solche Aussage nicht gültig.

$$y^{(3)} = \frac{d}{dt}L_f^2h(x) = L_f^3h(x) + L_gL_f^2h(x) \cdot u$$

$$\text{falls } \begin{cases} L_gL_f^2h(x) \neq 0 \quad \forall x \in \mathbb{R}^n \Rightarrow r = 3 \\ L_gL_f^2h(x) = 0 \quad \forall x \in \mathbb{R}^n \Rightarrow r > 3 \end{cases}$$

$$\vdots$$

$$y^{(j)} = L_f^j h(x) + L_g L_f^{j-1} h(x) \cdot u$$

$$\text{falls } \begin{cases} L_g L_f^{j-1} h(x) \neq 0 \quad \forall x \in \mathbb{R}^n \Rightarrow r = j \\ L_g L_f^{j-1} h(x) = 0 \quad \forall x \in \mathbb{R}^n \Rightarrow r > j \end{cases}$$

Nachdem bei Systemen mit Relativgrad r in der r -ten Ableitung $y^{(r)}$ das Eingangssignal u explizit enthalten ist, nicht jedoch in niedrigeren Ableitungen, lässt sich allgemein

$$y^{(r)}(t) = L_f^r h(x(t)) + L_g L_f^{r-1} h(x(t)) \cdot u(t) \quad (\text{A.48})$$

angeben. Aus dieser Darstellung folgt direkt eine formale Definition des Relativgrades:

Definition A.20 (Relativgrad am Punkt x_o) [26], [47, S. 137], [82, S. 386]
 Das nichtlineare SISO-System (A.47) hat an der Stelle $x_o \in \mathbb{R}^n$ den Relativgrad r , wenn ein $\varepsilon > 0$ existiert, so dass die Bedingungen

$$L_g L_f^j h(x) = 0 \quad \forall j \in \{0, 1, \dots, r-2\}, \quad \forall x \in B_\varepsilon^n(x_o) \quad (\text{A.49})$$

und

$$L_g L_f^{r-1} h(x_o) \neq 0 \quad (\text{A.50})$$

erfüllt sind.

Anmerkung

- (i) Bei linearen Systemen hängen die Faktoren $x \mapsto L_g L_f^j h(x)$ für $0 < j \leq r-1$ nicht vom Zustandsvektor $x \in \mathbb{R}^n$ ab. Auswerten dieser Ausdrücke ergibt

$$L_g L_f^j h(x) = cA^j b \quad \text{für alle } j \leq r-1 \quad (\text{A.51})$$

und zeigt, dass das Argument x keinen Einfluss besitzt und die Lie-Ableitungen damit konstante Werte annehmen. Aus diesem Grund weisen LTI-Systeme stets einen konstanten Relativgrad auf, der im gesamten Zustandsraum \mathbb{R}^n global definiert ist.

- (ii) Für nichtlineare Systeme dagegen kann durchaus eine Abhängigkeit vom Zustandsvektor bestehen, weswegen der Relativgrad prinzipiell variieren kann. Unter der Annahme, dass in der Folge

$$L_g L_f^0 h(x), L_g L_f^1 h(x), L_g L_f^2 h(x), \dots, L_g L_f^j h(x), \dots$$

der Ausdruck $L_g L_f^{r-1} h(x)$ das erste Element ist, das nicht identisch 0 ist, gibt es Bereiche im Zustandsraum, in welchen der Relativgrad $r \in \mathbb{N}$ vorliegt. Falls jedoch an einer Stelle $x_o \in \mathbb{R}^n$ der Term $L_g L_f^{r-1} h(x_o)$ verschwindet, wird dort die r -te Ableitung nicht mehr durch u beeinflusst. Folglich ändert sich der Relativgrad am Punkt x_o . In der Literatur [26] wird hierfür der Ausdruck „schwacher Relativgrad“ gebraucht.

- (iii) Der Relativgrad muss nicht notwendigerweise im gesamten Zustandsraum definiert sein. Falls der Systemausgang in $G \subseteq \mathbb{R}^n$ vom Eingangssignal gänzlich unbeeinflusst bleibt, tritt für $x \in G$ in keiner Ableitung u explizit auf. Weil in einem solchen Fall unendlich viele Ableitungen unabhängig vom Eingang sind, sprechen einige Autoren [70, S. 247] von „unendlichem Relativgrad“ $r = \infty$, andere [47, S. 137], [94, S. 244] alternativ auch von „undefiniertem Relativgrad“.
- (iv) Sofern ein Relativgrad $r \in \mathbb{N}$ vorliegt, gilt stets $r \leq n$ [94, S. 218]. Bei dynamischen Systemen, deren Ausgang in irgend einer Weise vom Eingangssignal abhängt, nimmt die Stellgröße möglicherweise bereits auf niedrigere, spätestens jedoch auf die n -te Ableitung von $y(t)$ Einfluss. Andernfalls, d.h. wenn $u(\cdot)$ auf Ableitungen von höherer Ordnung als n einwirkt, überschreitet die Ordnung des Systems den Wert n , oder der Eingang besitzt keinerlei Auswirkung auf den Ausgang. Abgesehen von diesem Fall, der durch $r = \infty$, bzw. r nicht definiert zu beschreiben ist, wird der Relativgrad durch die Systemordnung n daher nach oben hin beschränkt [82, S. 390].

□

Beispiel (inverses Pendel [95, S. 91f])

Ein aufrecht stehender, masseloser Stab der Länge $l > 0$ soll balanciert werden. Der Stab ist drehbar auf einem Schlitten gelagert und trägt an seinem oberen Ende die Punktmasse $m > 0$, die zusammen mit der Erdbeschleunigung $g \approx 9.81 \text{ m/s}^2$ eine zur Fahrbahnebene senkrechte Gewichtskraft $mg > 0$ ergibt. Als Steuergröße wird die Kraft $F \in \mathcal{L}_{lok}^1(\mathbb{R}_{\geq 0}; \mathbb{R})$ auf den Schlitten eingepreßt (siehe Abbildung A.2). Mit Hilfe

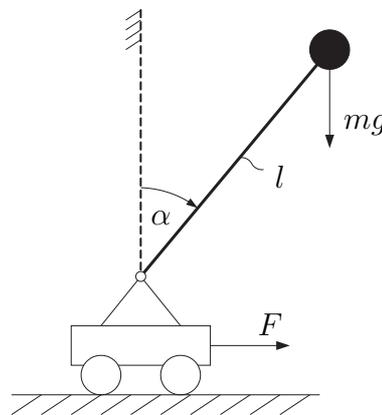


Abbildung A.2: Invertiertes Pendel.

des Drallsatzes kann die Momentenbilanz aus Drehbeschleunigungsmoment, Gravitationsmoment und Steuermoment aufgestellt werden, woraus sich unter Vernachlässigung

der Kreiselmomente die Differentialgleichung

$$ml^2\ddot{\alpha}(t) = mgl \sin(\alpha(t)) - l \cos(\alpha(t))F(t) \quad (\text{A.52})$$

ergibt. Wird die Steuergröße $F(\cdot)$ als Eingang festgelegt und der Winkel $\alpha(\cdot)$ als Systemausgang betrachtet, so folgt nach Einführen der Zustandsgrößen $x_1 := \alpha$ und $x_2 := \dot{\alpha}$ die Beschreibung:

$$\begin{aligned} \dot{x}_1 &= x_2, & x_1(0) &\in \mathbb{R} \\ \dot{x}_2 &= \frac{g}{l} \sin(x_1) - \frac{1}{ml} \cos(x_1) \cdot u, & x_2(0) &\in \mathbb{R} \\ y &= x_1 \end{aligned} \quad (\text{A.53})$$

Mit $x = (x_1, x_2)^T \in \mathbb{R}^2$ und durch Festlegung der Funktionen

$$f : x \mapsto \begin{pmatrix} x_2 \\ \frac{g}{l} \sin(x_1) \end{pmatrix}, \quad g : x \mapsto \begin{pmatrix} 0 \\ -\frac{1}{ml} \cos(x_1) \end{pmatrix}, \quad h : x \mapsto x_1 \quad (\text{A.54})$$

wird eine Systembeschreibung gemäß Gleichung (A.47) erreicht. Um den Relativgrad zu ermitteln, sind die Lie-Ableitungen

$$L_g L_f^0 h(x) = L_g h(x) = \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ -\frac{1}{ml} \cos(x_1) \end{pmatrix}^T = 0 \quad (\text{A.55})$$

$$L_f h(x) = x_2 \quad (\text{A.56})$$

$$L_g L_f^1 h(x) = \begin{pmatrix} 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ -\frac{1}{ml} \cos(x_1) \end{pmatrix}^T = -\frac{1}{ml} \cos(x_1) \quad (\text{A.57})$$

zu untersuchen. Die Kosinusfunktion besitzt Nullstellen bei $x_1 = (2j-1)\pi/2$ mit $j \in \mathbb{N}$. Bedingt durch die 2π -Periodizität kann das Argument auf den Bereich $\tilde{x}_1 \in [0, 2\pi[$ reduziert werden, wenn durch die Modulo-Division $\tilde{x}_1 := (x_1 \bmod 2\pi)$ gesetzt wird. Dadurch ergibt sich

$$L_g L_f^0 h(x) = 0 \quad \forall x \in \mathbb{R}^2 \quad \text{und} \quad L_g L_f^1 h(x) \neq 0 \quad \forall \tilde{x}_1 \notin \left\{ \frac{\pi}{2}, \frac{3\pi}{2} \right\} \quad (\text{A.58})$$

woraus sich der Relativgrad fast überall zu $r = 2$ errechnet. Nimmt der Zustand x_1 (bzw. der Winkel α) ein ungeradzahliges Vielfaches von $\pi/2$ an, erreicht der Kosinus in Gleichung (A.57) einen Nulldurchgang. Weil der Systemausgang von der gewählten Steuergröße aus in dieser Konfiguration nicht beeinflusst werden kann, ist dort der Relativgrad nicht definiert. Es handelt sich um diejenigen Winkel, für die sich das Pendel in horizontaler Lage befindet. Die Kraftwirkung der Steuergröße $F(\cdot)$ verläuft entlang der Verbindungslinie zwischen Lagerpunkt und Masse, so dass kein wirksamer Hebelarm auftritt. Folglich kann in diesem Zustand durch die Stellgröße $F(\cdot)$ kein Drehmoment aufgebracht werden, so dass keine Einflussmöglichkeit auf das Pendel besteht und somit die Steuerbarkeit verloren geht. Dies geht einher mit einer Erhöhung des Relativgrades auf $r = \infty$ (siehe Abbildung A.3) und resultiert aus der Tatsache, dass in der Differentialgleichung (A.52) die Stellgröße mit einer vom Winkel α abhängigen Kosinusfunktion gewichtet wird. \square

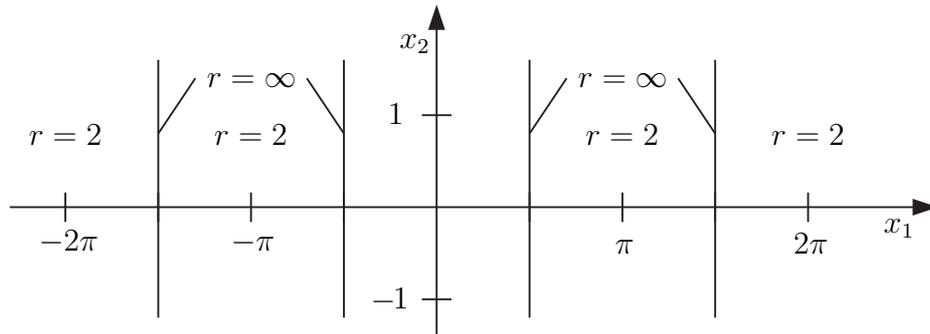


Abbildung A.3: Gebiete mit unterschiedlichem Relativgrad in der Phasenebene für ein invertiertes Pendel.

Am Beispiel des inversen Pendels zeigt sich, dass nichtlineare Systeme einen variablen Relativgrad aufweisen können. Diese Tatsache motiviert präzisere Definitionen des Relativgrades, deren Namensgebung in der Literatur allerdings nicht einheitlich ausfällt. Gleichzeitig verstehen verschiedene Autoren unter der selben Bezeichnung unterschiedliche Definitionen. Daher soll an dieser Stelle ein zusammenfassender Überblick gegeben werden.

Definition A.21 (gleichförmiger Relativgrad im Gebiet G) [15]

Das nichtlineare SISO-System (A.47) hat einen gleichförmigen Relativgrad r im Gebiet $G \subseteq \mathbb{R}^n$, wenn die Bedingungen

$$L_g L_f^j h(x) = 0 \quad \forall j \in \{0, 1, \dots, r-2\} \quad \text{und} \quad L_g L_f^{r-1} h(x) \neq 0 \quad (\text{A.59})$$

für alle Zustände $x \in G$ erfüllt sind.

Wird durch die Festlegung $G = \mathbb{R}^n$ das Gebiet G auf den gesamten \mathbb{R}^n erweitert, resultiert die in [26] verwendete Definition:

Definition A.22 (starker Relativgrad) [26]

Das nichtlineare SISO-System (A.47) hat einen starken Relativgrad r , wenn die Bedingungen

$$L_g L_f^j h(x) = 0 \quad \forall j \in \{0, 1, \dots, r-2\} \quad \text{und} \quad L_g L_f^{r-1} h(x) \neq 0 \quad (\text{A.60})$$

für alle Zustände $x \in \mathbb{R}^n$ erfüllt sind.

An dieser Stelle besteht in der Literatur eine mehrdeutige Namensgebung. Die Eigenschaft, die Saberi und Lin in [81] als starken Relativgrad bezeichnen, sei hier im Sinne einer eindeutigen Benennung als strenger Relativgrad wiedergegeben.

Definition A.23 (strenger Relativgrad) [81]

Das nichtlineare SISO-System (A.47) hat einen strengen Relativgrad r , wenn für alle $x \in \mathbb{R}^n$ die Bedingungen

$$L_g L_f^j h(x) = 0 \quad \forall j \in \{0, 1, \dots, r-2\} \quad (\text{A.61})$$

erfüllt sind und zusätzlich eine Konstante $m > 0$ existiert, die der Ungleichung

$$|L_g L_f^{r-1} h(x)| \geq m \quad \forall x \in \mathbb{R}^n \quad (\text{A.62})$$

genügt.

Nachdem die instantane Verstärkung $L_g L_f^{r-1} h(x)$ in der Gleichung für $y^{(r)}$ einen Verstärkungsfaktor für die Stellgröße u darstellt, ist durch die Schranke $m > 0$ eine Mindesteingriffsmöglichkeit auf das System durch die Stellgröße sichergestellt. Wohingegen Definition A.22 im gesamten Zustandsraum lediglich $|L_g L_f^{r-1} h(x)| > 0$ verlangt, wird in [81] die instantane Verstärkung durch die Bedingung $|L_g L_f^{r-1} h(x)| \geq m$ mit $m > 0$ von 0 wegbeschränkt. Aus diesem Grund ist der strenge Relativgrad eine restriktivere Eigenschaft als der starke Relativgrad.

Definition A.24 (schwacher Relativgrad) [26]

Sei mit $B_\varepsilon^n(x_o)$, $\varepsilon > 0$, eine Umgebung um einen Punkt $x_o \in \mathbb{R}^n$ im Zustandsraum gegeben. Das nichtlineare SISO-System (A.47) hat an der Stelle x_o den schwachen Relativgrad r , wenn für alle Zustände $x \in B_\varepsilon^n(x_o)$ die Bedingungen

$$\begin{aligned} L_g L_f^j h(x) &= 0 \quad \forall j \in \{0, 1, \dots, r-2\}, \\ L_g L_f^{r-1} h(x_o) &= 0 \quad \text{und} \quad L_g L_f^{r-1} h(x) \neq 0 \quad \forall x \in B_\varepsilon^n(x_o) \setminus \{x_o\} \end{aligned} \quad (\text{A.63})$$

erfüllt sind.

Ein Systemverhalten, das nach [26] mit dem Begriff „schwachen“ Relativgrad beschrieben wird, bezeichnen die Autoren in [47, S. 137] und [94, S. 244] als „undefinierten“ Relativgrad.

Bezugnehmend auf das Beispiel bleibt festzuhalten, dass ein inverses Pendel keinen starken oder strengen Relativgrad nach [26] bzw. [81] besitzt, da in der horizontalen Lage die entsprechende Lie-Ableitung verschwindet. Ein gleichförmiger Relativgrad existiert in allen Gebieten G_o ,

$$G_o \subseteq G := \mathbb{R}^2 \setminus \{(x_1, x_2)^T \in \mathbb{R}^2 \mid x_1 = (2j-1)\pi/2, j \in \mathbb{N}, x_2 \in \mathbb{R}\}, \quad (\text{A.64})$$

welche die horizontale Position des Pendels ausschließen. In dieser Lage, in der durch die Stellgröße kein Drehmoment aufgebracht werden kann, liegt laut [26] ein schwacher Relativgrad von $r = 2$ vor.

Anmerkung (Spezialfall LTI)

Der Relativgrad linearer zeitinvarianter Systeme (LTI) ist in Lehrbüchern [23, S. 489], [50, S. 537], [58, S. 102f], [58, S. 299] gewöhnlich als Differenz zwischen dem Grad des Nennerpolynomes und dem Grad des Zählerpolynomes der Übertragungsfunktion definiert, woraus sich die Namensgebung⁴⁾ ursprünglich ableitet. Um die Kompatibilität

⁴⁾ Teils sind in der Literatur auch die Bezeichnungen Differenzegrad oder Differenzordnung gebraucht.

der verallgemeinerten Definitionen des Relativgrades mit der Definition als Differenzgrad der Polynome in der Übertragungsfunktion zu zeigen, wird ein lineares System

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), & x(0) &\in \mathbb{R}^n \\ y(t) &= cx(t), & n &\in \mathbb{N} \end{aligned} \quad (\text{A.65})$$

mit $A \in \mathbb{R}^{n \times n}$, $b, c^T \in \mathbb{R}^n$ und Relativgrad $r \in \mathbb{N}$ zugrunde gelegt. Für eine formale Übereinstimmung mit (A.47) wird $f : x \mapsto Ax$, $g : x \mapsto b$ und $h : x \mapsto cx$ gesetzt, so dass die oben beschriebene Vorgehensweise zur Bestimmung des Relativgrades nichtlinearer Systeme ohne Abwandlung verwendbar ist. Diese ergibt für ein lineares System mit Relativgrad r die Aussagen

$$L_g L_f^j h(x) = cA^j b = 0 \quad \text{für alle } j < r - 1 \quad \text{und} \quad L_g L_f^{r-1} h(x) = cA^{r-1} b \neq 0, \quad (\text{A.66})$$

die zu

$$y^{(r)}(t) = cA^r x(t) + cA^{r-1} b \cdot u(t) \quad (\text{A.67})$$

führen. Mittels Laplace-Transformation der Zustandsbeschreibung errechnet sich die Übertragungsfunktion zu

$$F(s) = c(sI_n - A)^{-1} b = \frac{cb}{s} + \frac{cAb}{s^2} + \frac{cA^2 b}{s^3} + \dots = \sum_{j=1}^{\infty} \frac{cA^{j-1} b}{s^j}, \quad (\text{A.68})$$

die wegen (A.66) durch

$$F(s) = \sum_{j=r}^{\infty} \frac{cA^{j-1} b}{s^j} \quad (\text{A.69})$$

dargestellt werden kann. Daran zeigt sich, dass r exakt der Differenz der Grade von Nenner- und Zählerpolynom der Übertragungsfunktion entspricht, die Definition des Relativgrades für LTI-Systeme ist daher mit Definition A.20 konsistent. \square

A.7 Basiswechsel mit diffeomorphen Abbildungen

Der Begriff „Diffeomorphismus“ bezeichnet eine bijektive und stetig differenzierbare Abbildung mit stetig differenzierbarer Umkehrabbildung. Diese Eigenschaften sind für eine Koordinatentransformation bei dynamischen Systemen, d.h. für einen Basiswechsel, notwendig. Linearität wird dabei nicht gefordert. Insofern ermöglichen nichtlineare, diffeomorphe Abbildungen eine Verallgemeinerung von Ähnlichkeitstransformationen, die im Falle linearer Systeme oft Anwendung finden.

Definition A.25 (Diffeomorphismus) [3, S. 224], [94, S. 229]

Die in $D \subseteq \mathbb{R}^n$ erklärte Funktion $\Phi : D \rightarrow Y$, $Y \subseteq \mathbb{R}^n$, heißt Diffeomorphismus, wenn

(i) ihre Inverse $\Phi^{-1} : Y \rightarrow D$ existiert und

(ii) wenn $\Phi \in C^\infty(D; Y)$ und $\Phi^{-1} \in C^\infty(Y; D)$, so dass stetige partielle Ableitungen vorliegen.

Wenn sowohl Φ , als auch Φ^{-1} im gesamten Raum \mathbb{R}^n definiert sind, spricht man von globalem Diffeomorphismus, andernfalls von lokalem Diffeomorphismus.

Liegt – wie z.B. in der Zustandsbeschreibung (A.70) – die Differentialgleichung in x -Koordinaten vor, kann ein Basiswechsel im Grundsatz mit jedem beliebigen Diffeomorphismus $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ durch das Transformationsgesetz $z = \Phi(x)$ erfolgen. Die Verwendung einer zufällig gewählten Abbildungsvorschrift Φ bringt im Allgemeinen jedoch keinen Vorteil, da vorrangiges Ziel der Transformation sein wird, das System in vorgegebener Struktur darzustellen. Um diese zu erreichen, ist eine auf das System angepasste Transformation notwendig, die teils schwer zu finden ist. Vor allem mit steigender Komplexität des zu transformierenden Systems erschwert sich die Konstruktion eines geeigneten globalen Diffeomorphismus. Um den Aufwand zu begrenzen, beschränkt sich die Praxis in solchen Fällen gezwungenermaßen auf lokale Diffeomorphismen und untersucht das nichtlineare System in einer begrenzten, aber hinreichend großen Umgebung um einen gegebenen Arbeitspunkt $x_o \in \mathbb{R}^n$. Darüberhinaus kann die Verwendung eines lokalen Diffeomorphismus im System selbst begründet sein: falls kein global definierter Relativgrad vorliegt, existiert zum Beispiel keine global gültige Abbildung, die das System in Byrnes-Isidori-Normalform transformiert. Ob eine vorliegende Funktion Φ einen lokalen Diffeomorphismus an einer relevanten Stelle $x_o \in \mathbb{R}^n$ darstellt und damit für eine lokal gültige Koordinatentransformation verwendet werden kann, lässt sich durch folgenden Satz überprüfen:

Theorem A.26 [94, S. 229]

Die in $D \subseteq \mathbb{R}^n$ erklärte Funktion $\Phi : D \rightarrow \mathbb{R}^n$, $\Phi \in C^\infty(D; \mathbb{R}^n)$, stellt einen lokalen Diffeomorphismus dar, wenn die Jacobi-Matrix von $\Phi(x)$ an der Stelle $x = x_o \in D$ regulär ist, d.h. wenn $\det \frac{\partial \Phi}{\partial x} \Big|_{x_o} \neq 0$ gilt.

Beweis

Theorem A.26 ist eine direkte Folgerung [94, S. 229] aus dem Theorem über implizite Funktionen [50, S. 62], [6, S. 146ff]. ■

Um die allgemeine Vorgehensweise der Transformation zu erläutern, wird beispielhaft das in x -Koordinaten gegebene System

$$\begin{aligned} \dot{x}(t) &= f(x(t)) + g(x(t)) \cdot u(t), & x(0) &= x_o \in \mathbb{R}^n \\ y(t) &= h(x(t)) \end{aligned} \tag{A.70}$$

durch die invertierbare Transformationsvorschrift

$$\Phi : x \in \mathbb{R}^n \mapsto z \in \mathbb{R}^n \tag{A.71}$$

in z -Koordinaten überführt. Aus der Transformation (A.71) ergibt sich durch Differenzieren und Einsetzen der Differentialgleichung (A.70):

$$\dot{z} = \frac{\partial \Phi(x)}{\partial x} \cdot \dot{x} = \frac{\partial \Phi(x)}{\partial x} \cdot (f(x) + g(x) \cdot u) \tag{A.72}$$

Wird darin im Argument $x = \Phi^{-1}(z)$ substituiert, folgt:

$$\begin{aligned}\dot{z} &= \frac{\partial\Phi(\Phi^{-1}(z))}{\partial x} \cdot (f(\Phi^{-1}(z)) + g(\Phi^{-1}(z)) \cdot u) \\ y &= h(\Phi^{-1}(z))\end{aligned}\tag{A.73}$$

Zusammen mit den Funktionen

$$\tilde{f} : z \mapsto \frac{\partial\Phi(\Phi^{-1}(z))}{\partial x} f(\Phi^{-1}(z)), \quad \tilde{g} : z \mapsto \frac{\partial\Phi(\Phi^{-1}(z))}{\partial x} g(\Phi^{-1}(z)), \quad \tilde{h} : z \mapsto h(\Phi^{-1}(z))\tag{A.74}$$

resultiert daraus die Zustandsbeschreibung in z -Koordinaten:

$$\begin{aligned}\dot{z}(t) &= \tilde{f}(z(t)) + \tilde{g}(z(t)) \cdot u(t), & z(0) &= \Phi(x_o) \\ y(t) &= \tilde{h}(z(t))\end{aligned}\tag{A.75}$$

Analog zur bekannten Ähnlichkeitstransformation im Falle linearer Systeme kann durch diffeomorphe Funktionen ein nichtlineares dynamisches System in strukturell verschiedene Zustandsbeschreibungen gewandelt werden, wobei das charakteristische E/A-Verhalten des Systems unverändert bleibt.

A.8 Byrnes-Isidori-Normalform

Der folgende Abschnitt beschreibt eine Normalform, die in der Literatur unter der Bezeichnung Byrnes-Isidori-Normalform (BINF) geläufig ist und ein dynamisches System in zwei verkoppelte Subsysteme aufteilt. Vorteilhaft ist diese Normalform vor allem dann, wenn Eigenschaften wie der Relativgrad eines Systems oder das Verhalten der Nulldynamik untersucht werden müssen, was im Zusammenhang mit hochverstärkungsbasierten Reglern der Fall ist. Liegt ein dynamisches System in dieser Zustandsbeschreibung vor, so kann der Relativgrad anhand der Ordnung der Integratorkette im Vorwärtszweig direkt abgelesen werden. Die Nulldynamik⁵⁾ ist dagegen durch das rückkoppelnde Teilsystem bestimmt.

Bevor die Normalform selbst beschrieben wird, soll zuerst die Konstruktion einer Transformationsvorschrift betrachtet werden, die ein gegebenes dynamisches System durch einen Basiswechsel in BINF bringt. Um die folgenden Ausführungen allgemein zu halten, wird an keiner Stelle die Existenz eines global gültigen Abbildungsgesetzes gefordert, so dass auch Systeme mit zustandsabhängigem Relativgrad eingeschlossen sind. Kann lediglich ein lokaler Diffeomorphismus angegeben werden, wird demnach die Koordinatentransformation in der Nähe eines gegebenen Arbeitspunktes durchgeführt, wodurch die erhaltene Normalform nur dort Gültigkeit besitzt. Wenn dagegen – wie in der vorliegenden Arbeit – die Regelstrecke einen strikten Relativgrad aufweist und gleichzeitig ein globaler Diffeomorphismus gefunden werden kann, lässt sich der Gültigkeitsbereich der ermittelten Normalform auf den gesamten Zustandsraum ausdehnen.

⁵⁾ Entgegen der in der Literatur gängigen Bezeichnung ist der Begriff „Nulldynamik“ in der vorliegenden Arbeit in erweiterter Form verwendet, siehe hierzu die Anmerkung auf Seite 248.

A.8.1 Transformation nichtlinearer Systeme

Eine ausführliche Abhandlung über die Transformation nichtlinearer Systeme in BINF findet sich in [47, Kap. 4.1]. Daran angelehnt wird hier für die Strecke (A.70) mit der Ordnung $n \in \mathbb{N}$ exemplarisch eine Koordinatentransformation durchgeführt. Unter der Bedingung eines kausalen, steuer- und beobachtbaren Systems kann (zumindest) in einem ε -Ball $B_\varepsilon^n(x_o)$, $\varepsilon > 0$, um einen geeigneten Arbeitspunkt $x_o \in \mathbb{R}^n$ im Zustandsraum der Relativgrad $r \leq n$ angenommen werden. Folglich darf gemäß Abschnitt A.6 die Gültigkeit der Bedingung

$$L_g L_f^j h(x) = 0 \quad \forall j < r - 1 \quad \text{und} \quad L_g L_f^{r-1} h(x) \neq 0 \quad \forall x \in B_\varepsilon^n(x_o) \quad (\text{A.76})$$

vorausgesetzt werden. Die BINF wird mit der bijektiven Abbildungsvorschrift

$$\Phi : B_\varepsilon^n(x_o) \rightarrow Y := \Phi(B_\varepsilon^n(x_o)) \quad (\text{A.77})$$

durch die Transformation

$$z = \Phi(x) = \begin{pmatrix} \phi_1(x) \\ \phi_2(x) \\ \vdots \\ \phi_n(x) \end{pmatrix}, \quad \phi_i : B_\varepsilon^n(x_o) \rightarrow Y \subseteq \mathbb{R}, \quad i = 1, \dots, n \quad (\text{A.78})$$

erreicht, wobei die ersten r Elemente von $\Phi(x)$ durch die Lie-Ableitungen der Ordnung 0 bis $r - 1$

$$\begin{aligned} z_1 &= \phi_1(x) := L_f^0 h(x) = h(x) \\ z_2 &= \phi_2(x) := L_f h(x) \\ &\vdots \\ z_r &= \phi_r(x) := L_f^{r-1} h(x) \end{aligned} \quad \text{für } x \in B_\varepsilon^n(x_o) \quad (\text{A.79})$$

festgelegt werden. Es lässt sich nachweisen, dass bei einer solchen Wahl die r Gradienten der Funktionen $\phi_1(\cdot)$ bis $\phi_r(\cdot)$ linear unabhängig sind [47, S. 140f], [82, S. 389f], was bezüglich der geforderten Regularität⁶⁾ der Jacobi-Matrix von Φ bedeutsam ist.

Falls das zu transformierende System eine Nulldynamik enthält, gilt die strikte Ungleichung $r < n$. In diesem Fall werden die Funktionen $\phi_{r+1}(\cdot), \dots, \phi_n(\cdot)$ nicht durch Lie-Ableitungen vorgegeben, sondern können frei gewählt werden. Lediglich der Regularität der Jacobi-Matrix von Φ ist die Gestaltung der verbleibenden $n - r$ Elemente $\phi_{r+1}(\cdot), \dots, \phi_n(\cdot)$ unterworfen, so dass große Freiheiten erhalten bleiben. Isidori zeigt [47, S. 142], dass es insbesondere sogar möglich ist, diese Funktionen derart zu wählen, dass die Forderung

$$L_g \phi_j(x) = 0 \quad \forall r + 1 \leq j \leq n, \quad \forall x \in B_\varepsilon^n(x_o) \quad (\text{A.80})$$

⁶⁾ Nachdem es sich bei der Abbildung $z = \Phi(x)$ um einen Diffeomorphismus handeln muss, damit eine Koordinatentransformation in einer Nachbarschaft um den Punkt $x_o \in \mathbb{R}^n$ möglich ist, muss die Jacobi-Matrix von $\Phi(x)$ an der Stelle x_o vollen Rang besitzen.

erfüllt ist. Daraus resultiert eine besonders vorteilhafte Struktur der Normalform, der Systemeingang wirkt nicht auf die Nulldynamik im Rückwärtszweig ein. Falls die Nebenbedingung (A.80) nicht eingehalten wird, ist eine Transformation zwar dennoch möglich. Wie sich später zeigt, wird dann jedoch die Nulldynamik unerwünschterweise vom Systemeingang u beeinflusst.

Wird die Abbildung $\Phi : x \mapsto z$ in der oben beschriebenen Weise gewählt, besitzt diese eine reguläre Jacobi-Matrix und beschreibt daher eine ein-eindeutige Abbildung, was eine notwendige Voraussetzung für die Eignung als Koordinatentransformation darstellt. Dabei können die Funktionswerte $\phi_{r+1}(x_o), \dots, \phi_n(x_o)$ an der Stelle $x_o \in \mathbb{R}^n$ beliebig festgelegt werden.

Die Systembeschreibung in z -Koordinaten errechnet sich analog zur Vorgehensweise in Abschnitt A.7 aus der zeitlichen Ableitung der Abbildung $\Phi : x \mapsto z$. Für die ersten r Funktionen $\phi_j(x)$ mit $1 \leq j \leq r$ ergibt das komponentenweise Differenzieren nach der Zeit:

$$\begin{aligned} \dot{z}_1 &= \frac{\partial \phi_1(x)}{\partial x} \cdot \frac{dx}{dt} = \frac{\partial h(x)}{\partial x} \cdot \frac{dx}{dt} = L_f h(x) + \underbrace{L_g h(x)}_{=0} \cdot u = \phi_2(x) = z_2 \\ &\vdots \\ \dot{z}_{r-1} &= \frac{\partial \phi_{r-1}(x)}{\partial x} \cdot \frac{dx}{dt} = \frac{\partial (L_f^{r-2} h(x))}{\partial x} \cdot \frac{dx}{dt} = L_f^{r-1} h(x) = \phi_r(x) = z_r \\ \dot{z}_r &= L_f^r h(x) + L_g L_f^{r-1} h(x) \cdot u \end{aligned}$$

In der letzten Gleichung muss der ursprüngliche Zustandsvektor x durch den neuen Zustandsvektor z ausgedrückt werden, d.h. x ist durch $\Phi^{-1}(z)$ zu ersetzen. Zusammen mit den Abkürzungen

$$\begin{aligned} \alpha(z) &:= L_g L_f^{r-1} h(\Phi^{-1}(z)) & \forall z \in Y \\ \beta(z) &:= L_f^r h(\Phi^{-1}(z)) & \forall z \in Y \end{aligned} \quad (\text{A.81})$$

folgt schließlich für die Ableitung des r -ten Zustandes:

$$\dot{z}_r = \beta(z) + \alpha(z) \cdot u \quad (\text{A.82})$$

Prinzipiell erhält der Eingang $u(t)$ Einfluss auf die zeitliche Veränderung des Zustandes $z_r(t)$, solange die instantane Verstärkung $\alpha(z(t)) \neq 0$ erfüllt. Wird in Gleichung (A.81) die Rücktransformation $x_o = \Phi^{-1}(z_o)$ als Argument von h eingesetzt und mit Gleichung (A.76) verglichen, so folgt direkt

$$\alpha(z) \neq 0 \quad \forall z \in Y. \quad (\text{A.83})$$

Wegen der Festlegung (A.77) gibt Y den Gültigkeitsbereich der Transformation an, in dem sichergestellt ist, dass der Eingang u stets auf \dot{z}_r einwirkt. Für die Differentiation der restlichen Zustände z_j mit $r+1 \leq j \leq n$ gilt generell:

$$\dot{z}_j = \frac{\partial \phi_j(x)}{\partial x} (f(x) + g(x) \cdot u) = L_f \phi_j(x) + L_g \phi_j(x) \cdot u \quad (\text{A.84})$$

Daran ist zu erkennen, dass diese Zustände im Allgemeinen vom Eingangssignal $u(\cdot)$ direkt abhängen können. Sofern die Abbildung Φ derart gewählt ist, dass Bedingung (A.80) erfüllt wird, entfällt die Abhängigkeit vom Systemeingang und es verbleibt

$$\dot{z}_j = L_f \phi_j(x) \quad \text{für } r+1 \leq j \leq n. \quad (\text{A.85})$$

Der Eingang wirkt unter der Voraussetzung (A.80) nicht auf dasjenige Teilsystem ein, das durch die Zustandsgrößen z_{r+1} bis z_n charakterisiert wird. Auch in der Beschreibung dieses Systemteils wird der Zustandsvektor x durch z substituiert und die Abkürzungen

$$q_j(z) = L_f \phi_j(\Phi^{-1}(z)) \quad \text{für } r+1 \leq j \leq n \quad (\text{A.86})$$

eingeführt. Damit lautet schließlich die Zustandsbeschreibung des dynamischen Systems in lokal gültiger Normalform:

$$\begin{aligned} \dot{z}_1(t) &= z_2(t) \\ \dot{z}_2(t) &= z_3(t) \\ &\vdots \\ \dot{z}_{r-1}(t) &= z_r(t) \\ \dot{z}_r(t) &= \beta(z(t)) + \alpha(z(t)) \cdot u(t) \\ \dot{z}_{r+1}(t) &= q_{r+1}(z(t)) \\ &\vdots \\ \dot{z}_n(t) &= q_n(z(t)) \end{aligned} \quad (\text{A.87})$$

Der Systemausgang berechnet sich zu:

$$y(t) = h(x(t)) = \phi_1(x(t)) = \phi_1(\Phi^{-1}(z(t))) = z_1(t) \quad (\text{A.88})$$

In dieser Normalform setzt sich ein dynamisches System aus zwei verkoppelten Subsystemen zusammen, wie Abbildung A.4 zeigt. Eines der beiden Teilsysteme bildet eine Integratorkette mit den Zuständen z_1 bis z_r und liegt im Vorwärtszweig. Die Anzahl der Integratoren stimmt daher stets mit dem Relativgrad r des Gesamtsystems überein. Der Ausgang dieses Subsystems stellt gleichzeitig auch den Ausgang y des Gesamtsystems dar. Das zweite der beiden Teilsysteme wird als Nulldynamik bezeichnet

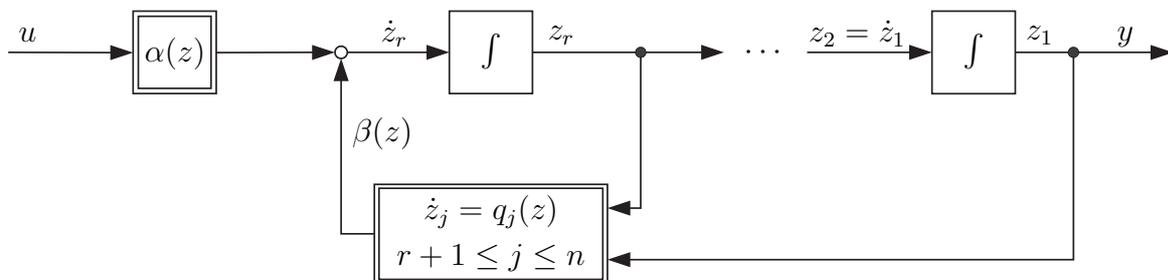


Abbildung A.4: Darstellung des nichtlinearen Systems (A.70) in Byrnes-Isidori-Normalform (BINF).

und befindet sich im Rückführzweig. Nachdem bereits r Integratoren im Vorwärtspfad enthalten sind, verbleiben für die Nulldynamik $m = n - r$ Integratoren, wenn n die Ordnung des Gesamtsystems beschreibt. Die Nulldynamik wird mit dem Ausgangssignal $y(\cdot)$ und mit den Zuständen der Integrator-kette beaufschlagt. Weil der Ausgang der Nulldynamik – additiv mit dem Eingang u des Gesamtsystems – in die Integrator-kette eingespeist wird, besteht dadurch über die Nulldynamik eine Rückkopplung der im Vorwärtszweig enthaltenen Zustände.

Als Besonderheit linearer Systeme ist anzuführen, dass bei der Festlegung der Transformation (A.78) stets hinreichend viele Freiheitsgrade bestehen, um eine Normalform (A.87) mit der Eigenschaft

$$\frac{\partial q_j}{\partial z_i} = 0 \quad \text{für alle } i = 2, \dots, r \text{ und für alle } j = r + 1, \dots, n \quad (\text{A.89})$$

zu erreichen. Die einzige Eingangsgröße des Teilsystems im Rückwärtspfad ist daher der Zustand $y = z_1$.

A.8.2 Transformation linearer Systeme

Im speziellen Falle linearer, zeitinvarianter Systeme (LTI) geben Morse [67, 68] bzw. Khalil [50, S. 537f] eine Vorgehensweise an, um die BINF auf einfachem Wege zu erhalten, d.h. ohne einen Diffeomorphismus $\Phi(\cdot)$ und damit die Funktionen $\phi_{r+1}(\cdot)$ bis $\phi_n(\cdot)$ explizit bestimmen zu müssen. Dabei wird das zu transformierende System in einer Beschreibung mittels Übertragungsfunktion

$$F(s) = g \cdot \frac{Z(s)}{N(s)}, \quad g \neq 0 \quad \text{mit} \quad \deg(Z) = m < \deg(N) = n \quad (\text{A.90})$$

angesetzt, wobei der Nennergrad $n \in \mathbb{N}$ den Zählergrad $m \in \mathbb{N}$ aus Gründen der Realisierbarkeit übersteigt. Desweiteren werden Zähler- und Nennerpolynom $Z(\cdot), N(\cdot) \in \mathbb{R}[s]$ als monisch und teilerfremd angenommen. Letztere Eigenschaft verhindert die Möglichkeit einer Pol-Nullstellen-Kürzung. Aus der Ordnung von $Z(\cdot)$ und $N(\cdot)$ ergibt sich der Relativgrad r zu

$$r = n - m \geq 1. \quad (\text{A.91})$$

Der Verstärkungsfaktor $g \in \mathbb{R}$ beschreibt die instantane Verstärkung und stellt den notwendigen Freiheitsgrad bereit, damit die Verwendung monischer Polynome keine Einschränkung bedeutet. Die Zerlegung in Nulldynamik und Integrator-kette errechnet sich aus einer Polynomdivision

$$\frac{N(s)}{Z(s)} = Q(s) + \frac{R(s)}{Z(s)}, \quad (\text{A.92})$$

durch die ein Quotientenpolynom $Q(\cdot)$ und ein Restpolynom $R(\cdot)$ entstehen. Mit den berechneten Polynomen $Q(\cdot)$ und $R(\cdot)$ kann, zusammen mit dem ursprünglichen Zählerpolynom $Z(\cdot)$, die Übertragungsfunktion $F(\cdot)$ durch die in Abbildung A.5 gezeigte

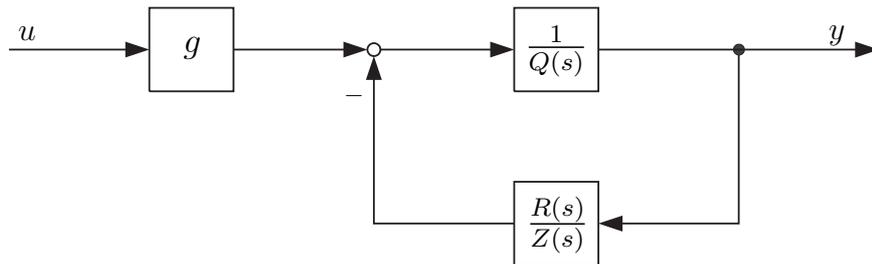


Abbildung A.5: Byrnes-Isidori-Normalform (BINF) für LTI-Systeme.

Struktur dargestellt werden. Auch hier ist eine Aufspaltung in zwei gekoppelte Subsysteme erfolgt. Die Äquivalenz beider Darstellungsweisen lässt sich durch Bestimmung der Übertragungsfunktion leicht nachvollziehen:

$$g \cdot \frac{Q^{-1}}{1 + Q^{-1}RZ^{-1}} = g \cdot \underbrace{\frac{1}{Q + RZ^{-1}}}_{NZ^{-1}} = g \cdot \frac{Z(s)}{N(s)} = F(s) \quad (\text{A.93})$$

Gleichung (A.93) zeigt die Übereinstimmung der Systemdarstellung aus Abbildung A.5 mit der ursprünglichen Übertragungsfunktion $F(\cdot)$. Die durch Q^{-1} festgelegte Vorwärtsdynamik lässt sich als lineare Regelungsnormalform (Integrator-kette mit Rückkopplung der Zustände auf den Eingang der Kette) realisieren. Der Systemteil RZ^{-1} führt den Systemausgang y auf den Vorwärtszweig zurück und stimmt deshalb mit dem als Null-dynamik bezeichneten Subsystem der BINF überein. Nachdem das Polynom $Z(\cdot)$ dort als Nenner auftritt, zeigt sich hier die in der Regelungstechnik bekannte Tatsache, dass die Übertragungsnullstellen alternativ als Pole des Rückführzweiges betrachtet werden können.

Wird der Divisionsalgorithmus explizit ausgeführt, so zeigt sich, dass im ersten Schritt ein Polynom der Ordnung n durch ein Polynom der Ordnung m geteilt wird. Folglich entsteht ein Quotient der Ordnung $n - m = r$. Bei jedem weiteren Divisionsschritt reduziert sich die Ordnung des Dividenden um mindestens 1, bis schließlich ein Polynom vom Grad m durch $Z(\cdot)$ zu teilen ist. Nach diesem letzten Divisionsschritt verbleibt ein Restpolynom $R(\cdot)$, das höchstens die Ordnung $m - 1$ erreichen kann. Je nach Konstellation der Polynomkoeffizienten kann auch $\deg(R) < m - 1$ auftreten. Allgemein lassen sich für die Ordnungen der Polynome $Q(\cdot)$ und $R(\cdot)$ die Aussagen

$$\deg(Q) = n - m = r \quad \text{und} \quad \deg(R) \leq m - 1 \quad (\text{A.94})$$

treffen. Ein Beispiel, bei dem ein Restpolynom reduzierter Ordnung auftritt, ist das System

$$F(s) = \frac{s^2 + s + 15}{s^3 + 8s^2 + 22s + 20} \quad (\text{A.95})$$

mit Nullstellen und Polen, die einen negativen Realteil aufweisen. Aus der Polynomdivision errechnet sich ein Quotient $Q(s) = s + 7$, der erwartungsgemäß die Ordnung $n - m = 3 - 2 = 1$ besitzt. Für das Restpolynom gilt $R(s) = -85$, also $\deg(R) = 0 < m - 1 = 1$. Allein die Tatsache, dass $Z(\cdot)$ und $N(\cdot)$ teilerfremde Hurwitzpolynome sind, schützt nicht vor einer Reduktion der Ordnung des Restpolynoms.

A.9 Nulldynamik, Minimalphasigkeit

Die Begriffe „Nulldynamik“ und „Minimalphasigkeit“ sind eng miteinander verwandt und werden deshalb in einem gemeinsamen Abschnitt erläutert. Bei linearen Systemen ist die Nulldynamik genau dann stabil, wenn Minimalphasigkeit vorliegt. Nachdem sich das Adjektiv „minimalphasig“ allerdings auf bestimmte Eigenschaften der Übertragungsnullstellen bezieht, und dadurch die Existenz einer Übertragungsfunktion voraussetzt, ist die Anwendung dieses Begriffs strenggenommen auf lineare Systeme begrenzt. Die Stabilität der Nulldynamik dagegen kann sowohl für lineare als auch für nichtlineare Systeme untersucht werden. Insofern lässt sich die Stabilität der Nulldynamik als Verallgemeinerung der Minimalphasigkeit ansehen. Mit dieser Begründung werden in der Literatur teils auch nichtlineare Systeme mit stabiler Nulldynamik abkürzend als „minimalphasig“ bezeichnet.

Definition A.27 (Minimalphasigkeit) [19, S. 135]

Das lineare dynamische System

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), & x(0) &\in \mathbb{R}^n, \quad n \in \mathbb{N} \\ y(t) &= cx(t) + du(t), & A &\in \mathbb{R}^{n \times n}, \quad b, c^T \in \mathbb{R}^n, \quad d \in \mathbb{R} \end{aligned} \quad (\text{A.96})$$

heißt genau dann minimalphasig, wenn

$$\det \begin{bmatrix} sI_n - A & -b \\ c & d \end{bmatrix} \neq 0 \quad \forall s \in \mathbb{C} \text{ mit } \operatorname{Re}\{s\} \geq 0. \quad (\text{A.97})$$

gilt.

Anmerkung

Für die Realisierung (A.96) berechnet sich die zugehörige Übertragungsfunktion aus der Schur-Formel zu:

$$F(s) = \frac{Z(s)}{N(s)} = \frac{\det \begin{bmatrix} sI_n - A & -b \\ c & d \end{bmatrix}}{\det[sI_n - A]} = g \cdot \frac{(s - q_1)(s - q_2) \cdots (s - q_m)}{(s - p_1)(s - p_2) \cdots (s - p_n)}, \quad (\text{A.98})$$

mit $g \neq 0$, $m \leq n$, $q_i \in \mathbb{C}, i = 1, \dots, m$, $p_j \in \mathbb{C}, j = 1, \dots, n$

Die Konsequenz der Minimalphasigkeit besteht offensichtlich darin, dass (A.98) im Zähler ein Hurwitzpolynom enthält. Folglich besitzt bei einem minimalphasigen System keine der Nullstellen $q_i, i = 1, \dots, m$ einen positiven Realteil, d.h. es gilt $\operatorname{Re}\{q_i\} < 0$ für alle $i = 1, \dots, m$. Die Realteile der Polstellen $p_j \in \mathbb{C}, j = 1, \dots, n$ beeinflussen nach Definition A.27 die Eigenschaft der Minimalphasigkeit nicht. \square

In diesem Punkt weicht Definition A.27 von der Festlegung ab, die in der linearen Regelungstechnik gebräuchlich ist. In [11, S. 249], [22, S. 147], [59, S. 283] oder [73,

S. 88] wird eine restriktivere Form angegeben, die nicht nur für die Nullstellen, sondern ebenfalls für die Polstellen positive Realteile ausschließt. Dort heißt das System (A.98) minimalphasig, wenn $F(\cdot)$ eine asymptotisch stabile Übertragungsfunktion beschreibt, d.h. wenn

$$\operatorname{Re}\{p_i\} < 0 \quad \text{für } i = 1, \dots, n \quad (\text{A.99})$$

gilt und gleichzeitig

$$\operatorname{Re}\{q_i\} < 0 \quad \text{für } i = 1, \dots, m \quad (\text{A.100})$$

erfüllt ist. Folglich sind nach der Definition in [11, S. 249], [22, S. 147], [59, S. 283] bzw. [73, S. 88] für minimalphasige Systeme Pole in der rechten Halbebene ausgeschlossen.

Dagegen kann nach Definition A.27 die Eigenschaft der Minimalphasigkeit auch für instabile Prozesse zutreffend sein. Besonders im Hinblick auf Regelkreise mit hohen Kreisverstärkungen ist eine solche Definition vorteilhaft. Für hochverstärkungsbasierte Regler ist die Stabilität der Nulldynamik ein wesentliches Kriterium für die Regelbarkeit einer Strecke, die Stabilität der Strecke selbst spielt hingegen keine Rolle.

Begründen lässt sich die Forderung nach Stabilität bei der Definition von Minimalphasigkeit in [22], [59] durch den in [79, S. 240] angeführten Sachverhalt. Beschreibt die rationale Übertragungsfunktion $F(\cdot)$ in (A.98) ein stabiles System, so ist diese in der abgeschlossenen rechten Halbebene $\bar{\mathbb{C}}_+$ holomorph und beschränkt. Demnach sind die Funktionswerte $F(s)$ für alle $s \in \mathbb{C}_+$ durch den Verlauf von $F(\cdot)$ über der imaginären Achse eindeutig bestimmt, d.h. für den Funktionswert $F(s_o)$ mit $s_o = \sigma_o + j\omega_o$, $\sigma_o > 0$, $\omega_o \in \mathbb{R}$ gilt:

$$F(s_o) = \frac{1}{\pi} \int_{-\infty}^{+\infty} F(j\omega) \frac{\sigma_o}{\sigma_o^2 + (\omega - \omega_o)^2} d\omega \quad (\text{A.101})$$

Bekanntermaßen zeichnen sich stabile, nichtminimalphasige Systeme durch die Besonderheit aus, dass deren Übertragungsfunktion in eine Komponente mit minimalphasigem Verhalten und in einen Allpass-Anteil zerlegt werden kann [79, S. 242]. Hierbei ist unter dem Begriff Allpass ein System mit der folgenden Eigenschaft zu verstehen:

Definition A.28 (Allpass)

Ein stabiles Übertragungssystem der Form (A.98) heißt Allpass, wenn ein konstanter Amplitudengang

$$|F(j\omega)| \equiv 1 \quad \forall \omega \in \mathbb{R} \quad (\text{A.102})$$

vorliegt.

Es ist leicht nachzuweisen [59], dass Allpässe eine symmetrische Pol-Nullstellen-Verteilung besitzen. Zu jeder Polstelle muss spiegelbildlich zur Imaginärachse eine Nullstelle existieren. Daher bedingt jeder Pol mit negativem Realteil eine Nullstelle mit positivem Realteil. Folglich sind Allpässe nichtminimalphasige Übertragungsglieder.

Die Definition des Allpasses verlangt, dass jede sinusoidale Anregung unabhängig von ihrer Frequenz mit dem Faktor 1 verstärkt wird. Dagegen ist keine Aussage über die

Signallaufzeit getroffen, deren Frequenzabhängigkeit zu einem nicht konstanten Phasengang führt.

Am Beispiel des Allpass-Systems

$$F_{AP}(s) = \frac{s - \sigma_o - j\omega_o}{s + \sigma_o - j\omega_o}, \quad \text{mit } \sigma_o > 0, \omega_o \in \mathbb{R} \quad (\text{A.103})$$

wird der Frequenzgang analysiert. Es ergibt sich

$$F_{AP}(j\omega) = \frac{-\sigma_o + j(\omega - \omega_o)}{+\sigma_o + j(\omega - \omega_o)}, \quad (\text{A.104})$$

woraus sich unmittelbar die Allpass-Bedingung $|F_{AP}(j\omega)| = 1$ für alle $\omega \in \mathbb{R}$ entnehmen lässt. Der Betragsgang zeigt somit einen konstanten Verlauf und unterscheidet sich nicht von demjenigen einer Verstärkung mit dem Faktor 1.

Der Phasenwinkel $\arg(F_{AP}(j\omega))$ errechnet sich aus der Differenz zwischen den Phasenwinkeln von Zähler und Nenner, die auf

$$\arg(F_{AP}(j\omega)) = \arg(-\sigma_o + j(\omega - \omega_o)) - \arg(\sigma_o + j(\omega - \omega_o)) = \pi - 2 \arctan\left(\frac{\omega - \omega_o}{\sigma_o}\right) \quad (\text{A.105})$$

führt. Sowohl die lineare Funktion $\omega \mapsto (\omega - \omega_o)/\sigma_o$ mit der positiven Steigung $1/\sigma_o > 0$, als auch die Arkustangensfunktion, wachsen monoton. Daher fällt der Phasengang monoton über der ω -Achse. Für $\omega \rightarrow \pm\infty$ lauten die Grenzwerte:

$$\lim_{\omega \rightarrow -\infty} \arg(F_{AP}(j\omega)) = 2\pi, \quad \lim_{\omega \rightarrow +\infty} \arg(F_{AP}(j\omega)) = 0 \quad (\text{A.106})$$

Daran ist ersichtlich, dass der Allpass $F_{AP}(s)$ über der ω -Achse eine globale Phasenabsenkung um -2π bewirkt. Im Gegensatz zum Allpass behält eine Verstärkung mit dem Faktor 1 den Phasenwinkel 0 bei.

Allpässe höherer Ordnung mit $n \in \mathbb{N}$ stabilen Polen und demzufolge n Nullstellen in der rechten Halbebene lassen sich aus dem Grundbaustein $F_{AP}(\cdot)$ mit den entsprechenden Werten für σ_o und ω_o zusammensetzen. Aus diesem Grund können die oben angestellten Überlegungen auf Allpässe beliebiger Ordnung übertragen werden, die folglich eine globale Phasendrehung von $-2\pi n$ verursachen.

Wird einem stabilen, minimalphasigen System ein Allpass in Serie hinzugefügt, bleibt dessen Betragsgang unverändert erhalten, der Einfluss des Allpasses ist ausschließlich auf die Phase beschränkt. Als Konsequenz hieraus geht die Eigenschaft verloren, dass aus dem gegebenen Betragsverlauf der zugehörige Phasengang konstruiert werden kann. Dem Betragsgang sind die Kenngrößen σ_o und ω_o des Allpasses nicht zu entnehmen, so dass die Funktion

$$\omega \mapsto \pi - 2 \arctan\left(\frac{\omega - \omega_o}{\sigma_o}\right) = \arg(F_{AP}(j\omega)) \quad (\text{A.107})$$

unbekannte Parameter enthält und daher nicht ausgewertet werden kann. Einzig die globale Phasendrehung lässt sich angeben, wenn die Ordnung eines vorhandenen Allpasses bekannt ist.

Aus diesem Grund lassen allpasshaltige und daher nichtminimalphasige Systeme keinen eindeutigen Schluss vom Betragsgang auf den Phasengang zu.

Stabile, minimalphasige Übertragungsglieder, die sich aus einer multiplikativen Kombination der Grundbausteine

$$F_Z(s) = s + \sigma_o - j\omega_o, \quad \text{mit } \sigma_o > 0, \omega_o \in \mathbb{R} \quad (\text{A.108})$$

und

$$F_N(s) = \frac{1}{s + \sigma_o - j\omega_o} = F_Z^{-1}(s) \quad (\text{A.109})$$

zusammensetzen, erlauben dagegen einen eindeutigen Schluss vom Betragsgang auf den Phasenverlauf. Weil Betrag und Phase einer holomorphen Funktion voneinander nicht unabhängig sind [22, S. 171], kann mit Ausnahme einer Konstanten $g > 0$ immer eineindeutig vom Phasen- auf den Amplitudenverlauf der Frequenzgangfunktion $F(j\omega)$ geschlossen werden.

Für die Amplitudengänge gilt

$$|F_Z(j\omega)| = \sqrt{\sigma_o^2 + (\omega - \omega_o)^2} = \frac{1}{|F_N(j\omega)|}, \quad (\text{A.110})$$

die Phasengänge ergeben sich zu:

$$\arg(F_Z(j\omega)) = \arctan\left(\frac{\omega - \omega_o}{\sigma_o}\right) = -\arg(F_N(j\omega)) \quad (\text{A.111})$$

Aus den Grenzübergängen $\omega \rightarrow \pm\infty$ folgt für $F_Z(s)$ die globale Phasendrehung von $\pi/2 - (-\pi/2) = \pi$ und von $-\pi/2 - \pi/2 = -\pi$ für $F_N(s)$. Daher entscheidet der Relativgrad des Systems über die globale Phasendrehung bei stabilen, minimalphasigen Systemen.

Das eigentümliche Verhalten nichtminimalphasiger Systeme lässt sich anhand von Modellen für technisch relevante Prozesse herausstellen. In diesem Zusammenhang sind unter anderem Wasserkraftwerke [59, S. 285f], [83, S. 118f], Dampferzeuger [59, S. 286] und Flugzeuge [13], [73, S. 201f] zu nennen. Beispielhaft sei unter diesem Aspekt das Verhalten von Flugzeugen auf Veränderung des Höhenruders herausgegriffen.

Beispiel (Reaktion von Flugzeugen auf Höhenruderänderungen)

In Tabelle A.3 und in Abbildung A.6 sind die physikalischen Größen und Koordinatensysteme festgelegt, die zur Darstellung des nichtminimalphasigen Verhaltens eines Flugzeuges benötigt werden.

Da Windströmungen die prinzipielle Reaktion des Flugzeugs auf eine Verstellung des Höhenruders nicht beeinflussen, darf die Bewegungsgeschwindigkeit der Luft in diesem Beispiel zur Vereinfachung unberücksichtigt bleiben,⁷⁾ weshalb

$$v_K = v_A \quad (\text{A.112})$$

⁷⁾ Wenn die Windgeschwindigkeit zu 0 angenommen wird, liegt ein Flug in unbewegter Luft vor. Dadurch stimmt die Bahngeschwindigkeit v_K mit der Fluggeschwindigkeit v_A überein, und die Richtung der anströmenden Luft wird durch den Vektor v_K beschrieben. Deshalb gleicht der Winkel zwischen x_F und v_K demjenigen zwischen x_F und der anströmenden Luft, der als Anstellwinkel α bezeichnet wird.

| | |
|-------------------------------|--|
| x_E, z_E | x -Achse bzw. z -Achse eines erdfesten Koordinatensystems |
| x_F, z_F | x -Achse bzw. z -Achse eines flugzeugfesten Koordinatensystems |
| $v_K \in \mathbb{R}^3$ | Bahngeschwindigkeit (Relativbewegung zwischen Flugzeug und Erde) |
| $v_A \in \mathbb{R}^3$ | Fluggeschwindigkeit (Relativbewegung zwischen Flugzeug und Luft) |
| $\Theta \in \mathbb{R}$ | Nickwinkel (Längsneigung) |
| $\alpha \in \mathbb{R}$ | Anstellwinkel |
| $\gamma \in \mathbb{R}$ | Bahnneigungswinkel |
| $\dot{\Theta} \in \mathbb{R}$ | Drehgeschwindigkeit um y_F -Achse (Nickgeschwindigkeit) |
| $\eta \in \mathbb{R}$ | Ruderwinkel am Höhenleitwerk |
| $m > 0$ | Flugzeugmasse |
| $J > 0$ | Massenträgheitsmoment um die y_F -Achse |

Tabelle A.3: Definition von physikalischen Größen und Koordinatensystemen für das Beispielsystem Flugzeug.

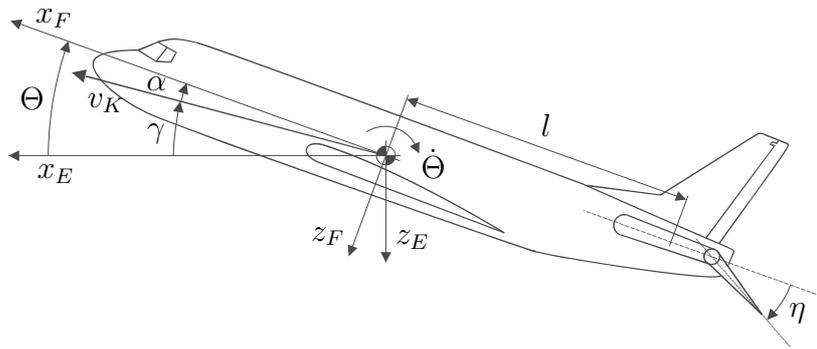


Abbildung A.6: Flugzeug als Beispiel für nichtminimalphasiges Verhalten mit zugehörigen Koordinatensystemen.

gilt. Ein gemäß Abbildung A.6 positiver Ausschlag η des Höhenruders vergrößert die Profilwölbung und steigert damit den aerodynamisch erzeugten Auftriebseffekt am Höhenleitwerk. Diese zusätzliche Auftriebskraft $F_H \in \mathbb{R}^3$ führt zunächst zu einer merklichen Beschleunigung des Flugzeuges in Richtung der negativen z_E -Achse, hebt das Flugzeug also an. Nachdem F_H jedoch nicht im Schwerpunkt angreift, sondern im Heckbereich, bewirkt die Auftriebserhöhung gleichzeitig auch ein negatives Nickmoment. Es tritt daher eine Nickbeschleunigung $\dot{\Theta} < 0$ ein. Dadurch wird eine Rotation entgegen der durch $\dot{\Theta}$ gekennzeichneten Richtung um die y_F -Achse verursacht (das Flugzeug nickt vorne ab, Θ verringert sich), in deren Folge über den Zusammenhang

$$\alpha = \Theta - \gamma \quad (\text{A.113})$$

auch der Anstellwinkel α verringert wird. Weil der am Flügel erzeugte Auftrieb proportional vom Anstellwinkel α abhängt, wird durch die eintretende Rotation der Gesamtauftrieb⁸⁾ abnehmen, wodurch die Flugbahn in der Folge nach unten abgeändert

⁸⁾ Für kleine Anstellwinkel $\alpha < 15^\circ$ gilt näherungsweise ein linearer Zusammenhang zwischen Anstellwinkel und Flügelauftrieb, siehe [13, S. 9].

wird.

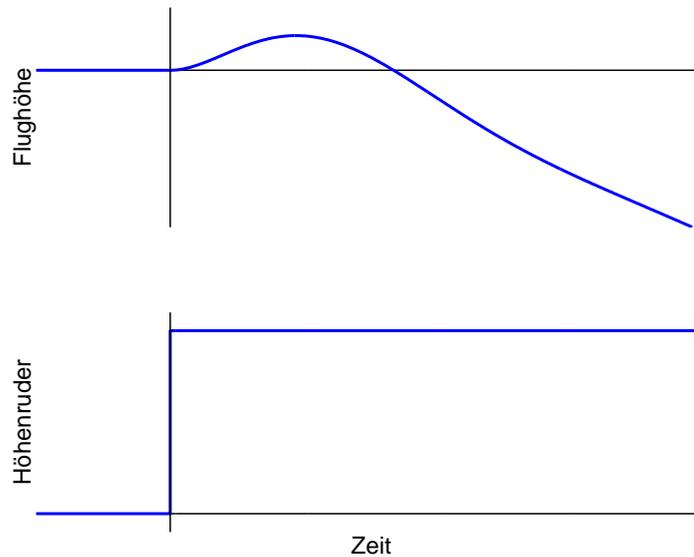


Abbildung A.7: Qualitative Sprungantwort eines Flugzeugs bei Verstellung des Höhenruders.

Ein qualitativer Verlauf der Flughöhe ist in Abbildung A.7 aufgezeichnet, wenn der Höhenrudderwinkel η sprunghaft verstellt wird. Bemerkenswert ist vor allem die anfängliche Richtungsumkehr: zunächst tritt ein Anstieg der Flughöhe ein, bevor das Flugzeug zu sinken beginnt. Der Grund für die anfängliche Richtungsumkehr liegt in der Tatsache verankert, dass zwei gegensätzliche Wirkungswege mit unterschiedlicher Geschwindigkeit im System vorhanden sind. Der schnelle Wirkungsweg ist das sofortige Auftreten einer Auftriebserhöhung F_H im Heckbereich, d.h. einer beschleunigenden Kraft entlang der negativen z_E -Achse. Daraus entsteht nach zweifacher Integration

$$a_E = \frac{F_H}{m} \Rightarrow v_E(t) = \int_0^t a_E(\tau) d\tau \Rightarrow z_E(t) = \int_0^t v_E(\tau) d\tau \quad (\text{A.114})$$

ein Zuwachs der Flughöhe z_E . Gleichzeitig bewirkt die Kraft F_H , die im Abstand $l > 0$ vom Schwerpunkt entfernt angreift, ein Drehmoment $M_H = F_H \cdot l$, das zu einer Nickbewegung führt. Auch hier ist eine zweifache Integration vorzunehmen, um daraus den Nickwinkel Θ zu erhalten:

$$\ddot{\Theta} = \frac{F_H \cdot l}{J} \Rightarrow \dot{\Theta}(t) = \int_0^t \ddot{\Theta}(\tau) d\tau \Rightarrow \Theta(t) = \int_0^t \dot{\Theta}(\tau) d\tau \quad (\text{A.115})$$

Nachdem allerdings der Nickwinkel Θ bzw. der Anstellwinkel α für die Auftriebskraft am Flügel verantwortlich ist, resultiert aus dem Abnicken des Flugzeuges über den Zusammenhang (A.113) ein Verlust des Flügelauftriebes, der eine Beschleunigung entlang der positiven z_E -Achse bewirkt, weil der reduzierte Auftrieb die Gewichtskraft nicht mehr vollständig kompensiert. Ab hier tritt der langsamere Wirkungsweg zu Tage. Die resultierende Kraftdifferenz, die sich aus der zweifachen Integration von F_H ergibt, besitzt zwar ein gegenteiliges Vorzeichen wie F_H , muss aber analog zu Gleichung (A.114) noch zwei mal integriert werden, um den Höhenverlust zu ermitteln.

Daraus zeigt sich, dass der Höhengewinn aus einer zweifachen Integration der Heckauftriebskraft F_H entsteht und daher schneller eintritt als der Höhenverlust, der aus einer vierfachen Integration von F_H resultiert. Es existieren somit zwei parallele Wirkungswege mit unterschiedlichem Vorzeichen aber auch unterschiedlicher Wirkungsgeschwindigkeit. Diese Eigenheit ist typisch für nicht-minimalphasige Systeme und stellt letztendlich den Grund für die anfängliche Richtungsumkehr dar. \square

Beispiel

Ein weiteres Beispiel (ohne direkten technischen Bezug) soll den Einfluss des Auskoppelvektors auf die Systemeigenschaft Minimalphasigkeit demonstrieren. Das in Regeleingangsnormalform gegebene LTI-System

$$\begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{pmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1.25 & -2.25 & -2 \end{bmatrix} \cdot \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \cdot u(t), \quad \begin{pmatrix} x_1(0) \\ x_2(0) \\ x_3(0) \end{pmatrix} \in \mathbb{R}^3$$

$$y(t) = c \begin{pmatrix} x_1(t) & x_2(t) & x_3(t) \end{pmatrix}^T, \quad c \in \mathbb{R}^3 \quad (\text{A.116})$$

hat eine Eigendynamik, welche durch die Pole bei $p = \{-1; -0.5 \pm j\}$ bestimmt wird. Die Sprungantwort für dieses System besitzt abhängig vom Auskoppelvektor c unterschiedliches Verhalten. Wird

$$c = c_1 := \begin{pmatrix} \frac{5}{4} & \frac{21}{8} & 1 \end{pmatrix} \quad (\text{A.117})$$

gesetzt, liegen die Nullstellen bei $q = \{-2; -5/8\}$ und weisen negative Realteile auf. Das System ist folglich minimalphasig und zeigt die Sprungantwort, die in Abbildung A.8 skizziert ist. Abändern des Auskoppelvektors auf

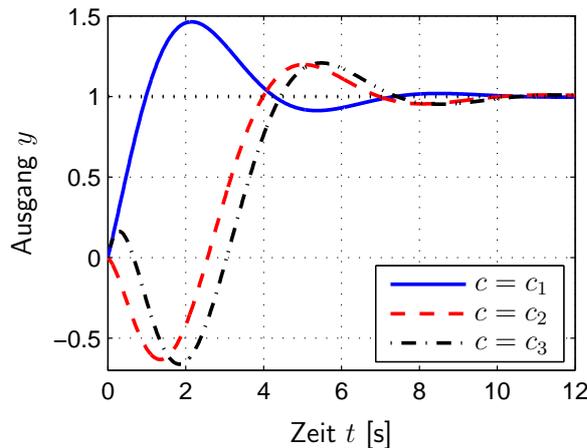


Abbildung A.8: Sprungantworten bei unterschiedlicher Auskopplung.

$$c = c_2 := \begin{pmatrix} \frac{5}{4} & -\frac{21}{8} & -\frac{1}{4} \end{pmatrix} \quad (\text{A.118})$$

verlegt eine der beiden Nullstellen $q \approx \{-10.96; 0.46\}$ in die rechte Halbebene. Das System wird nichtminimalphasig und antwortet anfänglich mit einer Richtungsumkehr. Die in Abbildung A.8 mit $c = c_2$ beschriftete Kurve entfernt sich zunächst vom positiven Endwert und läuft in negative Wertebereiche. Erst ab etwa 2.5 s verhält sich die Sprungantwort für $c = c_2$ qualitativ ähnlich wie jene für $c = c_1$.

Eine Mischung beider Antworten wird durch eine Auskopplung mit

$$c = c_3 := \begin{pmatrix} \frac{5}{4} & -\frac{25}{8} & \frac{5}{4} \end{pmatrix} \quad (\text{A.119})$$

erreicht. Hier liegen beide Nullstellen $q = \{2; 1/2\}$ in der rechte Halbebene. Zwar strebt die Sprungantwort von Beginn an in Richtung Endwert. Dennoch stellt sich zwischenzeitlich die für nichtminimalphasige Systeme typische Richtungsumkehr ein, bevor auch für $c = c_3$ der positive Endwert asymptotisch erreicht wird. \square

Abschließend sei angemerkt, dass eine Richtungsumkehr zwar charakteristisch für nichtminimalphasige Systeme ist, jedoch nicht in allen Fällen notwendigerweise eintreten muss. Bei Sprunganregung entfernt sich anfänglich ein lineares, asymptotisch stabiles System mit Relativgrad $r \geq 1$ genau dann vom Endwert, wenn eine ungerade Anzahl von Nullstellen mit positivem Realteil vorliegt [33]. Aus der Abwesenheit einer Richtungsumkehr in der Sprungantwort folgt daher nicht zwingend Minimalphasigkeit.

Für eine Verallgemeinerung der Minimalphasigkeit auf nichtlineare Systeme

$$\begin{aligned} \dot{x}(t) &= f(x(t)) + g(x(t)) \cdot u(t), & x(0) &= x_o \in \mathbb{R}^n \\ y(t) &= h(x(t)) \end{aligned} \quad (\text{A.120})$$

mit den Funktionen $f, g \in \mathcal{C}^l(\mathbb{R}^n; \mathbb{R}^n)$, $h \in \mathcal{C}^l(\mathbb{R}^n; \mathbb{R})$ und $l \in \mathbb{N}$ muss die Betrachtung der Übertragungsnullstellen aufgegeben und stattdessen die Nulldynamik analysiert werden.

Gemäß der in Abschnitt A.8 vorgestellten Normalform lässt sich ein dynamisches System in zwei Subsysteme zerlegen. Dasjenige im Rückführzweig ist einem dynamischen Verhalten unterworfen, das Isidori als „zero dynamics“ bezeichnet [47, Kap. 4.3], was in der deutschsprachigen Literatur gemeinhin als „Nulldynamik“ übersetzt wird. Sowohl die systemtheoretische Bedeutung der Nulldynamik als auch die Namensgebung erklären sich aus folgender Definition.

Definition A.29 (Nulldynamik) [97]

Die Nulldynamik beschreibt den Verlauf der Zustände von (A.120) für den Fall, dass die Ausgangsgröße $y(t)$ für ein geeignetes Eingangssignal $u(\cdot)$ und gegebene Anfangsbedingungen $x(0) \in \mathbb{R}^n$ identisch 0 ist. Die Gesamtheit aller Trajektorien $x(\cdot)$ unter der Restriktion $y(\cdot) = 0$ heißt Nulldynamik.

Die Nulldynamik ist damit die Bewegungsmöglichkeit des Systems, unter der Nebenbedingung $y(t) = 0$ für alle $t \geq 0$. Deren Berechnung erfordert die Bestimmung eines geeigneten Eingangssignales, so dass die Forderung

$$y(t) = \dot{y}(t) = \ddot{y}(t) = \dots = 0, \quad \forall t \geq 0 \quad (\text{A.121})$$

erfüllt ist. Die Lösbarkeit dieser Aufgabe⁹⁾ hängt in erheblichem Maße von den gegebenen Anfangswerten ab. Anhand einer Systembeschreibung in BINF tritt auf vereinfachte Weise zu Tage, für welche Anfangszustände $x(0)$ eine Ausgangsnulldynamik möglich ist.

Unter der Annahme eines Relativgrades $r \in \mathbb{N}$ geht aus Abschnitt A.6 hervor, dass die Forderung $y(t) = 0$ für alle $t \geq 0$ bzw. die daraus resultierende Gleichung (A.121) mit

$$h(x(t)) = L_f h(x(t)) = L_f^2 h(x(t)) = \dots = L_f^{r-1} h(x(t)) = 0 \quad (\text{A.122})$$

übereinstimmt. Nachdem die Ableitung $y^{(r)}$ durch den Systemeingang u direkt beeinflussbar ist, und damit durch eine geeignete Stellgröße stets bei 0 gehalten werden kann, müssen höhere Ableitungen nicht weiter beachtet werden. Die Transformationsvorschrift (A.79) enthält die Aussage, dass die Lie-Ableitungen $L_f^0 h(x)$ bis $L_f^{r-1} h(x)$ mit den Zustandsgrößen z_1 bis z_r übereinstimmen und daher in Gleichung (A.122) substituiert werden können. Daher erweist sich eine Aufspaltung des Zustandsvektors z als hilfreich, die der Aufteilung zwischen Dynamik im Vorwärtszweig und Dynamik im Rückführpfad entspricht. Wird der Zustandsvektor der Systembeschreibung (A.87) in

$$\xi := \begin{pmatrix} z_1 \\ \vdots \\ z_r \end{pmatrix} \in \mathbb{R}^r \quad \text{und} \quad \zeta := \begin{pmatrix} z_{r+1} \\ \vdots \\ z_n \end{pmatrix} \in \mathbb{R}^{n-r} \quad (\text{A.123})$$

partitioniert, so ist Gleichung (A.122) zur Aussage

$$\xi(t) = 0 \quad \forall t \geq 0 \quad (\text{A.124})$$

äquivalent. Daran ist erkennbar, dass sich die Aufgabenstellung für alle Anfangswerte $\xi(0) \neq 0$ unlösbar gestaltet. Ist hingegen Forderung (A.124) erfüllt, entfällt die Dynamik im Vorwärtszweig und es kann für die verkürzte Systemdynamik

$$\xi(t) = 0 \quad (\text{A.125})$$

$$\dot{\zeta}_r(t) = \beta(0, \zeta(t)) + \alpha(0, \zeta(t)) \cdot u(t) \equiv 0 \quad (\text{A.126})$$

$$\dot{\zeta}(t) = \begin{pmatrix} q_{r+1}(0, \zeta(t)) \\ \vdots \\ q_n(0, \zeta(t)) \end{pmatrix}, \quad \zeta(0) \in \mathbb{R}^{n-r} \quad (\text{A.127})$$

geschrieben werden. Folglich muss das Eingangssignal

$$u(\cdot) = -\frac{\beta(0, \zeta(\cdot))}{\alpha(0, \zeta(\cdot))} \quad (\text{A.128})$$

gewählt werden, das allein von den Zuständen des Subsystems im Rückführzweig abhängt. Der Verlauf der Zustandsgrößen im System wird in diesem Falle ausschließlich durch die Differentialgleichung (A.127) festgelegt. Offensichtlich ist nur für die

⁹⁾ Üblicherweise wird diese Aufgabe als „zeroing the output“ bezeichnet. In der deutschsprachigen Literatur ist die Übersetzung „Ausgangsnulldynamik“ zu finden.

Zustände ζ im Teilsystem des Rückführpfades eine Bewegung zulässig, die wegen der Nebenbedingung $y \equiv 0$ als „Nullodynamik“ bezeichnet wird.

Anmerkung

Nachdem bei Beachtung des Kontextes keine Verwechslungsmöglichkeit besteht, wird in dieser Arbeit mit einer geringfügigen Erweiterung des Begriffes nicht nur der Zustandsverlauf $\zeta(\cdot)$, sondern das gesamte Teilsystem, welches diese Zustände enthält, als Nullodynamik bezeichnet. \square

Die Systemdarstellung nach Abschnitt A.8.2 schafft für lineare Systeme eine Verbindung zwischen der Stabilität der Nullodynamik und der Eigenschaft der Minimalphasigkeit. In Abbildung A.5 fällt auf, dass die Nullodynamik als Nennerpolynom das ursprüngliche Zählerpolynom $Z(\cdot)$ erhält. Die Stabilität der Nullodynamik linearer Systeme wird daher vom Zählerpolynom der Übertragungsfunktion des Gesamtsystems bestimmt, was sich in folgendem Satz niederschlägt:

Hilfssatz A.30

Die Nullodynamik des linearen Systems $F(s) = Z(s)/N(s)$ mit $Z, N \in \mathbb{R}[s]$, $N(\cdot) \neq 0$, $\deg(Z) < \deg(N)$, ist genau dann asymptotisch stabil, wenn $Z(\cdot)$ ein Hurwitz-Polynom ist.

Nach der Anmerkung auf Seite 239 handelt es sich bei einem linearen System mit Hurwitzpolynom im Zähler um ein minimalphasiges System.

Minimalphasige Systeme besitzen somit eine stabile Nullodynamik, nichtminimalphasige Systeme eine instabile Nullodynamik. Dadurch wird die Äquivalenz von Minimalphasigkeit und stabiler Nullodynamik bzw. Nichtminimalphasigkeit und instabiler Nullodynamik offensichtlich.

Da für nichtlineare Differentialgleichungen die Laplace-Transformation nicht angewendet werden kann, existiert keine Übertragungsfunktion und damit auch kein Zählerpolynom. Übertragungsnullstellen, wie im Falle linearer Systeme, bestehen daher nicht. Dennoch lässt sich die Nomenklatur analog zu Satz A.30 auf nichtlineare Systeme erweitern:

Hilfssatz A.31

Ein nichtlineares dynamisches System der Form (A.120) heißt minimalphasig, wenn

$$y \equiv 0 \quad \Rightarrow \quad (x, u) \rightarrow 0 \quad (\text{A.129})$$

gilt. Andernfalls liegt ein nichtminimalphasiges System vor.

Um die systemtheoretische Bedeutung der Nullodynamik hervorzuheben, wird vorzugsweise die Systemdarstellung in BINF betrachtet. In der Darstellung (A.87) beinhaltet das als Nullodynamik bezeichnete Subsystem die Zustände z_{r+1} bis z_n und generiert das Signal $\beta(z(\cdot))$. Dieses wird mit $\alpha(z(\cdot)) \cdot u(\cdot)$ summiert und bestimmt die zeitliche Änderung der Zustandsgröße $z_r(t)$. Unter der Voraussetzung eines starken Relativgrades (d.h. $\alpha(z) \neq 0$ für alle $z \in \mathbb{R}^n$) besteht prinzipiell die Möglichkeit, durch

die Stellgröße $u(t) = -\beta(z(t))/\alpha(z(t))$ den Einfluss von $\beta(z(t))$ zu eliminieren. Bei minimalphasigen Systemen mit stabiler Nulldynamik ist hierzu ein beschränktes Eingangssignal $u(\cdot) = -\beta(z(\cdot))/\alpha(z(\cdot)) \in \mathcal{L}^\infty$ ausreichend, solange für ein beschränktes Ausgangssignal $y \in \mathcal{L}^\infty$ Sorge getragen wird.

Nichtminimalphasige Systeme mit ihrer instabilen Nulldynamik erfordern dagegen im Allgemeinen ein unbeschränktes Eingangssignal um deren Einfluss zu unterdrücken, selbst wenn der Systemausgang y beschränkt gehalten wird. Nichtminimalphasige Systeme besitzen daher die Besonderheit, dass ein unbeschränkt anwachsendes Eingangssignal im System „verschwinden“ kann, d.h. zu einem nicht unbeschränkt wachsendem Ausgangssignal führt.

Aus diesem Grund sind nichtminimalphasige Regelstrecken von hochverstärkungsbaierten Reglern nicht zu beherrschen. Solchen Reglern liegt stets die Annahme zu Grunde, dass die zeitliche Änderung des Streckenausganges \dot{y} durch geeignete Erhöhung der Stellgröße u frei beeinflussbar ist. Wenn jedoch ein beliebig großes Eingangssignal keine Veränderung des Ausganges bewirkt, ist die Annahme verletzt, was in der Folge Stabilitätsprobleme verursacht.

A.10 Essentielles Supremum

Das essentielle Supremum¹⁰⁾ einer Funktion wird hier aufgegriffen, da es für die Definition der ∞ -Norm für Signale benötigt wird. Im Gegensatz zum Supremum, das im gesamten Definitionsbereich einer Funktion gesucht wird, schließt die Suche nach dem essentiellen Supremum Teile des Definitionsbereiches mit Maß Null aus. Für die Charakterisierung solcher Teilbereiche ist das Lebesgue-Maß¹¹⁾ $\mu : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ ein grundlegendes Hilfsmittel.

Definition A.32 (Menge mit Maß Null) [95, S. 467]

Eine Menge $M \subset \mathbb{R}^n$ besitzt das Maß Null, d.h. $\mu(M) = 0$, wenn für jede beliebige Zahl $\varepsilon > 0$ eine abzählbare Vereinigungsmenge von Bällen $B_1, B_2, \dots \subset \mathbb{R}^n$ existiert, so dass

$$\sum_{i=1}^{\infty} \mu(B_i) < \varepsilon \quad (\text{A.130})$$

und gleichzeitig auch

$$M \subseteq \cup_{i=1}^{\infty} B_i \quad (\text{A.131})$$

erfüllt ist.

Werden Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ mit eindimensionalem Definitionsbereich betrachtet, kann als Maß μ einem Intervall $[a, b]$, $a \leq b \in \mathbb{R}$, dessen Länge zugeordnet werden und

¹⁰⁾ In der deutschsprachigen Literatur wird das essentielle Supremum teils auch als wesentliches Supremum bezeichnet.

¹¹⁾ Benannt nach dem französischen Mathematiker Henri Léon Lebesgue (1875 - 1941).

es gilt $\mu([a, b]) = b - a$. Ein einzelner Punkt des Definitionsbereiches (d.h. ein Intervall $I_o := [a, a]$) besitzt offensichtlich das Lebesgue-Maß $\mu(I_o) = 0$. Mit diesem Maß kann nun die Definition der essentiellen Beschränktheit erfolgen.

Definition A.33 (essentiell beschränkte Funktion, essentielle Schranke)

Eine stückweise stetige Funktion $f : D \rightarrow Y$, $D, Y \subseteq \mathbb{R}$, heißt essentiell beschränkt, wenn es eine Schranke $M \in \mathbb{R}$ gibt, so dass:

$$\mu(\{x \in D \mid |f(x)| > M\}) = 0 \quad (\text{A.132})$$

Eine solche Schranke M heißt essentielle Schranke.

Es ist festzustellen, dass eine essentiell beschränkte Funktion nicht zwangsläufig beschränkt ist. So ist zum Beispiel die Funktion

$$f : x \mapsto f(x) = \begin{cases} |\sin(x)| & \forall x \in \mathbb{R} \setminus \{0, \pm\pi, \pm2\pi, \dots\} \\ x^2 & \forall x \in \{0, \pm\pi, \pm2\pi, \dots\} \end{cases} \quad (\text{A.133})$$

mit $M = 1$ essentiell beschränkt, aber nicht beschränkt. Die Festlegung einer essentiellen Schranke ignoriert Teilintervalle mit Maß Null (in diesem Beispiel alle $x \in \{0, \pm\pi, \pm2\pi, \dots\}$), so dass eine essentiell beschränkte Funktion ihre essentielle Schranke überschreiten kann. Auf Intervallen mit einem Maß $\mu > 0$, wie z.B. $[a, b]$ mit $a < b$, muss dagegen die durch eine essentielle Schranke vorgegebene Grenze eingehalten werden.

Definition A.34 (essentielles Supremum)

Das essentielle Supremum wird durch die kleinstmögliche essentielle Schranke $M \in \mathbb{R}$ für die Funktion $f(\cdot)$ festgelegt:

$$\text{ess-sup } f = \inf \{M > 0 \mid M \text{ ist essentielle Schranke von } f\} \quad (\text{A.134})$$

Bedingt durch die Eigenschaften des essentiellen Supremums gilt allgemein

$$\sup_{x \in D} f(x) \geq \text{ess-sup}_{x \in D} f(x). \quad (\text{A.135})$$

Unter der Einschränkung auf stetige Funktionen werden einzelne „Ausreißerwerte“, die über einer essentiellen Schranke liegen, ausgeschlossen. Stetige Funktionen überschreiten eine gegebene Schranke entweder auf Intervallen mit einem Maß $\mu > 0$ oder die Funktion liegt vollständig unterhalb der Schranke. Daher kann es keine Teilintervalle im Definitionsbereich mit Maß Null geben, auf welchen die essentielle Schranke überstiegen wird. Somit stimmen im speziellen Falle stetiger Funktionen das Supremum und das essentielle Supremum überein. In Gleichung (A.135) ist daher das Gleichheitszeichen anzusetzen, wenn $f(\cdot) \in \mathcal{C}^0(D; \mathbb{R})$ eine stetige Funktion darstellt.

Desweiteren ist für stetige Funktionen $f(\cdot) \in \mathcal{C}^0(D; \mathbb{R})$, $D \subseteq \mathbb{R}$, die Aussage

$$|f(x)| \leq \text{ess-sup}_{x \in D} |f(x)| = \sup_{x \in D} |f(x)| \quad \forall x \in D \quad (\text{A.136})$$

zutreffend. Bei unstetigen Funktionen hingegen kann diese Ungleichung auf Intervallen mit Maß Null verletzt werden. Daher ist eine Einschränkung bezüglich des Gültigkeitsbereiches erforderlich:

$$|f(x)| \leq \operatorname{ess-sup}_{x \in D} |f(x)| \quad \text{für fast alle } x \in D \quad (\text{A.137})$$

Dadurch sind Intervalle mit Maß Null ausgeschlossen, weil der Ausdruck „für fast alle“ der Aussage „überall, außer auf Intervallen mit Maß Null“ gleicht.

A.11 Norm

Im Folgenden werden Normen für Vektoren, Matrizen und Funktionen betrachtet, die vertieft in [18, Kap. II] diskutiert sind.

A.11.1 Norm für Vektoren

Allgemein stellt die Norm eine Abbildung

$$\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0} \quad (\text{A.138})$$

dar, die jedem Element $x \in \mathbb{R}^n$, $n \in \mathbb{N}$, eines Vektorraumes eine nichtnegative reelle Zahl zuordnet und dabei drei wesentliche Eigenschaften erfüllt:

(i) positive Definitheit:

$$\|x\| \geq 0 \text{ für alle } x \in \mathbb{R}^n, \text{ wobei } \|x\| = 0 \Leftrightarrow x = 0$$

(ii) Homogenität:

$$\|ax\| = |a| \cdot \|x\| \text{ für jede skalare Konstante } a \in \mathbb{R} \text{ und für alle } x \in \mathbb{R}^n$$

(iii) Dreiecksungleichung:

$$\|x + y\| \leq \|x\| + \|y\|, \text{ für alle } x, y \in \mathbb{R}^n$$

Definition A.35 (*p*-Norm eines Vektors $x \in \mathbb{R}^n$)

Einem Vektor $x \in \mathbb{R}^n$ wird die p-Norm

$$\|x\|_p := \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}, \quad p \geq 1 \quad (\text{A.139})$$

zugewiesen, die den geometrischen Begriff der Länge eines Vektors verallgemeinert. Für $p = \infty$ wird

$$\|x\|_\infty := \max_i |x_i|, \quad i = 1, \dots, n \quad (\text{A.140})$$

definiert.

Durch diese Definition erfüllt die Norm o.g. Eigenschaften und stellt in gewisser Weise ein Maß für die Länge des Vektors dar. Es geschieht eine Summation sämtlicher Elemente in der p -ten Potenz mit anschließender Wurzelbildung. Für den Sonderfall $p = 2$ gibt die Norm direkt die geometrische Länge des Vektors x an.

Handelt es sich anstelle eines Vektors um eine skalare Größe $x \in \mathbb{R}$, so heben sich die Umkehrfunktionen Potenzierung und Wurzelbildung auf und die Norm entspricht exakt der Betragsbildung, unabhängig vom Wert p .

Theorem A.36 (Äquivalenz der Vektornorm) [50, S. 59]

Zu jedem gegebenen $p \in [1, \infty]$ und (möglicherweise davon abweichendem) $q \in [1, \infty]$ existieren zwei Konstante $m_1, m_2 > 0$, so dass die Ungleichung

$$m_1 \|x\|_q \leq \|x\|_p \leq m_2 \|x\|_q \quad (\text{A.141})$$

für alle $x \in \mathbb{R}^n$ zutrifft.

Eine Erweiterung für komplexe Vektoren findet sich in [18, S. 14]. Dort ist gezeigt, dass sämtliche p -Normen für $p \in [1, \infty]$ in \mathbb{C}^n äquivalent sind.

Vor allem die drei Normen

$$\begin{aligned} \|x\|_1 &= \sum_i |x_i| && \text{Manhattan-Norm/ Betragssummennorm} \\ \|x\|_2 &= \sqrt{\sum_i |x_i|^2} = \sqrt{x^T x} && \text{Euklidische Norm} \\ \|x\|_\infty &= \max_i |x_i| && \text{Tschebyschew-Norm/ Maximum-Norm} \end{aligned} \quad (\text{A.142})$$

finden in der praktischen Anwendung ihre Verwendung.

A.11.2 Norm für Matrizen

Weil eine Matrix-Vektor-Multiplikation mit der Matrix $A \in \mathbb{R}^{m \times n}$ eine lineare Abbildung vom n -dimensionalen Raum \mathbb{R}^n in den m -dimensionalen Raum \mathbb{R}^m darstellt, kann aus der Norm für Vektoren eine Norm für Matrizen abgeleitet (induziert) werden. Aus einer gegebenen Vektornorm $\|x\|_p$ ergibt sich die induzierte Matrixnorm durch:

Definition A.37 (induzierte Matrixnorm)

Für $p \in \{q \mid q \geq 1\} \cup \{\infty\}$ ist die induzierte Matrixnorm für eine Matrix $A \in \mathbb{R}^{m \times n}$ durch

$$\|A\|_p := \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} = \sup_{\|x\|_p \leq 1} \|Ax\|_p = \sup_{\|x\|_p=1} \|Ax\|_p \quad (\text{A.143})$$

gegeben.

Die induzierte Matrixnorm stellt damit ein Maß für die maximal mögliche Längenänderung eines Vektors $x \in \mathbb{R}^n$ dar, wenn dieser gemäß

$$y = Ax \quad (\text{A.144})$$

in den Vektor $y \in \mathbb{R}^m$ abgebildet wird. Die induzierte Matrix-Norm besitzt einerseits die auf Seite 251 genannten Eigenschaften der zugrunde liegenden Vektornorm als auch die folgenden Eigenschaften:

- (i) $\|Ax\|_p \leq \|A\|_p \cdot \|x\|_p \quad \forall x \in \mathbb{R}^n$
- (ii) $\|AB\|_p \leq \|A\|_p \cdot \|B\|_p$

Häufige Anwendung erfahren wie bei der Vektor-Norm besonders die drei Normen für $p = 1, 2, \infty$:

$$\begin{aligned} \|A\|_1 &= \max_j \sum_i |a_{ij}| && \text{maximale Spaltensumme} \\ \|A\|_2 &= \sqrt{\lambda_m(A^T A)} && \text{Spektralnorm} \\ \|A\|_\infty &= \max_i \sum_j |a_{ij}| && \text{maximale Zeilensumme} \end{aligned} \quad (\text{A.145})$$

Durch $\lambda_m(A^T A)$ wird in der Spektralnorm der maximale Eigenwert der Matrix $A^T A$ symbolisiert.

Auch im Falle der Matrizen-Norm besteht eine Äquivalenz der p -Norm. Es gilt für alle Matrizen $A \in \mathbb{R}^{n \times m}$ [71, S. 267]:

$$\begin{aligned} n^{-\frac{1}{2}} \|A\|_2 &\leq \|A\|_\infty \leq \sqrt{m} \|A\|_2 \\ m^{-\frac{1}{2}} \|A\|_2 &\leq \|A\|_1 \leq \sqrt{n} \|A\|_2 \\ n^{-1} \|A\|_1 &\leq \|A\|_\infty \leq m \|A\|_1 \end{aligned} \quad (\text{A.146})$$

Neben der induzierten Norm existieren weitere Definitionen für die Norm von Matrizen. Eine bekannte Norm ist beispielsweise

$$\|A\| := \max_{ij} |a_{ij}|. \quad (\text{A.147})$$

Hier ist die Norm als der größte Betrag der einzelnen Matrixelemente definiert. Häufig ist auch die Frobeniusnorm zu finden, die für Matrizen mit reellen Elementen

$$\|A\| := \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{\text{Spur}(A^T A)} \quad (\text{A.148})$$

lautet. Obwohl beide Normen die erforderlichen Eigenschaften für eine Norm erfüllen, handelt es sich hierbei nicht um induzierte Normen, da keine zugehörige Vektornorm existiert, von welcher sich diese Normen ableiten ließen.

Diese Tatsache wird für die Frobeniusnorm am Beispiel der Einheitsmatrix I_n offensichtlich. Die Frobeniusnorm lautet $\|I_n\| = \sqrt{n}$, wohingegen Gleichung (A.143) für alle $p \geq 1$ die p -Norm $\|I_n\|_p = \sup_{x \neq 0} \frac{\|I_n x\|_p}{\|x\|_p} = 1$ ergibt. Weil für $n > 1$ eine Diskrepanz zwischen der Frobeniusnorm und jeder p -Norm besteht, ist obige Aussage bestätigt.

Ebenso ist die in Gleichung (A.147) definierte Norm keine induzierte Norm, wie sich am Beispiel der Matrix

$$A = \begin{bmatrix} 0 & 2 \\ 0 & 1 \end{bmatrix} \quad (\text{A.149})$$

zeigen lässt. Zur Bestimmung der p -Norm wird in Gleichung (A.143) einer der beiden Vektoren $x^T = (0, \pm 1)$ eingesetzt. Daraus errechnet sich $\|A\|_p = \sqrt[p]{2^p + 1^p} > 2$ und ist damit für alle $p \geq 1$ strikt größer als die Norm $\|A\| = \max_{ij} |a_{ij}| = 2$.

A.11.3 Norm für Funktionen

Analog zur Vektornorm kann auch bei Signalen eine p -Norm eingeführt werden. Da Vektoren eine beschränkte Anzahl $n \in \mathbb{N}$ von Einzelelementen besitzen, eine Funktion dagegen über einer nicht abzählbaren Menge von Punkten (Kontinuum) definiert ist, muss anstelle der Summation eine Integration durchgeführt werden. Ähnlich wie bei Vektoren sind durch Variation des Parameters $p \geq 1$ unterschiedliche Definitionen für die Norm möglich, wobei für $p = \infty$ wiederum eine gesonderte Definition benötigt wird.

Definition A.38 (p -Norm für Funktionen) [31, S. 739]

Sei $D \subset \mathbb{R}$ ein Intervall und $(X, \|\cdot\|_X)$ ein Banach-Raum. Für $p \in [1, \infty[$ wird einer stückweise stetigen Funktion $x : D \rightarrow X$ die p -Norm

$$\|x(\cdot)\|_p := \left(\int_D \|x(t)\|_X^p dt \right)^{\frac{1}{p}} \quad (\text{A.150})$$

zugewiesen, bzw. für $p = \infty$ die Norm

$$\|x(\cdot)\|_\infty := \text{ess-sup}_{t \in D} \|x(t)\|. \quad (\text{A.151})$$

Falls $x : D \rightarrow \mathbb{R}$ eine skalare Funktion beschreibt, stimmt $\|x(t)\|$ im Integranden mit dem Absolutbetrag $|x(t)|$ überein. Wenn dagegen $x : D \rightarrow \mathbb{R}^n$ mit $n > 1$ eine vektorwertige Funktion symbolisiert, bedeutet $\|x(t)\|$ im Integranden nicht den Absolutbetrag einer skalaren Größe, sondern stellt eine beliebige Vektornorm (z.B. p -Norm) dar.

Es ist zu beachten, dass die Menge aller stückweise stetigen Funktionen $x : D \rightarrow X$ mit endlicher p -Norm zwar einen linearen, jedoch keinen normierten linearen Raum bildet [31, S. 739]. Der Grund hierfür ist, dass die in Definition A.38 angegebene Abbildung die erforderliche Eigenschaft der positiven Definitheit nicht erfüllt und daher lediglich eine Halbnorm darstellt. Aus $\|x(\cdot)\|_p = 0$ folgt nicht $x(t) = 0$ für alle $t \in D$, sondern nur $x(t) = 0$ für fast alle $t \in D$. Daher existieren unendlich viele verschiedene Funktionen $x_i : D \rightarrow X$, $i \in \mathbb{N}$, mit $\|x_i(\cdot)\|_p = 0$.

Um einen normierten Raum zu erhalten, werden deshalb Äquivalenzklassen von Funktionen eingeführt und durch das Symbol $[x(\cdot)]$ gekennzeichnet. In $[x(\cdot)]$ sind sämtliche Funktionen zusammengefasst, die dem Repräsentanten $x(\cdot)$ fast überall gleichen. Falls

$y(\cdot)$ der Äquivalenzklasse $[x(\cdot)]$ angehört, gilt $x(t) = y(t)$ für fast alle $t \in D$, d.h. die Menge $\{t \in D \mid x(t) \neq y(t)\}$ hat das Maß Null. Folglich stimmen die p -Normen $\|x(\cdot)\|_p = \|y(\cdot)\|_p$ überein.

Benannt nach Lebesgue werden Funktionenräume \mathcal{L}^p definiert, deren Elemente nicht die Funktionen selbst, sondern Äquivalenzklassen von Funktionen sind. Die Äquivalenzklassen bilden einen linearen Raum, für den $\|[x(\cdot)]\|_p = \|x(\cdot)\|_p$ eine Norm definiert. Die Zugehörigkeit einer Funktion (oder im Sinne der Regelungstechnik eines Signales) zum Lebesgue-Raum \mathcal{L}^p wird von der Existenz bzw. Beschränktheit der p -Norm des Signales bestimmt.

Definition A.39 (Lebesgue-Raum \mathcal{L}^p)

Der lineare Raum \mathcal{L}^p beinhaltet alle Lebesgue-messbaren Funktionen $x : D \rightarrow Y$ mit endlicher p -Norm

$$\mathcal{L}^p := \{x(\cdot) \mid \|x(\cdot)\|_p < \infty\}, \quad p \in [1, \infty] \quad (\text{A.152})$$

und heißt Lebesgue-Raum.

Dadurch werden diese Räume über das Lebesgue-Integral definiert und erhalten durch die p -Norm eine Metrik. Allgemein umfasst der jeweilige Funktionenraum \mathcal{L}^p für $p < \infty$ alle in p -ter Potenz absolut integrierbaren Funktionen. Es handelt sich dabei um Funktionen, deren Beträge in der p -ten Potenz eine Fläche mit beschränktem Inhalt umschließen. Der Raum \mathcal{L}^∞ dagegen beinhaltet alle essentiell beschränkten Funktionen.

Die p -Norm eines Signales berechnet sich aus der Integration über dessen gesamten Definitionsbereich. Häufig ist in der Literatur mit vorwiegend regelungstechnischem Hintergrund die Annahme gebräuchlich, dass bis zum Zeitpunkt $t = 0$ das vorliegende System im Nullpunkt verharrt und daher $x(t) = 0$ für alle $t < 0$ gilt. In diesem Sonderfall ergibt die Integration über negative Zeitpunkte keinen Beitrag. Daher wird für das zu betrachtende Signal der Definitionsbereich $D = \mathbb{R}_{\geq 0}$ angesetzt. Konsequenterweise wird deshalb z.B. in [46, S. 69] oder [50, S. 262] eine geringfügig abgewandelte Definition der Signalnorm angegeben, die untere Integrationsgrenze beginnt in diesem Fall bei $t = 0$. Dort wird die Norm durch

$$\|x(\cdot)\|_p := \left(\int_0^\infty |x(t)|^p dt \right)^{\frac{1}{p}}, \quad p \in [1, \infty[\quad (\text{A.153})$$

bzw. für $p = \infty$ durch

$$\|x(\cdot)\|_\infty := \sup_{t \in \mathbb{R}_{\geq 0}} |x(t)| \quad (\text{A.154})$$

dargestellt. Ebenfalls weicht [50, S. 262] von der hier angegebenen Definition bei der ∞ -Norm ab, indem dort das Supremum, anstelle des essentiellen Supremums, verwendet wird. Im Falle stetiger Signale besteht dabei jedoch kein Unterschied.

Eine Unterscheidung, ob sich die untere Integrationsgrenze bei $t = -\infty$ oder bei $t = 0$ befindet, wird im Weiteren nicht getroffen und ist kontextabhängig zu verstehen. In beiden Fällen wird vereinfachend stets das Symbol $\|x(\cdot)\|_p$ geschrieben.

Die p -Norm für Signale besitzt die für Normen erforderlichen Eigenschaften. Die Dreiecksungleichung ist auch hier gültig, wird jedoch im Falle von Signal-Normen als „Minkowski-Ungleichung“ bezeichnet.

Besonders die drei Normen

$$\begin{aligned} \|x(\cdot)\|_1 &= \int_0^\infty |x(t)| dt && \text{Betragsfläche} \\ \|x(\cdot)\|_2 &= \sqrt{\int_0^\infty |x(t)|^2 dt} && \text{quadratische Fläche} \\ \|x(\cdot)\|_\infty &= \text{ess-sup}_{t \in \mathbb{R}_{\geq 0}} |x(t)| && \text{essentielles Supremum} \end{aligned} \quad (\text{A.155})$$

sind häufig gebräuchlich, da diesen eine herausgehobene physikalische bzw. systemtheoretische Bedeutung zufällt.

Die Äquivalenz der Norm, die bei der p -Norm für Vektoren und Matrizen gilt, ist nicht ohne Einschränkungen auf die p -Norm für Signale übertragbar, wie das Beispiel der Sinusfunktion $x : t \mapsto \sin(t)$ zeigt. Wegen

$$\int_{-\infty}^{\infty} |\sin(\tau)| d\tau = \infty \quad \text{und} \quad \sqrt{\int_{-\infty}^{\infty} |\sin(\tau)|^2 d\tau} = \infty \quad (\text{A.156})$$

ist die Sinusfunktion weder integrierbar, noch quadratisch integrierbar. Daher existieren die Normen $\|x(\cdot)\|_1$ und $\|x(\cdot)\|_2$ nicht. Dennoch ist die Sinusfunktion beschränkt, die ∞ -Norm

$$\|x(\cdot)\|_\infty = \text{ess-sup}_{t \in \mathbb{R}} |\sin(t)| = 1 \quad (\text{A.157})$$

lässt sich im Gegensatz zur 1- bzw. 2-Norm angeben. Obwohl eine generelle Gleichwertigkeit der Normen nicht besteht, kann für reellwertige, skalare Funktionen dennoch eine Beziehung zwischen verschiedenen p -Normen hergestellt werden.

Hilfssatz A.40 [18, S. 17]

Sofern die reellwertige Funktion $x(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ sowohl integrierbar, als auch beschränkt ist und daher $x \in \mathcal{L}^1 \cap \mathcal{L}^\infty$ erfüllt, gilt:

$$x \in \mathcal{L}^p \quad \forall p \in [1, \infty] \quad (\text{A.158})$$

Beweis

Sei $I \subseteq \mathbb{R}$ eine (nicht notwendigerweise zusammenhängende, möglicherweise auch leere) Teilmenge der reellen Achse mit der Eigenschaft

$$I := \{t \in \mathbb{R} \mid |x(t)| \geq 1\}. \quad (\text{A.159})$$

Bedingt durch die Voraussetzung $x \in \mathcal{L}^1$ ist sichergestellt, dass $|x(\cdot)|$ über der reellen Achse eine Fläche mit endlichem Inhalt umschließt und daher das Lebesgue-Maß des Intervalles I endlich ist. Gleichzeitig ist $x(\cdot)$ eine beschränkte Funktion, so dass auch

$|x(\cdot)|^p$ für alle reellen¹²⁾ $p > 1$ beschränkt ist. Sowohl Integrationsintervall, als auch Integrand in

$$F_I := \int_I |x(t)|^p dt < \infty \quad (\text{A.160})$$

sind beschränkte Größen, weshalb $F_I < \infty$ eine endliche Fläche bezeichnet. Die Definition von I in Gleichung (A.159) stellt sicher, dass $|x(t)| < 1$ für alle $t \in \mathbb{R} \setminus I$ gilt. Folglich lässt sich die Ungleichung

$$F_{\mathbb{R} \setminus I} := \int_{\mathbb{R} \setminus I} |x(t)|^p dt < \int_{\mathbb{R} \setminus I} |x(t)| dt \leq \int_{-\infty}^{\infty} |x(t)| dt < \infty \quad (\text{A.161})$$

angeben. Wegen $\int_{-\infty}^{\infty} |x(t)|^p dt = F_I + F_{\mathbb{R} \setminus I} < \infty$ gilt $x \in \mathcal{L}^p$ für alle $p \in]1, \infty[$. ■

Im Falle absolut stetiger Funktionen $x \in \mathcal{AC}([0, \infty[; \mathbb{R}^n)$ ist folgender Zusammenhang gegeben:

Hilfssatz A.41

Sei $x(\cdot) : [0, \infty[\rightarrow \mathbb{R}^n$ eine absolut stetige Funktion, die für $p \in [1, \infty[$ die Voraussetzung

$$x(\cdot) \in \mathcal{L}^p([0, \infty[; \mathbb{R}^n) \quad (\text{A.162})$$

erfüllt. Falls ihre Ableitung für $q \in [1, \infty[$ der Bedingung

$$\dot{x}(\cdot) \in \mathcal{L}^q([0, \infty[; \mathbb{R}^n) \quad (\text{A.163})$$

genügt, gelten die Aussagen:

$$x(\cdot) \in \mathcal{L}^i([0, \infty[; \mathbb{R}^n) \quad \forall i \in [p, \infty[\quad \text{und} \quad \lim_{t \rightarrow \infty} x(t) = 0. \quad (\text{A.164})$$

Beweis

Siehe Beweis von Lemma 2.1.7 in [35, S. 17]. ■

A.12 Grundlegende Theoreme und Hilfssätze

A.12.1 Zusammenhang zwischen Minimalphasigkeit, Stabilität und Detektierbarkeit

Das LTI-System

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), & x(0) &\in \mathbb{R}^n \\ y(t) &= cx(t), & n \in \mathbb{N}, \quad A &\in \mathbb{R}^{n \times n}, \quad b, c^T \in \mathbb{R}^n \end{aligned} \quad (\text{A.165})$$

¹²⁾ Die Fälle $p = 1$ und $p = \infty$ müssen nicht bewiesen werden, sondern sind per Voraussetzung erfüllt.

heißt nach Definition A.27 minimalphasig, wenn die Aussage

$$\det \begin{bmatrix} sI_n - A & -b \\ c & 0 \end{bmatrix} \neq 0 \quad \forall s \in \mathbb{C} \text{ mit } \operatorname{Re}\{s\} \geq 0 \quad (\text{A.166})$$

zutrifft. Proposition 2.1.2 in [35, S. 10] stellt einen Zusammenhang zwischen Minimalphasigkeit und Stabilisierbarkeit bzw. Detektierbarkeit her.

Theorem A.42

Für das LTI-System (A.165) ist Minimalphasigkeit gleichbedeutend mit den Eigenschaften

(i) Stabilisierbarkeit

$$\operatorname{rang} \begin{bmatrix} sI_n - A & b \end{bmatrix} = n \quad \forall s \in \mathbb{C} \text{ mit } \operatorname{Re}\{s\} \geq 0 \quad (\text{A.167})$$

(ii) Detektierbarkeit

$$\operatorname{rang} \begin{bmatrix} sI_n - A \\ c \end{bmatrix} = n \quad \forall s \in \mathbb{C} \text{ mit } \operatorname{Re}\{s\} \geq 0 \quad (\text{A.168})$$

(iii) Die Übertragungsfunktion hat keine Nullstelle mit positivem Realteil.

Beweis

Siehe [35, S. 10]. ■

A.12.2 Berechnung der Übertragungsfunktion aus der RNF

Theorem A.43

Bei einem Zustandsraummodell

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), & x(0) &\in \mathbb{R}^n \\ y(t) &= cx(t), & n &\in \mathbb{N}, \end{aligned} \quad (\text{A.169})$$

das in RNF vorliegt und deshalb durch die Größen

$$A = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_0 & -a_1 & \dots & -a_{n-1} \end{bmatrix}, \quad b = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \quad \text{und} \quad c = \begin{pmatrix} c_0 \\ \vdots \\ c_{n-2} \\ c_{n-1} \end{pmatrix}^T \quad (\text{A.170})$$

$$a_j, c_j \in \mathbb{R} \quad \forall j = 0, \dots, n-1$$

beschrieben wird, bestimmen ausschließlich die Elemente des Auskoppelvektors $c^T \in \mathbb{R}^n$ die Koeffizienten des Zählerpolynoms $Z(\cdot) \in \mathbb{R}[s]$ in der zugehörigen Übertragungsfunktion.

Beweis

Weil die Übertragungsfunktion durch

$$F(s) = \frac{Z(s)}{N(s)} = \frac{\det \begin{bmatrix} sI - A & -b \\ c & 0 \end{bmatrix}}{\det(sI - A)} \tag{A.171}$$

errechnet werden kann, ist für das Zählerpolynom

$$Z(s) = \det \begin{bmatrix} sI - A & -b \\ c & 0 \end{bmatrix} = \det \left[\begin{array}{cccc|c} s & -1 & & & 0 \\ & & \ddots & & \vdots \\ & & & s & -1 \\ \hline a_0 & a_1 & \dots & s + a_{n-1} & -1 \\ c_0 & c_1 & \dots & c_{n-1} & 0 \end{array} \right] \tag{A.172}$$

zu entnehmen. Die Koeffizienten a_0, \dots, a_{n-1} entfallen, wenn die Determinante entlang der letzten Spalte entwickelt wird, woraus

$$Z(s) = \det \begin{bmatrix} s & -1 & & & \\ & & \ddots & & \\ & & & s & -1 \\ c_0 & c_1 & \dots & c_{n-1} & \end{bmatrix} = c_0 + c_1s + c_2s^2 + \dots + c_{n-2}s^{n-2} + c_{n-1}s^{n-1} \tag{A.173}$$

resultiert. ■

A.12.3 Gronwall-Bellman Ungleichung

Die Gronwall-Bellman Ungleichung wird bevorzugt bei einer Abschätzung der Wachstumsrate für den Ausgang oder die Norm des Zustandsvektors dynamischer Systeme herangezogen und ist zentraler Bestandteil bei der Betrachtung von Eindeutigkeit der Lösungen gewöhnlicher Differentialgleichungen.

Theorem A.44 (Gronwall-Bellman Ungleichung) [50, S. 63f]

Durch die Konstanten $a < b$, $a, b \in \mathbb{R}$ sei ein Kompaktum $D = [a, b] \subseteq \mathbb{R}$ vorgegeben, auf welchem die stetigen Funktionen $y : D \rightarrow \mathbb{R}$, $\lambda : D \rightarrow \mathbb{R}$ und $\mu : D \rightarrow \mathbb{R}_{\geq 0}$ definiert sind. Sofern in D die Ungleichung

$$y(t) \leq \lambda(t) + \int_a^t \mu(\tau)y(\tau)d\tau \quad \forall t \in D \tag{A.174}$$

erfüllt ist, gilt die Aussage

$$y(t) \leq \lambda(t) + \int_a^t \exp \left\{ \int_\tau^t \mu(s)ds \right\} \lambda(\tau)\mu(\tau)d\tau \quad \forall t \in D. \tag{A.175}$$

Falls λ eine Konstante ist, lässt sich die Abschätzung

$$y(t) \leq \lambda \exp \left\{ \int_a^t \mu(\tau) d\tau \right\} \quad (\text{A.176})$$

angeben. Ist zusätzlich μ eine Konstante, ergibt sich daraus durch Integration direkt

$$y(t) \leq \lambda e^{\mu(t-a)} \quad \forall t \in D. \quad (\text{A.177})$$

Beweis

Siehe [50, S. 63f]. ■

A.12.4 Geršgorin-Theorem

Theorem A.45 (Geršgorin) [55, S. 371ff]

Sei eine quadratische Matrix $A \in \mathbb{C}^{n \times n}$, $n \in \mathbb{N}$, gegeben durch die Elemente $a_{jk} \in \mathbb{C}$ mit $j, k = 1, 2, \dots, n$. Bezeichne

$$\rho_j = \sum_{k=1, k \neq j}^n |a_{jk}| \quad (\text{A.178})$$

den Geršgorin-Radius, der sich aus der Zeilensumme der Beträge berechnet, reduziert um das jeweilige Diagonalelement a_{jj} . Dann liegt jeder Eigenwert der Matrix A in mindestens einer der Kreisscheiben (Geršgorin-Scheiben)

$$\mathcal{B}_{\rho_j}(a_{jj}) := \{s \in \mathbb{C} \mid |a_{jj} - s| \leq \rho_j\}, \quad j = 1, 2, \dots, n \quad (\text{A.179})$$

in der komplexen s -Ebene. Eine Menge von m Kreisscheiben, die disjunkt zu den verbleibenden $n - m$ Scheiben ist, enthält genau m Eigenwerte der Matrix A .

Beweis

Siehe [55, S. 371ff]. ■

Anmerkung

Aus dem Geršgorin-Theorem folgt nicht notwendigerweise, dass in jeder Scheibe ein Eigenwert liegen muss. Ausschließlich Scheiben, die zu allen restlichen Scheiben disjunkt sind, enthalten genau einen Eigenwert. Bei einer Überlappung kann ggf. eine Geršgorin-Scheibe ein Gebiet beschreiben, in welchem sich kein Eigenwert befindet, wie folgendes Beispiel illustriert. Für die Matrix

$$A = \begin{bmatrix} 4.5 & 1 \\ 3 & 4 \end{bmatrix} \quad \Rightarrow \quad \sigma(A) = \{2.5, 6.0\} \quad (\text{A.180})$$

errechnen sich die Radien $\rho_1 = |a_{12}| = 1$ und $\rho_2 = |a_{21}| = 3$, die Mittelpunkte sind durch $M_1 = a_{11} = 4.5$ und $M_2 = a_{22} = 4$ gegeben. Wie die Abbildung A.9 zeigt, liegt keiner der beiden Eigenwerte in der Scheibe $\mathcal{B}_{\rho_1}(a_{11})$. □

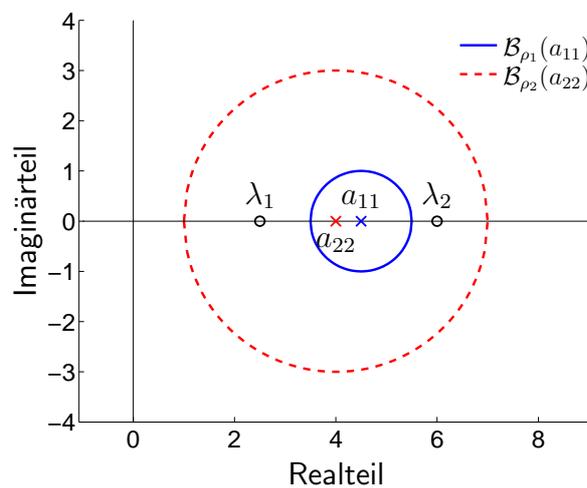


Abbildung A.9: Geršgorin-Scheiben und Eigenwerte für die Matrix A in der komplexen Ebene.

Literaturverzeichnis

- [1] ACKERMANN, J.: *Robuste Regelung – Analyse und Entwurf von linearen Regelungssystemen mit unsicheren physikalischen Parametern*. Springer–Verlag, Berlin, 1993.
- [2] AMANN, H., ESCHER, J.: *Analysis I*. Birkhäuser Verlag, Basel, 1998.
- [3] AMANN, H., ESCHER, J.: *Analysis II*. Birkhäuser Verlag, Basel, 1999.
- [4] AMANN, H., ESCHER, J.: *Analysis III*. Birkhäuser Verlag, Basel, 2001.
- [5] ANGERER, B., HINTZ, C., SCHRÖDER, D.: *Online identification of a nonlinear mechatronic system*. Control Engineering Practice, Vol. 12, No. 11, pp. 1465-1478, 2004.
- [6] APOSTOL, T.: *Mathematical Analysis – A Modern Approach to Advanced Calculus*. Addison–Wesley, Reading MA, 5ter Nachdruck, 1971.
- [7] ARMSTRONG-HÉLOUVRY, B., DUPONT, P., CANUDAS DE WIT, C.: *A Survey of Models, Analysis Tools and Compensation Methods for the Control of Machines with Friction*. Automatica, Vol. 30, No. 7, pp. 1083-1138, 1994.
- [8] ÅSTRÖM, K., WITTENMARK, B.: *Adaptive Control*. Addison–Wesley, Reading MA, 2nd ed, 1995.
- [9] ATASSI, A., KHALIL, H.: *A Separation Principle for the Stabilization of a Class of Nonlinear Systems*. IEEE Trans. on Automatic Control, Vol. 44, No. 9, pp. 1672-1687, 1999.
- [10] BLEISTEINER, G., v. MANGOLDT, W. (HRSG.): *Handbuch der Regelungstechnik*. Springer–Verlag, Berlin, 1961.
- [11] BODE, H.: *Network Analysis and Feedback Amplifier Design*. D. Van Nostrand Company, Inc., Toronto, 1952.
- [12] BOGACHEV, V., MAYER-WOLF, E.: *Some Remarks on Rademacher’s Theorem in Infinite Dimensions*. Potential Analysis, Vol. 5, No. 1, pp. 23-30, 1996.
- [13] BROCKHAUS, R.: *Flugregelung*. Springer–Verlag, Berlin, 2001.
- [14] BULLINGER, E.: *Adaptive λ -tracking for Systems with Higher Relative Degree*. Dissertation am Institut für Automatik, ETH Zürich No. 13858, 2001.

- [15] BYRNES, C., ISIDORI, A.: *Asymptotic Stabilization of Minimum Phase Nonlinear Systems*. IEEE Trans. on Automatic Control, Vol. 36, No. 10, pp. 1122-1137, 1991.
- [16] CANUDAS DE WIT, C., OLSSON, H., ÅSTRÖM, K., LISCHINSKY, P.: *A New Model for Control of Systems with Friction*. IEEE Trans. on Automatic Control, Vol. 40, No. 3, pp. 419-425, 1995.
- [17] DABROOM, A., KHALIL, H.: *Discrete-time implementation of high-gain observers for numerical differentiation*. Int. Journal of Control, Vol. 72, No. 17, pp. 1523-1537, 1999.
- [18] DESOER, C., VIDYASAGAR, M.: *Feedback Systems: Input-Output Properties*. Academic Press, New York, 1975.
- [19] DRĂGAN, V., HALANAY, A.: *Stabilization of Linear Systems*. Birkhäuser, Boston, 1999.
- [20] FEILER, M.: *Adaptive Control in the Presence of Disturbances*. <http://tumb1.biblio.tu-muenchen.de/publ/diss/ei/2004/feiler.pdf>, Dissertation am Lehrstuhl für Elektrische Antriebssysteme, TU München, 2004.
- [21] FEILER, M., WESTERMAIER, C., SCHRÖDER, D.: *Adaptive Speed Control of a Two-Mass System*. Proceedings of the 2003 IEEE Int. Conf. on Control Applications CCA, Istanbul, Türkei, pp. 1112-1117, Jun. 2003.
- [22] FÖLLINGER, O.: *Regelungstechnik: Einführung in die Methoden und ihre Anwendung*. Hüthig, Heidelberg, 1985.
- [23] FÖLLINGER, O.: *Regelungstechnik: Einführung in die Methoden und ihre Anwendung*. Hüthig, Heidelberg, 1994.
- [24] FRANCIS, B., WONHAM, W.: *The Internal Model Principle for Linear Multivariable Regulators*. Jour. Applied Mathematics & Optimization, Vol. 2, No. 2, pp. 170-194, 1975.
- [25] FRANCIS, B., WONHAM, W.: *The Internal Model Principle of Control Theory*. Automatica, Vol. 12, No. 5, pp. 457-465, 1976.
- [26] FUCHS, A., VARGHESE, M., MUKUNDAN, R.: *On Weak Relative Degree and the McGehee Transformation*. IEEE Trans. on Automatic Control, Vol. 38, No. 7, pp. 1147-1150, 1993.
- [27] GOLUB, G., VAN LOAN, C.: *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Maryland, 1983.
- [28] HACKL, C., SCHUSTER, H., WESTERMAIER, C., SCHRÖDER, D.: *Funnel-Control with Integrating Prefilter for Nonlinear, Time-varying Two-Mass Flexible Servo Systems*. Proc. of the 9th IEEE Int. Workshop on Advanced Motion Control, AMC, Istanbul, Türkei, pp. 456-461, Mrz. 2006.

- [29] HAHN, W.: *Stability of Motion*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Band 138, Springer-Verlag, Berlin, 1967.
- [30] HENRICHFREISE, H., WITTE, C.: *Beobachtergestützte nichtlineare Kompensation trockener Reibung in einem Positionierantrieb*. Automatisierungstechnik, Vol. 46, No. 3, pp. 128-135, 1998.
- [31] HINRICHSEN, D., PRITCHARD, A.J.: *Mathematical Systems Theory I – Modelling, State Space Analysis, Stability, and Robustness*. Springer-Verlag, Berlin, 2005.
- [32] HINTZ, C.: *Identifikation nichtlinearer mechatronischer Systeme mit strukturierten rekurrenten Netzen*. <http://tumb1.biblio.tu-muenchen.de/publ/diss/ei/2003/hintz.pdf>, Dissertation am Lehrstuhl für Elektrische Antriebssysteme, TU München, 2003.
- [33] HOAGG, J., BERNSTEIN, D.: *Nonminimum-Phase Zeros: Much to Do about Nothing*. IEEE Control Systems Magazine, Vol. 27, No. 3, pp. 45-57, 2007.
- [34] ILCHMANN, A.: *Non-Identifier-Based Adaptive Control of Dynamical Systems: A Survey*. IMA Journal of Mathematical Control & Information, Vol. 8, pp. 321-366, 1991.
- [35] ILCHMANN, A.: *Non-Identifier-Based High-Gain Adaptive Control*. Lecture Notes in Control and Information Sciences 189, Springer-Verlag, London, 1993.
- [36] ILCHMANN, A., OWENS, D., PRÄTZEL-WOLTERS, D.: *High-gain robust adaptive controllers for multivariable systems*. Systems & Control Letters, Vol. 8, No. 5, pp. 397-404, 1987.
- [37] ILCHMANN, A., RYAN, E.: *Universal λ -Tracking for Nonlinearly-Perturbed Systems in the Presence of Noise*. Automatica, Vol. 30, No. 2, pp. 337-346, 1994.
- [38] ILCHMANN, A., RYAN, E.: *Asymptotic tracking with prescribed transient behaviour for linear systems*. Int. Journal of Control, Vol. 79, No. 8, pp. 910-917, 2006.
- [39] ILCHMANN, A., RYAN, E.: *High-gain Control Without Identification: A Survey*. GAMM-Mitt., Vol. 31, No. 1, pp. 115-125, 2008.
- [40] ILCHMANN, A., RYAN, E.: *Performance funnels and tracking control*. zur Veröffentlichung eingereicht bei International Journal of Control, 2009.
- [41] ILCHMANN, A., RYAN, E., SANGWIN, C.: *Tracking with prescribed transient behaviour*. ESAIM: Control, Optimisation and Calculus of Variations, Vol. 7, pp. 471-493, 2002.
- [42] ILCHMANN, A., RYAN, E., TOWNSEND, P.: *Tracking control with prescribed transient behaviour for systems of known relative degree*. Systems & Control Letters, Vol. 55, No. 5, pp. 396-406, 2006.

- [43] ILCHMANN, A., RYAN, E., TOWNSEND, P.: *Tracking With Prescribed Transient Behaviour For Nonlinear Systems Of Known Relative Degree*. SIAM Journal on Control & Optimization, Vol. 46, No. 1, pp. 210-230, 2007.
- [44] ILCHMANN, A., SCHUSTER, H.: *PI-funnel control for two mass systems*. zur Veröffentlichung eingereicht bei IEEE Trans. on Automatic Control, 2008.
- [45] IOANNOU, P., KOKOTOVIĆ, P.: *Instability Analysis and Improvement of Robustness of Adaptive Control*. Automatica, Vol. 20, No. 5, pp. 583-594, 1984.
- [46] IOANNOU, P., SUN, J.: *Robust Adaptive Control*. Prentice-Hall, Upper Saddle River, NJ, 1996.
- [47] ISIDORI, A.: *Nonlinear Control Systems*. Springer-Verlag, London, 2001.
- [48] KAILATH, T.: *Linear Systems*. Prentice-Hall, Englewood Cliffs, 1980.
- [49] KAUFMAN, H., BAR-KANA, I., SOBEL, K.: *Direct Adaptive Control Algorithms: Theory and Applications*. Springer-Verlag, New York, 1994.
- [50] KHALIL, H.: *Nonlinear Systems*. Prentice-Hall, Upper Saddle River, New Jersey, 1996.
- [51] KHALIL, H.: *High-Gain Observers in Nonlinear Feedback Control*. in NIJMEIJER, H., FOSSEN, T., (EDS.): *New Directions in Nonlinear Observer Design*. Lecture Notes in Control and Information Sciences 244, Springer-Verlag, London, 1999.
- [52] KHARITONOV, V.: *Asymptotic stability of an equilibrium position of a family of systems of linear differential equations*. Differential Equations, Vol. 14, No. 3, pp. 26-35, 1979.
- [53] KÖNIGSBERGER, K.: *Analysis 1*. Springer-Verlag, Berlin, 1995.
- [54] KWAKERNAAK, H., SIVAN, R.: *Modern Signals and Systems*. Prentice Hall, Englewood Cliffs, 1991.
- [55] LANCASTER, P., TISMENETSKY, M.: *The Theory of Matrices*. Academic Press, Orlando, 1985.
- [56] LIBERZON, D., MORSE, A., SONTAG, E.: *Output-Input Stability and Minimum-Phase Nonlinear Systems*. IEEE Trans. on Automatic Control, Vol. 47, No. 3, pp. 422-436, 2002.
- [57] LUDYK, G.: *Theoretische Regelungstechnik 1: Grundlagen, Synthese linearer Regelungssysteme*. Springer-Verlag, Berlin, 1995.
- [58] LUDYK, G.: *Theoretische Regelungstechnik 2: Zustandsrekonstruktion, optimale und nichtlineare Regelungssysteme*. Springer-Verlag, Berlin, 1995.
- [59] LUNZE, J.: *Regelungstechnik 1: Systemtheoretische Grundlagen, Analyse und Entwurf einschleifiger Regelungen*. Springer-Verlag, Berlin, 2004.

- [60] LUNZE, J.: *Regelungstechnik 2: Mehrgrößensysteme, Digitale Regelung*. Springer-Verlag, Berlin, 2005.
- [61] LUTZ, H., WENDT, W.: *Taschenbuch der Regelungstechnik*. Verlag Harri Deutsch, Frankfurt a. Main, 2002.
- [62] MACIEJOWSKI, J.: *The changing face and role of CACSD*. Proceedings of the 2006 IEEE Conf. on Computer Aided Control Systems Design, CACSD, München, Deutschland, pp. 1-7, Okt. 2006.
- [63] MAREELS, I., VAN GILS, S., POLDERMAN, J., ILCHMANN, A.: *Asymptotic Dynamics in Adaptive Gain Control*. in Frank, P. (Hrsg.): *Advances in Control, Highlights of the 5th European Control Conference ECC*, Karlsruhe, pp. 29-63, Springer-Verlag, London, 1999.
- [64] MEYBERG, K., VACHENAUER, P.: *Höhere Mathematik 1: Differential- und Integralrechnung, Vektor- und Matrizenrechnung*. Springer-Verlag, Berlin, 1990.
- [65] MEYBERG, K., VACHENAUER, P.: *Höhere Mathematik 2: Differentialgleichungen, Funktionentheorie, Fourier-Analyse und Variationsrechnung*. Springer-Verlag, Berlin, 1991.
- [66] MILLER, D., DAVISON, E.: *An Adaptive Controller Which Provides an Arbitrarily Good Transient and Steady-State Response*. IEEE Trans. on Automatic Control, Vol. 36, No. 1, pp. 68-81, 1991.
- [67] MORSE, S.: *A Three-Dimensional Universal Controller for the Adaptive Stabilization of Any Strictly Proper Minimum-Phase System with Relative Degree Not Exceeding Two*. IEEE Trans. on Automatic Control, Vol. 30, No. 12, pp. 1180-1191, 1985.
- [68] MORSE, S.: *Simple Algorithms For Adaptive Stabilization*. in Byrnes, C., Kurzhanski, A. (Hrsg.): *Lecture Notes in Control and Information Sciences*, No. 105, pp.254-264, Springer-Verlag, Berlin, 1988.
- [69] NARENDRA, K., ANNASWAMY, A.: *Stable Adaptive Systems*. Prentice-Hall International Limited, London, 1989.
- [70] NIJMEIJER, H., VAN DER SCHAFT, A.: *Nonlinear Dynamical Control Systems*. Springer-Verlag, New York, 1996.
- [71] NOBLE, B., DANIEL, J.: *Applied linear algebra*. Prentice-Hall, Englewood Cliffs, New Jersey, 1988.
- [72] NUSSBAUM, R.: *Some remarks on a conjecture in parameter adaptive control*. Systems & Control Letters, Vol. 3, No. 5, pp. 243-246, 1983.
- [73] OPPELT, W.: *Kleines Handbuch technischer Regelvorgänge*. Verlag Chemie GmbH, Weinheim/Bergstr., 1972.

- [74] POLDERMAN, J., MAREELS, I.: *High gain adaptive control revisited: first and second order case*. Proceedings of the 38th Conference on Decision & Control, Phoenix, Arizona, Dez. 1999.
- [75] POLDERMAN, J., MAREELS, I.: *Two scale high gain adaptive control*. International Journal of Adaptive Control and Signal Processing, Vol. 18, No. 4, pp. 393-402, 2004.
- [76] POLZER, G., MEISSNER, F., FLEISCHER, G. (HRSG): *Grundlagen zu Reibung und Verschleiß*. VEB Deutscher Verlag für Grundstoffindustrie, Leipzig, 1979.
- [77] RADEMACHER, H.: *Über partielle und totale Differenzierbarkeit von Funktionen mehrerer Variablen und über die Transformation der Doppelintegrale*. Mathematische Annalen, Vol. 79, No. 4, pp. 340-359, 1919.
- [78] RAU, M.: *Nichtlineare modellbasierte prädiktive Regelung auf Basis lernfähiger Zustandsraummodelle*.
<http://mediatum2.ub.tum.de/doc/601571/document.pdf>,
 Dissertation am Lehrstuhl für Elektrische Antriebssysteme, TU München, 2003.
- [79] REINSCHKE, K.: *Lineare Regelungs- und Steuerungstheorie*. Springer-Verlag, Berlin, 2006.
- [80] ROHRS, C., VALAVANI, L., ATHANS, M., STEIN, G.: *Robustness of Continuous-Time Adaptive Control Algorithms in the Presence of Unmodeled Dynamics*. IEEE Trans. on Automatic Control, Vol. 30, No. 9, pp. 881-889, 1985.
- [81] SABERI, A., LIN, Z.: *Simultaneous Stabilization of a Family of SISO Nonlinear Systems via Output Feedback*. Proceedings of the American Control Conference, ACC, Baltimore, Maryland USA, pp. 952-953, Jun. 1994.
- [82] SASTRY, S.: *Nonlinear Systems: Analysis, Stability, and Control*. Springer-Verlag, New York, 1999.
- [83] SCHMIDT, G.: *Grundlagen der Regelungstechnik – Analyse und Entwurf linearer und einfacher nichtlinearer Regelungen sowie diskreter Steuerungen*. Springer-Verlag, Berlin, 1991.
- [84] SCHRÖDER, D. (HRSG.): *Intelligent Observer and Control Design for Nonlinear Systems*. Springer-Verlag, Berlin, 2000.
- [85] SCHRÖDER, D.: *Elektrische Antriebe – Regelung von Antriebssystemen*. Springer-Verlag, Berlin, 2001.
- [86] SCHRÖDER, D., SCHUSTER, H., WESTERMAIER, C.: *Mechatronic – Advanced Computational Intelligence*. Proceedings of the 7th IEEE Int. Conf. on Power Electronics and Drive Systems, PEDS, Bangkok, Thailand, pp. 994-1001, Nov. 2007.

- [87] SCHUSTER, H., WESTERMAIER, C., SCHRÖDER, D.: *High-Gain Control of Systems with Arbitrary Relative Degree: Speed Control for a Two Mass Flexible Servo System*. Proceedings of the 8th IEEE Int. Conf. on Intelligent Engineering Systems, INES, Cluj-Napoca, Rumänien, pp. 486-491, Sept. 2004.
- [88] SCHUSTER, H., WESTERMAIER, C., SCHRÖDER, D.: *High-Gain-Control for Systems with Zero Dynamics and Arbitrary Relative Degree*. Proceedings of the IEEE Int. Conf. on Mechatronics & Robotics, MechRob, Aachen, Deutschland, pp. 426-431, Sept. 2004.
- [89] SCHUSTER, H., WESTERMAIER, C., SCHRÖDER, D.: *Non-Identifier-Based Adaptive Tracking Control for a Two-Mass System*. Proceedings of the 10th IEEE Int. Conf. on Intelligent Engineering Systems, INES, London, Großbritannien, pp. 190-195, Jun. 2006.
- [90] SCHUSTER, H., WESTERMAIER, C., SCHRÖDER, D.: *Non-Identifier-Based Adaptive Speed Control for a Two-Mass Flexible Servo System: Consideration of Stability and Steady State Accuracy*. Proceedings of the 14th Mediterranean Conf. on Control and Automation, MED, Ancona, Italien, Jun. 2006.
- [91] SCHUSTER, H., WESTERMAIER, C., SCHRÖDER, D.: *Non-Identifier-Based Adaptive Control for a Mechatronic System Achieving Stability and Steady State Accuracy*. Proceedings of the 2006 IEEE International Conference on Control Applications, CCA, München, Deutschland, pp. 1819-1824, Okt. 2006.
- [92] SCHUSTER, H., WESTERMAIER, C., SCHRÖDER, D.: *Mechatronics and Intelligent Control*. Proceedings of the 2007 Korea-Germany Joint Symposium on Power Electronics, Pohang, Korea, pp. 1-6, Okt. 2007.
- [93] SCHWARZ, H.: *Nichtlineare Regelungssysteme: Systemtheoretische Grundlagen*. R. Oldenbourg Verlag, München, 1991.
- [94] SLOTINE, J., LI, W.: *Applied Nonlinear Control*. Prentice-Hall, Englewood Cliffs, New Jersey, 1991.
- [95] SONTAG, E.: *Mathematical Control Theory – Deterministic Finite Dimensional Systems*. Springer-Verlag, New York, 1998.
- [96] SONTAG, E.: *Adaptation and regulation with signal detection implies internal model*. Systems & Control Letters, Vol. 50, No. 2, pp. 119-126, 2003.
- [97] SVARICEK, F.: *Nullodynamik linearer und nichtlinearer Systeme: Definitionen, Eigenschaften und Anwendungen*. Automatisierungstechnik, Vol. 54, No. 7, pp. 310-322, 2006.
- [98] WILLEMS, J., BYRNES, C.: *Global Adaptive Stabilization In The Absence Of Information On The Sign Of The High Frequency Gain*. in Balakrishnan, A., Thoma, M. (Hrsg.): *Lecture Notes in Control and Information Sciences*, No. 62, pp. 49-57, Springer-Verlag, Berlin, 1984.

- [99] WONHAM, W.: *Towards an abstract internal model principle*. IEEE Trans. on Systems, Man & Cybernetics, Vol. 6, No. 11, pp. 735-740, 1976.
- [100] WONHAM, W.: *Linear Multivariable Control: a Geometric Approach*. Springer-Verlag, New York, 1979.