

12. Steirisches Seminar über Regelungstechnik und Prozeßautomatisierung

Tagungsband

Schloß Retzhof
Bildungshaus des Landes Steiermark

24. bis 27. September 2001

Herausgeber und Verleger:
Institut für Regelungstechnik
Technische Universität Graz
Inffeldgasse 16 c, A-8010 Graz
Tel.: +43 316 873 7021
Fax: +43 316 873 7028

ISBN 3-901439-04-8

Vorwort

Das Steirische Seminar über Regelungstechnik und Prozeßautomatisierung findet alle zwei Jahre im Schloß Retzhof, dem Bildungshaus des Landes Steiermark in der Nähe der südsteirischen Stadt Leibnitz, statt. Bei dieser Veranstaltung, die vom Institut für Regelungstechnik der Technischen Universität Graz durchgeführt wird, werden grundlegende Probleme der Regelungs- bzw. Automatisierungstechnik in einem fruchtbringenden Dialog zwischen Angehörigen der Universitäten und der Industrie erörtert. Dabei besteht für die Vortragenden die Möglichkeit, ihre Beiträge im Rahmen eines Tagungsbandes zu veröffentlichen. Den Autoren sei an dieser Stelle für die Sorgfalt bei der Erstellung ihrer Beiträge gedankt.

Graz, im September 2001

Inhaltsverzeichnis

Ein Beitrag zur algebraischen Riccati-Gleichung D. Dreyer (Berlin)	1
Stabilität von Regelkreisen in Feldbussystemen R. Tracht, Th. Eymann (Essen)	19
Beobachtung nichtlinearer Deskriptorsysteme J. Menke, F. Gausch (Paderborn)	29
Über das Filtern verrauschter Daten in hochgenauen Bondmaschinen H. Pristauz, M. Loinger (Radfeld)	48
Iso-Lyapunov Dynamics – a Challenge A. Weinmann (Wien)	61
Linearisierung von Lautsprechern mit AC-Verstärkern D. Fränken (Paderborn)	97
Die philosophische Vorgeschichte des geschlossenen Regelkreises D. Kalligeropoulos (Piraeus)	108
The Platonic Dichotomy and the Euclidean Anthyphairetic Method – The First Closed Loop Mathematical Algorithms S. Vasileiadou (London)	118
Automatische Steuerung der Ultrafiltrationsrate bei Dialysepatienten R. Pilgram, D. Schneditz (Graz)	130
Servicerobotik – Umgebungserfassung und Bewegungssteuerung für einen mobilen Manipulator T. Wösch, W. Neubauer (München)	141
Robuste Regelung eines parameterabhängigen Übertragungssystems S. Doczy (Graz)	155

Ein Beitrag zur algebraischen Riccati-Gleichung. Hier: Die einfachsten Aussagen zur Lösbarkeit der CARE mit semidefiniten Termen

Dietrich Dreyer
Einsteinufer 71B, D-10587 Berlin

Zusammenfassung

Hier geht es im wesentlichen um die Existenz von Lösungen der kontinuierlichen algebraischen Riccati-Gleichungen (CARE) $A^*X + XA + XRX + Q = 0$ und $A^*X + XA \mp XBB^*X + C^*C = 0$ mit $R = \mp BB^*$, $Q = C^*C$ (R und Q sind semidefinite Terme), und zwar um die Existenz von hermiteschen ($X^* = X$) und stabilisierenden Lösungen ($A + RX$ bzw. $A \mp BB^*X$ stabil) dieser speziellen quadratischen Matrizen-Gleichungen.¹ Daneben geht es um die Existenz von positiv definiten und stabilisierenden Lösungen der algebraischen Riccati-Ungleichung $A^*X + XA + XBB^*X + C^*C < 0$ und das Lemma von Kalman–Yakubovich–Popov.

Ausgehend von dem trivialen Fall im Eindimensionalen werden für die Lösungen dieser Gleichungen/Ungleichungen einfache Existenzsätze formuliert. Hierbei wird im Zusammenhang mit der Gleichung $A^*X + XA + XBB^*X + C^*C = 0$ und der Ungleichung $A^*X + XA + XBB^*X + C^*C < 0$ die Dissipativität des Matrizen ins Spiel gebracht, und es wird gezeigt, daß sich gewisse Dissipativitätsaussagen nahtlos in die bisher bekannten Aussagen über diese Gleichung und Ungleichung einordnen lassen. Wesentliche Teile dieser Betrachtung sind neu — so glaube ich wenigstens. Die Beweise zu allen diesen Aussagen werden hier jedoch nicht angegeben².

Abschließend wird darauf eingegangen, wo und wie die Aussagen über die algebraische Riccati-Gleichungen in der deutschsprachigen Lehrbuchliteratur ihre Anwendung finden, und es wird angemerkt, daß die zum Beweis der Aussagen dieser Arbeit bisher verwendete Technik vermutlich nicht dafür verwendet werden kann, um den Beweis für das Lemma von Kalman–Yakubovich–Popov auch in allgemeinen B^* - bzw. C^* -Algebren auszuführen.

1 Einführung

Es seien $A, B, C, R, Q \in \mathbb{C}^{n \times n}$ mit $R^* = R$ oder $R = \mp BB^*$, $Q^* = Q$ oder $Q = C^*C$ und

$$\mathcal{R}(X) := A^*X + XA + XRX + Q, \quad X \in \mathbb{C}^{n \times n} \quad (1.1)$$

für ein $n \in \mathbb{N}$. In jedem der Fälle heißt der Operator $\mathcal{R}$ *Riccati-Operator*, die Gleichung $\mathcal{R}(X) = 0$ *algebraische Riccati-Gleichung* und die Ungleichung $\mathcal{R}(X) < 0$ *algebraische Riccati-Ungleichung*. Wir werden nur für die in den Anwendungen hauptsächlich eine Rolle spielenden Lösungen X dieser Riccati-(Un)Gleichung(en) interessieren, die *hermitisch* und *stabilisierend* sind, d.h. für die gilt: $X^* = X$ und $A + QX$ bzw. $A \mp BB^*X$ stabil.

Die Existenz derartiger Lösungen der algebraischen Riccati-(Un)Gleichung(en) spielt heute in vielen Bereichen eine wesentliche Rolle, beispielsweise

- in der Kontrolltheorie, zum einen bei der linearen quadratisch optimalen Regelung und bei der linearen stochastische Filterung, heute zusammen oft H_2 -(optimal-)control genannt, und zum anderen bei der Stabilitätsuntersuchung und beim Entwurf robuster Regelungen, heute zusammen oft H_∞ -(optimal-)control genannt, siehe z.B. bei DULLERUD und PAGANINI in [6], bei IONESCU, OARĂ und WEISS in [11], bei LUDYK in [19], bei MÜLLER in [22] oder bei RAISCH in [24].³

¹ Für eine Matrix $M \in \mathbb{C}^{n \times n}$ bezeichnet M^* die Konjugiert-Transponierte dieser Matrix M .

² Die ausführlichen Beweise sind in dem gesonderten Umdruck [5] zusammengestellt, der beim Autor zu erhalten ist.

³ Als Referenz gebe ich nur Lehrbücher an, keine Originalarbeiten (außer der „historischen“ Arbeit [7]), da ich bei der Aufarbeitung der in den letzten 20 Jahren eingestandenen Literatur zur Kontrolltheorie noch nicht weiter vorgedrungen bin.

- in den Wirtschaftswissenschaften bei der Theorie der dynamischen Spiele, siehe z.B. bei BAŞAR in [1] und
- in der Mathematik bei der Faktorisierung von Polynommatrizen und bei der Untersuchung der Konvergenz von Iterationsverfahren, siehe z.B. bei MARTIN und AMMAR in [21].

In allen diesen Theorien ist die Existenz einer hermiteschen und stabilisierenden Lösung einer Riccati-(Un)Gleichung zu einer innerhalb dieser Theorie ganz wesentlichen Aussage äquivalent, man denke beispielsweise an die Existenz und Eindeutigkeit der Lösung des Problems der linearen quadratisch optimalen Regelung bei linearen zeitinvarianten kontinuierlichen Systemen der Form $\dot{x} = Ax + Bu$, $y = Cx$. Und zumeist werden diese Äquivalenzen dazu benutzt, um von der Existenz einer Lösung der Riccati-(Un)Gleichung auf die Gültigkeit einer wesentlichen Aussage innerhalb der betreffenden Theorie zu schließen. Umgekehrt kann über diese Äquivalenzen aber natürlich auch die Existenz einer Lösung der Riccati-(Un)Gleichung gesichert werden. So habe ich beispielsweise in [4] auf diesem Wege, über das Problem der linearen quadratisch optimalen Regelung, die Existenz einer positiv definiten Lösung der algebraischen Riccati-Ungleichung $A^*X + XA + XBB^*X + C^*C < 0$ gesichert, und zwar für beschränkte lineare Operatoren auf Hilberträumen anstelle der $n \times n$ Matrizen, und diese positiv definite Lösung dann innerhalb der Stabilitätstheorie von Differentialgleichungen weiter verwendet. Und bis vor ungefähr 20 Jahren gab es nur dieses Verfahren, oder das Verfahren mittels der Faktorisierung von Polynommatrizen, zum Nachweis der Existenz von Lösungen dieser algebraischen Riccati-(Un)Gleichung $A^*X + XA + XBB^*X + C^*C(<) = 0$, siehe z.B. bei POPOV in [23].

In den letzten 30 Jahren — zunächst nur für die Gleichung $A^*X + XA - XBB^*X + C^*C = 0$ und dann später, so innerhalb der letzten 20 Jahre, auch für die Gleichung $A^*X + XA + XBB^*X + C^*C = 0$ — ist ein weiteres Verfahren zur Sicherung von Lösungen der algebraischen Riccati-Gleichungen hinzugekommen und ausgebaut worden, das sogenannte „Eigenvektor-Verfahren“, in welchem im wesentlichen nur mit Hilfsmitteln aus der Linearen Algebra gearbeitet wird, siehe z.B. bei LAUBER in [16], bei LANCASTER und RODMAN in [14] und in [15] und bei DULLERUD und PAGANINI in [6].

Und im Rahmen der Durchsicht der Beweise der Sätze zu diesem heute allgemein verwendeten Eigenvektor-Verfahren zur Sicherung von Lösungen der algebraischen Riccati-Gleichungen, konnte ich einen wesentlichen Beweisschritt, und zwar einen wesentlichen Beweisschritt zum Nachweis der Existenz einer Lösung der speziellen algebraischen Riccati-Gleichung $A^*X + XA + XBB^*X + C^*C = 0$ mit gleichartig semidefiniten Termen, ganz deutlich vereinfachen (so glaube ich wenigstens), siehe [5].

Im Abschnitt 6 wird dann eine Eigenschaft für Matrizen ins Spiel gebracht, die ich in den Arbeiten [3] und [4] bereits benutzt habe, und die hier wieder *dissipativ* genannt werden soll. Hierzu folgendes: Liegt der numerische Bereich $\nu(M) := \{x^*Mx \mid x \in \mathbb{C}^n, x = 1\}$ einer Matrix $M \in \mathbb{C}^{n \times n}$ innerhalb der linken offenen komplexen Halbebene so hat dies unter Umständen starke Auswirkungen auf das Verhalten der Gleichungssysteme, in denen die Matrix M auftritt. Und in dieser Situation, wenn also der numerische Bereich $\nu(M)$ einer Matrix M innerhalb der linken offenen komplexen Halbebene liegt, werde ich die Matrix M einfach *dissipativ* nennen, so daß beispielsweise folgendes gilt:

- Ist die Matrix $A(t)$ für alle t dissipativ (genauer: es gibt ein $\varepsilon > 0$, so daß für alle t und alle $z \in \nu(A(t))$ die Ungleichung $\operatorname{Re}(z) < -\varepsilon$ gilt), dann ist die Ruhelage der Dgl. $\dot{x} = A(t)x$ global asymptotisch stabil, d.h. die Null-Lösung dieser Dgl. ist asymptotisch stabil im Ganzen.

Und auch im Zusammenhang mit der algebraischen Riccati-Gleichung $A^*X + XA + XBB^*X + C^*C = 0$ spielt diese Dissipativität eine entscheidende Rolle. Dies soll im Abschnitt 6 gezeigt werden, in dem der in der Arbeit [4] bereits ansatzweise angegebene Zusammenhang zwischen dieser Dissipativität der Matrizen und der algebraischen Riccati-Ungleichung $A^*X + XA + XBB^*X + C^*C < 0$, erstmalig ausformuliert wird (so glaube ich wenigstens).

Zuvor aber noch einige allgemeine Ausführungen zu diesem „neuen“ Eigenvektor-Verfahren, das eigentlich auf einer sehr alte Idee des Österreichers G. VON ESCHENBACH in [7] von 1898 beruht, daß sich nämlich die Lösungen gewisser nichtlinearer Gleichungen und bestimmte Unterräume von Vektorräumen eindeutig entsprechen.

So wird im Rahmen des Eigenvektor-Verfahrens dem obigen Riccati-Operator $\mathcal{R}$ die folgende Matrix zugeordnet:

$$H := \begin{bmatrix} A & R \\ -Q & -A^* \end{bmatrix}. \quad (1.2)$$

Diese Matrix $H \in \mathbb{C}^{2n \times 2n}$ wird *Hamilton-Matrix* genannt. Und bestimmte n -dimensionale Unterräume des Vektorraumes $\mathbb{C}^{2n}$, die bezüglich dieser Hamilton-Matrix invariant sind, die also durch diese

Hamilton-Matrix in sich abgebildet werden, entsprechen umkehrbar eindeutig den Lösungen der entsprechenden Riccati-Gleichung.

In dieser Arbeit werden — neben der *Linearen Algebra* — lediglich die Begriffe *stabil*, *steuerbar*, *stabilisierbar*, *beobachtbar* und *erkennbar* aus der Kontrolltheorie benötigt. Und bezüglich dieser Begriffen sei daran erinnert, daß von folgenden Erklärungen ausgegangen werden kann:

- Eine Matrix heißt *stabil*, wenn alle ihre Eigenwerte einen negativen Realteil besitzen,
- das Matrizenpaar (A, B) heißt *steuerbar*, wenn die $(n \times n^2)$ -Matrix $[B, AB, \dots, A^{n-1}B]$ den Rang n besitzt,
- das Matrizenpaar (A, B) heißt *stabilisierbar*, wenn eine Matrix X existiert, so daß die Matrix $A + BX$ stabil ist,
- das Matrizenpaar (A, C) heißt *beobachtbar*, wenn die $(n \times n^2)$ -Matrix $[C^*, A^*C^*, \dots, (A^*)^{n-1}C^*]$ den Rang n besitzt (d.h. wenn (A^*, C^*) steuerbar ist), und
- das Matrizenpaar (A, C) heißt *erkennbar*, wenn eine Matrix X existiert, so daß die Matrix $A + XC$ stabil ist (d.h. wenn (A^*, C^*) stabilisierbar ist).

Sofern in dieser Arbeit die Norm $\|v\|$ eines Vektors $v \in \mathbb{C}^n$ oder die Norm $\|M\|$ einer Matrix $M \in \mathbb{C}^{n \times n}$ auftritt, so ist hiermit immer die euklidische Norm des Vektors v bzw. die Spektralnrm der Matrix M gemeint, d.h. es ist $\|v\|^2 = v^*v$, und es ist $\|M\|^2$ das Maximum der Eigenwerte der Matrix M^*M , der maximale singuläre Wert der Matrix M zum Quadrat. Hierbei bezeichnet v^* bzw. M^* , wie allgemein üblich, den konjugiert-transponierten Vektor bzw. die konjugiert-transponierte Matrix.

In dieser Arbeit sollen für komplexe Zahlen a, b und c die folgenden allgemeinen Begriffe wie *hermitisch* usw. verwendet werden, um die spätere Umformulierung der für komplexe Zahlen geltenden Aussagen auf den allgemeineren Fall der komplexen Matrizen einfacher zu gestalten.

a hermitisch	$\iff a^* = a$, d.h. a reell
a positiv definit	$\iff a$ reell und $a > 0$
a positiv semidefinit	$\iff a$ reell und $a \geq 0$
a negativ semidefinit	$\iff a$ reell und $a \leq 0$
a stabil	$\iff \operatorname{Re}(a) < 0$
(a, b) steuerbar	$\iff b \neq 0$
(a, b) stabilisierbar	$\iff \operatorname{Re}(a) < 0$ oder $b \neq 0$
(a, c) beobachtbar	$\iff c \neq 0$
(a, c) erkennbar	$\iff \operatorname{Re}(a) < 0$ oder $c \neq 0$
a ohne rein imaginären Eigenwert	$\iff \operatorname{Re}(a) \neq 0$ (a nicht rein imaginär).

2 Die Riccati-Gleichung $A^*X + XA + XRX + Q = 0$

In diesem Abschnitt sind $A, R, Q \in \mathbb{C}^{n \times n}$ für $n \in \mathbb{N}$ mit $R^* = R, Q^* = Q$ und

$$\mathcal{R}(X) := A^*X + XA + XRX + Q, \quad X \in \mathbb{C}^{n \times n}, \quad (2.1)$$

$$H := \begin{bmatrix} A & R \\ -Q & -A^* \end{bmatrix} \in \mathbb{C}^{2n \times 2n}. \quad (2.2)$$

Oder es sind $a, r, q \in \mathbb{C}$ mit $r^* = r, q^* = q$ (d.h. r und q reell) und

$$\mathcal{R}(x) := a^*x + xa + xrx + q = 2\operatorname{Re}(a)x + rx^2 + q, \quad x \in \mathbb{C}, \quad (2.3)$$

$$H := \begin{bmatrix} a & r \\ -q & -a^* \end{bmatrix} \in \mathbb{C}^{2 \times 2}. \quad (2.4)$$

2.1 H und $\mathcal{R}(x) = 0$ für komplexe Zahlen

Die Existenz der in den Anwendungen zumeist interessieren hermitische und stabilisierende Lösungen der algebraischen Riccati-Gleichung $\mathcal{R}(x) = 0$ ist selbst in den komplexen Zahlen, d.h. für den Fall $n = 1$, keineswegs selbstverständlich. Wählen wir daher den folgenden Weg, der zunächst sehr umständlich erscheint, um zu einer Aussage über die Existenz von hermitischen und stabilisierenden Lösungen der algebraischen Riccati-Gleichung $\mathcal{R}(x) = 0$ zu gelangen.

Als erstes soll der Frage nachgegangen werden, in welcher Situation die Hamilton-Matrix H keine rein imaginären Eigenwerte besitzt. Wegen

$$\begin{aligned}\det(\lambda I - H) &= \det \begin{bmatrix} \lambda - a & -r \\ q & \lambda + a^* \end{bmatrix} = (\lambda - a)(\lambda + a^*) + rq = \lambda^2 - (a - a^*)\lambda - |a|^2 + rq \\ &= \lambda^2 - 2j \operatorname{Im}(a)\lambda - |a|^2 + rq = \lambda^2 - 2j \operatorname{Im}(a)\lambda + (j \operatorname{Im}(a))^2 - (\operatorname{Re}(a))^2 + rq \\ &= (\lambda - j \operatorname{Im}(a))^2 - ((\operatorname{Re}(a))^2 - rq)\end{aligned}\quad (2.5)$$

gilt für die Eigenwerte λ von H :

$$\lambda = \begin{cases} j \operatorname{Im}(a) \pm \sqrt{(\operatorname{Re}(a))^2 - rq} & \text{für } (\operatorname{Re}(a))^2 > rq \\ j \operatorname{Im}(a) & \text{für } (\operatorname{Re}(a))^2 = rq \\ j \operatorname{Im}(a) \pm j \sqrt{rq - (\operatorname{Re}(a))^2} & \text{für } (\operatorname{Re}(a))^2 < rq. \end{cases}\quad (2.6)$$

Also besitzt die Hamilton-Matrix H genau dann keine rein imaginären Eigenwerte, wenn die Ungleichung $(\operatorname{Re}(a))^2 > rq$ erfüllt ist.

Dieses Ergebnis ist im folgendem Hilfssatz festgehalten.

Hilfssatz 2.1. *Die folgenden beiden Aussagen (1) und (2) sind äquivalent.*

(1) $(\operatorname{Re}(a))^2 > rq$.

(2) H ist ohne rein imaginären Eigenwert.

Nun zu der Frage, in welcher Situation die algebraische Riccati-Gleichung $\mathcal{R}(x) = 2 \operatorname{Re}(a)x + rx^2 + q = 0$ hermitische und stabilisierende Lösungen besitzt. Es gilt bekanntlich:

$$\mathcal{R}(x) = 0 \iff x \in \begin{cases} \left\{ -\frac{\operatorname{Re}(a)}{r} \pm \frac{1}{r} \sqrt{(\operatorname{Re}(a))^2 - rq} \right\} & \text{für } r \neq 0, (\operatorname{Re}(a))^2 > rq \\ \left\{ -\frac{\operatorname{Re}(a)}{r} \right\} & \text{für } r \neq 0, (\operatorname{Re}(a))^2 = rq \\ \left\{ -\frac{\operatorname{Re}(a)}{r} \pm j \frac{1}{r} \sqrt{rq - (\operatorname{Re}(a))^2} \right\} & \text{für } r \neq 0, (\operatorname{Re}(a))^2 < rq \\ \left\{ -\frac{q}{2 \operatorname{Re}(a)} \right\} & \text{für } r = 0, \operatorname{Re}(a) \neq 0 \\ \emptyset & \text{für } r = 0, \operatorname{Re}(a) = 0, q \neq 0 \\ \mathbb{C} & \text{für } r = 0, \operatorname{Re}(a) = 0, q = 0. \end{cases}\quad (2.7)$$

Für die hermitischen und stabilisierenden Lösungen der Gleichung $\mathcal{R}(x) = 0$, d.h. für die komplexen Zahlen x_0 mit den drei Eigenschaften (a) $x_0^* = x_0$, d.h. x_0 reell, (b) $a + rx_0$ stabil, d.h. $\operatorname{Re}(a + rx_0) < 0$ und (b) $\mathcal{R}(x_0) = 0$ gilt hier sodann:

$$\begin{aligned}x_0^* &= x_0, a + rx_0 \text{ stabil, } \mathcal{R}(x_0) = 0 \\ \iff x_0 &= \begin{cases} -\frac{\operatorname{Re}(a)}{r} - \frac{1}{r} \sqrt{(\operatorname{Re}(a))^2 - rq} & \text{für } r \neq 0, (\operatorname{Re}(a))^2 > rq \\ -\frac{q}{2 \operatorname{Re}(a)} & \text{für } r = 0, \operatorname{Re}(a) < 0. \end{cases}\end{aligned}\quad (2.8)$$

Die algebraische Riccati-Gleichung besitzt also genau dann eine hermitische und stabilisierende Lösung, wenn gilt: $r \neq 0$ und $(\operatorname{Re}(a))^2 > rq$ oder $r = 0$ und $\operatorname{Re}(a) < 0$. Und wie dieser Herleitung zu ersehen ist, gibt es in diesem Fall genau nur eine derartige Lösung.

Nun sind aber die folgenden beiden Aussagen (1) und (2) äquivalent:

(1) Es gilt eine der beiden Bedingungen (α) oder (β):

(α) $r \neq 0$ und $(\operatorname{Re}(a))^2 > rq$.

(β) $r = 0$ und $\operatorname{Re}(a) < 0$.

(2) Es gelten die drei Aussagen (i), (ii) und (iii):

(i) H ist ohne rein imaginären Eigenwert.

(ii) (a, r) ist stabilisierbar.

(iii) r ist positiv oder negativ semidefinit.

Zusammenfassend kann daher gesagt werden:

Satz 2.1. *Ist (a, r) stabilisierbar, so sind die folgenden vier Aussagen (1)–(4) äquivalent.*

- (1) *Es gelten die folgenden beiden Bedingungen (i) und (ii):*
 - (i) *H ist ohne rein imaginären Eigenwert.*
 - (ii) *r ist positiv oder negativ semidefinit.*
- (2) *Es existiert genau eine hermitesche und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}(x) = 0$, d.h. es existiert genau eine komplexe Zahl $x_0 \in \mathbb{C}$ mit den folgenden drei Eigenschaften (a), (b) und (c):*
 - (a) $x_0^* = x_0$.
 - (b) $a + rx_0$ stabil.
 - (c) $\mathcal{R}(x_0) = a^*x_0 + x_0a + x_0rx_0 + q = 0$.
- (3) *Es existiert eine hermitesche und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}(x) = 0$, d.h. es existiert eine komplexe Zahl $x_0 \in \mathbb{C}$ mit den folgenden drei Eigenschaften (a), (b) und (c):*
 - (a) $x_0^* = x_0$.
 - (b) $a + rx_0$ stabil.
 - (c) $\mathcal{R}(x_0) = a^*x_0 + x_0a + x_0rx_0 + q = 0$.
- (4) *H ist ohne rein imaginären Eigenwert.*

Die hier angegebene Formulierung der Bedingung für die Existenz hermitescher und stabilisierender Lösungen der algebraischen Riccati-Gleichung $\mathcal{R}(x) = 0$ wirkt zunächst sicherlich sehr geschwollen. Es wird sich aber gleich zeigen, daß diese Formulierung auch für den allgemeineren Fall beim Arbeiten mit Matrizen beibehalten werden kann, im Gegensatz zu den Formulierungen, die die „Diskriminanten-Ungleichung“ $(\operatorname{Re}(a))^2 > rq$ enthalten würden. Später wird sich auch noch zeigen, daß die Bedingung (1,i) bzw. (4) „ H ist ohne rein imaginären Eigenwert“ für spezielle Matrizen R und Q leicht in eine sehr aussagekräftige Bedingung umformuliert werden kann, und daß in diesen Spezialfällen die Bedingung (1,ii) stets erfüllt ist.

2.2 H und $\mathcal{R}(X) = 0$ für Matrizen

In diesem Abschnitt geht es um den Satz 2.1, nun aber für $n \times n$ Matrizen für beliebige $n \in \mathbb{N}$.

Theorem 2.1. *Ist (A, R) stabilisierbar, so gilt für die folgenden vier Aussagen (1), (2), (3) und (4):*
(1) $\implies$ (2) $\implies$ (3) $\implies$ (4).

- (1) *Es gelten die folgenden beiden Bedingungen (i) und (ii):*
 - (i) *H ist ohne rein imaginären Eigenwert.*
 - (ii) *R positiv oder negativ semidefinit.*
- (2) *Es existiert genau eine hermitesche und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}(X) = 0$, d.h. es existiert genau eine Matrix $X_0 \in \mathbb{C}^{n \times n}$ mit den folgenden drei Eigenschaften (a), (b) und (c):*
 - (a) $X_0^* = X_0$.
 - (b) $A + RX_0$ stabil.
 - (c) $\mathcal{R}(X_0) = A^*X_0 + X_0A + X_0RX_0 + Q = 0$.
- (3) *Es existiert eine hermitesche und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}(X) = 0$, d.h. es existiert eine Matrix $X_0 \in \mathbb{C}^{n \times n}$ mit den folgenden drei Eigenschaften (a), (b) und (c):*
 - (a) $X_0^* = X_0$.
 - (b) $A + RX_0$ stabil.
 - (c) $\mathcal{R}(X_0) = A^*X_0 + X_0A + X_0RX_0 + Q = 0$.
- (4) *H ist ohne rein imaginären Eigenwert.*

Die Äquivalenzen aus dem Satz 2.1 sind hier für $n \times n$ Matrizen für beliebige $n \in \mathbb{N}$ zwar zu Implikationen abgeschwächt worden, in den wichtigen Spezialfällen aber, in denen R semidefinit ist, sind die vier Aussagen in diesem Theorem 2.1 wieder äquivalent. In diesem Fall ist es angebracht, die beiden äquivalenten Aussagen (2) und (3) zu einer Aussage der folgenden Form zusammen zu ziehen:

- (2) Es existiert (genau) eine hermitesche und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}(X) = 0$, d.h. es existiert (genau) eine Matrix $X_0 \in \mathbb{C}^{n \times n}$ mit den folgenden drei Eigenschaften (a), (b) und (c):
- (a) $X_0^* = X_0$.
 - (b) $A + RX_0$ stabil.
 - (c) $\mathcal{R}(X_0) = A^*X_0 + X_0A + X_0RX_0 + Q = 0$.

Zum Abschluß seien die Ergebnisse dieses Abschnittes in einer einfacher zu handhabenden Form zusammengefaßt.

Korollar 2.1. *Ist (A, R) stabilisierbar und R positiv oder negativ semidefinit, so sind die folgenden beiden Aussagen (1) und (2) äquivalent.*

- (1) *H ist ohne rein imaginären Eigenwert.*
- (2) *Es existiert (genau) eine hermitesche und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}(X) = 0$, d.h. es existiert (genau) eine Matrix $X_0 \in \mathbb{C}^{n \times n}$ mit den folgenden drei Eigenschaften (a), (b) und (c):*
 - (a) $X_0^* = X_0$.
 - (b) $A + RX_0$ stabil.
 - (c) $\mathcal{R}(X_0) = A^*X_0 + X_0A + X_0RX_0 + Q = 0$.

Die Aussagen dieses Abschnittes 2.2 sind — ebenso wie die in [5] hierzu angegebenen Beweise — als Standard anzusehen.

3 Die Gleichung $A^*X + XA - XBB^*X + C^*C = 0$

In diesem Abschnitt sind $A, B, C \in \mathbb{C}^{n \times n}$ für $n \in \mathbb{N}$ und

$$\mathcal{R}_d(X) := A^*X + XA - XBB^*X + C^*C, \quad X \in \mathbb{C}^{n \times n}, \quad (3.1)$$

$$H_d := \begin{bmatrix} A & -BB^* \\ -C^*C & -A^* \end{bmatrix} \in \mathbb{C}^{2n \times 2n}. \quad (3.2)$$

Oder es sind $a, b, c \in \mathbb{C}$ und

$$\mathcal{R}_d(x) := a^*x + xa - xbb^*x + c^*c = 2\operatorname{Re}(a)x - |b|^2x^2 + |c|^2, \quad x \in \mathbb{C}, \quad (3.3)$$

$$H_d := \begin{bmatrix} a & -bb^* \\ -c^*c & -a^* \end{bmatrix} = \begin{bmatrix} a & -|b|^2 \\ -|c|^2 & -a^* \end{bmatrix} \in \mathbb{C}^{2 \times 2}. \quad (3.4)$$

In diesem Abschnitt geht es also um die algebraische Riccati-Gleichung mit unterschiedlich semidefiniten Termen (dies sind die Terme $-BB^*$ bzw. $-XBB^*X$ und C^*C). Der Index „d“ steht übrigens für „different“.

3.1 H_d und $\mathcal{R}_d(x) = 0$ für komplexe Zahlen

In dem hier vorliegenden Spezialfall mit $r = -bb^* = -|b|^2$ und $q = c^*c = |c|^2$ kann für die Bedingung $(\operatorname{Re}(a))^2 > rq$ aus dem Hilfssatz 2.1 wegen $rq = -|b|^2|c|^2 = -|bc|^2$ die folgende Umformulierung vorgenommen werden:

$$\begin{aligned} (\operatorname{Re}(a))^2 > rq &\iff (\operatorname{Re}(a))^2 > -|bc|^2 \\ &\iff \operatorname{Re}(a) \neq 0 \text{ oder } (b \neq 0 \text{ und } c \neq 0) \\ &\iff \operatorname{Re}(a) \neq 0 \text{ oder } ((\operatorname{Re}(a) < 0 \text{ oder } b \neq 0) \text{ und } (\operatorname{Re}(a) < 0 \text{ oder } c \neq 0)) \\ &\iff \operatorname{Re}(a) \neq 0 \text{ oder } ((a, b) \text{ stabilisierbar und } (a, c) \text{ erkennbar}). \end{aligned} \quad (3.5)$$

Das für diesen Spezialfall erhaltenen Ergebnisse ist im folgenden Hilfssatz zusammengefaßt.

Hilfssatz 3.1. Die folgenden beiden Aussagen (1) und (2) sind äquivalent.

(1) Eine der folgenden beiden Bedingungen (α) oder (β) ist erfüllt:

(α) a ist ohne rein imaginären Eigenwert.

(β) (a, b) ist stabilisierbar, und (a, c) ist erkennbar.

(2) H_d ist ohne rein imaginären Eigenwert.

Mit diesem Hilfssatz 3.1 und dem Satz 2.1 folgt unmittelbar der folgende Satz, sofern beachtet wird, daß jede hermitesche und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}_d(x) = 0$ stets positiv semidefinit ist. Denn wegen $x_0^* = x_0$ und $\mathcal{R}_d(x_0) = 0$ ist

$$\begin{aligned} a^*x_0 + x_0a - x_0bb^*x_0 + c^*c &= (a - bb^*x_0)^*x_0 + x_0(a - bb^*x_0) + x_0bb^*x_0 + c^*c \\ &= 2\operatorname{Re}(a - bb^*x_0)x_0 + |b|^2x_0^2 + |c|^2 = 0, \end{aligned} \quad (3.6)$$

und wegen $\operatorname{Re}(a - bb^*x_0) < 0$ folgt:

$$x_0 = -\frac{|b|^2x_0^2 + |c|^2}{2\operatorname{Re}(a - bb^*x_0)} \begin{cases} > 0 & \text{für } |b|^2x_0^2 + |c|^2 > 0 \\ = 0 & \text{für } |b|^2x_0^2 + |c|^2 = 0. \end{cases} \quad (3.7)$$

Satz 3.1. Die folgenden drei Aussagen (1), (2) und (3) sind äquivalent.

(1) Eine der folgenden beiden Bedingungen (α) oder (β) ist erfüllt:

(α) a ist ohne rein imaginären Eigenwert, und (a, b) ist stabilisierbar.

(β) (a, b) ist stabilisierbar, und (a, c) ist erkennbar.

(2) Es gelten die beiden folgenden Bedingungen (i) und (ii):

(i) H_d ist ohne rein imaginären Eigenwert.

(ii) (a, b) ist stabilisierbar.

(3) Es existiert (genau) eine positiv semidefinite und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}_d(x) = 0$, d.h. es existiert (genau) eine komplexe Zahl $x_0 \in \mathbb{C}$ mit den folgenden drei Eigenschaften (a), (b) und (c):

(a) $x_0 \geq 0$.

(b) $a - bb^*x_0$ stabil.

(c) $\mathcal{R}_d(x_0) = a^*x_0 + x_0a - x_0bb^*x_0 + c^*c = 0$.

In den Anwendungen spielt dieser Satz 3.1 keine große Rolle (sicher wegen seiner Komplexität). In den Anwendungen wird beispielsweise zusätzlich gefordert, daß (a, c) beobachtbar sein soll. In diesem Fall ist $c \neq 0$, und jede hermitesche und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}_d(x) = 0$ ist zufolge der Überlegung vor dem Satz 3.1 stets positiv definit. Damit wird folgende einfachere Aussage erhalten:

Folgerung 3.1. Ist (a, c) beobachtbar, so sind die folgenden beiden Aussagen (1) und (2) äquivalent.

(1) (a, b) ist stabilisierbar.

(2) Es existiert (genau) eine positiv definite und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}_d(x) = 0$, d.h. es existiert (genau) eine komplexe Zahl $x_0 \in \mathbb{C}$ mit den folgenden drei Eigenschaften (a), (b) und (c):

(a) $x_0 > 0$.

(b) $a - bb^*x_0$ stabil.

(c) $\mathcal{R}_d(x_0) = a^*x_0 + x_0a - x_0bb^*x_0 + c^*c = 0$.

3.2 H_d und $\mathcal{R}_d(x) = 0$ für Matrizen

Zunächst folgt als Verallgemeinerung von Hilfssatz 3.1 eine hinreichende Bedingung für Hamilton-Matrizen H_d ohne rein imaginären Eigenwert.

Lemma 3.1. Für die folgenden beiden Aussagen (1) und (2) gilt: (1) $\implies$ (2).

(1) Eine der folgenden beiden Bedingungen (α) oder (β) ist erfüllt:

- (α) A ist ohne rein imaginären Eigenwert.
- (β) (A, B) ist stabilisierbar, und (A, C) ist erkennbar.

(2) H_d ist ohne rein imaginären Eigenwert.

Mit diesem Lemma 3.1 und dem Korollar 2.1 folgt als Verallgemeinerung von Satz 3.1 das folgende Theorem, sofern beachtet wird, daß jede hermitische und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}_d(X) = 0$ stets positiv semidefinit ist.

Theorem 3.1. Für die folgenden drei Aussagen (1), (2) und (3) gilt: (1) $\implies$ (2) und (2) $\iff$ (3).

(1) Eine der folgenden beiden Bedingungen (α) oder (β) ist erfüllt:

- (α) A ist ohne rein imaginären Eigenwert, und (A, B) ist stabilisierbar.
- (β) (A, B) ist stabilisierbar, und (A, C) ist erkennbar.

(2) Es gelten die folgenden beiden Bedingungen (i) und (ii):

- (i) H_d ist ohne rein imaginären Eigenwert.
- (ii) (A, B) ist stabilisierbar.

(3) Es existiert (genau) eine positiv semidefinite und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}_d(X) = 0$, d.h. es existiert (genau) ein $X_0 \in \mathbb{C}^{n \times n}$ mit den folgenden drei Eigenschaften (a), (b) und (c):

- (a) $X_0^* \geq 0$.
- (b) $A - BB^*X_0$ stabil.
- (c) $\mathcal{R}_d(X_0) = A^*X_0 + X_0A - X_0BB^*X_0 + C^*C = 0$.

Und wird weiter beachtet, daß für (A, C) beobachtbar jede hermitische und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}_d(X) = 0$ stets positiv definit ist, so wird in Verallgemeinerung von Folgerung 3.1 das folgende Korollar erhalten.

Korollar 3.1. Ist (A, C) beobachtbar, so sind die folgenden beiden Aussagen (1) und (2) äquivalent.

(1) (A, B) ist stabilisierbar.

(2) Es existiert (genau) eine positiv definite und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}_d(X) = 0$, d.h. es existiert (genau) eine Matrix $X_0 \in \mathbb{C}^{n \times n}$ mit den folgenden drei Eigenschaften (a), (b) und (c):

- (a) $X_0 > 0$.
- (b) $A - BB^*X_0$ stabil.
- (c) $\mathcal{R}_d(X_0) = A^*X_0 + X_0A - X_0BB^*X_0 + C^*C = 0$.

Einzelne Teile aus dem Theorem 3.1 oder dem Korollar 3.1 werden in den deutschsprachigen Lehrbüchern zur Regelungstechnik (Kontrolltheorie) zumeist herangezogen, wenn es um die Existenz von Lösungen der algebraischen Riccati-Gleichung geht.

Die Aussagen dieses Abschnittes 3.2 sind — ebenso wie die in [5] hierzu angegebenen Beweise — als Standard anzusehen.

4 Die Gleichung $A^*X + XA + XBB^*X + C^*C = 0$ und die Kontraktivität des Tripels (A, B, C)

In diesem Abschnitt sind $A, B, C \in \mathbb{C}^{n \times n}$ für $n \in \mathbb{N}$ und

$$\mathcal{R}_s(X) := A^*X + XA + XBB^*X + C^*C, \quad X \in \mathbb{C}^{n \times n}, \quad (4.1)$$

$$H_s := \begin{bmatrix} A & BB^* \\ -C^*C & -A^* \end{bmatrix} \in \mathbb{C}^{2n \times 2n}. \quad (4.2)$$

Oder es sind $a, b, c \in \mathbb{C}$ und

$$\mathcal{R}_s(x) := a^*x + xa + xbb^*x + c^*c = 2\operatorname{Re}(a)x + |b|^2x^2 + |c|^2, \quad x \in \mathbb{C}, \quad (4.3)$$

$$H_s := \begin{bmatrix} a & bb^* \\ -c^*c & -a^* \end{bmatrix} = \begin{bmatrix} a & |b|^2 \\ -|c|^2 & -a^* \end{bmatrix} \in \mathbb{C}^{2 \times 2}. \quad (4.4)$$

In diesem Abschnitt geht es also um die algebraische Riccati-Gleichung mit gleichartig semidefiniten Termen (dies sind die Terme BB^* bzw. XBB^*X und C^*C). Der Index „s“ steht übrigens für „similar“. Untersucht werden soll, wie die Existenz einer Lösung dieser algebraischen Riccati-Gleichung $\mathcal{R}_s(X) = 0$ mit einer der beiden Eigenschaften der folgenden Definition zusammenhängt.

Definition 4.1.

- (1) Das Matrizen­tripel (A, B, C) heißt *schwach kontraktiv*, wenn die Matrix A ohne rein imaginäre Eigenwerte ist, und wenn für alle $\omega \in \mathbb{R}$ gilt: $\|C(j\omega I - A)^{-1}B\| < 1$.
- (2) Das Matrizen­tripel (A, B, C) heißt *kontraktiv*, wenn die Matrix A stabil ist, und wenn für alle $\omega \in \mathbb{R}$ gilt: $\|C(j\omega I - A)^{-1}B\| < 1$.

Vor allem die kontraktiven Matrizen­tripel (A, B, C) spielen in der Stabilitätstheorie eine große Rolle. Um dies zu verdeutlichen, sei das folgende Regelungssystem der Abbildung 1 mit dem unstrukturierten unbestimmten Parameter $K(t)$ mit $\|K(t)\| \leq 1$ für alle t betrachtet.

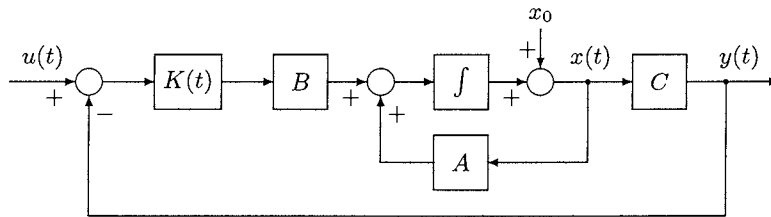


Abbildung 1: Strukturbild eines Regelungssystems

Das Verhalten dieses Regelungssystem wird durch das folgende Gleichungssystem beschrieben:

$$\begin{aligned} \dot{x}(t) &= (A - BK(t)C)x(t) + BK(t)u(t), \quad x(0) = x_0, \quad \|K(t)\| \leq 1, \\ y(t) &= Cx(t), \quad t \in \mathbb{R}. \end{aligned} \quad (4.5)$$

Und die Funktion $\omega \mapsto G(j\omega) := C(j\omega I - A)^{-1}B$ ist hier gerade der Frequenzgang des offenen Kreises (ohne den Faktor $K(t)$ mit $\|K(t)\| \leq 1$), so daß durch die Kontraktivität des Matrizen­tripels (A, B, C) gesichert wird, daß

- wegen der Stabilität von A der offene Regelkreis stabil ist, und
- wegen $\|C(j\omega I - A)^{-1}B\| < 1$ für alle $\omega \in \mathbb{R}$ und wegen $\|K(t)\| \leq 1$ für alle $t \in \mathbb{R}$ die „Kreisverstärkung“ des offenen Regelkreises kleiner als 1 ist.

Also ist mehr oder weniger anschaulich klar, daß die folgende Proposition gilt (zumindest für die Implikation (2) $\implies$ (3), für das sogenannte *Kreiskriterium*, sollte dies klar sein).

Proposition 4.1. Für die folgenden Aussagen (1), (2) und (3) gilt: (1) $\implies$ (2) $\implies$ (3).

- (1) Die Ruhelage des Systems gemäß (4.5) ist global asymptotisch stabil, und für alle $\omega \in \mathbb{R}$ ist die Matrix $C(j\omega I - A)^{-1}B$ normal.
- (2) Das Matrizen­tripel (A, B, C) ist kontraktiv, d.h. die Matrix A ist stabil, und für alle $\omega \in \mathbb{R}$ ist $\|C(j\omega I - A)^{-1}B\| < 1$.
- (3) Die Ruhelage des Systems gemäß (4.5) ist global asymptotisch stabil.

Ein Beweis für die Gültigkeit dieser Proposition ist beispielsweise in der Arbeit [4] zu finden. Der hier angegebenen „Einbettung“ der Aussage (2) ist zu ersehen, daß für den Fall, daß die Matrix $G(j\omega) = C(j\omega I - A)^{-1}B$ für alle $\omega \in \mathbb{R}$ normal ist, die Ruhelage des Systems gemäß (4.5) **genau dann global asymptotisch stabil ist**, wenn das Matrizen­tripel (A, B, C) kontraktiv ist. Und in den wichtigen Spezialfällen

- eines entkoppelten Regelkreises, $G(j\omega)$ ist eine Diagonalmatrix, und noch spezieller,
- eines Einfach-Regelkreises, die Matrizen B und C enthalten nur in einer Spalte bzw. Zeile von Null verschiedene Elemente,

kann die Stabilitätsbedingung „das Matrizen­tripel (A, B, C) ist kontraktiv“ daher nicht enger gefaßt werden, wenn weiterhin die globale asymptotische Stabilität der Ruhelage des Systems gemäß (4.5) gesichert werden soll (woran man natürlich denkt, wenn man nur das „klassische Kreiskriterium“ der obigen Implikation (2) $\implies$ (3) vor Augen hat).

4.1 H_s und $\mathcal{R}_s(x) = 0$ für komplexe Zahlen

In dem hier vorliegenden Spezialfall mit $r = bb^* = |b|^2$ und $q = c^*c$ kann die Bedingung $(\operatorname{Re}(a))^2 > rq$ aus dem Hilfssatz 2.1 wegen $rq = |b|^2|c|^2 = |cb|^2$ wie folgt umformuliert werden:

$$\begin{aligned}
 (\operatorname{Re}(a))^2 > rq &\iff (\operatorname{Re}(a))^2 > |cb|^2 \\
 &\iff \operatorname{Re}(a) \neq 0 \text{ und } |\operatorname{Re}(a)| > |cb| \\
 &\iff \operatorname{Re}(a) \neq 0 \text{ und } |j\omega - a| > |cb| \text{ für alle } \omega \in \mathbb{R} \\
 &\iff \operatorname{Re}(a) \neq 0 \text{ und } |c(j\omega - a)^{-1}b| < 1 \text{ für alle } \omega \in \mathbb{R} \\
 &\iff a \text{ nicht rein imaginär und } |c(j\omega - a)^{-1}b| < 1 \text{ für alle } \omega \in \mathbb{R}.
 \end{aligned} \tag{4.6}$$

Die für diesen Spezialfall erhaltenen Ergebnisse sind in dem folgenden Hilfssatz zusammengefaßt.

Hilfssatz 4.1. *Für die folgenden vier Aussagen (1), (2), (3) und (4) gilt: (1) $\implies$ (2) $\implies$ (3) $\implies$ (4). Die Aussagen (2), (3) und (4) sind sogar äquivalent.*

- (1) (a, b) ist stabilisierbar, und H_s ist ohne rein imaginären Eigenwert.
- (2) a und H_s ist ohne rein imaginären Eigenwert.
- (3) (a, b, c) ist schwach kontraktiv, d.h. a ist ohne rein imaginären Eigenwert, und für alle $\omega \in \mathbb{R}$ ist $|c(j\omega - a)^{-1}b| < 1$.
- (4) H_s ist ohne rein imaginären Eigenwert.

Dieser Hilfssatz 4.1 ist die Stelle, an der die sogenannte Kreisbedingung mit der Übertragungsfunktion des offenen Regelkreises zum ersten mal innerhalb der Theorie der Riccati-Gleichung auftritt und ins Spiel gebracht wird. Ohne die Hamilton-Matrix war dies früher nur mittels der Laplace- bzw. Fourier-Transformation möglich, und dieser Weg war viel schwerer zu begehen. Und wie der Herleitung (4.6) zu ersehen ist, handelt es sich bei der Kreisbedingung um eine Umformulierung der Diskriminanten-Bedingung $(\operatorname{Re}(a))^2 > |cb|^2$.

Für $r = bb^* = |b|^2$ und $q = c^*c = |c|^2$ gilt natürlich: $r \geq 0$, $r = 0 \iff b = 0$, (a, bb^*) stabilisierbar $\iff (a, b)$ stabilisierbar. Mit dem Satz 2.1 und dem Hilfssatz 4.1 folgt sodann der folgende Hilfssatz:

Hilfssatz 4.2. *Ist (a, b) stabilisierbar, so sind die folgenden drei Aussagen (1), (2) und (3) äquivalent.*

- (1) (a, b, c) ist schwach kontraktiv, d.h. a ist ohne rein imaginären Eigenwert, und für alle $\omega \in \mathbb{R}$ ist $|c(j\omega - a)^{-1}b| < 1$.
- (2) $(a$ und $) H_s$ ist ohne rein imaginären Eigenwert.
- (3) Es existiert (genau) eine hermitesche und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}_s(x) = 0$, d.h. es existiert (genau) eine komplexe Zahl $x_0 \in \mathbb{C}$ mit den folgenden drei Eigenschaften (a), (b) und (c):
 - (a) $x_0^* = x_0$.
 - (b) $a + bb^*x_0$ stabil.
 - (c) $\mathcal{R}_s(x_0) = a^*x_0 + x_0a + x_0bb^*x_0 + c^*c = 0$.

Dieser Hilfssatz 4.2 liefert den folgenden einfacher zu handhabenden Satz.

Satz 4.1 (Bounded Real Lemma). *Die folgenden drei Aussagen (1), (2) und (3) sind äquivalent.*

- (1) (a, b, c) ist kontraktiv, d.h. a ist stabil, und für alle $\omega \in \mathbb{R}$ ist $|c(j\omega - a)^{-1}b| < 1$.
- (2) a ist stabil, und H_s ist ohne rein imaginären Eigenwert.
- (3) Es existiert (genau) eine positiv semidefinite und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}_s(x) = 0$, d.h. es existiert (genau) eine komplexe Zahl $x_0 \in \mathbb{C}$ mit den folgenden drei Eigenschaften (a), (b) und (c):
 - (a) $x_0 \geq 0$.
 - (b) $a + bb^*x_0$ stabil.
 - (c) $\mathcal{R}_s(x_0) = a^*x_0 + x_0a + x_0bb^*x_0 + c^*c = 0$.

Dieser Satz 4.1 wird hin und wieder schon als „Lemma von Kalman–Yakubovich–Popov“ bezeichnet. Ich möchte diesen Namen aber erst im nächsten Abschnitt 5 im Zusammenhang mit einer algebraischen Riccati–Ungleichung verwenden, so wie es zumeist üblich ist.

Natürlich ist die Aussage (1) in diesem Satz 4.1 zu der folgenden Aussage (1') äquivalent:

(1') Für alle $s \in \mathbb{C}$ mit $s \notin \sigma(a)$ und $\operatorname{Re}(s) \geq 0$ ist $|c(s - a)^{-1}b| < 1$.

Und der Name „Bounded Real Lemma“ soll nun dadurch zustande kommen, weil diese Aussage (1') als **Bounded Real Statement** bezeichnet wird.

4.2 H_s und $\mathcal{R}_s(X) = 0$ für Matrizen

Hier folgt nun die für das Folgende wesentliche Aussage, für die wie ich glaube, den Beweis der wesentlichen Implikation (3) $\implies$ (4) ganz entscheidend vereinfacht zu haben (siehe [5]).

Lemma 4.1. Für die folgenden vier Aussagen (1), (2), (3) und (4) gilt: (1) $\implies$ (2) $\implies$ (3) $\implies$ (4).

- (1) (A, B) ist stabilisierbar, und H_s ist ohne rein imaginären Eigenwert.
- (2) A und H_s ist ohne rein imaginären Eigenwert.
- (3) (A, B, C) ist schwach kontraktiv, d.h. A ist ohne rein imaginären Eigenwert, und für alle $\omega \in \mathbb{R}$ ist $\|C(j\omega I - A)^{-1}B\| < 1$.
- (4) H_s ist ohne rein imaginären Eigenwert.

Im Rahmen der Stabilitätstheorie wird stets A stabil (oder zumindest (A, B) stabilisierbar) vorausgesetzt, so daß die vier Aussagen dieses Lemmas äquivalent sind. Mit dem Korollar 2.1 folgt daher wegen $BB^* \geq 0$ sofort das folgende Lemma anstelle des Hilfssatzes 4.2.

Lemma 4.2. Ist (A, B) stabilisierbar, so sind die folgenden drei Aussagen (1), (2) und (3) äquivalent.

- (1) (A, B, C) ist schwach kontraktiv, d.h. A ist ohne rein imaginären Eigenwert, und für alle $\omega \in \mathbb{R}$ ist $\|C(j\omega I - A)^{-1}B\| < 1$.
- (2) $(A$ und) H_s ist ohne rein imaginären Eigenwert.
- (3) Es existiert (genau) eine hermitesche und stabilisierende Lösung der algebraischen Riccati–Gleichung $\mathcal{R}_s(X) = 0$, d.h. es existiert (genau) eine Matrix $X_0 \in \mathbb{C}^{n \times n}$ mit den folgenden drei Eigenschaften (a), (b) und (c):
 - (a) $X_0^* = X_0$.
 - (b) $A + BB^*X_0$ stabil.
 - (c) $\mathcal{R}_s(X_0) = A^*X_0 + X_0A + X_0BB^*X_0 + C^*C = 0$.

Einfacher zu handhaben ist das folgende Theorem, das analog zum Satz 4.1 entstanden gedacht werden kann.

Theorem 4.1 (Bounded Real Lemma). Die folgenden drei Aussagen (1), (2) und (3) sind äquivalent:

- (1) (A, B, C) ist kontraktiv, d.h. A ist stabil, und für alle $\omega \in \mathbb{R}$ ist $\|C(j\omega I - A)^{-1}B\| < 1$.
- (2) A ist stabil, und H_s ist ohne rein imaginären Eigenwert.
- (3) Es existiert (genau) eine positiv semidefinite und stabilisierende Lösung der algebraischen Riccati–Gleichung $\mathcal{R}_s(X) = 0$, d.h. es existiert (genau) eine Matrix $X_0 \in \mathbb{C}^{n \times n}$ mit den folgenden drei Eigenschaften (a), (b) und (c):
 - (a) $X_0^* \geq 0$.
 - (b) $A + BB^*X_0$ stabil.
 - (c) $\mathcal{R}_s(X_0) = A^*X_0 + X_0A + X_0BB^*X_0 + C^*C = 0$.

Die Aussagen dieses Abschnittes 4.2 sind als Standard anzusehen, ebenso die in [5] hierzu angegebenen Beweise, ausgenommen lediglich der Beweis der Implikation (3) $\implies$ (4) von Lemma 4.1.

5 Die Ungleichung $A^*X + XA + XBB^*X + C^*C < 0$

In diesem Abschnitt geht es um die algebraische Riccati–Ungleichung mit gleichartig semidefiniten Termen. Es gelten daher weiter die Bezeichnungen aus dem vorhergehenden Abschnitt 4.

5.1 Die Riccati–Ungleichung $\mathcal{R}_s(x) < 0$ für komplexe Zahlen

Als erstes soll eine Möglichkeit geschaffen werden, um die algebraische Riccati–Gleichung $\mathcal{R}_s(x) = 0$ betreffenden Aussagen auf die algebraische Riccati–Ungleichung $\mathcal{R}_s(x) < 0$ zu übertragen.

Hilfssatz 5.1. *Die folgenden beiden Aussagen (1) und (2) sind äquivalent.*

(1) *Es existiert eine positiv semidefinite und stabilisierende Lösung der algebraischen Riccati–Gleichung $\mathcal{R}_s(x) = 0$, d.h. es existiert eine komplexe Zahl $x_0 \in \mathbb{C}$ mit den folgenden drei Eigenschaften (a), (b) und (c):*

(a) $x_0 \geq 0$.

(b) $a + bb^*x_0$ stabil.

(c) $\mathcal{R}_s(x_0) = a^*x_0 + x_0a + x_0bb^*x_0 + c^*c = 0$.

(2) *Es existiert eine positiv definite und stabilisierende Lösung der algebraischen Riccati–Ungleichung $\mathcal{R}_s(x) < 0$, d.h. es existiert eine komplexe Zahl $x_1 \in \mathbb{C}$ mit den folgenden drei Eigenschaften (a), (b) und (c):*

(a) $x_1 > 0$.

(b) $a + bb^*x_1$ stabil.

(c) $\mathcal{R}_s(x_1) = a^*x_1 + x_1a + x_1bb^*x_1 + c^*c < 0$.

Dies ist wie folgt einzusehen: Für (1) und auch für (2) ist wegen (a) und (b) a stabil. Also liegt für die Funktion $x \mapsto y = \mathcal{R}_s(x) = 2\operatorname{Re}(a)x + |b|^2x^2 + |c|^2$ einer der vier Fälle der Abbildung 2 vor.

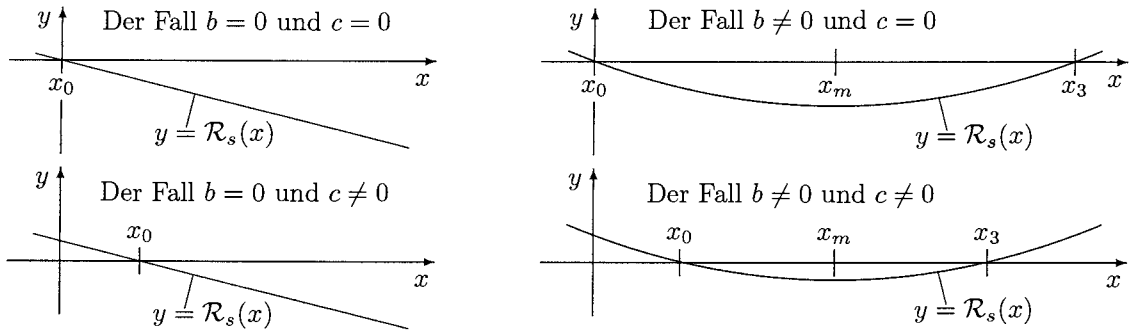


Abbildung 2: Die reelle Funktion $x \mapsto y = \mathcal{R}_s(x)$ für a stabil.

In den Fällen mit $b \neq 0$ ist nur darauf zu achten, daß x_0 und x_1 links von x_m liegen. Denn wegen

$$\frac{d}{dx}\mathcal{R}_s(x_m) = 2\operatorname{Re}(a) + 2|b|^2x_m = 2\operatorname{Re}(a + bb^*x_m) = 0 \quad (5.1)$$

ist $a + bb^*x_0$ und $a + bb^*x_1$ nur hierfür stabil.

Ebenso ist die Gültigkeit des folgenden Hilfssatzes einzusehen.

Hilfssatz 5.2. *Die folgenden beiden Aussagen (1) und (2) sind äquivalent.*

(1) *Es existiert eine positiv definite und stabilisierende Lösung der algebraischen Riccati–Ungleichung $\mathcal{R}_s(x) < 0$, d.h. es existiert eine komplexe Zahl $x_1 \in \mathbb{C}$ mit den folgenden drei Eigenschaften (a), (b) und (c):*

(a) $x_1 > 0$.

(b) $a + bb^*x_1$ stabil.

(c) $\mathcal{R}_s(x_1) = a^*x_1 + x_1a + x_1bb^*x_1 + c^*c < 0$.

- (2) Es existiert eine positiv definite Lösung der algebraischen Riccati-Ungleichung $\mathcal{R}_s(x) < 0$, d.h. es existiert eine komplexe Zahl $x_2 \in \mathbb{C}$ mit den folgenden drei Eigenschaften (a), (b) und (c):
- (a) $x_2 > 0$.
 - (b) $\mathcal{R}_s(x_2) = a^*x_2 + x_2a + x_2bb^*x_2 + c^*c < 0$.

Aufgrund der Aussage dieses Hilfssatzes 5.2 wird in den Sätzen, in denen es um die Existenz positiv definiter und stabilisierenden Lösungen der algebraischen Riccati-Ungleichung $\mathcal{R}_s(x) < 0$ geht, zumeist nur von der Existenz positiv definiter Lösungen der algebraischen Riccati-Ungleichung $\mathcal{R}_s(x) < 0$ gesprochen. In diesem Sinne folgt der für die Anwender so wichtige folgende Satz, der lediglich eine Zusammenfassung der Aussagen des Satzes 4.1 und der beiden Hilfssätze 5.1 und 5.2 darstellt.

Satz 5.1 (Lemma von Kalman–Yakubovich–Popov). Die folgenden beiden Aussagen (1) und (2) sind äquivalent.

- (1) (a, b, c) ist kontraktiv, d.h. a ist stabil, und für alle $\omega \in \mathbb{R}$ ist $|c(j\omega - a)^{-1}b| < 1$.
- (2) Es existiert eine positiv definite Lösung der algebraischen Riccati-Ungleichung $\mathcal{R}_s(x) < 0$, d.h. es existiert eine komplexe Zahl $x_2 \in \mathbb{C}$ mit den folgenden zwei Eigenschaften (a) und (b):
 - (a) $x_2 > 0$.
 - (b) $\mathcal{R}_s(x_2) = a^*x_2 + x_2a + x_2bb^*x_2 + c^*c < 0$.

5.2 Die Riccati-Ungleichung $\mathcal{R}_s(X) < 0$ für Matrizen

Zunächst folgt als Verallgemeinerung von Hilfssatz 5.1 das Lemma 5.1, nun aber zunächst nur mit der Implikation (1) $\implies$ (2) anstelle der Äquivalenz (1) $\iff$ (2). Die Äquivalenz wird sich später „von alleine“ ergeben.

Lemma 5.1. Für die folgenden beiden Aussagen (1) und (2) gilt: (1) $\implies$ (2).

- (1) Es existiert eine positiv semidefinite und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}_s(X) = 0$, d.h. es existiert eine Matrix $X_0 \in \mathbb{C}^{n \times n}$ mit den folgenden drei Eigenschaften (a), (b) und (c):
 - (a) $X_0^* \geq 0$.
 - (b) $A + BB^*X_0$ stabil.
 - (c) $\mathcal{R}_s(X_0) = A^*X_0 + X_0A + X_0BB^*X_0 + C^*C = 0$.
- (2) Es existiert eine positiv definite und stabilisierende Lösung der algebraischen Riccati-Ungleichung $\mathcal{R}_s(X) < 0$, d.h. es existiert eine Matrix $X_1 \in \mathbb{C}^{n \times n}$ mit den folgenden drei Eigenschaften (a), (b) und (c):
 - (a) $X_1 > 0$.
 - (b) $A + BB^*X_1$ stabil.
 - (c) $\mathcal{R}_s(X_1) = A^*X_1 + X_1A + X_1BB^*X_1 + C^*C < 0$.

Der obige Hilfssatz 5.2 soll hier jetzt aber nicht für Matrizen ausformuliert werden. Hier soll ein anderer Weg zum Lemma von Kalman–Yakubovich–Popov eingeschlagen werden, der für Matrizen einfacher gangbar ist, und der später auch die Aussage des Hilfssatzes 5.2 für Matrizen liefern wird.

Lemma 5.2. Für die folgenden beiden Aussagen (1) und (2) gilt: (1) $\implies$ (2).

- (1) Es existiert eine positiv definite Lösung der algebraischen Riccati-Ungleichung $\mathcal{R}_s(X) < 0$, d.h. es existiert eine Matrix $X_2 \in \mathbb{C}^{n \times n}$ mit den folgenden zwei Eigenschaften (a) und (b):
 - (a) $X_2 > 0$.
 - (b) $\mathcal{R}_s(X_2) = A^*X_2 + X_2A + X_2BB^*X_2 + C^*C < 0$.
- (2) (A, B, C) ist kontraktiv, d.h. A ist stabil, und für alle $\omega \in \mathbb{R}$ ist $\|C(j\omega I - A)^{-1}B\| < 1$.

Eine Zusammenfassung der Aussagen der beiden letzten Lemmata mit der Aussage von Theorem 4.1 liefert schließlich:

Theorem 5.1 (Lemma von Kalman–Yakubovich–Popov). Die folgenden beiden Aussagen (1) und (2) sind äquivalent.

- (1) (A, B, C) ist kontraktiv, d.h. A ist stabil, und für alle $\omega \in \mathbb{R}$ ist $\|C(j\omega I - A)^{-1}B\| < 1$.
- (2) Es existiert eine positiv definite Lösung der algebraischen Riccati-Ungleichung $\mathcal{R}_s(X) < 0$, d.h. es existiert eine Matrix $X_2 \in \mathbb{C}^{n \times n}$ mit den folgenden zwei Eigenschaften (a) und (b):
- (a) $X_2 > 0$.
- (b) $\mathcal{R}_s(X_2) = A^*X_2 + X_2A + X_2BB^*X_2 + C^*C < 0$.

Man beachte, in diesem Theorem 5.1 ist die folgende Aussage zur algebraischen Riccati-Ungleichung $\mathcal{R}_s(X) < 0$ enthalten: Gibt es eine positiv definite Lösung X_2 dieser Ungleichung, dann gibt es auch eine positiv definite und stabilisierende Lösung X_1 dieser Ungleichung.

Eine Zusammenfassung der in diesem Abschnitt erhaltenen Aussagen mit denjenigen aus dem vorhergehenden Abschnitt 4 soll erst im Abschnitt 7 erfolgen.

Die Aussagen dieses Abschnittes 5.2 sind — ebenso wie die in [5] hierzu angegebenen Beweise — als Standard anzusehen.

6 Die Gleichung $A^*X + XA + XBB^*X + C^*C = 0$ und die Dissipativität des Tripels (A, B, C)

In diesem Abschnitt geht es nochmals um die algebraische Riccati-Gleichung und Ungleichung mit gleichartig semidefiniten Termen. Es gelten daher weiter die Bezeichnungen aus dem Abschnitt 4.

In den Arbeiten [3] und [4] habe ich im Zusammenhang mit der Betrachtung des numerischen Bereiches von Elementen einer B-Algebra den Begriff der Dissipativität verwendet. Bei einer endlich dimensionalen B*-Algebra, speziell bei $\mathbb{C}^{n \times n}$ mit der konjugierten Transponierung als Involution und der Spektralnrm, kann gesagt werden:

Definition 6.1.

- (1) Eine Matrix $M \in \mathbb{C}^{n \times n}$ heißt *dissipativ*, wenn für alle $x \in \mathbb{C}^n$ mit $x \neq 0$ gilt: $\operatorname{Re}(x^*Mx) < 0$.
- (2) Das Matrizen Tripel (A, B, C) heißt *dissipativ*, wenn eine positiv definite Matrix $Q \in \mathbb{C}^{n \times n}$ existiert, so daß $Q(A + BKC)$ oder $Q(A - BKC)$ für alle $K \in \mathbb{C}^{n \times n}$ mit $\|K\| \leq 1$ dissipativ ist.

In den Arbeiten [3] und [4] sind verschiedene andere Formulierungen für die Dissipativität in Gebrauch. Für die endlich dimensionale B*-Algebra $\mathbb{C}^{n \times n}$ sind diese anderen Formulierungen der Dissipativität jedoch zu der in der Definition 6.1 gegebenen äquivalent.

In der Arbeit [3] habe ich ausführlich untersucht, wie die Stabilität der linearen zeitveränderlichen Differentialgleichung $\dot{x}(t) = A(t)x(t)$ mit der Dissipativität von $A(t)$ zusammenhängt. Und in der Arbeit [4] habe ich unter anderem gezeigt, daß die Existenz einer positiv definiten Lösung der algebraischen Riccati-Ungleichung $\mathcal{R}_s(X) < 0$ hinreichend für die Dissipativität des Matrizen Tripels (A, B, C) ist. Hier soll nun gezeigt werden, daß die Dissipativität von (A, B, C) sogar notwendig und hinreichend für die Existenz einer hermiteschen und stabilisierenden Lösung der Gleichung $\mathcal{R}_s(X) = 0$ sowie für die Existenz einer positiv definiten Lösung der Ungleichung $\mathcal{R}_s(X) < 0$ ist.

Motiviert wird die Einführung des Begriffs der Dissipativität wieder durch die Betrachtung des Regelungssystem der Abbildung 1.

Ist das Matrizen Tripel (A, B, C) dissipativ, d.h. existiert eine positiv definite Matrix Q , so daß $\operatorname{Re}(x^*Q(A + BKC)x) < 0$ für alle $K \in \mathbb{C}^{n \times n}$ mit $\|K\| \leq 1$ und alle $x \in \mathbb{C}^n$ mit $x \neq 0$ gilt, so wird durch die positiv definite Matrix Q mittels $V(x) := x^*Qx$ eine Lyapunov-Funktion zur Sicherung der globalen asymptotischen Stabilität der Ruhelage des Systems gemäß (4.5) zur Verfügung gestellt. Denn ist $u(t) = 0$ für alle t , so gilt für alle t :

$$\begin{aligned}
\frac{d}{dt}V(x(t)) &= \frac{d}{dt}x(t)^*Qx(t) = \dot{x}(t)^*Qx(t) + x(t)^*Q\dot{x}(t) \\
&= ((A - BK(t)C)x(t))^*Qx(t) + x(t)^*Q(A - BK(t)C)x(t) \\
&= (x(t)^*Q(A - BK(t)C)x(t))^* + (x(t)^*Q(A - BK(t)C)x(t)) \\
&= 2\operatorname{Re}(x(t)^*Q(A - BK(t)C)x(t)).
\end{aligned} \tag{6.1}$$

Also gilt für alle t mit $\|K(t)\| \leq 1$ und $x(t) \neq 0$: $V(x(t)) > 0$ und $\frac{d}{dt}V(x(t)) < 0$. Damit ist die globale asymptotische Stabilität der Ruhelage des Systems gemäß (4.5) nachgewiesen.

Und besitzt das System gemäß (4.5) eine globale asymptotisch stabile Ruhelage, und ist die Matrix $A + BKC$ für alle $K \in \mathbb{C}^{n \times n}$ mit $\|K\| \leq 1$ normal, so ist das Matrizen-tripel (A, B, C) dissipativ, wie in der Arbeit [4] bewiesen wurde. Damit ist die folgende Proposition gültig.

Proposition 6.1. *Für die folgenden drei Aussagen (1), (2) und (3) gilt: (1) $\implies$ (2) $\implies$ (3).*

- (1) *Die Ruhelage des Systems gemäß (4.5) ist global asymptotisch stabil, und für alle $K \in \mathbb{C}^{n \times n}$ mit $\|K\| \leq 1$ ist die Matrix $A + BKC$ normal.*
- (2) *Das Matrizen-tripel (A, B, C) ist dissipativ, d.h. es existiert eine positiv definite Matrix $Q \in \mathbb{C}^{n \times n}$, so daß die Matrix $Q(A + BKC)$ für alle Matrizen $K \in \mathbb{C}^{n \times n}$ mit $\|K\| \leq 1$ dissipativ ist.*
- (3) *Die Ruhelage des Systems gemäß (4.5) ist global asymptotisch stabil.*

Dieses Stabilitätskriterium entspricht in seiner Form genau der Proposition 4.1, dem Stabilitätskriterium im Zusammenhang mit der Kontraktivität des Matrizen-tripels (A, B, C) . Und analog zu dort, ist hier der „Einbettung“ der Aussage (2) zu ersehen, daß für den Fall, daß die Matrix $A + BKC$ für alle $K \in \mathbb{C}^{n \times n}$ mit $\|K\| \leq 1$ normal ist, die Ruhelage des Systems gemäß (4.5) **genau dann global asymptotisch stabil ist**, wenn das Matrizen-tripel (A, B, C) dissipativ ist. Hier ist es allerdings nicht so einfach möglich, einige wesentliche Beispiele derart anzugeben, daß die „Normalitätsbedingung“ mehr oder weniger automatisch erfüllt ist. Trotzdem scheint es angebracht zu sein, den Zusammenhang der Bedingung „das Matrizen-tripel (A, B, C) ist dissipativ“ mit der Lösbarkeit der Gleichung $\mathcal{R}_s(X) = 0$ und der Ungleichung $\mathcal{R}_s(X) < 0$ zu klären zu versuchen.

6.1 Die Dissipativität von (a, b, c) für komplexe Zahlen

Für komplexe Zahlen $m \in \mathbb{C}$ sind die Begriffe dissipativ und stabil äquivalent. Denn $\operatorname{Re}(m) < 0$ ist genau dann erfüllt, wenn für alle $x \in \mathbb{C}$ mit $x \neq 0$ gilt: $\operatorname{Re}(x^*mx) = |x|^2 \operatorname{Re}(m) < 0$.

Betrachte nun unter der Voraussetzung a stabil, d.h. $\operatorname{Re}(a) < 0$, die Diskriminanten-Bedingung $(\operatorname{Re}(a))^2 > |bc|^2$ für die Existenz reeller und stabilisierender Lösungen der Riccati-Gleichung $\mathcal{R}_s(x) = 0$, d.h. die Bedingung dafür, daß die Hamilton-Matrix H_s ohne rein imaginären Eigenwert ist. Wird beachtet, daß mit der Abkürzung $K_1 := \{k \in \mathbb{C} \mid |k| \leq 1\}$ gilt

$$|bc| = \max_{k \in K_1} \operatorname{Re}(bkc) \geq \operatorname{Re}(bkc) \quad \text{für alle } k \in K_1, \quad (6.2)$$

so gilt unter der Voraussetzung $\operatorname{Re}(a) < 0$:

$$\begin{aligned} (\operatorname{Re}(a))^2 > |bc|^2 &\iff |\operatorname{Re}(a)| > |bc| \iff |\operatorname{Re}(a)| - |bc| > 0 \iff -|\operatorname{Re}(a)| + |bc| < 0 \\ &\iff \operatorname{Re}(a) + |bc| < 0 \iff \operatorname{Re}(a) + \operatorname{Re}(bkc) < 0 \text{ für alle } k \in K_1 \\ &\iff \operatorname{Re}(a + bkc) < 0 \text{ für alle } k \in K_1 \\ &\iff \text{Es existiert ein } q > 0 \text{ mit } \operatorname{Re}(q(a + bkc)) < 0 \text{ für alle } k \in K_1 \\ &\iff \text{Es existiert ein } q > 0 \text{ mit } \operatorname{Re}(x^*q(a + bkc)x) < 0 \text{ für alle } k \in K_1 \\ &\quad \text{und alle } x \neq 0 \\ &\iff (a, b, c) \text{ ist dissipativ.} \end{aligned} \quad (6.3)$$

Wegen der Herleitung (4.6) und dem Satz 5.1 folgt daher, da die Dissipativität von (a, b, c) natürlich die Stabilität von a impliziert:

Hilfssatz 6.1. *Die folgenden drei Aussagen (1), (2) und (3) sind äquivalent.*

- (1) *Es existiert eine positiv definite Lösung der algebraischen Riccati-Ungleichung $\mathcal{R}_s(x) < 0$, d.h. es existiert eine komplexe Zahl $x_2 \in \mathbb{C}$ mit den folgenden zwei Eigenschaften (a) und (b):*
 - (a) $x_2 > 0$.
 - (b) $\mathcal{R}_s(x_2) = a^*x_2 + x_2a + x_2bb^*x_2 + c^*c < 0$.
- (2) *(a, b, c) ist dissipativ, d.h. es existiert eine positiv definite komplexe Zahl $q \in \mathbb{C}$, so daß die komplexe Zahl $q(a + bkc)$ für alle komplex Zahlen $k \in \mathbb{C}$ mit $|k| \leq 1$ dissipativ ist.*
- (3) *(a, b, c) ist kontraktiv, d.h. a ist stabil, und für alle $\omega \in \mathbb{R}$ ist $|c(j\omega - a)^{-1}b| < 1$.*

6.2 Die Dissipativität von (A, B, C) für Matrizen

Das folgende Lemma 6.1 entspricht dem letzten Hilfssatz 6.1.

Lemma 6.1. *Die folgenden drei Aussagen (1), (2) und (3) sind äquivalent.*

- (1) *Es existiert eine positiv definite Lösung der algebraischen Riccati-Ungleichung $\mathcal{R}_s(X) < 0$, d.h. es existiert eine Matrix $X_2 \in \mathbb{C}^{n \times n}$ mit den folgenden zwei Eigenschaften (a) und (b):*
 - (a) $X_2 > 0$.
 - (b) $\mathcal{R}_s(X_2) = A^*X_2 + X_2A + X_2BB^*X_2 + C^*C < 0$.
- (2) *(A, B, C) ist dissipativ, d.h. es existiert eine positiv definite Matrix $Q \in \mathbb{C}^{n \times n}$, so daß die Matrix $Q(A + BKC)$ für alle Matrizen $K \in \mathbb{C}^{n \times n}$ mit $\|K\| \leq 1$ dissipativ ist.*
- (3) *(A, B, C) ist kontraktiv, d.h. A ist stabil, und für alle $\omega \in \mathbb{R}$ ist $\|C(j\omega I - A)^{-1}B\| < 1$.*

In diesem Lemma 6.1 sind lediglich die neuen Gesichtspunkte zusammengefaßt, die aufgrund der Betrachtung der Dissipativität des Matrizen Tripels (A, B, C) in die Theorie der algebraischen Riccati-Gleichung $\mathcal{R}_s(X) = 0$ (bzw. algebraische Riccati-Ungleichung $\mathcal{R}_s(X) < 0$) eingebracht werden. Eine Zusammenfassung dieser Gesichtspunkte mit den Aussagen der vorhergehenden Abschnitte 4 und 5 soll erst im nächsten Abschnitt 7 erfolgen.

Die Aussagen von Lemma 6.1 sind — ebenso wie die in [5] hierzu angegebenen Beweise — als neu anzusehen (so glaube ich jedenfalls). Lediglich die Gültigkeit der Implikation (1) $\implies$ (2) von Lemma 6.1 wurde bereits in der Arbeit [4] bewiesen.

7 Die Ergebnisse für den Fall gleichartiger Terme

Nun zu dem Theorem, in dem die Aussagen der Abschnitte 4.2, 5.2 und 6.2 zusammengefaßt werden.

Theorem 7.1. *Für $A, B, C \in \mathbb{C}^{n \times n}$ sind die folgenden sechs Aussagen (1) – (6) äquivalent.*

- (1) *(A, B, C) ist dissipativ, d.h. es existiert eine positiv definite Matrix $Q \in \mathbb{C}^{n \times n}$, so daß die Matrix $Q(A + BKC)$ für alle Matrizen $K \in \mathbb{C}^{n \times n}$ mit $\|K\| \leq 1$ dissipativ ist.*
- (2) *(A, B, C) ist kontraktiv, d.h. A ist stabil, und für alle $\omega \in \mathbb{C}$ ist $\|C(j\omega I - A)^{-1}B\| < 1$.*
- (3) *A ist stabil, und H_s ist ohne rein imaginären Eigenwert.*
- (4) *Es existiert (genau) eine positiv semidefinite und stabilisierende Lösung der algebraischen Riccati-Gleichung $\mathcal{R}_s(X) = 0$, d.h. es existiert (genau) eine Matrix $X_0 \in \mathbb{C}^{n \times n}$ mit den folgenden drei Eigenschaften (a), (b) und (c):*
 - (a) $X_0 \geq 0$.
 - (b) $A + BB^*X_0$ stabil.
 - (c) $\mathcal{R}_s(X_0) = A^*X_0 + X_0A + X_0BB^*X_0 + C^*C = 0$.
- (5) *Es existiert eine positiv definite und stabilisierende Lösung der algebraischen Riccati-Ungleichung $\mathcal{R}_s(X) < 0$, d.h. es existiert eine Matrix $X_1 \in \mathbb{C}^{n \times n}$ mit den folgenden drei Eigenschaften (a), (b) und (c):*
 - (a) $X_1 > 0$.
 - (b) $A + BB^*X_1$ stabil.
 - (c) $\mathcal{R}_s(X_1) = A^*X_1 + X_1A + X_1BB^*X_1 + C^*C < 0$.
- (6) *Es existiert eine positiv definite Lösung der algebraischen Riccati-Ungleichung $\mathcal{R}_s(X) < 0$, d.h. es existiert eine Matrix $X_2 \in \mathbb{C}^{n \times n}$ mit den folgenden zwei Eigenschaften (a) und (b):*
 - (a) $X_2 > 0$.
 - (b) $\mathcal{R}_s(X_2) = A^*X_2 + X_2A + X_2BB^*X_2 + C^*C < 0$.

Für einzelne Aussagen in diesem Theorem 7.1 sind in der Literatur viele äquivalente Formulierungen zu finden. So können beispielsweise die positiv reellen Funktionen oder die linearen Matrizen-Ungleichungen ins Spiel gebracht werden.

In diesem Theorem 7.1 sind die grundlegenden Ergebnisse, die im Zusammenhang mit der algebraischen Riccati-Gleichung und der Stabilitätstheorie (bzw. der Analyse und dem Entwurf robuster Regelungen)

eine Rolle spielen, in sehr kompakter Form zusammengestellt. Wobei ich glaube, daß die hier erfolgte Einbeziehung der Dissipativität des Matrizen Tripels (A, B, C) wirklich neu ist. In der Arbeit [4] konnte ich nämlich nur die Implikation $(6) \implies (1)$, nicht aber die Implikation von (1) auf eine der Aussagen $(2) - (6)$ beweisen.

Dieses Theorem 7.1 war ursprünglich nur für die Gewinnung von Stabilitätsaussagen für zeitveränderliche Systeme gedacht (siehe die Ausführungen zu Beginn der Abschnitte 4 und 6). Aber natürlich kann aus diesem Theorem vieles weiteres gefolgert werden. So wird beispielsweise die Äquivalenz der beiden Aussagen (2) und (3) zur Bestimmung der H_∞ -Norm eines Übertragungssystems verwendet. Mit den Abkürzungen

$$\|G\|_\infty := \max_{\omega \in \mathbb{R}} \|G(j\omega)\| \quad \text{mit} \quad G(j\omega) := C(j\omega I - A)^{-1}B \quad \text{für} \quad \omega \in \mathbb{R}, \quad (7.1)$$

$$H_{s,\gamma} := \begin{bmatrix} A & \gamma^{-2}BB^* \\ -C^*C & -A^* \end{bmatrix} \quad \text{für} \quad \gamma > 0 \quad \text{und} \quad (7.2)$$

$$\gamma_0 := \inf\{\gamma \mid H_{s,\gamma} \text{ ohne rein imaginären Eigenwert}\} \quad (7.3)$$

gilt nämlich die Aussage: Ist A stabil, so ist $\|G\|_\infty = \gamma_0$. Diese Aussage kann zufolge Lemma 4.1 sogar zu folgendem Theorem 7.2 verallgemeinert werden.

Theorem 7.2. *Ist A ohne rein imaginären Eigenwert, so ist $\|G\|_\infty = \gamma_0$.*

8 Schlußbemerkungen

In den deutschen Lehrbüchern zur Regelungstheorie/Kontrolltheorie kommt nahezu ausschließlich nur die algebraische Riccati-Gleichung der Form $\mathcal{R}_d(X) = 0$ des Abschnitts 3 mit unterschiedlich semidefiniten Termen zum Einsatz. Siehe in [8], [12], [16], [17], [18], [19], [20], [22], [24] und [25]. Lediglich in [22] und [24] spielt auch die algebraische Riccati-Gleichung der Form $\mathcal{R}_s(X) = 0$ des Abschnitts 4 eine gewisse Rolle. Die algebraische Riccati-Ungleichung der Form $\mathcal{R}_s(X) < 0$ des Abschnitts 5 mit gleichartig semidefiniten Termen und das Lemma von Kalman-Yakubovich-Popov habe ich in diesen Lehrbüchern nicht gefunden.

In dieser Arbeit wurden nur algebraische Riccati-Gleichungen der Form $A^*X + XA + XRX + Q = 0$ mit semidefinitem R und hermischem Q betrachtet, im Abschnitt 3 mit R negativ und Q positiv semidefinit (unterschiedlich semidefinite Terme) und in den Abschnitten 4 bis 7 mit R und Q positiv semidefinit (gleichartig semidefinite Terme). Wirklich wesentlich war aber nur, daß R und Q im Abschnitt 3 vom unterschiedlichen Typ und in den Abschnitten 4 bis 7 vom gleichen Typ semidefinit waren. Sollten dagegen auch für algebraische Riccati-Gleichungen/Ungleichungen bei indefinitem (weder positiv noch negativ semidefinitem) R analoge Ergebnisse gewonnen werden, so wäre viel weiter auszuholen, und die Arbeit würde statt der hier vorliegenden 18 Seiten sehr schnell einen Umfang von mehreren hundert Seiten annehmen (siehe z.B. bei LANCASTER und RODMAN in [15]). In [22] und [24] kommen derartig allgemeinere Gleichungen teilweise zum Einsatz.

Und zum Schluß zu der Frage, ob sich die Beweise der hier angegebenen Lemmata und Theoreme — zumindest prinzipiell — auch in einer beliebigen B^* - bzw. C^* -Algebra ausführen lassen, ob also das „Eigenvektor-Verfahren zur Lösung der algebraischen Riccati-Gleichung“ auch in linearen Räumen mit nicht endlicher Dimension ausgeführt werden kann. Die Beweise der Lemmata 5.1 und 5.2 würden beispielsweise überhaupt keine Schwierigkeit bereiten. Ganz anders sieht es aber bei den Beweisen zum grundlegenden Theorem 2.1 und zum grundlegenden Lemma 4.1 aus. In diesen Beweisen wird bisher ganz wesentlich mit Eigenwerten und Eigenvektoren von Matrizen gearbeitet. Und vermutlich lassen sie sich nicht in beliebige B^* - bzw. C^* -Algebren übertragen. Ich sehe bisher nämlich keine Möglichkeit hierfür.

Literatur

- [1] T. Başar, „Generalized Riccati Equations in Dynamical Games“, in [2].
- [2] S. Bittanti, A. J. Laub, J. C. Willems, *The Riccati Equation*, Springer-Verlag, Berlin, 1991.

- [3] D. Dreyer, *Stabilitätskriterien für Differentialgleichungen in B - und in B^* -Algebren*, Diplomarbeit, Fachbereich Mathematik der TU-Berlin, Berlin, 1975.
- [4] D. Dreyer, *Die absolute Stabilität linearer Differentialgleichungen in B - und in B^* -Algebren*, Dissertation, Fachbereich Mathematik der TU-Berlin, Berlin, 1978.
- [5] D. Dreyer, „Ein Beitrag zur algebraischen Riccati-Gleichung. Hier: Die Beweise der Aussagen eines Vortrages“, Umdruck, Berlin, 2001.
- [6] G. E. Dullerud, F. Paganini, *A Course in Robust Control Theory: A Convex Approach*, Springer-Verlag, New York, 2000.
- [7] G. von Escherich, „Die zweite Variation der einfachen Integrale“, *Wiener Sitzungsberichte*, 107, 1191–1250, 1889.
- [8] A. H. Glattfelder, W. Schauffelberger, *Lineare Regelsysteme: Eine Einführung mit MATLAB*, vdf Hochschulverlag, Zürich, 1997.
- [9] R. A. Horn, C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, U.K., 1985.
- [10] R. A. Horn, C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, U.K., 1991.
- [11] V. Ionescu, C. Oară, M. Weiss, *Generalized Riccati Theory and Robust Control: A Popov Function Approach*, Wiley & Sons, Chichester, 1999.
- [12] H. W. Knobloch, H. Kwakernaak, *Lineare Kontrolltheorie*, Springer-Verlag, Berlin, 1985.
- [13] V. Kučera, „Algebraic Riccati Equation: Hermitian and Definite Solutions“, in [2].
- [14] P. Lancaster, L. Rodman, „Solutions of the Continuous and Discrete Time Algebraic Riccati Equation: A Review“, in [2].
- [15] P. Lancaster, L. Rodman, *Algebraic Riccati Equations*, Clarendon Press, Oxford, 1995.
- [16] A. J. Laub, „Invariant Subspace Methods for Numerical Solution of Riccati Equations“, in [2].
- [17] G. Ludyk, *CAE von Dynamischen Systemen: Analyse, Simulation, Entwurf von Regelungssystemen*, Springer-Verlag, Berlin, 1990.
- [18] G. Ludyk, *Theoretische Regelungstechnik 1: Grundlagen, Synthese linearer Regelungssysteme*, Springer-Verlag, Berlin, 1995.
- [19] G. Ludyk, *Theoretische Regelungstechnik 2: Zustandsrekonstruktion, optimale und nichtlineare Regelungssysteme*, Springer-Verlag, Berlin, 1995.
- [20] J. Lunze, *Regelungstechnik 2: Mehrgrößensysteme, Digitale Regelung*, Springer-Verlag, Berlin, 1997.
- [21] C. F. Martin, G. Ammar, „The Geometry of the Matrix Riccati Equation and Associated Eigenvalue Methods“, in [2].
- [22] K. Müller, *Entwurf robuster Regelungen*, Teubner, Stuttgart, 1996.
- [23] V. M. Popov, *Hyperstability of Control Systems*, Springer-Verlag, Berlin, 1973.
- [24] J. Raisch, *Mehrgrößenregelung im Frequenzbereich*, Oldenbourg, München, 1994.
- [25] H. Schlitt, *Regelungstechnik: Physikalisch orientierte Darstellung fachübergreifender Prinzipien*, Vogel, Würzburg, 1993.
- [26] A. Shayman, „A Geometric View of the Matrix Riccati Equation“, in [2].
- [27] H. L. Trentelman, J. C. Willems, „The Dissipation Inequality and the Algebraic Riccati Equation“, in [2].
- [28] Fuzhen Zhang, *Matrix Theory: Basic Results and Techniques*, Springer-Verlag, New York, 1999.

Stabilität von Regelkreisen in Feldbussystemen

R. Tracht, Th. Eymann

Universität Essen

Durch Nutzung von seriellen Bussystemen (Feldbussystemen) kann der Hardware-Aufwand für verteilte Automatisierungssysteme gravierend reduziert werden. Allerdings muss dann berücksichtigt werden, dass zeitvariable Verzögerungen bei der Datenübertragung auftreten. Feldbussysteme haben sich für die Realisierung von Steuerungen allgemein durchgesetzt. Wenn der Feldbus auch für die Regelung genutzt wird, kann dies zu Stabilitätsproblemen führen. Im folgenden Beitrag wird zunächst auf die Ursachen der Verzugszeiten bei der Datenübertragung eingegangen. Es wird dann für lineare zeitinvariante Regelstrecken eine lineare Matrix-Ungleichung angegeben, die es gestattet, Grenzwerte für die Datenübertragungszeiten auszurechnen. Es zeigt sich, dass zusätzlich die Änderungsgeschwindigkeit der Verzugszeiten begrenzt werden muss. Die angegebenen hinreichenden Bedingungen können für den Reglerentwurf genutzt werden. Eine Überprüfung des Regelkreisverhaltens durch Simulationsstudien muss dann Bestandteil des Reglerentwurfs sein. Dazu wird abschließend ein Simulationstool vorgestellt.

1. Einleitung

Die Automatisierung von Anlagen basiert heutzutage auf "intelligenten" Geräten mit eingebauten Mikrocontrollern, die nicht mehr zentral in einem Schaltschrank lokalisiert sind, sondern in der Anlage verteilt sind. Für die Steuerungsaufgaben stehen maßgeschneiderte Module zur Verfügung, die einen Großteil der Aufgaben an Ort und Stelle erledigen. Nach wie vor erfolgt jedoch ein Datenaustausch zwischen den einzelnen Teilsystemen. Anstelle von parallelen Verbindungen zu einem zentralen Automatisierungsgerät wird die Feldbustechnik eingesetzt, die zu einer gravierenden Reduzierung der Verdrahtungskosten führt. Darüber hinaus werden die Möglichkeiten zur Systemüberwachung und zur Systemerweiterung erheblich verbessert.

Die Kommunikation über den Feldbus erfolgt seriell. Je nach Bustyp kann ein Busknoten (z. B. ein Sensor) angesprochen werden oder aber auch selbständig auf den Feldbus zugreifen. Im

Gegensatz zu parallel verdrahteten Systemen hängt jetzt jedoch die Zeit für die Übertragung der Daten von der Busauslastung ab.

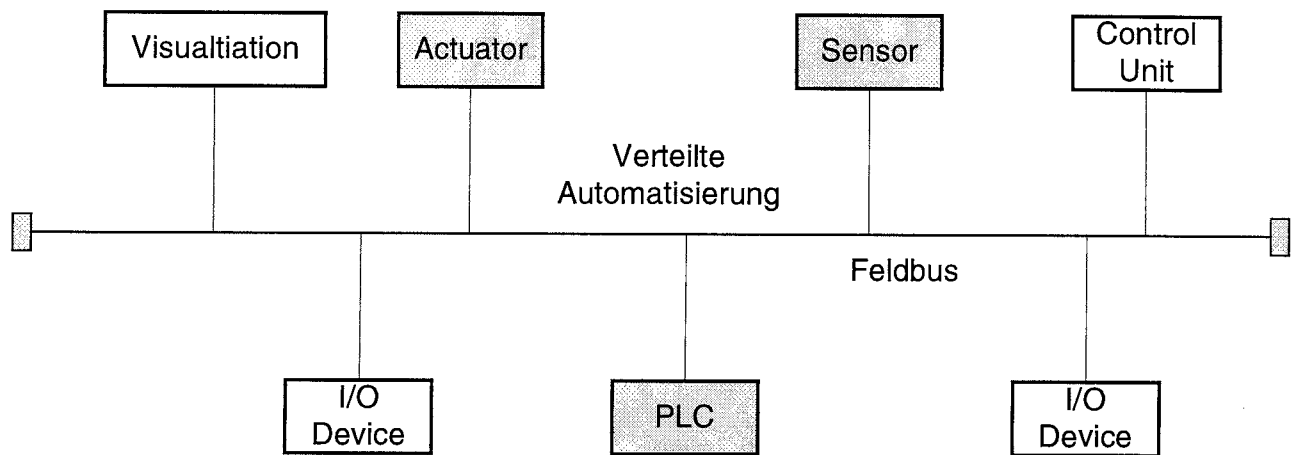


Bild 1: Feldbussystem

In vielen Anwendungen muss innerhalb einer vorgegebenen Zeit reagiert werden. Dies setzt natürlich voraus, dass auch die Datenübertragung innerhalb vorgegebener Grenzen erfolgt. Vor allem für zyklische Busanforderungen, wie sie beispielsweise bei der Regelung auftreten, müssen Zeitverzögerungen in engen Grenzen liegen, da sonst instabiles Regelkreisverhalten auftreten kann. Während bei Steuerungsaufgaben überwiegend binäre Variablen verarbeitet werden, hat man bei der Regelung im allgemeinen kontinuierliche Größen zu berücksichtigen. Die Regelgröße wird mit einem Sensor erfasst und an den Regler weitergeleitet. Im Regler werden mit einem geeigneten Regelalgorithmus die Stellgrößen berechnet und an die Aktoren übermittelt [1].

Die Erfassung der Messgrößen erfolgt ebenso zyklisch wie die Aufschaltung der Stellgrößen. Die gemessenen Größen müssen digitalisiert werden und mit einem geeigneten Datentelegramm über den Feldbus an den Mikrocontroller, der die Regelung übernimmt, übertragen werden. Dementsprechend wird nach der Berechnung der Stellgröße vom Regler ein Datentelegramm an den Aktor verschickt. Das Feldbussystem muss für jeden Regelkreis um diese drei Knoten erweitert werden.

Für gleichzeitig ablaufende Steuerungsaufgaben müssen außerdem digitale Eingänge und digitale Ausgänge zur Verfügung stehen.

Zur Beurteilung des Zeitverhaltens müssen außer den Aktoren, den Sensoren und der Datenübertragung die dynamischen Eigenschaften der zu regelnden Anlage mit berücksichtigt werden. Die Dynamik der Regelkreise kann durch ein geeignetes mathematisches Modell analysiert werden. Das Stabilitätsverhalten wird wesentlich durch die Zeitverzögerungen als Folge der Datenübertragung beeinflusst.

Im Folgenden wird daher zunächst untersucht, in welcher Größenordnung die Zeitverzögerungen bei den beiden Feldbustypen mit der größten Verbreitung liegen. Es wird dann ein Kriterium angegeben, das eine hinreichende Bedingung für die Stabilität von Regelkreisen in Feldbussystemen sicherstellt. Abschließend wird auf die Simulation von Feldbussystemen eingegangen.

Maßgebend für die Übertragungszeiten von Daten auf einem Feldbus ist die Art des Buszugriffs. Im nächsten Abschnitt wird daher zunächst auf die Buszugriffsverfahren eingegangen.

2. Buszugriffsverfahren

Im Unterschied zu einer Punkt zu Punkt-Verbindung muss bei einem Bussystem jeder Knoten die Möglichkeit erhalten, innerhalb einer vorgegebenen Zeit Daten zu übertragen. Der genaue Zeitpunkt, wann dies geschieht, kann entweder zentral durch einen „Master“ vorgegeben werden (Master-Slave-System), oder vom Knoten selbst initiiert werden (Multi-Master-System). Beim zuerst genannten Bustyp wird in einer Liste (der Polling-Liste) festgelegt, in welcher Reihenfolge die Buszuteilung an die einzelnen Knoten erfolgt. Beim Multi-Master-System muss dagegen durch die Wahl von Prioritäten vorgegeben werden, welcher Knoten bei gleichzeitigem Mehrfachzugriff bevorrechtigt ist. Oft werden beide Prinzipien miteinander kombiniert. Mehrere Master-Busknoten erhalten in diesem Fall nacheinander die Buszugriffsberechtigung und versorgen jeweils eine Gruppe von Slaves.

Typische Vertreter für die Multi-Master-Systeme sind die CAN-Bus basierten Feldbussysteme (z.B. Device-Net) [2,3]. Die Priorität wird hier für Nachrichten vergeben. Jeder Busknoten erhält eine Liste mit „Identifiern“ von Nachrichten und kann damit erkennen, ob eine Nachricht für ihn bestimmt ist. Je kleiner der Identifier ist, um so höher ist die Priorität der Nachricht. In einer Nachricht (Data Frame) können beim CAN-Bus maximal 64 Bytes übertragen werden. Neben dem Identifier und den Nachrichten-Bytes werden Steuerbits und Sicherungsbits (CRC-Bits) übertragen. Die Übertragungszeit ist abhängig von der Busbelastung. Eine maximale Übertragungszeit kann nur für die höchstprioräre Nachricht angegeben werden. Sie liegt bei 154 Mikrosekunden, wenn eine Bit-Übertragungsrate von 1Mbit/s genutzt wird. Da jeder Knoten innerhalb einer Bitzeit entscheiden muss, ob die von

ihm gesendete Nachricht von niedriger Priorität als eine eventuell gleichzeitig gesendete Nachricht ist, kann die Buslänge bei dieser Bit-Übertragungsrate maximal 100 m lang sein.

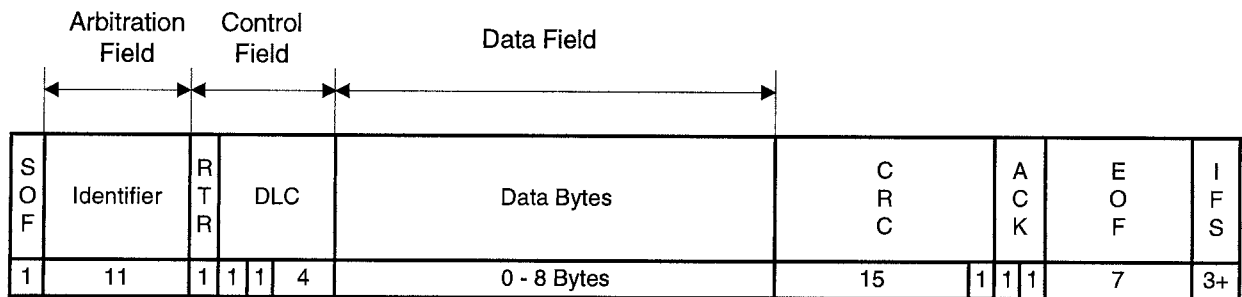


Bild 2: CAN-Bus Data-Frame

Beim Profibus werden Datentelegramme an die Slaves verschickt. In einem Datentelegramm können bis zu 246 Datenbytes verschickt werden. Daneben muss die Anzahl der Datenbytes, die Zieladresse, die Quelladresse und Information zur Datensicherung mitübertragen werden.

SD	LE	LEr	SD	DA	SA	FC	DSAP	SSAP	DU	FCS	ED
68h	1 Byte	1	1	1	1	1	3Ch	3Eh	1 ... 246 Bytes	1 Byte	16h

Bild 3: Profibus-Telegramm

Wenn z.B. ein Sensor aufgefordert wird, Messwerte zu übermitteln, sind zwei Telegramme erforderlich. Zur Übertragungszeit kommt dann noch die Wartezeit, die vom Platz des Knoten in der Polling-Liste abhängt. Die Bitübertragungsrate kann beim Profibus sehr viel höher sein als beim CAN-Bus (z.B. 12 Mbit/s). Die Übertragungszeit für ein Datentelegramm liegt dann wie beim CAN-Bus in der Größenordnung von 150-200 Mikrosekunden.

3. Regelkreis-Strukturen

Wenn die Datenübertragung innerhalb eines Regelkreises über einen Feldbus erfolgen soll – also der Sensor, der Regler und der Aktor jeweils ein Knoten ist – dann ergibt sich für einen Abtastregelkreis die in Bild 4 dargestellte Struktur [4]. Der Feldbus stellt also im wesentlichen ein Totzeitglied mit veränderlichen Totzeiten dar. Wenn ein mathematisches Modell der Regelstrecke bekannt ist und innerhalb der Totzeit keine Störungen auftreten,

dann kann ein Prädiktor entworfen werden, der bei bekannter Totzeit den aktuellen Systemzustand für die Stellgrößenberechnung liefert. Gerätetechnisch kann dies z.B. so realisiert werden, dass Sensor und Regler mit einer synchronisierten Uhr versehen werden und der Sensor bei jedem Messwert einen Zeitstempel mit überträgt. Aus der Zeitdifferenz zwischen der aktuellen Zeit im Regler und der im Zeitstempel angegebenen Zeit kann die im Prädiktor einzusetzende Zeit bestimmt werden. Die Totzeit zwischen Regler und Aktor kann auf diese Weise allerdings nicht kompensiert werden. Wenn auch der Aktor über eine synchronisierte Uhr verfügt, kann statt dessen bei der Stellgrößenberechnung eine Verzögerung von einem Abtastschritt einkalkuliert werden. Erfolgt die Datenübertragung innerhalb eines Abtastschrittes, dann kann der Aktor – falls er hinreichend intelligent ist – die Stellgröße unverzüglich aufschalten.

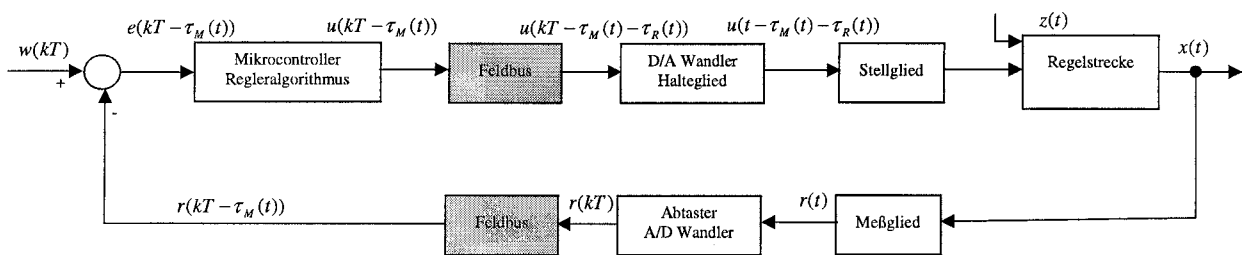


Bild 4: Abtastregelkreis

Um ohne die genannten gerätetechnischen Bedingungen eine Aussage über die Stabilität machen zu können, wird angenommen, dass die Abtastzeit hinreichend klein ist, so dass von einer quasikontinuierlichen Regelung ausgegangen werden kann. Außerdem wird vorausgesetzt, dass die Regelstrecke durch ein lineares mathematisches Modell beschrieben werden kann. In diesem Fall kann die Datenübertragung durch ein einziges Totzeitglied nachgebildet werden. Man erhält damit die in Bild 5 dargestellte Regelkreisstruktur.

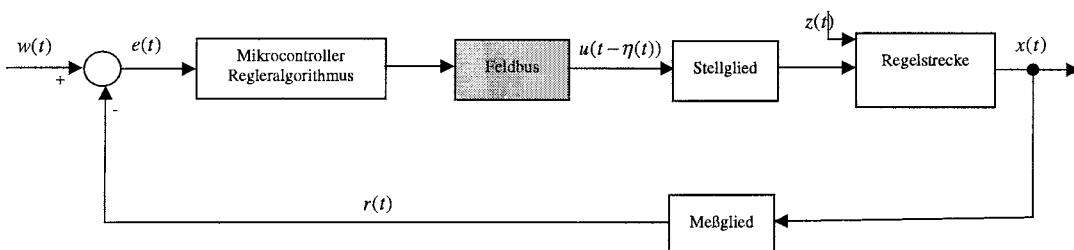


Bild 5: Ersatzsystem

4. Stabilitätsanalyse

Unter den oben genannten Voraussetzungen kann die Regelstrecke durch ein lineares Differentialgleichungssystem mit verzögerten Eingangsgrößen beschrieben werden.

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

Nimmt man an, dass ein Zustandsregler

$$u(t) = -Fx(t)$$

für die Regelung eingesetzt wird, dann ergibt sich ein homogenes lineares System:

$$\dot{x}(t) = Ax(t) - BFx(t - \eta(t))$$

Störungen werden durch den Anfangszustand berücksichtigt.

Die Stabilitätsanalyse dieses Systems kann mit Hilfe der direkten Methode von Ljapunow erfolgen [4]. Die Ljapunow-Funktion muss neben dem aktuellen Systemzustand die Dynamik des Systems innerhalb des durch die Totzeit vorgegebenen Zeitintervalls mitberücksichtigen. Einen Zusammenhang zwischen dem aktuellen Systemzustand und dem verzögerten Systemzustand erhält man aus der Systemgleichung:

$$x(t - \eta(t)) = x(t) - \int_{-\eta(t)}^0 \dot{x}(t + \theta) d\theta$$

Dieser Zusammenhang kann dazu genutzt werden um die Ljapunow-Funktion zu erweitern. Man erhält dann eine Lyapunow-Krasovskii-Funktion, die zusätzlich zu der üblichen quadratischen Form Integral-Terme enthält:

$$V(x, t) = V_0(x, t) + V_1(x, t) + V_2(x, t) + V_3(x, t)$$

mit

$$V_0(x, t) = x^T(t)Px(t)$$

$$V_1(x, t) = d_1 \int_{-\eta(t)}^0 \int_{t+\theta}^t x^T(s) A^T A x(s) ds d\theta$$

$$V_2(x, t) = d_2 \int_{-\eta(t)}^0 \int_{t-\eta(t+\theta)+\theta}^t x^T(s) F^T B^T B F x(s) ds d\theta$$

$$V_3(x, t) = d_3 \int_{-\eta(t)}^0 \int_{t+\theta}^t x^T(s) F^T B^T B F x(s) ds d\theta$$

In den Integralgrenzen erscheint die variable Totzeit. Die Ableitung der Lyapunow-Krasovskii-Funktion enthält daher neben der Lyapunowgleichung noch Terme, die durch eine maximale Totzeit und durch den Maximalwert der Ableitung der Totzeit nach der Zeit abgeschätzt werden können:

$$\begin{aligned} \dot{V}(x, t) &< (A - BF)X + X(A - BF)^T \\ &+ \eta^* \left(\frac{\alpha}{1 - 2\eta^*} + \frac{\beta}{1 - 2\eta^* + \dot{\eta}^{*2}} \right) B F F^T B^T \\ &+ \alpha^{-1} \eta^* X A^T A X + 2\beta^{-1} \eta^* X F^T B^T B F X \end{aligned}$$

mit $X = P^{-1}$, α und β sind freie Parameter.

Ergebnis ist eine lineare Matrix-Ungleichung,

$$\begin{bmatrix} (A - BF)X + X(A - BF)^T + \eta^* \left(\frac{\alpha}{1 - 2\eta^*} + \frac{\beta}{1 - 2\eta^* + \eta^{*2}} \right) BFF^T B^T & \eta^* XA^T & 2\eta^* XF^T B^T \\ \eta^* AX & -\alpha\eta^* I & 0 \\ 2\eta^* BFX & 0 & -2\beta\eta^* I \end{bmatrix} < 0$$

die mit vorhandenen Programmen (z.B. der Matlab-LMI-Toolbox) untersucht werden kann. (In der Literatur sind in den letzten Wochen ähnliche Bedingungen vorgestellt worden [5]). Besitzt die Matrix-Ungleichung eine Lösung, dann ist das System für den gewählten Zustandsregler und den gefundenen Grenzen für die maximale Totzeit und deren Ableitung stabil. Ausgehend davon kann nun der Regler modifiziert werden um das Regelkreisverhalten zu verbessern [4,6].

Falls keine Lösung gefunden wird, muß der ein anderer Regler mit geringeren Anforderung gewählt werden werden.

5. Simulation von Feldbussystemen

Die Stabilitätsanalyse liefert nur Information über die maximal zulässige Totzeit und deren Änderungsgeschwindigkeit also über die Grenzen der Busbelastung. Der Einfluss einer speziellen Busbelastung auf das Zeitverhalten in Feldbussystemen kann nur bedingt vorab berechnet werden. In vielen Fällen handelt es sich bei den auftretenden Ereignissen um stochastische Prozesse, für die nur Verteilungen bekannt sind. Andererseits ist es natürlich außerordentlich wünschenswert, bereits bei der Systementwicklung den Einfluss von zufällig auftretenden Busanforderungen auf das Systemverhalten beurteilen zu können. Für die Entwicklung von Regelkreisen werden in solchen Fällen Simulationsexperimente durchgeführt. Dadurch können auch Extremsituationen, die in der Realität zur Zerstörung des Systems führen würden, untersucht werden. Es liegt nahe, die vorhandenen Möglichkeiten so zu erweitern, dass Feldbuseigenschaften mit in die Simulation einbezogen werden können.

Ein leistungsfähiges Entwurfswerkzeug, das über umfangreiche Möglichkeiten zur Entwicklung von Simulationsmodellen und zur Visualisierung der Simulationsergebnisse verfügt, ist das Simulationssystem MATLAB/SIMULINK. Dieses System wurde durch eine Toolbox mit Simulationsblöcken für die drei verbreitetsten Feldbussysteme, nämlich der CAN-Bus, der Profibus und der Interbus-S, erweitert [7].

Während bei Steuerungsaufgaben überwiegend binäre Variablen verarbeitet werden, hat man bei der Regelung im allgemeinen kontinuierliche Größen zu berücksichtigen. Diese Größen

müssen digitalisiert werden und wie beschrieben mit einem geeigneten Datentelegramm über den Feldbus an den Mikrocontroller, der die Regelung übernimmt, übertragen werden. Dementsprechend wird nach der Berechnung der Stellgröße vom Regler ein Datentelegramm an den Aktor verschickt. Das Feldbussystem muss für jeden Regelkreis um diese drei Knoten erweitert werden.

Für gleichzeitig ablaufende Steuerungsaufgaben müssen außerdem digitale Eingänge und digitale Ausgänge zur Verfügung stehen.

Zur Beurteilung des Zeitverhaltens müssen außer den Aktoren, den Sensoren und der Datenübertragung die dynamischen Eigenschaften der zu regelnden Anlage mit berücksichtigt werden. Die Dynamik der Regelkreise kann durch ein geeignetes mathematisches Modell analysiert werden. Das Stabilitätsverhalten wird wie beschrieben durch die Zeitverzögerungen als Folge der Datenübertragung beeinflusst. Daher muss die Quantisierung und die Umsetzung der parallel anliegenden physikalischen Größen in serielle Datenströme für die Simulation nachgebildet werden. Datenübertragung und Systemdynamik sind gemeinsam zu untersuchen.

Simulation von CAN-Bus-Systemen muss die Simulink-Blockbibliothek so erweitert werden, dass die Einflüsse der verschiedenen Komponenten in den Simulationsmodellen berücksichtigt werden können. Der wichtigste Simulationsblock dient natürlich zur Nachbildung der Datenströme auf den CAN-Bus. Dazu wird ein Telegramm aufgebaut, das dem CAN-Standard entsprechend aus einem Identifier aus den maximal 8 Byte Daten und aus Informationen zur sicheren Übertragung – dem CRC-Byte – besteht. Das so gebildete Nachrichtenobjekt wird in ein Signal umgeformt, das dem Signal in einem realen CAN-System gleicht. Aus diesem Signal werden dann die Parameter, die für die Simulation des Restsystems erforderlich sind, ermittelt. Dazu gehören insbesondere die Zeitdauer der Datenübertragung und der Zeitpunkt, zu dem die Daten zur Verfügung stehen. Neben dem Datentelegramm stehen noch Telegramme zur Nachbildung von Fehlersituationen zur Verfügung. Außerdem kann über spezielle Telegramme von einem Knoten Information von einem anderen Knoten angefordert werden.

Um das Signal für die Datenübertragung generieren zu können, muss die Darstellung kontinuierlicher Größen bekannt sein. In realen Automatisierungssystemen werden Analog/Digitalwandler mit 10 Bit oder 12 Bit Wortbreite eingesetzt. Entsprechende Blöcke stehen für die Ankopplung des Feldbussystems an die analoge Welt auch in der Feldbus-Toolbox zur Verfügung. Für ereignisgesteuerte Simulationen wurde ein CAN-Interrupt-

Controller bereitgestellt. Zur Visualisierung der Signale auf den CAN-Bus steht ein Bus-Monitor-Block zur Verfügung. Mit diesen Blöcken ist es im Prinzip möglich, das Feldbussystem zu simulieren. Im allgemeinen wird man jedoch nur das Simulationsmodell für einen Ausschnitt des Gesamtsystems nachbilden können. Um die Auswirkungen des Restsystems auf die Datenübertragung abschätzen zu können, wurden Datenquellen mit vorgegebenen stochastischen Eigenschaften entwickelt. Damit können Nachrichtenobjekte mit einer vorgegebenen Priorität und einer vorgegebenen Wiederholrate erzeugt werden. Die Wiederholung der Nachrichten kann entweder zyklisch, gleichverteilt, exponentialverteilt oder Poissonverteilt erfolgen.

Mit den beschriebenen Simulationsblöcken steht eine Simulink-Bibliothek zur Verfügung, die es erlaubt, ein „integriertes“ Simulationsmodell zu entwickeln und die leistungsfähigen Algorithmen von Simulink für die Simulation zu nutzen, um dann die Ergebnisse mit dem Simulink-Visualisierungssystem darzustellen.

Viele der oben beschriebenen Blöcke können durch Modifikation auch für die anderen beiden Bussysteme, die mit der Feldbus-Toolbox simuliert werden können, genutzt werden. Beim Profibus muss das hybride Verfahren für den Buszugriff, das aus einem Master-Master-Kommunikationsverfahren und aus einem Master-Slave-Verfahren besteht, nachgebildet werden. Die für die Simulation wichtigen Parameter von Slave-Knoten sind in den Gerätestammdateien gespeichert. Diese vom Hersteller zu liefernden Daten können vom Nutzer mit einer Matlabfunktion interaktiv festgelegt werden und für den Simulinkblock, durch den der Profibus- DP-Master nachgebildet wird, bereitgestellt werden. Der Interbus-S-Simulationsblock bildet im wesentlichen ein Schieberegister nach, durch das die Übertragungsdauer festgelegt wird. Die Anwendung in Simulink entspricht im wesentlichen derjenigen der beiden anderen Bussysteme.

6. Zusammenfassung und Ausblick

Feldbussysteme haben sich für die Realisierung von Steuerungen in verteilt automatisierten Systemen durchgesetzt. Es liegt nahe, diese Technik auch für die Regelung zu nutzen. Es muss dann allerdings berücksichtigt werden, dass variable Totzeiten in den Regelkreis eingeführt werden. Wie stark sich die serielle Datenübertragung auf die Totzeiten auswirkt hängt vom Buszugriffsverfahren ab. Für die beiden wichtigsten Feldbustypen wurde jeweils das Prinzip des verwendeten Verfahrens vorgestellt.

Unter der Voraussetzung, dass die Regelstrecke durch ein lineares zeitinvariantes mathematisches Modell beschrieben wird, kann mit Hilfe der direkten Methode von Ljapunow eine Matrix-Ungleichung hergeleitet werden. Damit ist es möglich, die Grenzen für die maximale Totzeit und deren Änderungsgeschwindigkeit zu berechnen.

Diese Ungleichung kann als Basis für den Reglerentwurf dienen. Ein Simulationstool, das ebenfalls vorgestellt wurde, ermöglicht Simulationsstudien zur Beurteilung des Regelkreisverhaltens.

Da die angegebene Ungleichung vom Riccati-Typ ist, soll in weiteren Untersuchungen überprüft werden, ob eine Systematisierung des Reglerentwurfs möglich ist.

Literatur

- [1] Bushnell, L. G.: Networks and Control, IEEE Control System Magazine, February 2001
- [2] Lian, Feng-Li, Moyne, J.R. and Tilbury, D..M.: Performance Evaluation of Control Networks, IEEE Control System Magazine, February 2001
- [3] Tracht, R., Eymann Th.: The Influence of Fieldbus Systems on Measurement and Control Timing, ISIST '99 Proceedings, China Metrology Press 1999
- [4] Eymann, Th.: Reglerentwurf für Regelkreise in Feldbussystemen, VDI Fortschritts-Bericht Nr. 886, März 2001
- [5] Kim, Jin-Hoon: Delay and Its Time-Derivative Dependent Robust Stability of Time-Delayed Linear Systems with Uncertainty, IEEE Transactions on Automatic Control Vol.46, No.5, May 2001
- [6] Zhang, W., Branicky, M.S., Phillips, S.M.: Stability of Networked Control Systems, IEEE Control System Magazine, February 2001
- [7] Eymann Th., Tracht R.: Feldbus-Toolbox- ein Werkzeug zur Simulation von Feldbussystemen, at-Automatisierungstechnik 48 (2000)

Beobachtung nichtlinearer Deskriptorsysteme

J. Menke
F. Gausch
Universität Paderborn

Kurzfassung

In diesem Beitrag wird ein Ansatz zur vollständigen Beobachtung nichtlinearer Deskriptorsysteme vorgestellt. Der Ansatz basiert direkt auf der semi-expliziten Form eines Deskriptorsystems und ermöglicht unter bestimmten Voraussetzungen die Beobachtung aller Zustands- und Deskriptorvariablen. Die Herleitung beginnt bei der Beobachtung linearer Deskriptorsysteme und bildet damit die Grundlage zur Beobachtung nichtlinearer Deskriptorsysteme, die mit Hilfe differenzial-geometrischer Methoden linearisiert werden.

1 Einführung

Der hier vorgestellte Ansatz bietet die Möglichkeit, einen vollständigen Beobachter für ein nichtlineares Deskriptorsystem zu entwerfen. Der Ansatz basiert auf der Theorie über Beobachter für lineare Systeme im Zustandsraum und der exakten Linearisierung auf der Basis differenzial-geometrischer Methoden nach [4]. Die Theorie zum Beobachterentwurf linearer Systeme im Zustandsraum wird zunächst auf lineare Deskriptorsysteme in semi-expliziter Form erweitert. Im Anschluss folgt eine Modifizierung des Verfahrens, um auch nichtlineare Deskriptorsysteme beobachten zu können. Dabei entsteht ein nichtlineares System, das mit den Verfahren zur exakten Linearisierung linearisiert wird. Für das entstandene lineare System wird dann ein Beobachter angegeben, der alle Zustands- und Deskriptorvariablen liefert. Der Beobachtungsfehler wird dabei durch eine lineare Differenzialgleichung beschrieben.

Deskriptorsysteme entstehen z.B. bei der Modellbildung technischer Prozesse, wenn es sich um die mathematische Beschreibung der Dynamik von Systemen handelt, die aus mehreren Teilsystemen zusammengesetzt sind. Ein Koppeln von dynamischen Teilsystemen führt auf ein differenzial-algebraisches System, wie das folgende Beispiel zeigen soll.

1.1 Beispiel

Das mechanische System in Bild 1.1 setzt sich aus zwei schwingungsfähigen Teilsystemen zusammen, die über eine Feder gekoppelt sind. Auf die Massen m_1 und m_2 wirkt eine Koppelkraft, die durch $F = c_3(x_2 - x_1)$ gegeben ist.

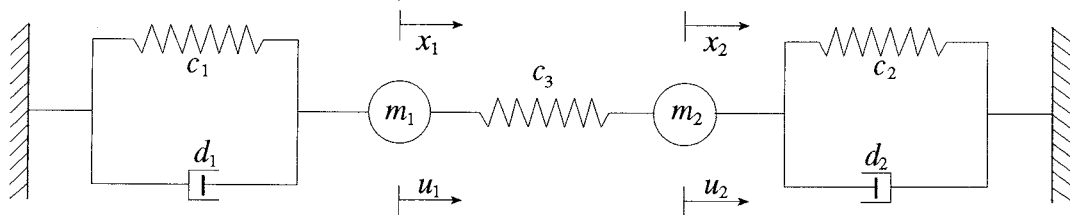


Bild 1.1: Mechanisches System

Die mathematische Beschreibung der Teilsysteme und die Berücksichtigung der Kopplung führt zu zwei Differenzialgleichungen und einer algebraischen Gleichung, mit der die Kopplung beschrieben wird. Das so entstandene mathematische Modell in Deskriptorform lautet

$$\begin{aligned} m_1 \ddot{x}_1 &= -d_1 \dot{x}_1 - c_1 x_1 + F + u_1 \\ m_2 \ddot{x}_2 &= -d_2 \dot{x}_2 - c_2 x_2 - F + u_2 \\ 0 &= F - c_3 \cdot (x_2 - x_1) \end{aligned} \quad (1.1)$$

bzw. in einer Zustandsdarstellung mit dem Zustandsvektor $\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix}$ und der Deskriptorvariablen z_1 , die die Koppelfkraft F kennzeichnet:

$$\begin{aligned} \dot{x}_1 &= x_3 \\ \dot{x}_2 &= x_4 \\ \dot{x}_3 &= -\frac{c_1}{m_1} x_1 - \frac{d_1}{m_1} x_3 + \frac{z_1}{m_1} + \frac{u_1}{m_1} \\ \dot{x}_4 &= -\frac{c_2}{m_2} x_2 - \frac{d_2}{m_2} x_4 - \frac{z_1}{m_2} + \frac{u_2}{m_2} \\ 0 &= z_1 - c_3 \cdot (x_2 - x_1) \end{aligned} \quad (1.2)$$

Wird die Deskriptorvariable mit Hilfe der algebraischen Gleichung eliminiert, so entsteht ein gewöhnliches Differenzialgleichungssystem der Ordnung vier. Diese Anzahl entspricht zugleich der minimalen Anzahl an Zustandsvariablen. Die Federkraft F taucht nicht mehr auf. Im Gegensatz zu diesem Beispiel bereitet das Eliminieren von Deskriptor-Variablen oft erheblichen Aufwand. Aus diesem Grund liegt es nahe, regelungstechnische Verfahren und Methoden direkt auf Deskriptorsysteme anwenden zu können. Das geschieht in diesem Beitrag durch einen Beobachterentwurf für bestimmte Klassen nichtlinearer Deskriptorsysteme [2], [3].

2 Kenngrößen von Deskriptorsystemen

Im weiteren Verlauf werden semi-explizite Deskriptorsysteme der folgenden Form betrachtet:

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{a}(\mathbf{x}, \mathbf{z}) + \mathbf{B}(\mathbf{x}, \mathbf{z}) \cdot \mathbf{u} \\ \mathbf{0} &= \mathbf{g}(\mathbf{x}, \mathbf{z}, \mathbf{u}) \\ \mathbf{y} &= \mathbf{c}(\mathbf{x}, \mathbf{z})\end{aligned}\tag{2.1}$$

Das System (2.1) besitzt n Differenzialgleichungen in den n Zustandsgrößen $\mathbf{x}$, sowie p algebraische Gleichungen, denen p Deskriptorvariablen $\mathbf{z}$ zugeordnet werden. Weiter sind m Eingangsgrößen und q Ausgangsgrößen vorhanden. Die Dimension n wurde in [5] als *generalisierte Ordnung* eingeführt. Ist eine Elimination der Deskriptorvariablen mit Hilfe der algebraischen Gleichung möglich, so ergibt dies ein Differenzialgleichungssystem mit einer *minimalen Anzahl* $\tilde{n}$ an Zustandsgrößen mit $\tilde{n} \leq n$. Eine weitere wichtige Kenngröße ist der *Index* eines Deskriptorsystems. Der Index gibt an, wie oft die algebraischen Gleichungen nach der Zeit differenziert werden müssen, um ein explizites Differenzialgleichungssystem in $\dot{\mathbf{z}}$ zu erhalten. Die Bestimmung des Index erfolgt mit Hilfe des Shuffle- bzw. des modifizierten Shuffle-Algorithmus [3].

Eine weitere Kenngröße ist die *Regularität* eines Deskriptorsystems. Ein Deskriptorsystem wird als regulär bezeichnet, wenn der Shuffle-Algorithmus in höchstens $n + 1$ Schritten terminiert. Ist das System regulär, so lässt sich ein Differenzialgleichungssystem in $\dot{\mathbf{z}}$ und $\dot{\mathbf{x}}$ angeben. Terminiert der Shuffle-Algorithmus nicht nach $n + 1$ Schritten, so liegt ein singuläres Deskriptorsystem vom Index unendlich vor. Technische Systeme besitzen die Eigenschaft der Regularität.

Bei einem System mit Index $k = 1$ werden die Kopplungen durch Beziehungen zwischen Zustands- und Deskriptorvariablen beschrieben. Ein Verlust an unabhängigen Bewegungsmöglichkeiten liegt nicht vor. Ist hingegen der Index $k \geq 2$, so existiert mindestens eine algebraische Beziehung nur zwischen den Zustandsvariablen. Dies hat einen Verlust an unabhängigen Bewegungsmöglichkeiten der Zustandsgrößen zur Folge und es gilt für die generalisierte Ordnung n und die minimale Ordnung $\tilde{n}$ der Zusammenhang $\tilde{n} < n$.

Es folgt noch eine weitere Kenngröße, die *Realisierbarkeit* eines Deskriptorsystems. Ein reguläres Deskriptorsystem wird als realisierbar bezeichnet, wenn der Shuffle-Algorithmus auf ein Differenzialgleichungssystem führt, das höchstens von der ersten Ableitung der Eingangsgrößen abhängt, nicht aber von höheren Ableitungen. Bei realisierbaren Systemen hängt das entsprechende minimale System nur von den Eingangsgrößen selber ab. Nicht realisierbare Systeme besitzen differenzierendes Verhalten.

3 Beobachtung linearer Deskriptorsysteme

Betrachtet werden lineare Deskriptorsysteme der folgenden Form:

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}_x \mathbf{x} + \mathbf{A}_z \mathbf{z} + \mathbf{B} \mathbf{u} \\ \mathbf{0} &= \mathbf{G}_x \mathbf{x} + \mathbf{G}_z \mathbf{z} + \mathbf{H} \mathbf{u} \\ \mathbf{y} &= \mathbf{C}_x \mathbf{x} + \mathbf{C}_z \mathbf{z} + \mathbf{D} \mathbf{u}\end{aligned}\tag{3.1}$$

Das System (3.1) besitzt wieder n Differenzialgleichungen in den n Zustandsgrößen $\mathbf{x}$, sowie p algebraische Gleichungen, denen p Deskriptorvariablen $\mathbf{z}$ zugeordnet werden, sowie m Eingangsgrößen $\mathbf{u}$ und q Ausgangsgrößen $\mathbf{y}$. Für diese Klasse von Systemen wird ein Beobachter gesucht, der eine lineare Fehlerdynamik aufweist. Dazu wird zunächst ein System angegeben, das aus dem betrachteten Deskriptorsystem (3.1) mit Voraussetzung der Regularität und Realisierbarkeit hervorgeht. Werden die Zustands- und Deskriptorvariablen zum Vektor $\mathbf{w}^T = [\mathbf{x}^T \mathbf{z}^T]$ zusammengefaßt, so ergibt sich formal das System:

$$\begin{aligned}\dot{\mathbf{w}} &= \tilde{\mathbf{A}} \mathbf{w} + \tilde{\mathbf{B}} \mathbf{u} + \tilde{\mathbf{F}} \dot{\mathbf{u}} \\ \mathbf{y} &= \tilde{\mathbf{C}} \mathbf{w} + \tilde{\mathbf{D}} \mathbf{u}\end{aligned}\tag{3.2}$$

Da die Regularität des Deskriptorsystems (3.1) vorausgesetzt wird, kann durch wiederholtes Ableiten der algebraischen Gleichungen nach der Zeit ein Differenzialgleichungssystem in $\dot{\mathbf{z}}$ berechnet werden. Die Voraussetzung der Realisierbarkeit hat zur Folge, dass höchstens die erste Ableitung der Eingangsgröße auftauchen kann.

Formal handelt es sich bei dem System (3.2) um ein lineares Differenzialgleichungssystem, für das mit Hilfe bekannter Methoden ein Beobachter (3.3) entworfen werden kann.

$$\begin{aligned}\dot{\hat{\mathbf{w}}} &= \tilde{\mathbf{A}} \hat{\mathbf{w}} + \tilde{\mathbf{B}} \mathbf{u} + \tilde{\mathbf{F}} \dot{\mathbf{u}} + \tilde{\mathbf{L}} (\hat{\mathbf{y}} - \mathbf{y}) \\ \hat{\mathbf{y}} &= \tilde{\mathbf{C}} \hat{\mathbf{w}} + \tilde{\mathbf{D}} \mathbf{u}\end{aligned}\quad \text{mit} \quad \hat{\mathbf{w}}^T = [\hat{\mathbf{x}}^T \hat{\mathbf{z}}^T]\tag{3.3}$$

Hier stellt sich nun die Frage, ob ein Vektor $\hat{\mathbf{w}}$ tatsächlich sinnvolle Schätzwerte für die Zustands- und Deskriptorvariablen des Systems (3.1) liefert. Leider kann dies im allgemeinen nicht sichergestellt werden, wenn für den Index des Systems $k \geq 2$ gilt. Der Grund liegt in der Vernachlässigung versteckter impliziter Zwangsbedingungen. Der Beobachter muss aber sicherstellen, dass alle Zwangsbedingungen berücksichtigt werden.

In jedem Deskriptorsystem mit Index $k \geq 2$ sind $k - 1$ weitere Zwangsbedingungen implizit enthalten. Diese können durch Ableiten der algebraischen Gleichung nach der Zeit sowie durch elementare Umformungen bestimmt werden, was im Folgenden angedeutet sei:

$$\mathbf{0} = \frac{d}{dt}(\mathbf{G}_x \mathbf{x} + \mathbf{G}_z \mathbf{z} + \mathbf{H} \mathbf{u})\tag{3.4}$$

Damit das System einen Index $k > 1$ besitzen kann, muss offensichtlich die Matrix $\mathbf{G}_z$ singulär sein; nimmt man nun der Einfachheit halber an, dass $\mathbf{G}_z = \mathbf{0}$ ist und beachtet, dass dann wegen der Realisierbarkeit auch $\mathbf{H} = \mathbf{0}$ gelten muss, so lautet die erste zeitliche Ableitung der algebraischen Gleichung in (3.1)

$$\begin{aligned}
0 &= G_x \dot{x} \\
0 &= G_x A_x x + G_x A_z z + G_x B u,
\end{aligned} \tag{3.5}$$

womit die erste implizit im System (3.1) enthaltene Zwangsbedingung gefunden ist.

Ist die Matrix $G_x A_z$ regulär, so ist auch die einzige implizite Zwangsbedingung gefunden worden. Andernfalls werden durch weiteres Ableiten nach der Zeit weitere implizite Zwangsbedingungen explizit zugänglich. Dieses Verfahren lässt sich insgesamt $k - 1$ mal durchführen. Danach sind alle Zwangsbedingungen explizit als Gleichungen vorhanden.

Ein Deskriptorsystem vom Index k besitzt insgesamt k Zwangsbedingungen, die mit Hilfe des beschriebenen Verfahrens explizit als algebraische Gleichungen zugänglich werden. Jede Zwangsbedingung wird damit durch eine algebraische Gleichung, die auch vektorieller Natur sein kann, beschrieben. Ein Beobachter kann also mit Hilfe seiner geschätzten Zustands- und Deskriptorvariablen den Fehler jeder Zwangsbedingung berechnen. Um diese Fehler zu berücksichtigen, wird die Ausgangsgleichung des Beobachters (3.3) mit den Zwangsbedingungen erweitert¹.

Für eine solche Erweiterung der Ausgangsgleichung ist es zweckmäßig, von den algebraischen Gleichungen in ihrer skalaren Form auszugehen:

$$\begin{aligned}
0 &= g_l = g_{xl}^T x + g_{zl}^T z + h_l^T u \\
&\quad \vdots \\
0 &= g_p = g_{xp}^T x + g_{zp}^T z + h_p^T u
\end{aligned} \quad G_x = \begin{bmatrix} g_{xl}^T \\ \vdots \\ g_{xp}^T \end{bmatrix} \quad G_z = \begin{bmatrix} g_{zl}^T \\ \vdots \\ g_{zp}^T \end{bmatrix} \quad H = \begin{bmatrix} h_l^T \\ \vdots \\ h_p^T \end{bmatrix} \tag{3.6}$$

Kennzeichnet k_i den Index, der zur i -ten algebraischen Gleichung gehört [3], dann sind in (3.7) alle Zwangsbedingungen, die aus der i -ten algebraischen Gleichung mit $k_i \geq 2$ hergeleitet werden können, berechnet.

$$\begin{aligned}
0 &= g_i = g_{xi}^T x \\
0 &= \frac{dg_i}{dt} = g_{xi}^T A_x x \\
&\quad \vdots \\
0 &= \frac{d^{k_i-1} g_i}{dt^{k_i-1}} = g_{xi}^T A_x^{k_i-1} x + g_{xi}^T A_x^{k_i-2} A_z z + g_{xi}^T A_x^{k_i-2} B u
\end{aligned} \tag{3.7}$$

Nun werden die so erhaltenen Zwangsbedingungen zu einem Vektor v_i zusammengefasst:

1. Allerdings kann es vorkommen, dass bei dieser Erweiterung eventuell Zwangsbedingungen mehrfach berücksichtigt werden. Dies führt aber in der Regel nicht zu Problemen.

$$v_i = \begin{bmatrix} g_{xi}^T x \\ g_{xi}^T A_x x \\ \vdots \\ g_{xi}^T A_x^{k_i-1} x + g_{xi}^T A_x^{k_i-2} A_z z + g_{xi}^T A_x^{k_i-2} B u \end{bmatrix} = v_i(x, z, u) \quad \begin{matrix} v_i \in \mathcal{R}^{k_i} \\ k_i \geq 2 \end{matrix} \quad (3.8)$$

Gilt hingegen $\mathbf{g}_{zi} \neq \mathbf{0}$, ist also in der i-ten algebraischen Gleichung wenigstens eine Deskriptorvariable enthalten, so ist der Index $k_i = l$ und der Vektor $\mathbf{v}_i$ lautet:

$$\mathbf{v}_i = \mathbf{g}_{xi}^T \mathbf{x} + \mathbf{g}_{zi}^T \mathbf{z} + \mathbf{h}_i^T \mathbf{u} \quad k_i = I \quad (3.9)$$

Jetzt kann ein System angegeben werden, das in der Lage ist, die Zustands- und Deskriptorvariablen des Systems (3.1) unter Berücksichtigung aller Zwangsbedingungen zu schätzen.

$$\begin{aligned} \hat{\mathbf{w}} &= \tilde{\mathbf{A}}\hat{\mathbf{w}} + \tilde{\mathbf{B}}\mathbf{u} + \tilde{\mathbf{F}}\dot{\mathbf{u}} + \tilde{\mathbf{L}}\left(\begin{bmatrix} \hat{\mathbf{y}} \\ \hat{\mathbf{v}} \end{bmatrix} - \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix}\right) \\ \hat{\mathbf{y}} &= \tilde{\mathbf{C}}\hat{\mathbf{w}} + \tilde{\mathbf{D}}\mathbf{u} \\ \hat{\mathbf{v}}^T &= \left[\mathbf{v}_l^T(\hat{\mathbf{w}}, \mathbf{u}) \dots \mathbf{v}_p^T(\hat{\mathbf{w}}, \mathbf{u}) \right] \end{aligned} \quad (3.10)$$

Der Vektor $\hat{\mathbf{v}}$ besitzt die Dimension $(k_I + \dots + k_p)$. Der Vektor $\mathbf{y}$ wird im Beobachter mit dem Nullvektor $\mathbf{0}$ erweitert, da das reale zu beobachtende System naturgemäß seine Zwangsbedingungen mit $\mathbf{v} = \mathbf{0}$ erfüllt.

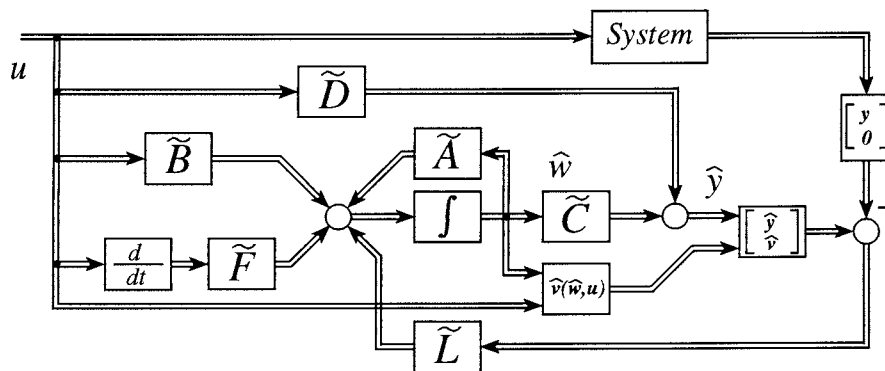


Bild 3.1: System mit Beobachter im linearen Fall

Die geschätzten Zwangsbedingungen $\hat{\mathbf{v}}$ lassen sich auch als Linearkombination mit Hilfe einer Matrix $\tilde{\mathbf{M}}$ und den Variablen $\hat{\mathbf{w}}$ und $\mathbf{u}$ darstellen als

$$\hat{\mathbf{v}} = \tilde{M} \begin{bmatrix} \hat{\mathbf{w}} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \tilde{M}_1 & \tilde{M}_2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{w}} \\ \mathbf{u} \end{bmatrix} \quad (3.11)$$

Abschließend bleibt noch zu zeigen, dass der Schätzfehler des Beobachters $e = w - \hat{w}$ zeitlich abklingt, also

$$\lim_{t \rightarrow \infty} e(t) = 0 \quad (3.12)$$

gilt. Dazu bildet man die Differenzialgleichung des Schätzfehlers e und berücksichtigt, dass für das reale System die Gleichung

$$0 = \begin{bmatrix} \tilde{M}_1 & \tilde{M}_2 \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix} \quad (3.13)$$

eingesetzt wird. Es ergibt sich damit:

$$\begin{aligned} \dot{e} &= \dot{w} - \dot{\hat{w}} \\ &= \tilde{A}e + \tilde{L} \begin{bmatrix} \tilde{C}e \\ \hat{v}(\hat{w}, u) - 0 \end{bmatrix} \\ &= \tilde{A}e + \tilde{L} \begin{bmatrix} \tilde{C}e \\ \begin{bmatrix} \tilde{M}_1 & \tilde{M}_2 \end{bmatrix} \begin{bmatrix} \hat{w} \\ u \end{bmatrix} - \begin{bmatrix} \tilde{M}_1 & \tilde{M}_2 \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix} \end{bmatrix} \\ &= \tilde{A}e + \tilde{L} \begin{bmatrix} \tilde{C} \\ \tilde{M}_1 \end{bmatrix} e \\ &= \left[\tilde{A} + \tilde{L} \begin{bmatrix} \tilde{C} \\ \tilde{M}_1 \end{bmatrix} \right] e \end{aligned} \quad (3.14)$$

Wie in (3.14) zu erkennen ist, lassen sich die Eigenwerte der Matrix

$$\left[\tilde{A} + \tilde{L} \begin{bmatrix} \tilde{C} \\ \tilde{M}_1 \end{bmatrix} \right] \quad (3.15)$$

mit Hilfe von $\tilde{L}$ beliebig vorgeben, wenn das Paar $\tilde{A}, \begin{bmatrix} \tilde{C}^T & \tilde{M}_1^T \end{bmatrix}^T$ beobachtbar ist. Daraus folgt, dass (3.12) erfüllt werden kann.

3.1 Beispiel

Es wird das mechanische System mit Federkopplung nach Bild 1.1 betrachtet.

Das System soll zunächst in der Form nach Gl. (3.1) dargestellt werden, wobei als Ausgangsgröße $y = x_1$ gewählt wird. Das ergibt:

$$n = 4, p = 1, m = 2, q = 1$$

$$\begin{aligned} A_x &= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{c_1}{m_1} & 0 & -\frac{d_1}{m_1} & 0 \\ 0 & -\frac{c_2}{m_2} & 0 & -\frac{d_2}{m_2} \end{bmatrix} & A_z &= \begin{bmatrix} 0 \\ 0 \\ \frac{1}{m_1} \\ -\frac{1}{m_2} \end{bmatrix} & B &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{1}{m_1} & 0 \\ 0 & \frac{1}{m_2} \end{bmatrix} \\ G_x &= \begin{bmatrix} c_3 & -c_3 & 0 & 0 \end{bmatrix} & G_z &= I & H &= \begin{bmatrix} 0 & 0 \end{bmatrix} \\ C_x &= \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} & C_z &= 0 & D &= \begin{bmatrix} 0 & 0 \end{bmatrix} \end{aligned} \quad (3.16)$$

Zunächst muss der Index der ersten und einzigen algebraischen Gleichung berechnet werden.

$$\begin{aligned} g &= g_0 = 0 = z_I - c_3 \cdot (x_2 - x_I) \\ g_I &= \dot{z}_I - c_3(\dot{x}_2 - \dot{x}_I) = \dot{z}_I - c_3(x_4 - x_3) \quad 1. \text{ Schritt: } \dot{z} \text{ taucht auf} \\ \dot{z}_I &= c_3(x_4 - x_3) \end{aligned} \quad (3.17)$$

Nach dem ersten Schritt taucht die Ableitung auf. Die Gleichung bzw. das System besitzt deshalb den Index $k_I = 1$. Weitere implizite Zwangsbedingungen sich deshalb nicht enthalten. Der Vektor $\mathbf{v}$ ist dem zufolge eine skalare Größe und berechnet sich nach Gleichung (3.8) zu

$$\mathbf{v} = \mathbf{G}_x \mathbf{x} + \mathbf{G}_z \mathbf{z} + \mathbf{H} \mathbf{u} = c_3(x_I - x_2) + z_I \quad (3.18)$$

Mit $\hat{\mathbf{w}}^T = [\hat{x}_I \ \hat{x}_2 \ \hat{x}_3 \ \hat{x}_4 \ \hat{z}_I]$ kann der Beobachter (3.10) angegeben werden:

$$\begin{aligned} \dot{\hat{w}}_1 &= \hat{w}_3 \\ \dot{\hat{w}}_2 &= \hat{w}_4 \\ \dot{\hat{w}}_3 &= -\frac{c_1}{m_1} \hat{w}_1 - \frac{d_1}{m_1} \hat{w}_3 + \frac{1}{m_1} \hat{w}_5 + \frac{1}{m_1} u_1 \\ \dot{\hat{w}}_4 &= -\frac{c_2}{m_2} \hat{w}_2 - \frac{d_2}{m_2} \hat{w}_4 - \frac{1}{m_2} \hat{w}_5 + \frac{1}{m_2} u_2 \\ \dot{\hat{w}}_5 &= -c_3 \hat{w}_3 + c_3 \hat{w}_4 \end{aligned} + \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \\ L_{31} & L_{32} \\ L_{41} & L_{42} \\ L_{51} & L_{52} \end{bmatrix} \begin{bmatrix} \hat{w}_1 - y \\ c_3 \hat{w}_1 - c_3 \hat{w}_2 + \hat{w}_5 \end{bmatrix} \quad (3.19)$$

Als nächster Schritt folgt die Überprüfung der Beobachtbarkeit. Dazu muss die Matrix

$$\mathbf{Q}_B = \left[\begin{bmatrix} \tilde{\mathbf{C}} \\ \tilde{\mathbf{M}}_I \end{bmatrix}^T \left(\begin{bmatrix} \tilde{\mathbf{C}} \\ \tilde{\mathbf{M}}_I \end{bmatrix} \tilde{\mathbf{A}} \right)^T \left(\begin{bmatrix} \tilde{\mathbf{C}} \\ \tilde{\mathbf{M}}_I \end{bmatrix} \tilde{\mathbf{A}}^2 \right)^T \left(\begin{bmatrix} \tilde{\mathbf{C}} \\ \tilde{\mathbf{M}}_I \end{bmatrix} \tilde{\mathbf{A}}^3 \right)^T \left(\begin{bmatrix} \tilde{\mathbf{C}} \\ \tilde{\mathbf{M}}_I \end{bmatrix} \tilde{\mathbf{A}}^4 \right)^T \right]^T \quad (3.20)$$

den Höchststrang $rg(\mathbf{Q}_B) = 5$ besitzen - vgl. Gl. (3.15). Werden die folgenden Zahlenwerte für die Parameter

$$m_1 = 0,2; c_1 = 0,1; d_1 = 0,05; m_2 = 0,2; c_2 = 0,1; d_2 = 0,05; c_3 = 100$$

eingesetzt, so ergibt sich für den Rang von $\mathbf{Q}_B$

$$rg(\mathbf{Q}_B) = rg \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 100 & -100 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 100 & 0 & -100 \\ -0,5 & 0 & -0,25 & 0 & 5 \\ -50 & 0 & 9975 & -10000 & 500 \\ 0,125 & 0 & -500,44 & 500 & -1,25 \\ 4987,5 & 5000 & -52543,8 & 52500 & 99875 \\ 250,22 & -250 & 250,23 & -250 & -5002,2 \\ 26272 & -26250 & -9979350 & 9974375 & -520219 \end{bmatrix} = 5 \quad (3.21)$$

Das System ist somit vollständig beobachtbar.

Mit der Korrekturmatriix $\tilde{\mathbf{L}}$ für eine asymlotisch stabile Fehlerdynamik

$$\tilde{\mathbf{L}}^T = \begin{bmatrix} 13,5 & -13,24 & -936,81 & 936,92 & -2673,93 \\ -0,06 & 0,06 & 4,68 & -4,68 & 13,37 \end{bmatrix} \quad (3.22)$$

und den Anfangswerten

$$\begin{aligned} \mathbf{x}^T(t=0) &= [0,01 \ 0,01 \ 0 \ 0] \\ \hat{\mathbf{w}}^T(t=0) &= [0 \ 0 \ 0 \ 0 \ 0] \end{aligned} \quad (3.23)$$

ergeben sich mit den Eingangsgrößen

$$\begin{aligned} u_1 &= 0,1\sigma(t) \\ u_2 &= 0 \end{aligned} \quad (3.24)$$

die im Bild 3.2 und im Bild 3.3 gezeigten Systemreaktionen:

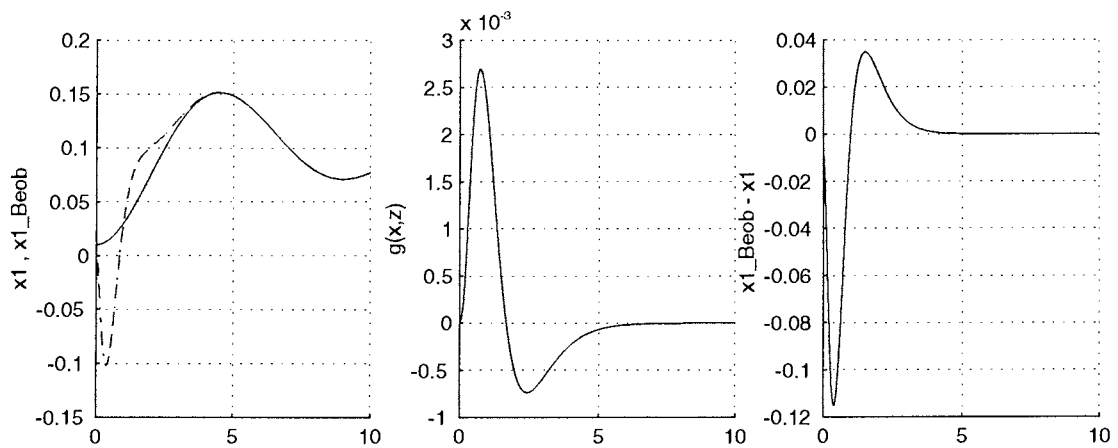


Bild 3.2: Reaktionen des Systems und Beobachter

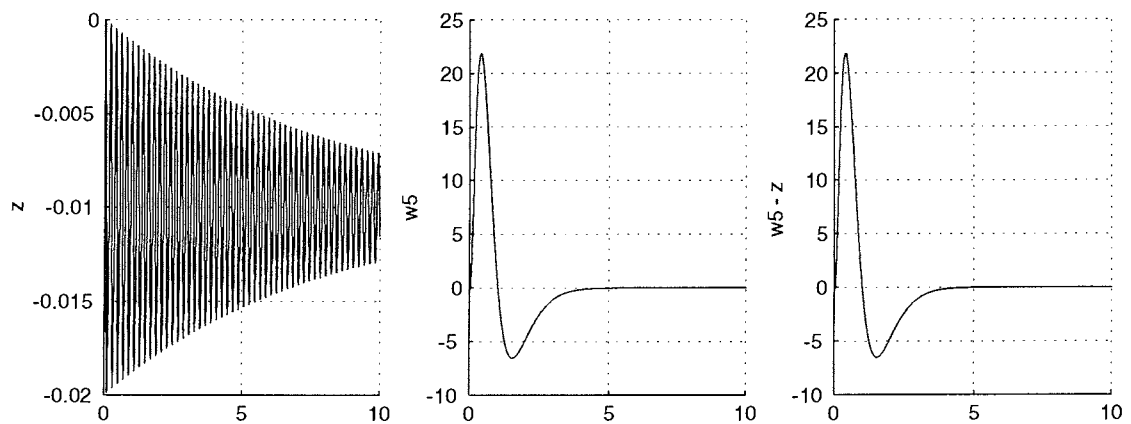


Bild 3.3: Deskriptorvariable des Systems und des Beobachters

Wie zu erkennen ist, besitzt die beobachtete Deskriptorvariable $\hat{w}_5$ (im Bild 3.3 mit w_5 bezeichnet) zu Beginn einen großen Fehler. Werden Anfangswerte verwendet, die keine so grossen Unterschiede zwischen Beobachter und System aufweisen, so ist der Fehler der beobachteten Deskriptorvariable deutlich kleiner.

4 Beobachtung nichtlinearer Deskriptorsysteme

Betrachtet werden reguläre realisierbare nichtlineare Deskriptorsysteme der folgenden Form:

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{a}(\mathbf{x}, \mathbf{z}) + \mathbf{B}(\mathbf{x}, \mathbf{z})\mathbf{u} \\ \mathbf{0} &= \mathbf{g}(\mathbf{x}, \mathbf{z}, \mathbf{u}) \\ \mathbf{y} &= \mathbf{c}(\mathbf{x}, \mathbf{z})\end{aligned}\tag{4.1}$$

Das System (4.1) besitzt wieder n Differenzialgleichungen in den n Zustandsgrößen $\mathbf{x}$, sowie p algebraische Gleichungen, denen p Deskriptorvariablen $\mathbf{z}$ zugeordnet werden, sowie m Eingangsgrößen $\mathbf{u}$ und q Ausgangsgrößen $\mathbf{y}$. Für dieses System wird ein vollständiger Beobachter gesucht. Dabei wird zunächst ein Verfahren angegeben, das es ermöglicht, einen nichtlinearen Beobachter für die Zustands- und Deskriptorvariablen unter Berücksichtigung aller im System (4.1) enthaltenen Zwangsbedingungen zu entwerfen. Dieser nichtlineare Beobachter wird anschließend linearisiert und dann gezeigt, dass der Schätzfehler eine lineare Fehlerdynamik aufweist.

4.1 Nichtlinearer Beobachter

Zu Beginn wird analog zu Kapitel 3 ein System angegeben, das aus (4.1) berechnet werden kann.

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{a}(\mathbf{x}, \mathbf{z}) + \mathbf{B}(\mathbf{x}, \mathbf{z})\mathbf{u} \\ \dot{\mathbf{z}} &= \tilde{\mathbf{g}}(\mathbf{x}, \mathbf{z}, \mathbf{u}, \dot{\mathbf{u}}) \\ \mathbf{y} &= \mathbf{c}(\mathbf{x}, \mathbf{z})\end{aligned}\tag{4.2}$$

Da reguläre realisierbare Deskriptorsysteme vorausgesetzt werden, ist sichergestellt, dass maximal die erste Ableitung der Eingangsgröße in der Differenzialgleichung für $\dot{\mathbf{z}}$ vorkommt. Wie bereits erwähnt wurde, ist dieses System noch nicht in der Lage als Beobachter zu dienen, da die impliziten Zwangsbedingungen u.U. nicht berücksichtigt werden. Deshalb wird in (4.2) der Ausgangsvektor wie im linearen Fall um alle Zwangsbedingungen erweitert. Die Berechnung der impliziten Zwangsbedingungen erfolgt wieder mit Hilfe des Shuffle- Algorithmus, also durch wiederholtes Ableiten der einzelnen algebraischen Gleichungen nach der Zeit, bis die Ableitung mindestens einer Deskriptorvariable explizit auftaucht. Diese Vorgehensweise soll hier zunächst gezeigt werden.

Dazu wird der Vektor $\mathbf{g}$ in seine skalaren Gleichungen zerlegt (vgl. die Ausführungen zu (3.6) und (3.7)).

$$\mathbf{g}(\mathbf{x}, \mathbf{z}, \mathbf{u}) = \begin{bmatrix} g_1(\mathbf{x}, \mathbf{z}, \mathbf{u}) \\ \dots \\ g_p(\mathbf{x}, \mathbf{z}, \mathbf{u}) \end{bmatrix} = \begin{bmatrix} 0 \\ \dots \\ 0 \end{bmatrix}\tag{4.3}$$

Für diese Gleichungen werden die impliziten Zwangsbedingungen wie folgt berechnet.

$$\begin{aligned}
0 &= g_{i,0}(\mathbf{x}, \mathbf{z}, \mathbf{u}) \\
0 &= \frac{dg_{i,0}}{dt} = \frac{\partial g_{i,0}}{\partial \mathbf{x}} \dot{\mathbf{x}} + \frac{\partial g_{i,0}}{\partial \mathbf{z}} \dot{\mathbf{z}} + \frac{\partial g_{i,0}}{\partial \mathbf{u}} \dot{\mathbf{u}} = g_{i,1} \quad \text{sei} \quad \frac{\partial g_{i,0}}{\partial \mathbf{z}} = \mathbf{0} \Rightarrow \frac{\partial g_{i,0}}{\partial \mathbf{u}} = \mathbf{0} \\
&\dots \\
0 &= \frac{d}{dt} g_{i,k_i-1} = \frac{\partial}{\partial \mathbf{x}} g_{i,k_i-1} \dot{\mathbf{x}} + \frac{\partial}{\partial \mathbf{z}} g_{i,k_i-1} \dot{\mathbf{z}} + \frac{\partial}{\partial \mathbf{u}} g_{i,k_i-1} \dot{\mathbf{u}} = g_{i,k_i} \quad \text{sei} \quad \frac{\partial g_{i,k_i-1}}{\partial \mathbf{z}} \neq \mathbf{0}
\end{aligned} \tag{4.4}$$

Werden die letzten Ableitungen in (4.4) wie folgt zusammengefasst, so kann $\dot{\mathbf{z}}$ berechnet werden,

$$\begin{bmatrix} \frac{\partial}{\partial \mathbf{z}} g_{1,k_1-1} \\ \dots \\ \frac{\partial}{\partial \mathbf{z}} g_{p,k_p-1} \end{bmatrix} \dot{\mathbf{z}} = \begin{bmatrix} -\frac{\partial}{\partial \mathbf{x}} g_{1,k_1-1} \dot{\mathbf{x}} - \frac{\partial}{\partial \mathbf{u}} g_{1,k_1-1} \dot{\mathbf{u}} \\ \dots \\ -\frac{\partial}{\partial \mathbf{x}} g_{p,k_p-1} \dot{\mathbf{x}} - \frac{\partial}{\partial \mathbf{u}} g_{p,k_p-1} \dot{\mathbf{u}} \end{bmatrix} \tag{4.5}$$

wenn die Matrix (4.6) regulär ist.

$$\begin{bmatrix} \frac{\partial}{\partial \mathbf{z}} g_{1,k_1-1} \\ \dots \\ \frac{\partial}{\partial \mathbf{z}} g_{p,k_p-1} \end{bmatrix} \tag{4.6}$$

Alle Zwangsbedingungen aus der i-ten algebraischen Gleichung werden zu dem Vektor $\mathbf{v}_i$ zusammengefasst.

$$\mathbf{v}_i = \begin{bmatrix} g_{i,0}(\mathbf{x}, \mathbf{z}, \mathbf{u}) \\ g_{i,1}(\mathbf{x}, \mathbf{z}, \mathbf{u}) \\ \vdots \\ g_{i,k_i-1}(\mathbf{x}, \mathbf{z}, \mathbf{u}) \end{bmatrix} \quad \mathbf{v}_i \in \mathbb{R}^{k_i} \quad k_i \geq 1 \tag{4.7}$$

Ist die Regularität von (4.6) erfüllt, kann mit (4.7) ein System angegeben werden, das zunächst als Ansatz für einen Beobachter betrachtet wird. Werden die Zustands- und Deskriptorvariablen zu dem Vektor

$$\hat{\mathbf{w}}^T = \begin{bmatrix} \hat{\mathbf{x}}^T & \hat{\mathbf{z}}^T \end{bmatrix} \tag{4.8}$$

zusammengefasst, dann besitzt dieses System die folgende Form:

$$\dot{\hat{\mathbf{w}}} = \begin{bmatrix} \mathbf{a}(\hat{\mathbf{w}}) + \mathbf{B}(\hat{\mathbf{w}})\mathbf{u} \\ \tilde{\mathbf{g}}(\hat{\mathbf{w}}, \mathbf{u}, \dot{\mathbf{u}}) \end{bmatrix} \quad \hat{\mathbf{y}} = \begin{bmatrix} \mathbf{c}(\hat{\mathbf{w}}) \\ \hat{\mathbf{v}} \end{bmatrix} \quad \text{mit} \quad \hat{\mathbf{v}}^T = \begin{bmatrix} \mathbf{v}_1^T(\hat{\mathbf{w}}, \mathbf{u}) & \dots & \mathbf{v}_p^T(\hat{\mathbf{w}}, \mathbf{u}) \end{bmatrix}^T \tag{4.9}$$

Es stellt sich nun die Frage, ob das System (4.9) zumindest lokal um einen Punkt w_r beobachtbar ist und wie der Schätzfehler bestimmt werden kann. Diese Fragen werden im folgenden beantwortet, allerdings ist dazu die Struktur des Beobachters (4.9) weiter einzuschränken. Es wird vorausgesetzt, dass die Vektorfunktion $\tilde{g}$ nicht von $\hat{u}$ abhängt und in u affin ist und die Zwangsbedingungen v nicht von u abhängen. Darüber hinaus werden aus Gründen der Einfachheit nur Eingrößensysteme betrachtet.

$$\dot{\hat{w}} = \begin{bmatrix} a(\hat{w}) + b(\hat{w})u \\ \tilde{g}_1(\hat{w}) + \tilde{g}_2(\hat{w})u \end{bmatrix} \quad \hat{y} = \begin{bmatrix} c(\hat{w}) \\ \hat{v}(\hat{w}) \end{bmatrix} \quad (4.10)$$

4.2 Eingangs- Zustandslinearisierung

Es soll zunächst allgemein erläutert werden, wie nichtlineare Systeme mit Hilfe von nichtlinearen Zustandstransformationen und -rückkopplungen in lineare System überführt werden können. Dabei werden Systeme, die der Form (4.10) entsprechen, untersucht. Ein System, wie es in (4.10) angegeben ist, besitzt die unten angegebene allgemeine Form (4.11). Im Folgenden wird eine Variable z verwendet, die eine Transformation des Zustandsvektors x kennzeichnet und nicht mit den Deskriptorvariablen zu verwechseln ist.

$$\dot{x} = a(x) + b(x) \cdot u \quad (4.11)$$

Für dieses System wird eine Zustandstransformation und -rückkopplung der Form

$$z = \phi(x) \quad \text{und} \quad u = \alpha(x) + \beta(x) \cdot v \quad (4.12)$$

gesucht, die das System (4.11) in die Form

$$\dot{z} = A_L z + b_L v \quad (4.13)$$

transformiert. Unter der Voraussetzung, dass das nichtlineare System linearisierbar ist, sollen sich die Systemmatrix A_L bzw. der Eingangsvektor b_L wie folgt ergeben:

$$A_L = \begin{bmatrix} 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ & & \dots & & \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad b_L = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (4.14)$$

Die Ordnung des linearen Systems sei n . Die Struktur des linearisierten Systems nach Bild (4.1) ist eine Kette von Integratoren, deren Ausgänge gleich den transformierten Zustandsvariablen z sind.

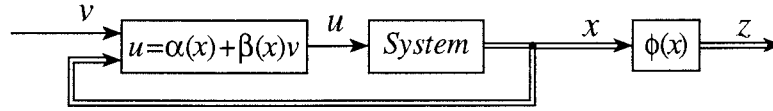


Bild 4.1: Das linearisierte System

Welche Voraussetzungen an das nichtlineare System gestellt werden müssen, wird in dem Theorem 1 angegeben [1], [4]. Der Operator $ad_a^k b$ steht für die k-fach wiederholte Anwendung der Lie-Klammer $[a, b]$.

Theorem 1:

Das nichtlineare System (4.11) ist genau dann linearisierbar, wenn eine Region Ω im Zustandsraum existiert, in der die folgenden Bedingungen erfüllt sind:

- Steuerbarkeit: Das Vektorfeld $\{b, ad_a b, \dots, ad_a^{n-1} b\}$ ist linear unabhängig in Ω .
- Involutivität: Die Menge $\{b, ad_a b, \dots, ad_a^{n-2} b\}$ ist involutiv in Ω .

Ist dieses Theorem erfüllt, so ist das nichtlineare System (4.11) in Ω in das lineare System (4.13), (4.14) überführbar.

4.3 Beobachter mit linearer Fehlerdynamik

Gegeben ist das System, das wieder der Beobachterstruktur (4.10) entspricht, als Eingrößensystem aber nur die skalare Ausgangsgröße y besitzt (die weiteren Elemente im Ausgangsvektor (4.10) sind die vom Beobachter berechneten Zwangsbedingungen des Systems):

$$\dot{x} = a(x) + b(x) \cdot u \quad y = c(x) \quad (4.15)$$

Es wird vorausgesetzt, dass (4.15) linearisierbar ist, also ein $z = \phi(x)$ und $u = \alpha(x) + \beta(x)v$ existiert, die das System (4.15) in die Form

$$\dot{z} = A_L z + b_L v \quad y = c^T z \quad (4.16)$$

transformiert. Dabei sei die Ausgangsgleichung linear in den transformierten Zuständen und (A_L, c^T) ein beobachtbares Paar. Dann ist für das System (4.16) ein Beobachter der Form

$$\dot{\hat{z}} = A_L \hat{z} + b_L v + L(\hat{y} - y) \quad (4.17)$$

angebbbar, dessen Fehler $\tilde{z} = \hat{z} - z = \hat{z} - \phi(x)$ über Gleichung

$$\dot{\tilde{z}} = (A_L + Lc^T)\tilde{z} \quad (4.18)$$

bestimmt wird. Trivialerweise muss für diese lineare Fehlerdynamik vorausgesetzt werden, dass beim linearisierten System (4.16) die Ausgangsgröße linear von den Zuständen z abhängen. Darüberhinaus ist leicht erkennbar, dass das System (4.16) mit der Matrix A_L nach (4.14) genau dann beobachtbar ist, wenn mit $c^T = [c_1 \dots c_n]$ gilt:

$$c_1 \neq 0 \quad (4.19)$$

4.4 Struktur des Systems mit Beobachter

Das Bild 4.2 zeigt, wie der Beobachter mit dem System in Verbindung mit der Zustands- und -rückkopplung das zu realisierende Gesamtsystem ergibt.

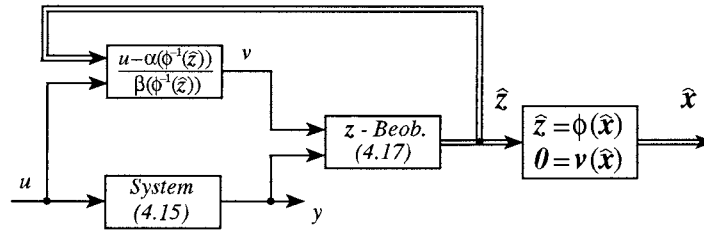


Bild 4.2: Struktur der Systems mit Beobachter

Über die nichtlinearen Gleichungen

$$v = \frac{u - \alpha(\phi^{-1}(\hat{z}))}{\beta(\phi^{-1}(\hat{z}))} \quad \text{und} \quad \hat{x} = F(\hat{z}) \quad (4.20)$$

kann der Beobachter eingebunden werden. $F(\hat{z})$ sind dabei die aufgelösten Gleichungen im Bild 4.2. Der geschätzte Zustand $\hat{x}$ des Beobachters - das sind die Schätzwerte für die Zustands- und Deskriptorvariablen des Deskriptorsystems - steht zur Verfügung.

4.5 Beispiel

Der Einsatz des Beobachters für ein nichtlineares System in Deskriptorform sei am Beispiel des flexiblen Gelenkmechanismus [1] gezeigt.

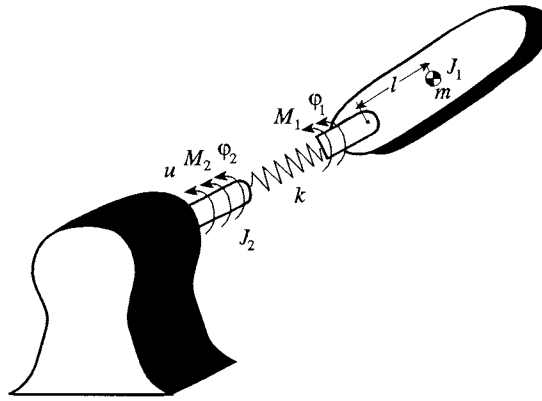


Bild 4.3: Flexibler Gelenkmechanismus (M. Spong)

Die Bewegungsgleichungen für das zusammengesetzte System nach Bild 4.3 lauten:

$$\begin{aligned}
 J_1 \ddot{\phi}_1 &= -mgl \sin \phi_1 + M_1 \\
 J_2 \ddot{\phi}_2 &= M_2 + u \\
 M_1 &= k(\phi_2 - \phi_1) \\
 M_2 &= k(\phi_1 - \phi_2)
 \end{aligned} \tag{4.21}$$

Werden die beiden Vektoren

$$\mathbf{x} = [\phi_1 \ \dot{\phi}_1 \ \phi_2 \ \dot{\phi}_2]^T \quad \text{und} \quad \mathbf{z} = [M_1 \ M_2]^T \tag{4.22}$$

entsprechend gewählt, so ergibt sich das System in Deskriptorform

$$\begin{aligned}
 \dot{x}_1 &= x_2 \\
 \dot{x}_2 &= -\frac{mgl}{J_1} \sin x_1 + \frac{1}{J_1} z_1 & 0 &= -z_1 + k(x_3 - x_1) \\
 \dot{x}_3 &= x_4 & 0 &= z_2 + k(x_3 - x_1) \\
 \dot{x}_4 &= \frac{1}{J_2} z_2 + \frac{1}{J_2} u
 \end{aligned} \tag{4.23}$$

Das System besitzt zwei algebraische Gleichungen, deren Indizes berechnet werden müssen.

$$\begin{aligned}
 0 &= g_1 = -z_1 + k(x_3 - x_1) & 0 &= g_2 = z_2 + k(x_3 - x_1) \\
 0 &= \dot{g}_1 = -\dot{z}_1 + k(\dot{x}_3 - \dot{x}_1) & 0 &= \dot{g}_2 = \dot{z}_2 + k(\dot{x}_3 - \dot{x}_1) \\
 \dot{z}_1 &= k(x_4 - x_2) & \dot{z}_2 &= -k(x_4 - x_2) \\
 \Rightarrow k_1 &= 1 & \Rightarrow k_2 &= 1
 \end{aligned} \tag{4.24}$$

Mit dem Zustandsvektor

$$\hat{\mathbf{w}} = [\hat{x}_1 \hat{x}_2 \hat{x}_3 \hat{x}_4 \hat{z}_1 \hat{z}_2]^T \quad (4.25)$$

kann der nichtlineare Beobachter (4.10) angegeben werden:

$$\begin{aligned} \dot{\hat{w}}_1 &= \hat{w}_2 \\ \dot{\hat{w}}_2 &= -\frac{mgl}{J_1} \sin \hat{w}_1 + \frac{l}{J_1} \hat{w}_5 \\ \dot{\hat{w}}_3 &= \hat{w}_4 & \hat{y}_1 &= \hat{w}_1 \\ \dot{\hat{w}}_4 &= \frac{l}{J_2} \hat{w}_6 + \frac{l}{J_2} u & \hat{y}_2 &= -\hat{w}_5 - k\hat{w}_1 + k\hat{w}_3 \\ & & \hat{y}_3 &= \hat{w}_6 - k\hat{w}_1 + k\hat{w}_3 \\ \dot{\hat{w}}_5 &= -k\hat{w}_2 + k\hat{w}_4 \\ \dot{\hat{w}}_6 &= k\hat{w}_2 - k\hat{w}_4 \end{aligned} \quad (4.26)$$

Gemäß Abschnitt 4.2 sind zunächst die Zustandstransformation und -rückkopplung, also die Funktionen $\phi(\mathbf{x})$, $\alpha(\mathbf{x})$ und $\beta(\mathbf{x})$ zu berechnen und dann ist für das linearisierte System gemäß Abschnitt 4.3 die Korrekturmatrix $\mathbf{L}$ zu ermitteln. Diese Berechnungen sollen hier aber nicht ausgeführt werden.

Mit den Zahlenwerten für die Parameter

$$m = 5; \quad J_1 = 0,3; \quad J_2 = 0,05; \quad l = 0,25; \quad k = 5000; \quad g = 9,81 \quad (4.27)$$

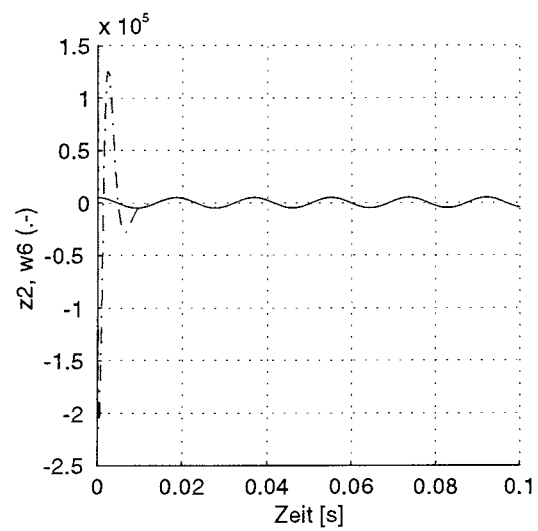
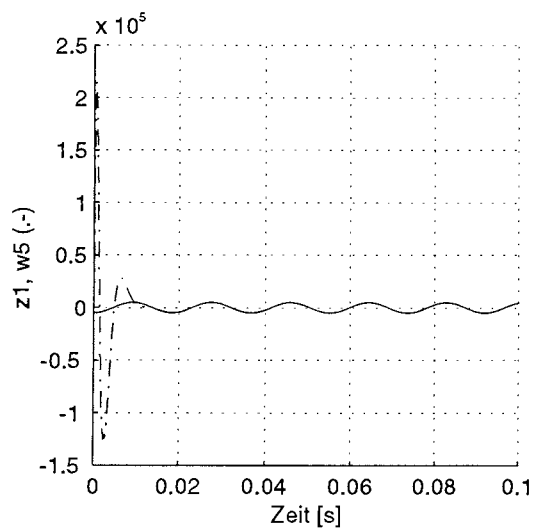
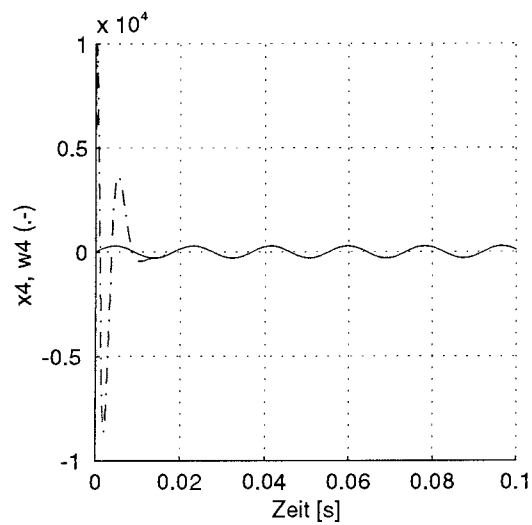
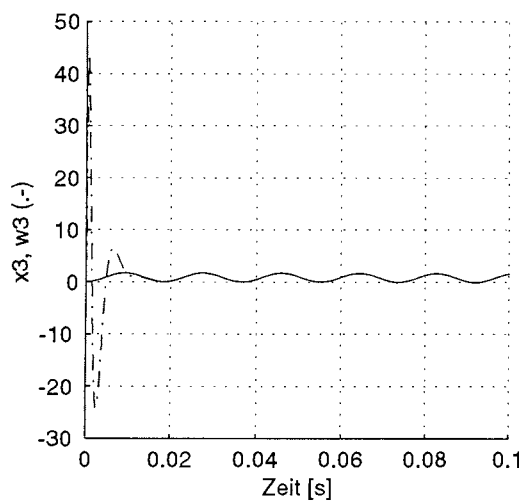
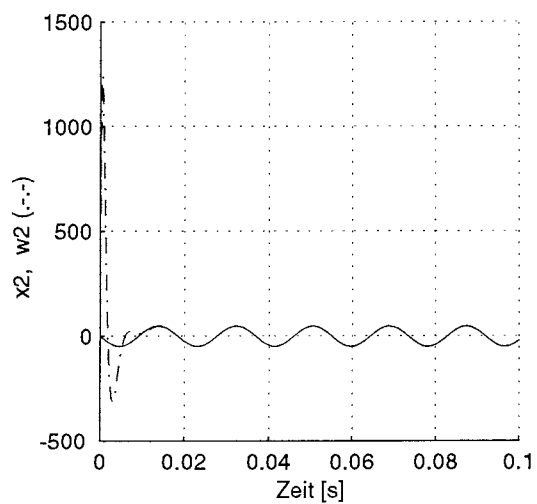
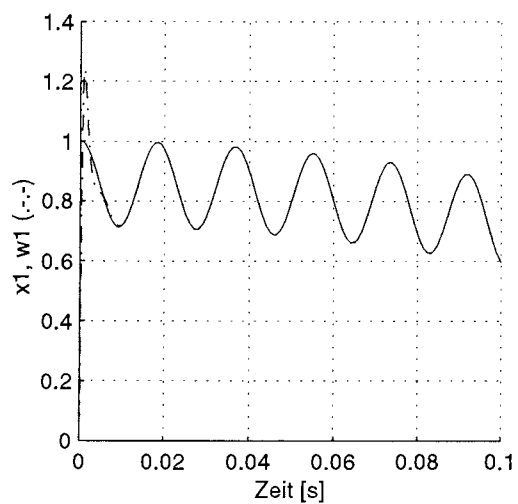
und mit allen Eigenwerten der Systemmatrix des Beobachterfehlers bei

$$s = -1000 \quad (4.28)$$

erhält man mit den Anfangswerten des Systems

$$\mathbf{x}(0) = [1 \ 0 \ 0 \ 0] \quad (4.29)$$

bei verschwindender Eingangsgröße u die folgenden Simulationsergebnisse, wobei die vom Beobachter geschätzten Größen jeweils gestrichelt gezeichnet sind.



5 Literatur:

- [1] Jean-Jacques E. Slotine, Weiping Li: Applied Nonlinear Control
- [2] Gausch, F.: Linearisierung und Entkopplung von Mehrgrößensystemen mit impliziten Ausgangsgleichungen, e&i, 365 - 371, 7/8/1993
- [3] Pia Müller: Linearisierung und Entkopplung von Deskriptorsystemen, Dissertation D 14-151 Universität Paderborn 2000.
- [4] Isidori, A: Nonlinear Control Systems, 2nd Edition, Springer Verlag, Berlin 1989
- [5] Verghese, G.C., Levy, B.C., Kailath, T., A Generalized State-Space for Singular Systems, IEEE Transactions on Automatic Control, Vol. AC - 26, August 1981

Über das Filtern verrauschter Daten in hochgenauen Bondmaschinen

(Hugo Pristauz / Michael Loinger)

Abstrakt

In hochgenauen Bondmaschinen mit Platziergenauigkeiten im $10\mu\text{m}$ Bereich und darunter ist die periodische Temperaturkalibrierung zur Kompensation thermischer Driften eine Schlüsseltechnologie. Dabei führt die Maschine regelmäßige Kalibriermessungen durch, die jedoch produktionsbedingt nicht zu exakt periodischen Zeitabständen stattfinden. Da diese Messungen verrauscht sind, spielt die Filterung der Meßdaten eine entscheidende Rolle. Dabei treten bei der Auswahl geeigneter Filtermethoden zwei Herausforderungen auf: Zum einen werden zeitdiskrete Filter mit variabler Abtastzeit benötigt. Zum anderen liefern Filter bei rampenförmigen Eingangssignalen (bedingt durch die Temperaturdrift) einen Schleppfehler, den es zu vermeiden gilt. Einen Zugang zu beiden Herausforderungen findet man in diesem Beitrag.

1 Einführung

1.1 Chip&Wire versus Flip-Chip

Die höchsten Platziergenauigkeiten bei der Nacktchipbestückung im Rahmen der elektronischen Assemblierung (electronic packaging) werden heute von sogenannten "Flip-Chip-Bondern" verlangt. Im Vergleich dazu werden beim konventionellen Chip&Wire-Bonden Nacktchips mit der aktiven Seite nach oben im sogenannten *Die-Attach-Schritt* auf ein Substrat gebondet und im Nachfolgeschritt die elektrischen Verbindungen sequentiell vom Wire-Bonder Draht für Draht gezogen werden (Abb. 1a).

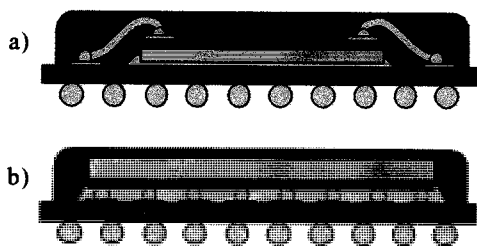


Abb. 1: Chip&Wire-Package (a) und Flip-Chip-Package (b)

Beim Flip-Chip-Bonden hingegen werden die Chips mit der aktiven Seite nach unten (geflippt) auf das

Substrat aufgesetzt, und die elektrischen Verbindungen durch Simultankontaktierung mit Hilfe sogenannter Bumps (Lotkugeln) durchgeführt (Abb. 1b), die in Maskenprozessen auf Waferebene hergestellt werden (Abb. 2a,b). Vor dem Aufsetzen des gebumpten Chips auf das Substrat werden die Bumps in einen Fluxfilm getaucht. In einem nachfolgenden Batch-Reflow-Prozess findet die eigentliche Kontaktierung statt, bei der Chip und Substrat mit Hilfe der Bumps simultan verlötet werden (Abb. 2c).

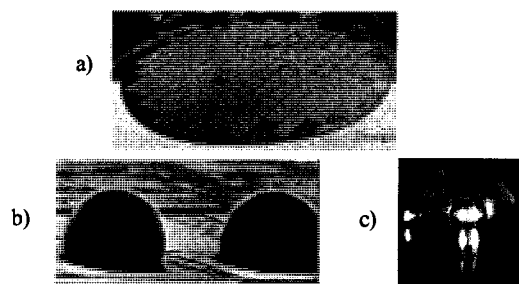


Abb. 2: a) Gesägter Wafer auf Waferfolie; b) Lotbumps am Wafer; c) Flip-Chip-Bond nach dem Reflow-Prozess

Experten sind sich einig, dass dieses Flip-Chip-Verfahren in nicht mehr allzu langer Zeit aufgrund vieler Vorteile (u.a. bessere elektrischen Hochfrequenzeigenschaften, billigere Simultankontaktierung bei hohen Anschlusszahlen, ...) der Chip&Wire Technik qualitativ und quantitativ überlegen sein.

1.2 Maschinenfähigkeit

Ein Treiber für die Platziergenauigkeit der Bondmaschine ist der Bump-Pitch. Man ist sich einig darüber, dass der Bump-Pitch aufgrund der wachsenden Komplexität der Chips und der fortschreitenden Miniaturisierung weiter sinken wird, womit die Platziergenauigkeit in den Mikrometer und Submikrometer-Bereich getrieben wird. Derzeit liegen die erwarteten Platziergenauigkeiten von Flip-Chip-Bondern bei $\pm 10\mu\text{m}@C_{pk}1.0$. Eine Feinheit bei der Bewertung der Maschinenfähigkeit ist darin zu sehen, dass der Bestückprozess stochastischer Natur ist und üblicherweise als normalverteilter Gauss-Prozess mit

Standardabweichung s und Mittelwertfehler μ beschrieben wird. Als Maß für die Maschinenfähigkeit wird der sogenannte Maschinenfähigkeitskoeffizient C_{mk} angesetzt,

$$C_{mk} := \frac{\Delta - |\mu|}{3\sigma}$$

wobei Δ die Platziertoleranz (hier: $10\mu\text{m}$) bezeichnet. Eine Genauigkeitsangabe $\pm 10\mu\text{m}@C_{mk}1.0$ ist so zu verstehen, dass der stochastische Prozess zunächst so kontrolliert sein muss, dass stationäre Bedingungen vorliegen, und die Platziergenauigkeit eine entsprechend hohe Qualität aufweisen muss, so dass eine Maschinenfähigkeitskennzahl $C_{mk} = 1$ gewährleistet werden kann. Es ist wichtig zu verstehen, dass diese Forderung nur dann erfüllt werden kann, wenn sowohl die Streuung des Prozesses (Standardabweichung s) klein als auch der Prozess gut zentriert ist (Mittelwertfehler μ gering). Bei einer Platziergenauigkeitsspezifikation $\pm 10\mu\text{m}@C_{mk}1.0$ ergeben sich die praktischen Forderungen (Faustregel): $s, \mu < \Delta/4 = 2.5\mu\text{m}$. Für Teilprozesse wie Bilderkennungsfunktionen gelten die praktischen Forderungen $s, \mu < 0.5\mu\text{m}$.

Um sich die Problematik bei der periodischen Temperaturkalibrierung zu vergegenwärtigen, stelle man sich eine Bondmaschine vereinfacht nach Abb. 3 vor. Dargestellt ist eine aufschauende Kamera, die am Grundgestell befestigt ist, und die Aufgabe hat, die relative Lage der Chips am Bondkopf (mittels Saugnadel gehalten) möglichst exakt zu vermessen. Dazu bringt die Maschine das an dem in $x/y/z$ beweglichen Bondkopf befindliche Chip über die aufschauende Kamera.

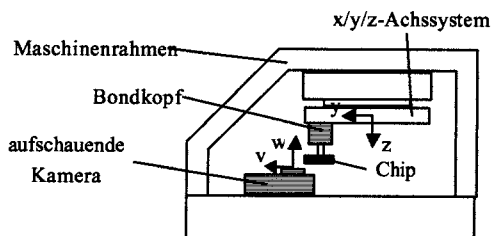


Abb. 3: Schematischer Aufbau einer Bondmaschine

Eine Bestimmung der exakten Lage des Chips am Bondkopf (Koordinatensystem $x/y/z$) ist nur dann möglich, wenn die relative Lage des Koordinatensystems der aufschauenden Kamera

($u/v/w$) zur aktuellen Lage des Bondkopfkoordinatensystems ($x/y/z$) genau bekannt ist. Diese relative Lage kann durch Kalibrierung z.B. so bestimmt werden, dass die aufschauende Kamera eine am Bondkopf befindliche Kalibriermarke vermisst.

Aufgrund der Temperaturdriften (z.B. unterschiedliche Ausdehnungen im Maschinenrahmen) ändert sich diese relative Lage der Koordinatensysteme mit der Zeit zueinander. Als Abhilfe kann die Kalibrierprozedur zu regelmäßigen Zeitabständen wiederholt werden. Bedingt durch Produktionseingriffe (z.B. Produktionsstops) kann aber niemals ein konstantes Abtastintervall gewährleistet werden. Damit hat man es mit zeitdiskreten Messdaten eines kontinuierlichen stochastischen Prozesses zu tun, deren Abtastzeitpunkte nicht äquidistant sind. Würde man ohne Filterung der Messdaten eine Kalibrierung durchführen, so würde sich aufgrund des Messrauschens ein Kalibrierfehler ergeben, der einen Mittelwertfehler bei der Platzierung jenseits der geforderten Vorgaben nach sich ziehen würde.

Deshalb wird als Abhilfe eine Messdatenfilterung durchgeführt. Allerdings tritt dann ein weiteres Problem auf, dass nämlich bei der Filterung rampenförmiger Trends signale, wie dies bei Drift-Erscheinungen üblich ist, ein Schleppfehler entsteht, der den Mittelwert wiederum verfälscht.

Die Herausforderung stellt sich also folgendermassen dar:

Zur Kompensation von Temperaturdrift-Erscheinungen gilt es Filteralgorithmen zu finden, die einerseits auf Messdaten mit nicht äquidistantem Abtastintervall anwendbar sind, und keine signifikanten Schleppfehler verursachen!

2 Filterung von nicht äquidistant abgetasteten Meßsignalen

2.1 Aufgabenstellung

Gegeben seien $N+1$ Zeitpunkte t_k und $N+1$ zugehörige Werte eines abgetasteten Signals:

$$t_0, t_1, \dots, t_N$$

$$u_0, u_1, \dots, u_N$$

Gesucht ist ein Filter zur Berechnung zeitdiskreter Werte eines gefilterten Signals

$$y_0, y_1, \dots, y_N$$

mit folgenden Eigenschaften:

- Das Filter soll kausal sein (d.h. dass zur Berechnung der Ausgangsgröße keine zukünftigen Werte der Eingangsgröße benötigt werden) und Tiefpaßeigenschaften haben.
- Für äquidistante Abtastung $t_k = k T_a$ soll sich das Filter wie ein zeitdiskreter Tiefpaß 1. Ordnung mit Zeitkonstante T verhalten.
- Der Verstärkungsfaktor des Filters soll gleich 1 sein, d.h. im stationären Fall $u_k = u_\infty$ soll gelten $y_k = y_\infty = u_\infty$.
- Bei kleinem Abtastintervall $t_k - t_{k-1} \ll T$ soll das gefilterte Signal y_k näherungsweise dem vorhergehenden Filtersignal y_{k-1} entsprechen (d.h. $y_k \approx y_{k-1}$).
- Bei großem Abtastintervall $t_k - t_{k-1} \gg T$ soll das gefilterte Signal y_k dem neuen Eingangssignal u_k entsprechen (d.h. $y_k \approx u_k$).

Bemerkungen

Die letzten beiden Eigenschaften sind speziell bei der Temperaturkalibrierung hochgenauer Bondmaschinen wichtig, da sich die Kalibrierintervalle zeitlich ändern können. Bei einer periodischen Kalibrierung in einer laufenden Produktion sind die Kalibrierintervalle eher kurz. Um den Einfluß der Streuung zu minimieren, muß das Filter möglichst viel Information aus der Historie mit einbeziehen, was u.a. bedeutet, dass die Abweichung $y_k - y_{k-1}$ aufgrund der Stetigkeit nicht allzu groß sein soll. Ist die Maschine hingegen im Produktionsstart oder im Wiederanlauf nach einer

längeren Produktionspause, so dürfen historische Werte nicht mehr als relevant betrachtet werden. Der Filterausgang y_k sollte dann möglichst dem Meßwert u_k entsprechen.

2.2 Algorithmus

Das Filter berechnet aus der zu den Zeitpunkten (t_k) ermittelten Eingangsfolge (u_k) die Ausgangsfolge (y_k) .



Sei y_{k-1} der zuletzt aktualisierte Ausgangswert des Filters zum Zeitpunkt t_{k-1} . Dann berechnet man aus der aktuellen Eingangsgröße u_k zum Zeitpunkt t_k die neue gefilterte Größe y_k nach folgendem Algorithmus:

$$a_k = e^{-\frac{t_k - t_{k-1}}{T}}$$

$$y_k = a_k \cdot y_{k-1} + (1 - a_k) \cdot u_k$$

Dabei ist T die Filterzeitkonstante. Für $k = 0$ setze man $a_0 = 0$.

Bemerkungen

- Das Filter hat alle oben geforderten Eigenschaften a) bis e).
- Zusätzlich hat das Filter die angenehme Eigenschaft, dass die Impulsantwort, also die Antwort auf die Eingangsfolge $u_0(0) = 1, u_1(t_1) = u_2(t_2) = u_3(t_3) = \dots = 0$, als Abtastung der Trägerfunktion $y_k = y(t_k) = e^{-t_k/T}$ entsteht.

Überprüfung der geforderten Eigenschaften

a,b) Tiefpaßeigenschaft: Dazu nimmt man äquidistante Abtastzeitpunkte der Form $t_k = k T_a$ an. Es wird dann $a_k = e^{-T_a/T} = \text{const} = a$, und es entsteht die Filterdifferenzengleichung eines linearen zeitinvarianten zeitdiskreten Tiefpaßfilters:

$$y_k = a y_{k-1} + (1-a) u_k$$

c) **Stationäres Verhalten:** Zur Berechnung des stationären Verhaltens setze man $u_k = u_\infty$ und $y_k = y_\infty$. Es folgt dann

$$y_\infty = a_k y_\infty + (1-a_k) u_\infty$$

$$y_\infty (1-a_k) = (1-a_k) u_\infty$$

$$y_\infty = u_\infty$$

woraus sich ein Verstärkungsfaktor $V = y_\infty / u_\infty = 1$ unabhängig von den Abtastzeitpunkten ergibt.

d) **Kleine Abtastintervalle:** Für kleine Abtastintervalle $\Delta t_k := t_k - t_{k-1} \ll T$ berechnet man

$$a_k = e^{-\Delta t_k / T} = 1 - \frac{\Delta t_k}{T} + \frac{1}{2!} \left(\frac{\Delta t_k}{T} \right)^2 -$$

$$\frac{1}{3!} \left(\frac{\Delta t_k}{T} \right)^3 + \dots \approx 1$$

$$y_k = a_k y_{k-1} + (1-a_k) u_k \approx 1 y_{k-1} + (1-1) u_k$$

$$y_k \approx y_{k-1}$$

e) **Große Abtastintervalle:** Für große Abtastintervalle $\Delta t_k := t_k - t_{k-1} \gg T$ berechnet man

$$e^{-\Delta t_k / T} \ll 1$$

$$y_k = a_k y_{k-1} + (1-a_k) u_k$$

$$y_k = e^{-\Delta t_k / T} y_{k-1} + (1-e^{-\Delta t_k / T}) u_k \approx 0 y_{k-1} + (1-0) u_k$$

$$y_k \approx u_k$$

f) **Trägerfunktion:** Der Beweis läuft über vollständige Induktion. Zu zeigen ist, dass bei einer impulsförmigen Eingangsgröße $u_0(0) = 1$, $u_1(t_1) = u_2(t_2) = u_3(t_3) = \dots = 0$, die Filterausgangsgrößen als Abtastung der Trägerfunktion $y_k = y(t_k) = e^{-t_k / T}$ entstehen. Man zeigt zuerst die Gültigkeit für $t_0 = 0$, $u(t_0) = 1$.

$$a_0 = 0$$

$$y_0 = a_0 y_{-1} + (1-a_0) u_0 = u_0 = 1 = e^{-t_0 / T}$$

Ausgehend von $y(t_{k-1}) = e^{-t_{k-1} / T}$ ist dann für $k > 0$

durch Anwendung der Filtergleichung zu zeigen, dass auch $y_k = y(t_k) = e^{-t_k / T}$ einem Abtastwert der Trägerfunktion entspricht.

$$y_k = e^{-(t_k - t_{k-1}) / T} y_{k-1} + (1 - e^{-(t_k - t_{k-1}) / T}) \cdot 0$$

$$(u_k = 0 \text{ für } k > 0)$$

$$y_k = e^{-(t_k - t_{k-1}) / T} \cdot e^{-t_{k-1} / T} = e^{-t_k / T}$$

2.3 Ergebnisse

Abb. 4 zeigt das Verhalten des Filters bei gleichmäßiger Abtastung, wobei das Eingangssignal unverrauscht ist. Man erkennt gleichzeitig eine Schwäche des Filters, dass nämlich eine Art Schleppfehler entsteht. In Abb. 5 wird das Abtastintervall variiert. Man erkennt deutlich, dass das Filter gut mit der Variation des Abtastintervalls zurecht kommt.

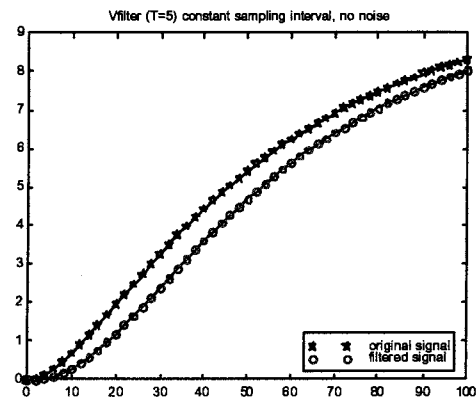


Abb. 4: Gleichmäßige Abtastung, Signal ohne Rauschen

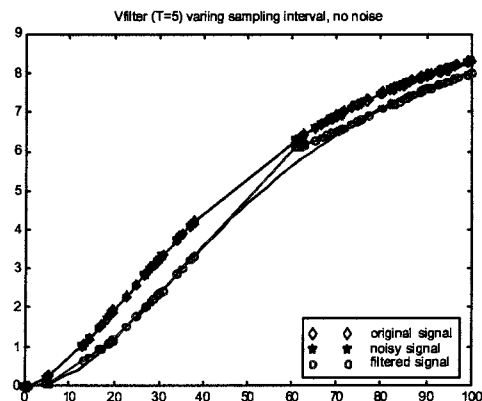


Abb. 5: Variables Abtastintervall, Signal ohne Rauschen

Im Vergleich dazu ist in Abb. 6 die Antwort eines konventionellen Filters dargestellt, bei dem so getan

wird, als ob das Abtastintervall konstant wäre. Im Zeitraum 40 bis 60 kommen keine neuen Informationen. Man sieht deutlich, dass das Filter zu sehr historische Werte mit einbezieht, unwissentlich, dass eine beträchtliche Pause bis zur nächsten Abtastung aufgetreten ist.

Die Stärke des hier vorgestellten Filteralgorithmus ist in Abb. 7 erkennbar. Die längere Pause bei der Abtastung wirkt sich nicht mehr negativ auf das Filtersignal aus. Das generelle Nachhinken muß jedoch nach wie vor in Kauf genommen werden.

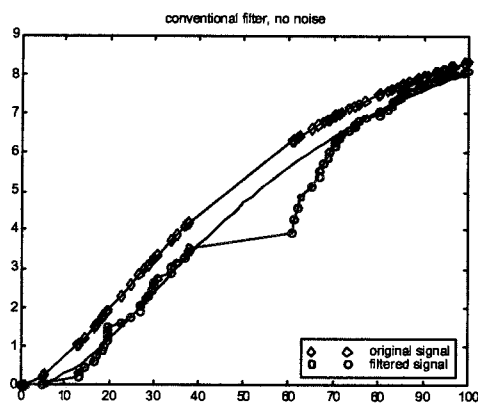


Abb. 6: Schwäche eines konventionelles Filter

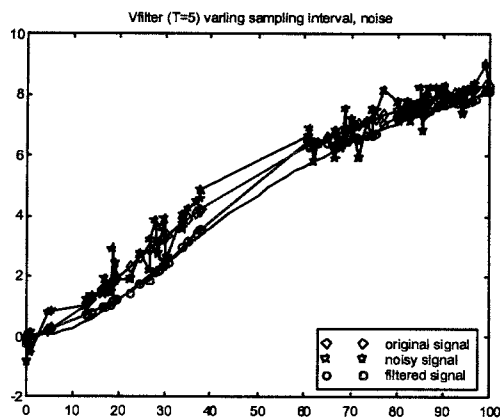


Abb. 7: Vfilter angewandt auf ein verrauschtes Signal

2.4 MATLAB-Funktion VFILTER

```
function y = vfilter(t,u,T)
%
% VFILTER Special filter for non equal sampling
% intervals
%
% y = vfilter(t,u,T)
%
% Arguments:
% y: vector [y0 y1 ... yN] (output sequence)
% t: vector [t0 t1 ... tN] (sampling time stamps)
% u: vector [u0 u1 ... uN] (input sequence)
% T: sampling time constant
%
% Example:
% t = sort(rand(1,100));
% t = t - min(t); t = t/max(t)*100;
% U = 1 + 1/48*exp(-t/2) + (1+1/48)*exp(-t/50);
% u = U + 0.05*randn(size(U));
% y = vfilter(t,u,2);
%
% y = 0*u; % dimension y
% t = t(:); y = y(:); % make vectors
%
% if length(t) ~= length(u)
% error('need equal dimension for arg1, arg2!');
% end
%
% y(1) = u(1);
% for (k=2:length(t))
% dt = t(k) - t(k-1);
% a = exp(-dt/T);
% y(k) = a*y(k-1) + (1-a)*u(k);
% end
% eof
```

3 Kausale Filter ohne Schleppfehler

3.1 Problematik

Zur Verdeutlichung der Problematik wird ein zeitkontinuierlicher Tiefpaß 1. Ordnung betrachtet, der auf die Rampe $u(t) = t$ angewendet wird.

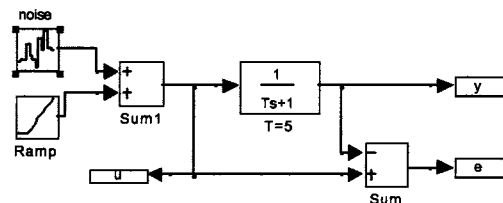


Abb. 8: Filterung eines verrauschten Signals

Die Filterdifferentialgleichung lautet:

$$T \cdot dy/dt + y = u$$

Für Abtastzeit $T_a = 1$ und Filterzeitkonstante $T = 5$ ist

die Filterantwort y und der Schleppfehler

$$e := u - y$$

in Abb. 9 dargestellt.

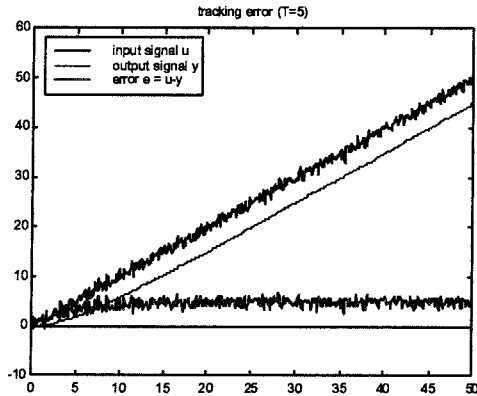


Abb. 9: Auftretender Schleppfehler (e)

Als Modifikation des Filters kann folgendermaßen vorgegangen werden: Man filtert den Fehler e (z.B. mit einem gleichartigen Filter) und addiert den gefilterten Fehler f zum gefilterten Eingangssignal v . Das Strukturbild dazu ist in Abb. 10 dargestellt, das Ergebnis in Abb. 11. Dabei wird das vorher mit y bezeichnete Signal jetzt mit v bezeichnet.

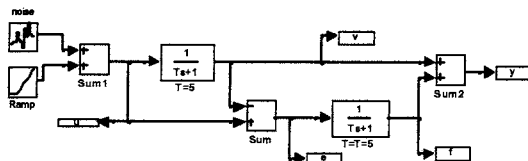


Abb. 10: Modifizierte Filterstruktur

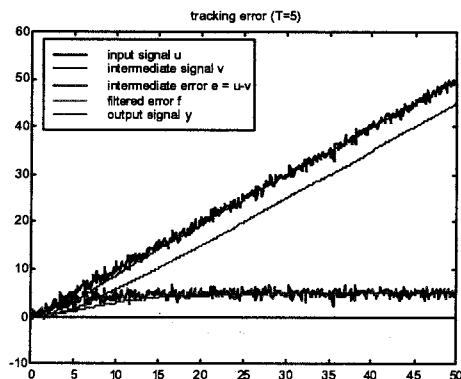
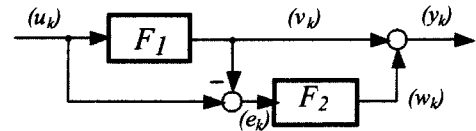


Abb. 11: Antwort des modifizierten Filters

Diese Idee läßt sich auch in den zeitdiskreten Bereich übertragen:

Twin-Filter Methodik

Zur Filterung einer Folge (u_k) zu gegebenen Zeitpunkten (t_k) gehe man wie folgt vor:



- 1) Berechne die Zwischengröße $(v_k) = F_1 \bullet (u_k)$
- 2) Berechne die Fehlerfolge $(e_k) = (u_k - v_k)$
- 3) Berechne die gefilterte Fehlerf. $(w_k) = F_2 \bullet (e_k)$
- 4) Gefiltertes Signal: $(y_k) = (v_k + w_k)$

Dabei wird davon ausgegangen, dass der Verstärkungsfaktor des Filters gleich 1 ist.

3.2 Theoretische Analyse

Der Einfachheit halber wird die Analyse für den zeitkontinuierlichen Bereich unter Zuhilfenahme der Laplace-Transformation durchgeführt. Eine Übertragung in den zeitdiskreten Bereich ist jedoch durch Anwendung der z-Transformation in geradliniger Weise möglich.

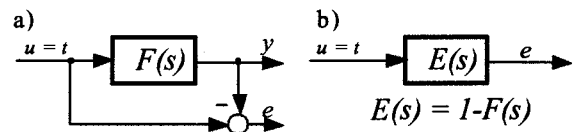


Abb. 12: Zur Darstellung des Schleppfehlers

Dazu betrachte man die Abb. 12a, welche zeigt, wie das Signal y aus Filterung des Eingangssignals u entsteht, und wie der auftretende Schleppfehler $e = u - y$ berechnet wird.

Unter Anwendung der Laplace-Transformation erhält man im Bildbereich:

$$\bar{e}(s) = E(s) \cdot \bar{u}(s) = (1 - F(s)) \cdot \bar{u}(s)$$

Soll der Schleppfehler verschwinden, muß gelten

$$e_{\infty} := \lim_{t \rightarrow \infty} e(t) = 0$$

Nachdem $u(t) = t$ bzw. $\bar{u}(s) = 1/s^2$ erhält man unter Anwendung des Grenzwertsatzes der Laplace-Transformation

$$e_{\infty} = \lim_{s \rightarrow 0} s \cdot E(s) \bar{u}(s) = \lim_{s \rightarrow 0} s E(s) \frac{1}{s^2} = \lim_{s \rightarrow 0} \frac{E(s)}{s}.$$

Dies bedeutet, dass $E(s)$ (mindestens) eine doppelte Nullstelle bei $s=0$ aufweisen muß, wenn das Filter schleppfehlerfrei sein soll.

Anstelle der *Rampenantwort* der Fehler-Übertragungsfunktion $E(s)$ kann man äquivalent die *Sprungantwort* der um einen Integrator erweiterten Übertragungsfunktion

$$S(s) := \frac{E(s)}{s}$$

untersuchen. Diese muß dann mindestens eine einfache Nullstelle bei $s=0$ aufweisen.

Für ein schleppfehlerfreies zeitkontinuierliches Filter mit Übertragungsfunktion $F(s)$ muß gelten:

$E(s) = 1 - F(s)$ hat (mindestens) eine Doppelnullstelle bei $s = 0$

bzw. $S(s) = E(s)/s$ hat (mindestens) eine einfache Nullstelle bei $s = 0$!

Analog erhält man durch Anwendung der z-Transformation für ein zeitdiskretes Filter:

Für ein schleppfehlerfreies zeitdiskretes Filter mit z-Übertragungsfunktion $F(z)$ muß gelten:

$E(z) = 1 - F(z)$ hat (mindestens) eine Doppelnullstelle bei $z = 1$

bzw. $S(z) = E(z) \cdot z/(z-1)$ hat (mindestens) eine einfache Nullstelle bei $z = 1$!

Beispiel: Ein zeitkontinuierliches Tiefpaßfilter mit Verstärkungsfaktor 1 hat die Übertragungsfunktion

$$F(s) = \frac{1}{sT + 1}$$

Die Fehlerübertragungsfunktion lautet:

$$E(s) = 1 - F(s) = \frac{sT + 1 - 1}{sT + 1} = \frac{sT}{sT + 1}$$

Da $E(s)$ nur eine einfache Nullstelle bei $s = 0$ hat, tritt folgender Schleppfehler auf:

$$e_{\infty} = \lim_{s \rightarrow 0} \frac{E(s)}{s} = \lim_{s \rightarrow 0} \frac{sT}{s(sT + 1)} = T$$

3.3 TwinFilter

Man betrachte nun ein Filter, das nach Abb. 13 aufgebaut ist, wobei die Teilübertragungsfunktionen $F_1(s)$ und $F_2(s)$ als stabile Übertragungsfunktionen mit Verstärkungsfaktoren 1 vorausgesetzt werden, d.h. $V_1 = F_1(0) = 1$ bzw. $V_2 = F_2(0) = 1$.

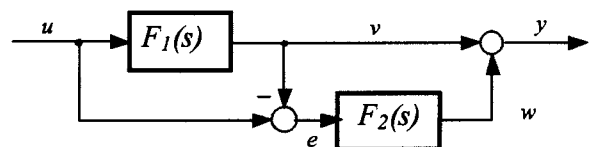


Abb. 13: Zeitkontinuierliche TwinFilter-Struktur

Für die Filtergesamtübertragungsfunktion $F(s) = \bar{y}(s)/\bar{u}(s)$ gilt dann:

$$\begin{aligned} F(s) &= F_1(s) + F_2(s) \cdot (1 - F_1(s)) = \\ &= F_1(s) - F_1(s)F_2(s) + F_2(s) \end{aligned}$$

Die Fehlerübertragungsfunktion errechnet sich:

$$\begin{aligned} E(s) &= 1 - F(s) = 1 - F_1(s) - F_2(s) + F_1(s)F_2(s) \\ E(s) &= (1 - F_1(s)) \cdot (1 - F_2(s)) \end{aligned}$$

Setzt man $E_1(s) := 1 - F_1(s)$ und $E_2(s) := 1 - F_2(s)$, so sieht man aufgrund der Tatsache, dass die Verstärkungsfaktoren eins sind, dass sowohl $E_1(s)$ als auch $F_2(s)$ eine Nullstelle bei $s = 0$ hat.

$$\begin{aligned} E_1(0) &= 1 - F_1(0) = 1 - V_1 = 1 - 1 = 0 \\ E_2(0) &= 1 - F_2(0) = 1 - V_2 = 1 - 1 = 0 \end{aligned}$$

Folglich gilt mit

$$E(s) = (1-F_1(s)) \cdot (1-F_2(s)) = E_1(s)E_2(s),$$

dass $E(s)$ mindestens eine Doppelnulstelle bei $s = 0$ hat, also das Filter mit Übertragungsfunktion $F(s)$ schleppfehlerfrei ist.

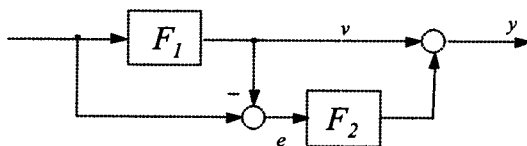
Zusammenfassung

TwinFilter-Struktur

Seien F_1 und F_2 stabile Filter mit Verstärkungsfaktor 1. Dann kann ein schleppfehlerfreies Filter wie folgt zusammengesetzt werden.

$$F = F_1 + F_2 - F_1F_2$$

Dies gilt sowohl für den zeitkontinuierlichen als auch zeitdiskreten Bereich. Ein Strukturbild für das zusammengesetzte Filter ist nachfolgend dargestellt.



Beispiel: zeitkontinuierliches Filter

Sei $F_1(s) = F_2(s) = 1/(sT+1)$ mit $T = 5$. Dann lautet die zusammengesetzte Filterübertragungsfunktion

$$F(s) = 2/(sT+1) - 1/(sT+1)^2 = \frac{2sT+1}{(sT+1)^2}$$

bzw. die Fehlerübertragungsfunktion

$$E(s) = 1 - F(s) = 1 - \frac{2sT+1}{(sT+1)^2} = \frac{s^2T^2}{(sT+1)^2}$$

Die integrierte Fehlerübertragungsfunktion $S(s)$ lautet:

$$S(s) = E(s)/s = \frac{sT^2}{(sT+1)^2}$$

Der Amplitudengang von $F(j\omega)$ und $S(j\omega)$ ist in Abb. 14 dargestellt. In Abb. 15 ist die Sprungantwort und der Schleppfehler des Filters dargestellt.

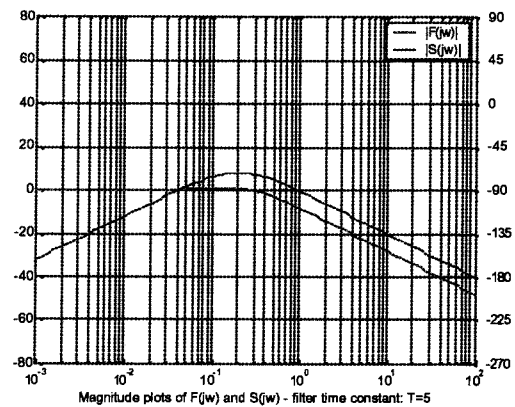


Abb. 14: Amplitudengänge von $F(s)$ und $S(s)$

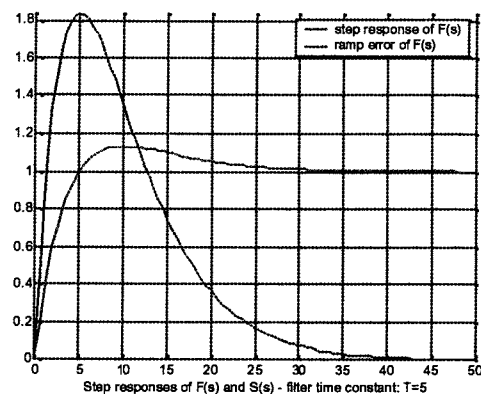


Abb. 15: Sprungantwort und Schleppfehler des Filters

4 TwinFilter-Ansätze

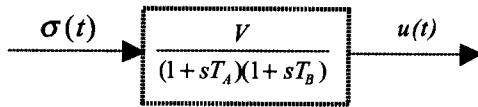
Nachfolgend werden verschiedene Ansätze für ein zeitkontinuierliches TwinFilter untersucht:

- auf Basis von Tiefpässen 1. Ordnung
- auf Basis von Tiefpässen 2. Ordnung mit Grad-differenz 2
- auf Basis von Tiefpässen 2. Ordnung mit Grad-differenz 1

Interessant dabei ist immer der Amplitudengang der Filterübertragungsfunktion $F(s)$ sowie der integrierten Fehlerübertragungsfunktion $S(s)$ im Frequenzbereich, sowie im Zeitbereich die Sprungantworten der beiden Übertragungsfunktionen, welche die Filtersprungantwort (Sprungantwort von $F(s)$) und den Rampenschleppfehler (Sprungantwort von $S(s)$) darstellen.

Als weiteren Test kann man sich die Antwort auf

glattes s-förmiges verrauschtes Signal ansehen. Ein solches Testsignal ist z.B. durch die Sprungantwort zweier hintereinandergeschalteter Tiefpässe 1.Ordnung mit Zeitkonstanten T_A und T_B gegeben.



Die Sprungantwort des Systems in geschlossener Form lautet:

$$u(t) = V \cdot \left(1 + \frac{T_A}{T_B - T_A} e^{-t/T_A} + \frac{T_B}{T_A - T_B} e^{-t/T_B} \right)$$

Für $V=10$, $T_A = 10$ und $T_B = 50$ ist diese Testfunktion in Abb. 16 dargestellt.

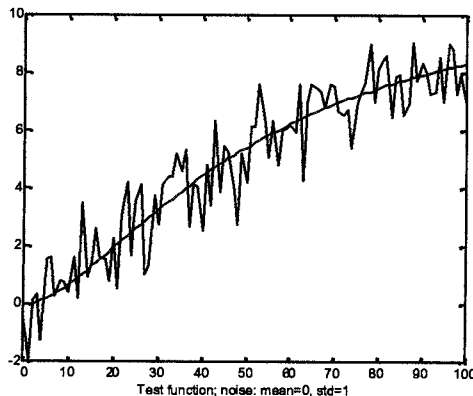


Abb. 16: Glatte Testfunktion (mit/ohne Rauschen)

4.1 Twinfilter auf Basis von Tiefpässen 1. Ordnung

Man setze

$$F_1(s) = \frac{1}{1 + sT_1}, \quad F_2(s) = \frac{1}{1 + sT_2}$$

und bilde

$$F(s) = F_1(s) + F_2(s) + F_1(s)F_2(s)$$

$$S(s) = (1 - F(s)) / s$$

Als konkrete Zahlenwerte wähle man $T = 5$ und setze $T_1 = T/\sqrt{k}$, $T_2 = T \cdot \sqrt{k}$ und untersuche die Filtercharakteristik für die Werte: $k = 1, 2, 5, 10, 20$. Beachte, dass

$$k = T_2 / T_1.$$

Man erkennt, dass mit steigender Spreizung der beiden Filterzeitkonstanten $k = T_2 / T_1$ das Maximum des

Schleppfehlers reduziert werden kann, die Transiente des Schleppfehlers jedoch langsamer abklingt (auf die Vergrößerung von T_2 zurückzuführen).

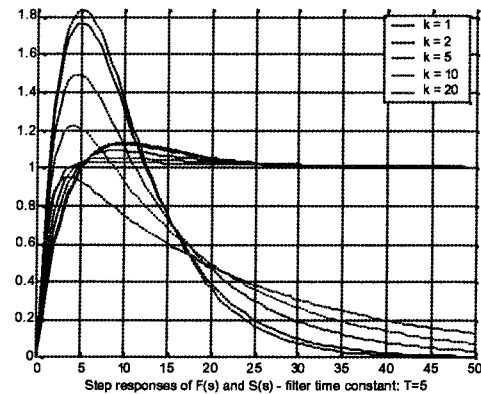


Abb. 17: Sprungantworten und Schleppfehler

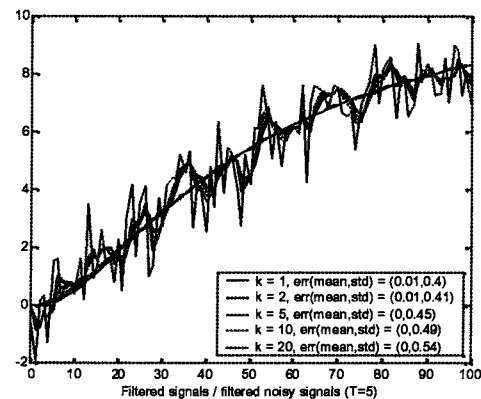


Abb. 18: Filterverhalten ($T=5$)

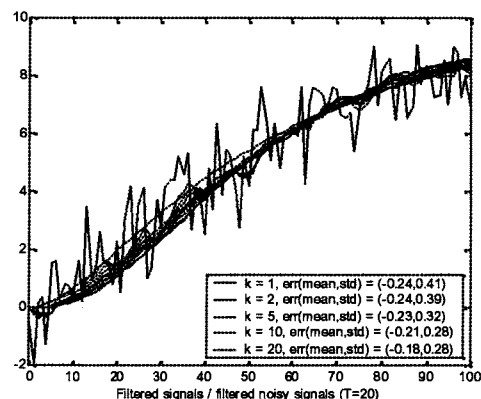


Abb. 19: Filterverhalten ($T=20$)

Abb. 18 zeigt das Filterverhalten. Man erkennt, dass bei ungespreizten Zeitkonstanten die Standardabweichung geringer ist ($\sigma=0.4$ bei $k=1.0$) als bei großer

Sprenzung ($\sigma=0.54$ bei $k=20$). Bei Vergrößerung der Filterzeitkonstante um den Faktor 4 zeigt sich das Verhalten in genau umgekehrter Charakteristik ($\sigma=0.41$ bei $k=1.0$ bzw. $\sigma=0.28$ bei $k=20$)

4.2 Twinfilter auf Basis von Tiefpässen 2. Ordnung, Graddifferenz 2

Man setze

$$F_1(s) = \frac{1}{1 + 2dT_s + T^2s^2}, \quad F_2(s) = F_1(s)$$

und bilde

$$F(s) = F_1(s) + F_2(s) + F_1(s)F_2(s)$$

$$S(s) = (1 - F(s)) / s$$

Als konkrete Zahlenwerte wähle man $T = 5$ und setze für den Dämpfungsparameter d die Werte:

$d = 0.4, 0.6, 0.7, 0.8, 1.0, 1.2$

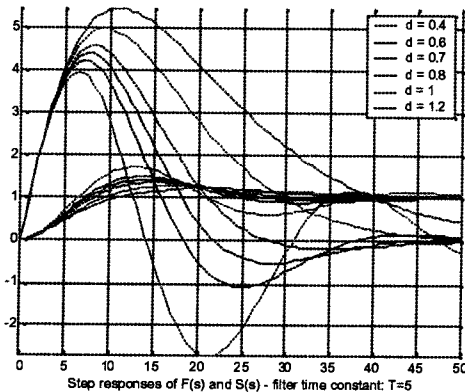


Abb. 20: Sprungantworten und Schleppfehler

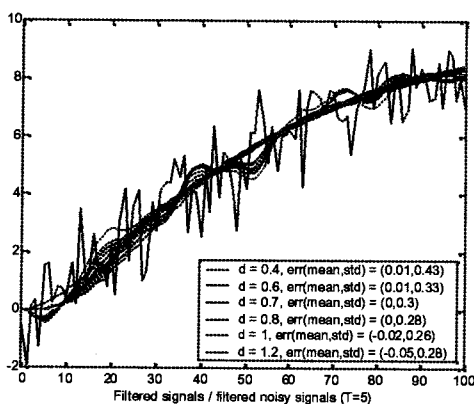


Abb. 21: Filterverhalten (T=5)

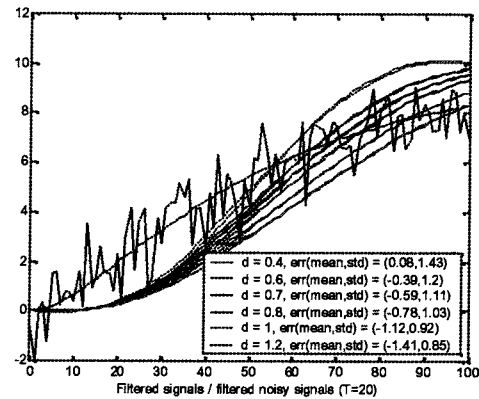


Abb. 22: Filterverhalten (T=20)

Das Filter neigt prinzipiell zu einem größeren Schleppfehler (Abb. 20). Bei richtig abgestimmter Zeitkonstante wirkt sich dies allerdings nicht wesentlich aus, und man erhält durch die steile Abschneidcharakteristik des Filters (40dB/Dekade) eine geringe Standardabweichung des Fehlers ($\sigma \leq 0.3$ für $d \geq 0.7$ - Abb. 21). Bei Vergrößerung der Filterzeitkonstante um den Faktor 4 kommt allerdings der große Schleppfehler voll zum Tragen (Abb. 22).

4.3 Twinfilter auf Basis von Tiefpässen 2. Ordnung, Graddifferenz 1

Man setze

$$F_1(s) = \frac{1 + sT}{1 + 2dT_s + T^2s^2}, \quad F_2(s) = F_1(s)$$

und bilde

$$F(s) = F_1(s) + F_2(s) + F_1(s)F_2(s)$$

$$S(s) = (1 - F(s)) / s$$

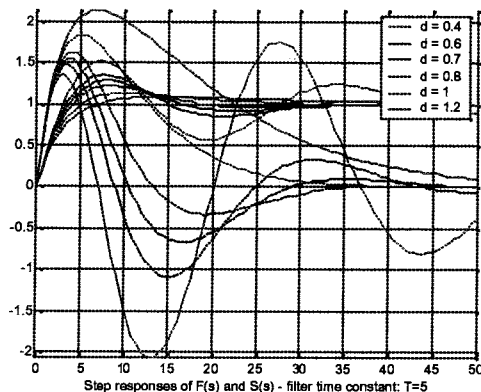


Abb. 23: Sprungantworten und Schleppfehler

Als konkrete Zahlenwerte wähle man $T = 5$ und setze für den Dämpfungsparameter d die Werte:

$d = 0.4, 0.6, 0.7, 0.8, 1.0, 1.2$

Man beachte, dass für $d = 1$ aufgrund der Pol/Nullstellenkürzung ein normaler Tiefpaß 1. Ordnung als Basis für das zweistufige TwinFilter vorliegt.

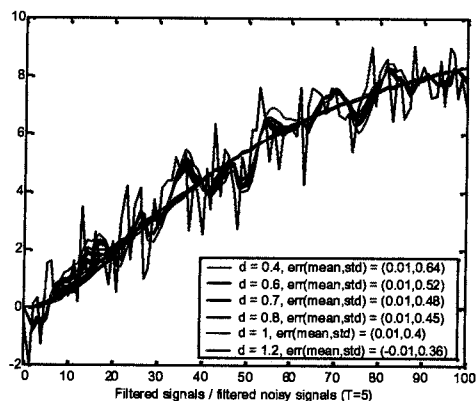


Abb. 24: Filterverhalten (T=5)

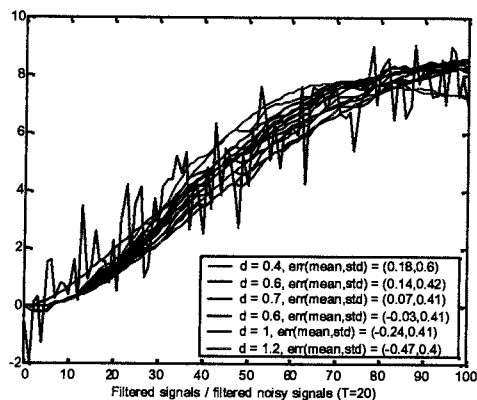


Abb. 25: Filterverhalten (T=20)

5 Vergleich mit schleppfehler-behafteten Filtern

5.1 Twinfilter auf Basis von Tiefpässen 1. Ordnung

Man setze

$$F(s) = \frac{1}{1 + sT}$$

Als konkrete Zahlenwerte wähle man $T = 1, 2, 5, 10, 20$.

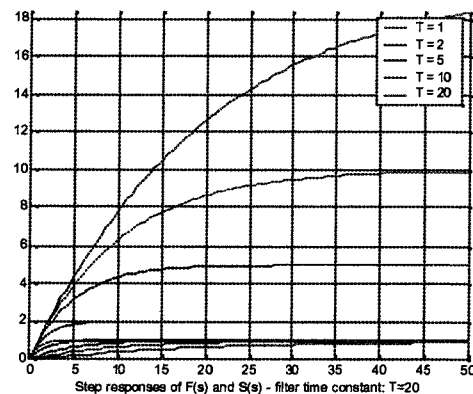


Abb. 26: Sprungantworten und Schleppfehler

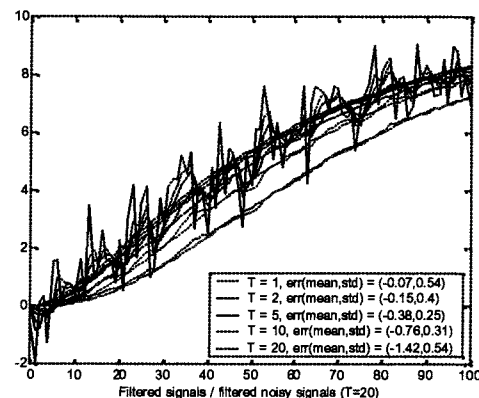


Abb. 27: Filterverhalten

5.2 Tiefpaß 2. Ordnung, Graddiff. 1

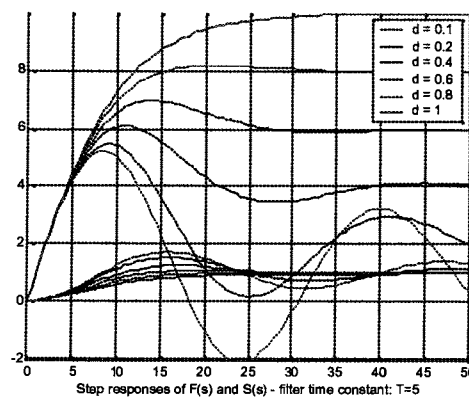


Abb. 28: Sprungantworten und Schleppfehler

Man setze

$$F(s) = \frac{1 + sT}{1 + 2dT s + T^2 s^2},$$

wähle als konkrete Zahlenwerte $T = 5$ und setze für den Dämpfungsparameter d die Werte:

$d = 0.4, 0.6, 0.7, 0.8, 1.0, 1.2$

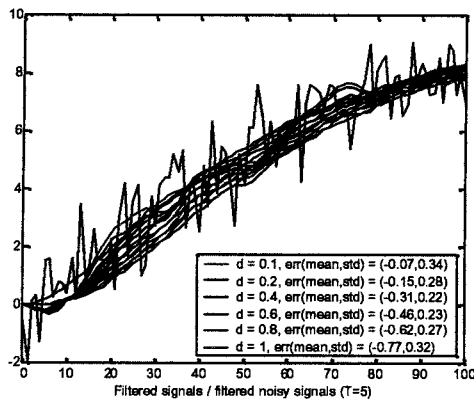


Abb. 29: Filterverhalten (T=5)

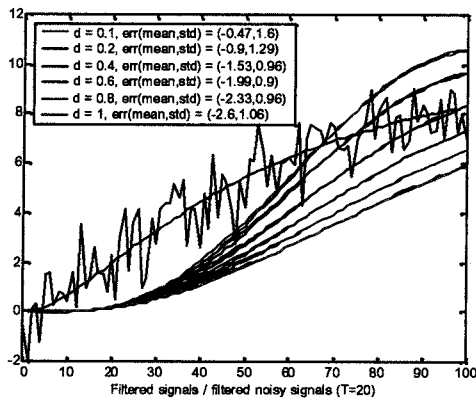


Abb. 30: Filterverhalten (T=20)

5.3 Tiefpaß 2. Ordnung, Graddifferenz 1

Man setze

$$F(s) = \frac{1 + sT}{1 + 2dT s + T^2 s^2},$$

wähle als konkrete Zahlenwerte $T = 5$ und setze für den Dämpfungsparameter d die Werte:

$d = 0.4, 0.6, 0.7, 0.8, 1.0, 1.2$

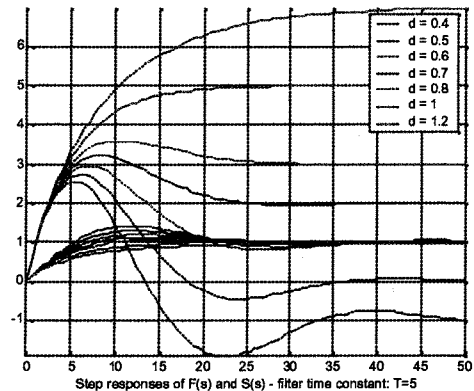


Abb. 31: Sprungantworten und Schleppfehler

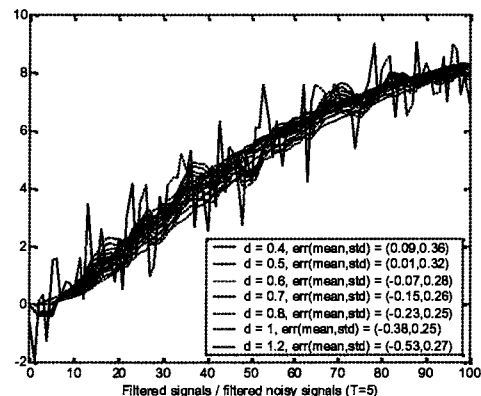


Abb. 32: Filterverhalten (T=5)

Das Interessante ist, dass beim Dämpfungsfaktor $d = 0.5$ eine Nullstelle von $S(s)$ bei $s = 0$ was nichts anderes heißt, als dass das Filter schleppfehlerfrei ist.

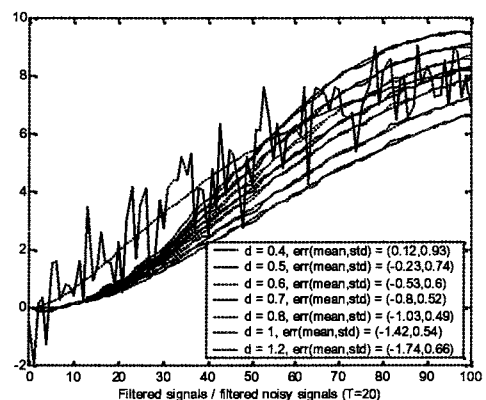


Abb. 33: Filterverhalten (T=20)

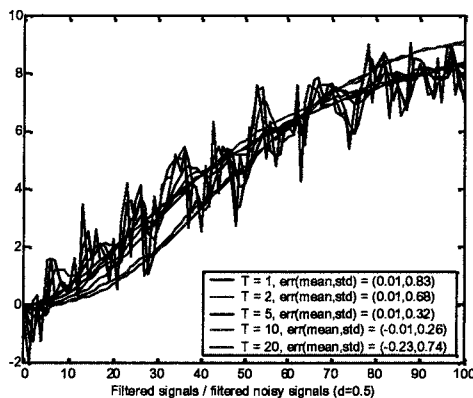


Abb. 34: Filterverhalten ($d=0.5$, T variabel)

Die entsprechend guten Filtereigenschaften dieses Spezialfalles sind in Abb. 34 noch einmal dargestellt, wobei die Dämpfung konstant ist und die Filterzeitkonstante variiert wird.

Man fragt sich, ob durch Hintereinanderschalten zweier solcher Filter, d.h.

$$F^*(s) = F(s)^2$$

das Filterverhalten verbessert wird. Laut Theorie fällt der Amplitudengang ab der Eckfrequenz $1/T$ mit 40dB/Dekade. Die Abb. 35 zeigt jedoch, dass aufgrund des größeren Schleppfehlers das Gesamtverhalten verschlechtert wird.

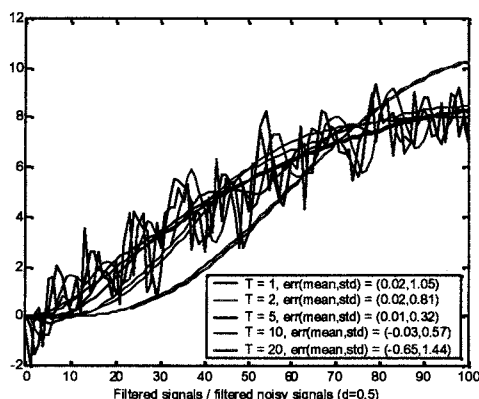


Abb. 35: Filterverhalten von $F(s)^2$ bei Dämpfung $d=0.5$ und variabler Zeitkonstante

6 Zusammenfassung

Es wurde eine Methode zur Filterung nicht äquidistant abgetasteter Signale vorgestellt, die speziell bei der periodischen Temperaturkalibrierung große Vorteile bringt. Das hier dargestellte Filter 1. Ordnung ist einfach und robust und wird in Bondmaschinen eingesetzt. Die Idee der Methode ist jedoch auch für Filter höherer Ordnung auf geradlinige Weise erweiterbar, wenn man anstelle der Exponentialfunktion die Transitionsmatrix ins Spiel bringt.

Weiters wurde eine zweistufige Filterstruktur vorgestellt, die einen bleibenden Schleppfehler als Antwort auf eine Eingangsrampe vermeidet. Diese Struktur wurde anhand zeitkontinuierlicher Filter anhand von Beispielen näher untersucht, eine Erweiterung auf zeitdiskrete Filter ist in geradliniger Weise möglich. Unter anderem kann das Filter auch zeitdiskret mit variabler Abtastperiode entsprechend dem Ansatz von [1] aufgebaut werden.

Fallbeispiele zeigen, dass ein Entwurf auf Basis eines Filters 1. Ordnung robust gegenüber Parameterunsicherheiten ist. Bei einem Spreizungsfaktor der beiden Zeitkonstanten 1..5 kann eine Reduktion der Standardabweichung in der Praxis um den Faktor 3 erreicht werden. Bei Basisfiltern 2. Ordnung werden teilweise bessere Ergebnisse erreicht, das Filter ist jedoch nicht so robust gegenüber Parameterunsicherheiten des Prozesses.

As Nebenergebnis wurde im Rahmen der Falluntersuchungen noch ein Filter 2.Ordnung mit Graddifferenz eins gefunden, das kein TwinFilter ist, aber ebenfalls keinen Schleppfehler aufweist. Dieses Filter scheint ebenfalls sehr robuste Eigenschaften zu haben.

7 Literatur

- [1] Pristauz H., Loinger M.: Filter für nicht äquidistante Abtastsignale. Datacon AT-Bericht 2000/10.
- [2] Pristauz H., Loinger M.: TwinFilter – Ein Ansatz zur Synthese schleppfehlerfreier kausaler Filter. Datacon AT-Bericht 2000/11.

Iso-Lyapunov Dynamics – a Challenge

Alexander Weinmann, ÖVE, Senior Member IEEE
Vienna University of Technology, Institute of Automatic Control
Gusshausstrasse 27-29/375, A-1040 Vienna / Austria
Phone: +43 1 58801 37500, Fax: +43 1 58801 37599
email: weinmann@iert.tuwien.ac.at

Manuscript received June 12, 2001

Keywords: Lyapunov derivative, nonlinear MIMO controller, continuous-time and discrete-time controller

Abstract: *A general nonlinear MIMO controller concept is presented achieving preselected Lyapunov derivative structure. Linear and nonlinear plants are considered as well as continuous and discrete time. Limits for the actuating variable are included.*

1 Introduction

Industrial plants are often characterized by nonlinear equations. Hence, nonlinear control is a wide field of application-oriented system theory (Böcker, J. et al., 1986; Qu, Z., 1998; Slotine, J.J.E., and Li, W., 1991).

In the control literature, Lyapunov-based methods play an important role and several authors contributed to its theory and application. Some of these contributions are outlined in what follows:

Stabilization and tracking with unified Lyapunov framework is given by Jiang, Z.P., 1999. The construction of Lyapunov functions is emphasized in Jankovic, M., et al., 1996; McConley, M.W., 2000. Saturated control laws were developed with Lyapunov stability theory by Robinett, R.D.,

et al., 1997. For time-varying parameter uncertainties, nonquadratic Lyapunov functions are considered; polyhedral functions were introduced as Lyapunov candidates (Blanchini, F., 1995). Receding horizon and Lyapunov function control theory was combined by Sznajder, M., et al., 2000. Using a nonlinear digital controller, stability properties of an equilibrium in Lyapunov sense was derived in Hu, B., and Michel, A.N., 2000.

Remarkable improvement of system performance was achieved by Lyapunov-type state-feedback controller (Yang, 1998). DeCarlo, R.A., and Drakunov, S., 1998 used a unified Lyapunov setting for continuous and discrete-time sliding mode control. An example for guaranteed cost control of uncertain systems using quadratic Lyapunov functions is given in Ugrinovskii, V.A., and Petersen, I.R., 2000. A Lyapunov-based control algorithm is used for set-point control of the acrobot (Zergerogh, et al., 1999).

Radial basis function neural network training was organized with Lyapunov theory by Feng, G., 1996. In neural network control, Lyapunov stability arguments were used for stabilizing the system (Kosmatopoulos, E.B., et al., 1998). Fuzzy logic controllers were synthesized using Lyapunov

punov functions by Tsai, C-Y., and Li, T-H.S., 1996. Neural network-based controllers with feedback linearization were designed for achieving specific tracking performance with Lyapunov functions (Yesildirek, A., and Lewis, F.L., 1994). Similarly, model reference adaptive control was presented in Lee, J.K., et al., 1995.

This paper attempts to render a challenge of rather unusual controller assumptions based on Lyapunov theory. But in the past, even very strange suppositions could produce results beneficial from the viewpoint of practitioners.

The main idea of this paper is to find an actuating variable for a system such that a prespecified Lyapunov derivative is guaranteed.

The motivation for the idea of iso-Lyapunov dynamics is to handle a large-range operation irrespective if the plant under control is linear or nonlinear. Nonlinearities should not be treated separately and not be compensated for.

The control algorithm should be simple from the analytical viewpoint even if it should turn out demanding with respect to computational effort. Since powerful low-cost computers are available, computation is not a problem.

The opportunities for a control engineer are enhanced by the iso-matrical controller in any case.

This paper is organized as follows: Single-input single-output nonlinear systems and multiple-input multiple-output systems are considered with iso-Lyapunov continuous-time dynamics. Minimum norm of $\mathbf{u}$ is preferred. Then, iso-Lyapunov linear MIMO systems are carried out. The Lyapunov function is also augmented by the system deviation or repalced by passivity function as a derivative. Hard constraints on $\mathbf{u}$ are also taken into consideration.

The investigations are repeated as far as

discrete-time systems are concerned. Desired behaviour results for maximum (negative) derivative of the Lyapunov function. Reference signals are also included when iso-Lyapunov dynamics is extended to iso-matrical controllers. Auxiliary linear controllers are used for assisting iso-matrical controllers.

Robust control is taken into consideration in various attempts: maximum permissible uncertainty and worst-case perturbation.

A very comprehensive appendix is provided with eleven sections to offer several additional cases and examples for discussion.

2 Main Idea

The main idea of this paper is to determine an actuating variable $\mathbf{u}(t)$ to a system which guarantees a prespecified Lyapunov derivative $\dot{V}$. At any time t of operation, for a preselected Lyapunov function V the derivative $\dot{V}$ with respect to time should be determined by $-\mathbf{x}^T \mathbf{Q} \mathbf{x}$ with fixed matrix $\mathbf{Q} > 0$. This is termed as the iso-Lyapunov condition $C_{IL} = 0$ (Fig. 1).

In Fig. 1 the state variable $\mathbf{x}(t)$ of the plant excites two block elements $\mathbf{x}^T \mathbf{P} \mathbf{x}$ and $-\mathbf{x}^T \mathbf{Q} \mathbf{x}$. The question is: Which actuating variable $\mathbf{u}(t)$ is required in order to determine $e_L(t) \equiv 0 \forall t$?

The key idea is to start with conditions for guaranteed stability in the sense of Lyapunov and only then to continue by including control system operation. Unexpectedly, there is the eventuality of a denominator inner product zero and the actuating variables growing up to infinity. However, there are four possibilities of avoiding undue consequences and constraining the input signal $\mathbf{u}(t)$:

- Either to minimize $\mathbf{u}^T \mathbf{R} \mathbf{u}$ subject to

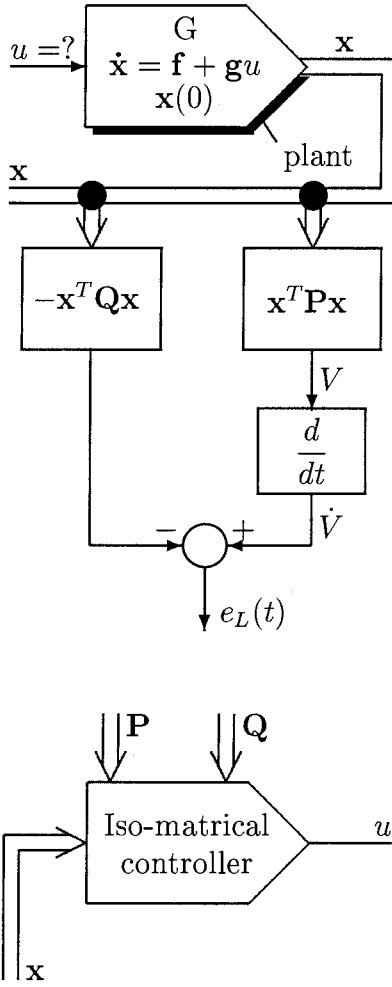


Figure 1: Iso-matrical controller to obtain $e_L(t) \equiv 0 \forall t$

the condition C_{IL} given by the paragraph above,

- or to replace $C_{IL} = 0$ by $C_{IL} \rightarrow \max_u$ in order to accelerate the motion even if it is subject to an upper norm limit of u .
- An alternative possibility is given by an energy rate term for $\dot{V}$.
- Hard limits for $u_i(t)$.

Intentionally, the iso-matrical controller is a nonlinear one, although there are basic well-known facts: (i) LQ problem: The Riccati controller cannot be improved by any nonlinear controller. (ii) If an LTI controller does not robustly stabilize a plant then there does not exist a nonlinear NLTI controller, i.e., increasing complexity does not help.

3 Scalar System

Assume the nonlinear plant

$$\dot{x} = f(x) + g(x)u \quad (1)$$

with the Lyapunov function

$$\begin{aligned} V(x) &= \frac{1}{2}x^T P x, \quad P > 0 \\ \dot{V}(x) &= x^T P (f + gu) < -x^T Q x, \quad Q > 0 \\ x^T P f + x^T P g u &< -x^T Q x. \end{aligned} \quad (2)$$

For the scalar factor $x^T P g > 0$ one has a condition for the actuating variable u

$$u < -\frac{x^T(Qx + Pf)}{x^T P g}; \quad (3)$$

for $x^T P g < 0$

$$u > -\frac{x^T(Qx + Pf)}{x^T P g}. \quad (4)$$

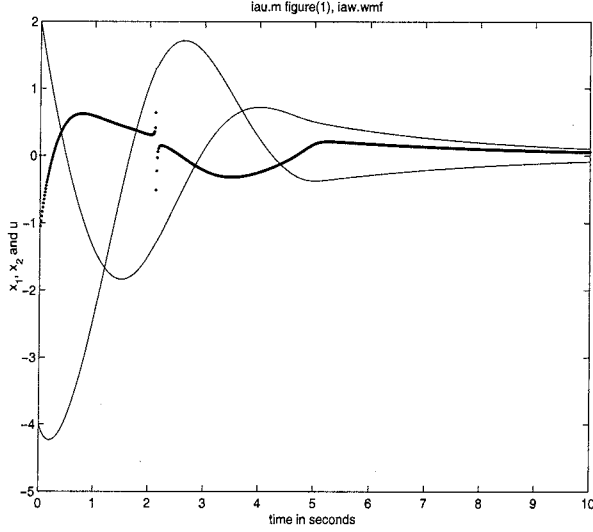


Figure 2: Variables $\mathbf{x}(t)$ and $u(t)$

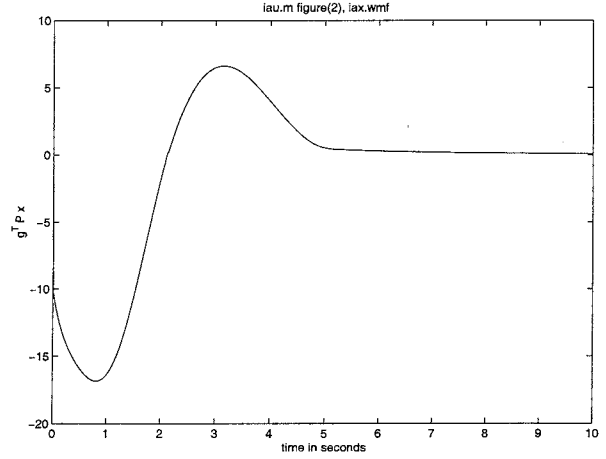


Figure 3: Denominator $\mathbf{x}^T \mathbf{P} \mathbf{g}$ versus time

Replacing the inequality sign by an equality sign, the result for both cases is

$$u = -\frac{\mathbf{x}^T (\mathbf{Q} \mathbf{x} + \mathbf{P} \mathbf{f})}{\mathbf{x}^T \mathbf{P} \mathbf{g}}. \quad (5)$$

Alternatively, replacing $-\mathbf{x}^T \mathbf{Q} \mathbf{x}$ by $-\alpha_0 \mathbf{x}^T \mathbf{x}$, or $-a \mathbf{x}^T \mathbf{P} \mathbf{x}$ ($a > 0$), then

$$u = -\frac{\mathbf{x}^T (\alpha_0 \mathbf{x} + \mathbf{P} \mathbf{f})}{\mathbf{x}^T \mathbf{P} \mathbf{g}} \quad \text{or} \quad (6)$$

$$u = -\frac{\mathbf{x}^T \mathbf{P} (\mathbf{f} + a \mathbf{x})}{\mathbf{x}^T \mathbf{P} \mathbf{g}}; \quad (7)$$

or use some general $\mathbf{Q}(\mathbf{x})$, see Fig. 44.

Example 1. Scalar nonlinear system:
Suppose

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} x_2 \\ -x_1 - 0.25x_2 - 0.2x_1^2 \end{pmatrix} \quad (8)$$

$$\mathbf{g}(\mathbf{x}) = \begin{pmatrix} 0.04x_1^2 + 1 \\ 0.1x_1^2 + 2 \end{pmatrix} \quad (9)$$

and $\mathbf{P} = \text{diag}\{4 \ 2\}$, $\mathbf{Q} = \mathbf{I}_2$. Considering Eq.(5), the results are given in Figs. 2, 3, 4. $\square$

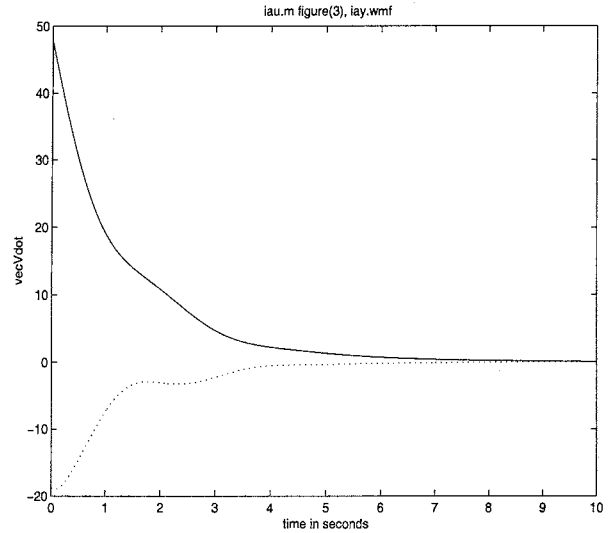


Figure 4: Lyapunov function $V(t)$ and $\dot{V}(t)$

3.1 Assessing the Maximum u

The denominator expression $\mathbf{x}^T \mathbf{P} \mathbf{g}$ can turn out very low and has to be assessed as far as its amplitude is concerned. Presuppose that even for the non-constant and symmetric matrix $\mathbf{W}(\mathbf{x})$ the relations

$$\min_{\mathbf{x}} \sigma_{\min}[\mathbf{W}] \leq \frac{\mathbf{x}^T \mathbf{W} \mathbf{x}}{\|\mathbf{x}\|_F^2} \leq \max_{\mathbf{x}} \sigma_{\max}[\mathbf{W}] \quad (10)$$

are true. Replace $\mathbf{G}_x \mathbf{x} \triangleq \mathbf{g}(x)$, $\mathbf{F}_x \mathbf{x} \triangleq \mathbf{f}(x)$ and assume $\mathbf{P} \mathbf{G}_x$ symmetric. Then for nonsingular matrices $\mathbf{G}_x$ there is the following limit: For Eq.(166) (if $\mathbf{P} \mathbf{G}_x$ symmetric) $|u| < u_0$ where

$$u_0 = \frac{|\mathbf{x}^T (\mathbf{Q} + \mathbf{P} \mathbf{F}_x) \mathbf{x}|}{|\mathbf{x}^T \mathbf{P} \mathbf{G}_x \mathbf{x}|} \leq \frac{\sigma_{\max}[\mathbf{Q} + \mathbf{P} \mathbf{F}_x]}{\sigma_{\min}[\mathbf{P} \mathbf{G}_x]} \quad (11)$$

If $\mathbf{g}(x)$ happens to contain constant elements then $\mathbf{G}_x$ contains expressions $\frac{1}{x_i}$, and elements of $\mathbf{G}_x$ in the denominator might tend to infinity.

4 Multivariable System Minimization of the Norm of the Actuator Variable

Assume the multivariable plant

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{G}(\mathbf{x}) \mathbf{u} \quad \mathbf{x} \in \mathcal{R}^n, \mathbf{u} \in \mathcal{R}^m \quad (12)$$

Then, $V(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{P} \mathbf{x}$ and $\mathbf{P} = \mathbf{P}^T > 0$ demands

$$\dot{V}(\mathbf{x}) = \mathbf{x}^T \mathbf{P} \dot{\mathbf{x}} = \mathbf{x}^T \mathbf{P} \mathbf{f} + \mathbf{x}^T \mathbf{P} \mathbf{G} \mathbf{u} \triangleq -\mathbf{x}^T \mathbf{Q} \mathbf{x} \quad (13)$$

for $\mathbf{Q} > 0$. In comparison to the case of scalar input u , using the abbreviation $\mathbf{b}_1^T \triangleq \mathbf{x}^T \mathbf{P} \mathbf{G}$, Eq.(13) is a scalar condition on $\mathbf{u}$ in the multivariable case, i.e.,

$$\mathbf{b}_1^T \mathbf{u} = -q \quad \text{where} \quad q = \mathbf{x}^T \mathbf{P} \mathbf{f} + \mathbf{x}^T \mathbf{Q} \mathbf{x} \quad (14)$$

One can impose an additive assumption in order to solve Eq.(13) with respect to $\mathbf{u}$. Assuming $\frac{1}{2} \mathbf{u}^T \mathbf{R} \mathbf{u} \rightarrow \min_{\mathbf{u}}$,

$$\frac{1}{2} \mathbf{u}^T \mathbf{R} \mathbf{u} + \mu (\mathbf{b}_1^T \mathbf{u} + q) \rightarrow \min_{\mathbf{u}} \quad (15)$$

is achieved. Differentiating with respect to $\mathbf{u}$ yields

$$\mathbf{R} \mathbf{u} + \mu \mathbf{b}_1 = \mathbf{0} \rightsquigarrow \mathbf{u} = -\mu \mathbf{R}^{-1} \mathbf{b}_1 \quad (16)$$

Combining with Eq.(14), the result is

$$\mathbf{u} = -\frac{q}{\mathbf{b}_1^T \mathbf{R}^{-1} \mathbf{b}_1} \mathbf{R}^{-1} \mathbf{b}_1 \quad (17)$$

or

$$\mathbf{u} = -\frac{\mathbf{x}^T (\mathbf{P} \mathbf{f} + \mathbf{Q} \mathbf{x})}{\mathbf{x}^T \mathbf{P} \mathbf{G} \mathbf{R}^{-1} \mathbf{G}^T \mathbf{P} \mathbf{x}} \mathbf{R}^{-1} \mathbf{G}^T \mathbf{P} \mathbf{x} \quad (18)$$

Stepping back the result of Eq.(18) to the scalar case yields the same result as Eq.(166) because in Eq.(18) the expression $\mathbf{x}^T \mathbf{P} \mathbf{G}$ for $\mathbf{G} = g$ is a scalar and can be canceled. Thus, Eq.(18) turns out as Eq.(166).

For comparison with Hamilton Jacobi Bellman equation, see Appendix C.

In HJBe the integral of sum squared errors is minimized versus $\mathbf{u}$; the isomatrixial controller determines the minimum of $\mathbf{u}^T \mathbf{R} \mathbf{u}$ and predetermined $\dot{V}$ at each instant t .

4.1 Assessing the Norm of the Actuating Variable

The maximum norm of $\mathbf{u}$ can be assessed in Eq.(18) by the gain factor g

$$\begin{aligned} g &\triangleq \frac{\|\mathbf{u}\|_F}{\|\mathbf{x}\|_F} \leq \max_{\mathbf{x}} \{ \sigma_{\max}[\mathbf{P} \mathbf{F}_x + \mathbf{Q}] \\ &\times \|\mathbf{R}^{-1} \mathbf{G}^T \mathbf{P}\|_F \} \\ &\times \left[\min_{\mathbf{x}} \{ \sigma_{\min}[\mathbf{P} \mathbf{G} \mathbf{R}^{-1} \mathbf{G}^T \mathbf{P}] \} \right]^{-1} \end{aligned} \quad (19)$$

In the special cases above, $\sigma_{\min}$ could have been replaced by $\lambda_{\min}$. For examples see Appendix E.

5 Linear SISO System

For the assumptions $\mathbf{f}(\mathbf{x}) \triangleq \mathbf{A}\mathbf{x}$ and $\mathbf{g}(\mathbf{x}) \triangleq \mathbf{b}$ and for $\mathbf{Q}$ given, then $\mathbf{P}$ results from

$$\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} = -\mathbf{Q} . \quad (20) \quad \text{Based on Eq.(18)}$$

Only stable matrices $\mathbf{A}$ are admitted, no integrator in $\mathbf{A}$ can be included.

The actuating variable is

$$u = -\frac{\mathbf{x}^T(\mathbf{Q} + \mathbf{P}\mathbf{A})\mathbf{x}}{\mathbf{x}^T\mathbf{P}\mathbf{b}} = \frac{\mathbf{x}^T\mathbf{A}^T\mathbf{P}\mathbf{x}}{\mathbf{x}^T\mathbf{P}\mathbf{b}} . \quad (21)$$

Remark: Select $\mathbf{Q}$ such that $\mathbf{A}^T\mathbf{P} = \mathbf{Q} + \mathbf{P}\mathbf{A} = -\mathbf{p}\mathbf{q}^T$ and $\mathbf{P}\mathbf{b} = \mathbf{p}$ then $u = \mathbf{q}^T\mathbf{x}$.

Proof:

$$\mathbf{P} = -\mathbf{A}^{-T}\mathbf{p}\mathbf{q}^T \quad (22)$$

$$\mathbf{p} = \mathbf{P}\mathbf{b} = -\mathbf{A}^{-T}\mathbf{p}\mathbf{q}^T\mathbf{b} = -(\mathbf{q}^T\mathbf{b})\mathbf{A}^{-T}\mathbf{p} \quad (23)$$

$$\mathbf{A}^T = -\mathbf{q}^T\mathbf{b}\mathbf{I} = \mathbf{A} . \quad (24)$$

Only if $\mathbf{A} = -\mathbf{q}^T\mathbf{b}\mathbf{I}$ (hardly realizable) then $u = \mathbf{q}^T\mathbf{x}$ is a linear controller and no infinity positions arise in

$$u = \frac{\mathbf{x}^T\mathbf{p}\mathbf{q}^T\mathbf{x}}{\mathbf{x}^T\mathbf{p}} . \quad (25)$$

5.1 Prefilter

With the intention $y = \mathbf{c}^T\mathbf{x} \rightarrow y_1$ and based on Eq.(5), the set-up

$$u = -\frac{\mathbf{x}^T(\mathbf{Q} + \mathbf{P}\mathbf{A})\mathbf{x}}{\mathbf{x}^T\mathbf{P}\mathbf{b}} + vy_1 \quad (26)$$

yields the steady-state

$$\dot{\mathbf{x}}_f = \mathbf{A}\mathbf{x}_f + \mathbf{b}\frac{\mathbf{x}_f^T\mathbf{A}^T\mathbf{P}\mathbf{x}_f}{\mathbf{x}_f^T\mathbf{P}\mathbf{b}} + \mathbf{b}vy_1 = \mathbf{0} . \quad (27)$$

The steady-state condition

$$\mathbf{c}^T\mathbf{x}_f - y_1 = 0 , \quad (28)$$

corresponds to two equations in v and $\mathbf{x}_f$.

6 Linear MIMO System. Minimization of the Norm of $\mathbf{u}$

$$\mathbf{f}(\mathbf{x}) \triangleq \mathbf{A}\mathbf{x} , \quad \mathbf{G}(\mathbf{x}) \triangleq \mathbf{B} , \quad \mathbf{u}^T\mathbf{R}\mathbf{u} \rightarrow \min_{\mathbf{u}} , \quad (29)$$

$\mathbf{P}$ results from $\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} = -\mathbf{Q}$. Then,

$$\mathbf{u} = \frac{\mathbf{x}^T\mathbf{A}^T\mathbf{P}\mathbf{x}}{\mathbf{x}^T\mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{P}\mathbf{x}}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{P}\mathbf{x} \quad (30)$$

$$g = \frac{\|\mathbf{u}\|_F}{\|\mathbf{x}\|_F} \leq \frac{\sigma_{\max}[\mathbf{A}^T\mathbf{P}]\sigma_{\max}[\mathbf{R}^{-1}\mathbf{B}^T\mathbf{P}]}{\sigma_{\min}[\mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{P}]} . \quad (31)$$

Select $\mathbf{R}$ for limiting the actuator variable and designing the closed-loop control.

Remark: For $\mathbf{R} = \alpha\mathbf{I}_2$ the result g is independent of α .

7 Lyapunov Function Augmented by System Deviation

Let the reference be termed y_{ref} and the output of the system be

$$y(t) = \mathbf{c}^T\mathbf{x}(t) + \int_0^t \mathbf{c}_1^T\mathbf{x}(t)dt . \quad (32)$$

(In *Bang, H., 1994*, even control inputs were taken into account in the Lyapunov function.) Now the Lyapunov function V is set up as

$$V = 0.5\mathbf{x}^T\mathbf{P}\mathbf{x} + 0.5(\mathbf{c}^T\mathbf{x} + \int_0^t \mathbf{c}_1^T\mathbf{x}dt - y_{ref})^2 . \quad (33)$$

The first differential quotient with respect to t is

$$\dot{V} = \mathbf{x}^T\mathbf{P}(\mathbf{f} + \mathbf{G}\mathbf{u}) + h_1(\mathbf{c}^T\dot{\mathbf{x}} + \mathbf{c}_1^T\mathbf{x}) = -\mathbf{x}^T\mathbf{Q}\mathbf{x} \quad (34)$$

where

$$h_1 \triangleq \mathbf{c}^T \mathbf{x} + \int_0^t \mathbf{c}_1^T \mathbf{x} dt - y_{ref} = -e. \quad (35)$$

Evaluating yields

$$-\mathbf{b}_1^T \mathbf{u} = q \quad (36)$$

where two abbreviations are used, i.e.,

$$\mathbf{b}_1 \triangleq \mathbf{G}^T(h_1 \mathbf{c} + \mathbf{P} \mathbf{x}) \quad (37)$$

and

$$q \triangleq (\mathbf{x}^T \mathbf{P} + h_1 \mathbf{c}^T) \mathbf{f} + (h_1 \mathbf{c}_1^T + \mathbf{x}^T \mathbf{Q}) \mathbf{x}. \quad (38)$$

Finally,

$$\mathbf{u} = -\frac{q}{\mathbf{b}_1^T \mathbf{R}^{-1} \mathbf{b}_1} \mathbf{R}^{-1} \mathbf{b}_1. \quad (39)$$

Including the system error $e(t) = -h_1(t)$, there results

$$\mathbf{u} = -\frac{q}{\mathbf{b}_1^T \mathbf{R}^{-1} \mathbf{b}_1} \mathbf{R}^{-1} \mathbf{b}_1 + v \begin{pmatrix} 0 \\ e \end{pmatrix} \quad (40)$$

where v is a constant gain.

8 Norm Constraint on $\mathbf{u}$

Now, the equation $\mathbf{b}_1^T \mathbf{u} + q = 0$ is replaced by the command $\mathbf{b}_1^T \mathbf{u} + q \rightarrow \max_{\mathbf{u}}$ subject to the condition of limited norm of $\mathbf{u}$, i.e.,

$$\mathbf{b}_1^T \mathbf{u} + q + \mu \frac{1}{2} \mathbf{u}^T \mathbf{R} \mathbf{u} \rightarrow \max_{\mathbf{u}} \quad (41)$$

$$\mathbf{b}_1^T + \mu \mathbf{R} \mathbf{u} = 0 \rightsquigarrow \mathbf{u} = -\mathbf{R}^{-1} \mathbf{b}_1 \frac{1}{\mu} \quad (42)$$

where $\mathbf{b}_1^T = \mathbf{x}^T \mathbf{P} \mathbf{G}$.

Limiting yields

$$0.5 \mathbf{u}^T \mathbf{R} \mathbf{u} = 0.5 U_0^2 \rightsquigarrow \mu = \frac{1}{U_0} \sqrt{\mathbf{b}_1^T \mathbf{R}^{-1} \mathbf{b}_1}. \quad (43)$$

The actuating variable

$$\mathbf{u} = -\mathbf{R}^{-1} \mathbf{b}_1 \frac{U_0}{\sqrt{\mathbf{b}_1^T \mathbf{R}^{-1} \mathbf{b}_1}} \quad (44)$$

operates as an on-off control and runs into a sliding mode.

For a SISO system we have $b_1 = \mathbf{x}^T \mathbf{P} \mathbf{b}$, $\mathbf{R} = r$.

9 Lyapunov Function Proportional to the Squared Deviation

Suppose a very different

$$V \triangleq 0.5 p (y_{ref} - \mathbf{c}^T \mathbf{x})^2, \quad (45)$$

then from Eq.(45)

$$\dot{V} = -p (y_{ref} - \mathbf{c}^T \mathbf{x}) \mathbf{c}^T \dot{\mathbf{x}} \triangleq -(\mathbf{x} - \mathbf{x}_f)^T \mathbf{Q} (\mathbf{x} - \mathbf{x}_f) \quad (46)$$

and

$$u = -\frac{\mathbf{f}^T \mathbf{c} - (\mathbf{x} - \mathbf{x}_f)^T \mathbf{Q} (\mathbf{x} - \mathbf{x}_f) / [p (y_{ref} - \mathbf{c}^T \mathbf{x})]}{\mathbf{g}^T \mathbf{c}}. \quad (47)$$

10 Lyapunov Derivative Proportional to the Deviation

Assume

$$V = 0.5 \mathbf{x}^T \mathbf{P} \mathbf{x} \quad (48)$$

and

$$\dot{V} = - (y_{ref} - \mathbf{c}^T \mathbf{x})^2 q \quad \text{or} \quad (49)$$

$$- (\mathbf{c}^T \mathbf{x}_f - \mathbf{c}^T \mathbf{x}) q. \quad (50)$$

Then, for the first alternative above

$$u = -\frac{(y_{ref} - \mathbf{c}^T \mathbf{x})^2 q + \mathbf{x}^T \mathbf{P} \mathbf{f}}{\mathbf{x}^T \mathbf{P} \mathbf{g}}. \quad (51)$$

Depending on $\mathbf{c}$, this might be an assumption corresponding to a semidefinite $\mathbf{Q}$.

There are three conditions for the steady-state (subscript f)

$$\dot{\mathbf{x}}_f = \mathbf{f}(\mathbf{x}_f) - \mathbf{g}(\mathbf{x}_f) \frac{\mathbf{x}_f^T \mathbf{P} \mathbf{f}(\mathbf{x}_f)}{\mathbf{x}_f^T \mathbf{P} \mathbf{g}(\mathbf{x}_f)} = 0 \quad (52)$$

$$\mathbf{c}^T \mathbf{x}_f = y_{ref} \quad (53)$$

$$\mathbf{P} > 0. \quad (54)$$

For a given system $\mathbf{f}$, $\mathbf{g}$, $\mathbf{c}$ there are $n + 1$ equations for the unknowns $\mathbf{x}_f$ and $\mathbf{P}$. Pre-assuming that the equations are solvable this leads to a set-point-dependent $\mathbf{P}$. The equations are solved to obtain $\mathbf{P}$ for a given setpoint. Depending on the nonlinearities in $\mathbf{f}$, $\mathbf{g}$ and y_{ref} , the solution need not be unique. Assumptions for $\mathbf{P} = \mathbf{P}(\mathbf{x})$ and $q = q(\mathbf{x})$ are possible.

11 Transient Disturbance

Suppose the plant and a controller from Eq.(5)

$$\dot{\mathbf{x}} = \mathbf{f} + \mathbf{g}u + \mathbf{l}w_d \quad (55)$$

$$\dot{\mathbf{x}} = \mathbf{f} - \mathbf{g} \frac{\mathbf{x}^T(\mathbf{Q}\mathbf{x} + \mathbf{P}\mathbf{f})}{\mathbf{x}^T\mathbf{P}\mathbf{g}} + \mathbf{l}w_d. \quad (56)$$

For $\dot{V} = -\mathbf{x}^T\mathbf{Q}\mathbf{x} \rightarrow 0$, $\mathbf{x}$ must vanish. This leads to the requirement of a transient disturbance with steady-state zero; unless one selects a superposition of classical and iso-matricial controller.

12 Iso-matricial Controller

Suppose the Lyapunov function V and $\dot{V}$ oriented at a final state variable $\mathbf{x}_f$ of a linear SISO system. In spite of expecting a sliding mode for some time interval, suppose $u_f = 0$. (Equilibria-dependent Lyapunov functions were also used in *Leonessa, A., et al., 1998.*) The plant obeys

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u + \mathbf{b}vy_{ref}. \quad (57)$$

From steady-state $\dot{\mathbf{x}} = 0$ the result $\mathbf{x}_f = -\mathbf{A}^{-1}\mathbf{b}vy_{ref}$ is obtained. Then,

$$V = 0.5(\mathbf{x} - \mathbf{x}_f)^T\mathbf{P}(\mathbf{x} - \mathbf{x}_f) \quad (58)$$

$$\dot{V} \triangleq -(\mathbf{x} - \mathbf{x}_f)^T q(\mathbf{x} - \mathbf{x}_f) \quad (59)$$

$$(\mathbf{x} - \mathbf{x}_f)^T\mathbf{P}\dot{\mathbf{x}} = -(\mathbf{x} - \mathbf{x}_f)^T q(\mathbf{x} - \mathbf{x}_f) \quad (60)$$

$$(\mathbf{x} - \mathbf{x}_f)^T\mathbf{P}(\mathbf{A}\mathbf{x} + \mathbf{b}u + \mathbf{b}vy_{ref}) = -(\mathbf{x} - \mathbf{x}_f)^T q(\mathbf{x} - \mathbf{x}_f) \quad (61)$$

$$\mathbf{u} = -\frac{(\mathbf{x} - \mathbf{x}_f)^T[q(\mathbf{x} - \mathbf{x}_f) + \mathbf{P}(\mathbf{A}\mathbf{x} + \mathbf{b}vy_{ref})]}{(\mathbf{x} - \mathbf{x}_f)^T\mathbf{P}\mathbf{b}}. \quad (62)$$

13 Discrete-time Systems

Sampled-data systems are characterized by the transition matrix $\Phi(t) = e^{\mathbf{A}t}$ for fixed $t = T$ sampling period. Assuming zero-order hold, the input matrix is

$$\Psi(T) = \mathbf{A}^{-1}[\mathbf{e}^{\mathbf{A}T} - \mathbf{I}]\mathbf{B} \quad (63)$$

and the system equations are

$$\mathbf{x}(k+1) = \Phi(T)\mathbf{x}(k) + \Psi(T)\mathbf{u}(k) \quad (64)$$

$$\mathbf{y}(k+1) = \mathbf{C}\mathbf{x}(k+1). \quad (65)$$

Consider a Lyapunov function V (*Vidal, P., 1969*)

$$V = 0.5\mathbf{x}^T(k+1)\mathbf{P}\mathbf{x}(k+1) \text{ and } \quad (66)$$

$$\Delta V = 0.5\mathbf{x}^T(k+1)\mathbf{P}\mathbf{x}(k+1) - 0.5\mathbf{x}^T(k)\mathbf{P}\mathbf{x}(k) \quad (67)$$

$$\Delta V = 0.5\mathbf{x}^T(\Phi^T\mathbf{P}\Phi - \mathbf{P})\mathbf{x} + \mathbf{u}^T\Psi^T\mathbf{P}\Phi\mathbf{x} + 0.5\mathbf{u}^T\Psi^T\mathbf{P}\Psi\mathbf{u}. \quad (68)$$

Demanding $\Delta V = -0.5\mathbf{x}^T(k)\mathbf{Q}\mathbf{x}(k)$ and abbreviating

$$h_1(\mathbf{x}) \triangleq 0.5\mathbf{x}^T(\Phi^T\mathbf{P}\Phi - \mathbf{P} + \mathbf{Q})\mathbf{x} \quad (69)$$

$$\mathbf{h}_2(\mathbf{x}) \triangleq \Psi^T\mathbf{P}\Phi\mathbf{x} \quad (70)$$

$$\mathbf{H}_1 \triangleq \Psi^T\mathbf{P}\Psi, \quad (70)$$

there results

$$h_1 + \mathbf{u}^T\mathbf{h}_2 + 0.5\mathbf{u}^T\mathbf{H}_1\mathbf{u} = 0. \quad (71)$$

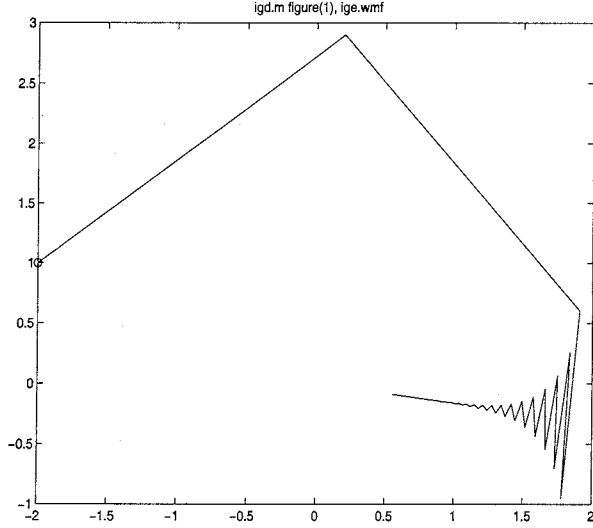


Figure 5: Phase plot $\mathbf{x}(kT)$ for plus sign

Optimization requires

$$0.5\mathbf{u}^T \mathbf{R} \mathbf{u} + \mu(h_1 + \mathbf{u}^T \mathbf{h}_2 + 0.5\mathbf{u}^T \mathbf{H}_1 \mathbf{u}) \rightarrow \min_{\mathbf{u}} \quad (72)$$

$$\mathbf{u} = -\left(\frac{\mathbf{R}}{\mu} + \mathbf{H}_1\right)^{-1} \mathbf{h}_2. \quad (73)$$

There is no danger of a zero denominator.

For comparison: Demanding $\max_{\mathbf{u}} \Delta V$ leads to the well-known result

$$\mathbf{u} = -(\Psi^T \mathbf{P} \Psi)^{-1} \Psi^T \mathbf{P} \Phi \mathbf{x} = -\mathbf{H}_1^{-1} \mathbf{h}_2 \quad (74)$$

which is obtained for $\mathbf{R} = \mathbf{0}$.

Supposing $\mathbf{u} = \mathbf{K} \mathbf{x}$ in Eq.(68) would lead to a Riccati equation in $\mathbf{K}$ with nonsquare matrices which is not solvable.

Example 2. SISO case of Eq.(71):

In the scalar case, from Eq.(71)

$$u = (-h_2 \pm \sqrt{h_2^2 - 2h_1 H_1})/H_1 \quad (75)$$

is achieved.

In the following figures the results of Eq.(75) are depicted for both signs of the square root. In Figs. 5 and 6 the phase plots, in Figs. 7 and 8 $\mathbf{x}(kT)$ and $u(kT)$; in Fig. 9 $V(kT)$ for both cases yielding an equivalent characteristic. $\square$

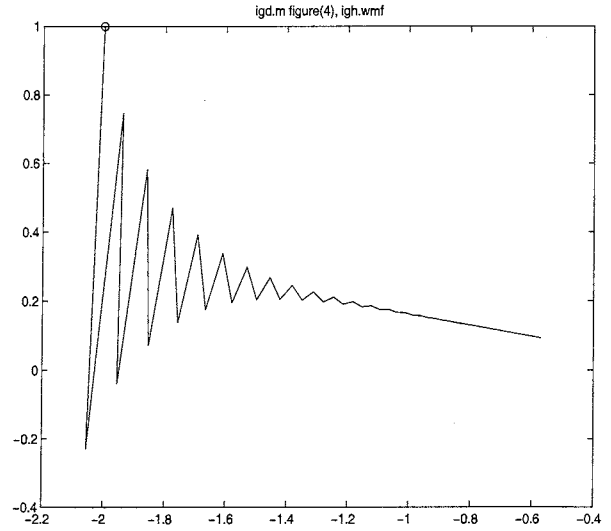


Figure 6: Phase plot $\mathbf{x}(kT)$ for minus sign

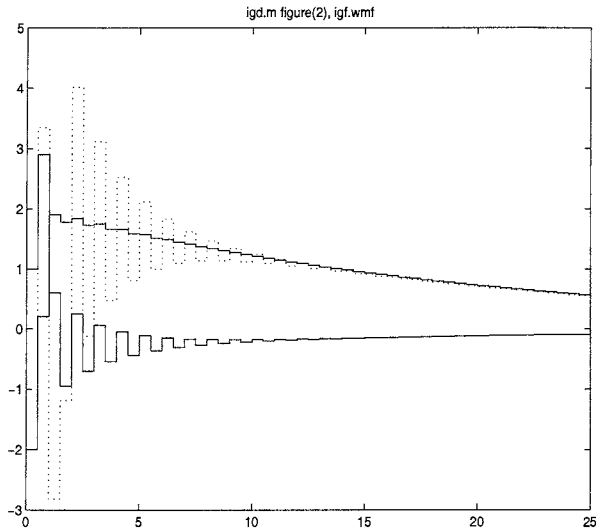


Figure 7: $\mathbf{x}(kT)$ and $u(kT)$ for plus sign

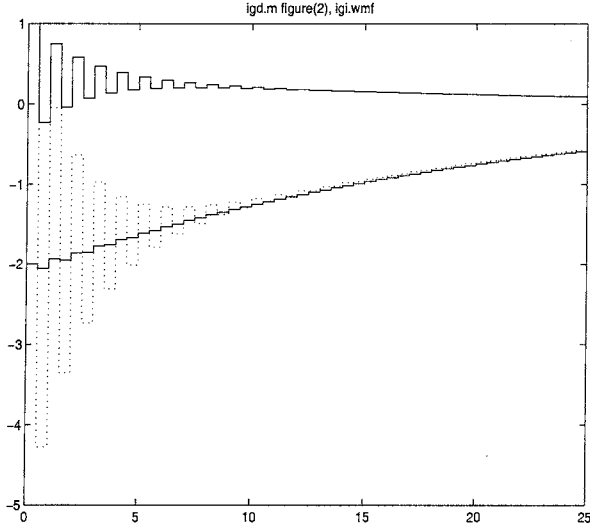


Figure 8: $\mathbf{x}(kT)$ and $u(kT)$ for minus sign

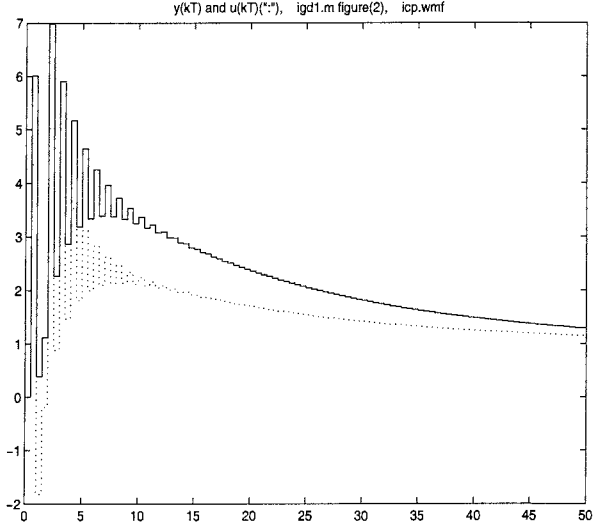


Figure 10: Sampled-data input u and output y of the controlled system

13.1 Discrete-time System with Reference

In the derivation with Eq.(68) replace $-\mathbf{x}^T \mathbf{Q} \mathbf{x}$ by $(\mathbf{x} - \mathbf{x}_f)^T \mathbf{Q} (\mathbf{x} - \mathbf{x}_f)$. Assuming a SISO system

$$\mathbf{x}(k+1) = \Phi(T)\mathbf{x}(k) + \Psi(T)u(k) + \Psi_1 y_{ref} \quad (76)$$

$$\mathbf{y}(k+1) = \mathbf{c}^T \mathbf{x}(k+1) + d \cdot u, \quad (77)$$

the results are

$$h_1(\mathbf{x}) \triangleq 0.5\mathbf{x}^T(\Phi^T \mathbf{P} \Phi - \mathbf{P} + \mathbf{Q})\mathbf{x} + 0.5y_{ref}^2 \Psi_1^T \mathbf{P} \Psi_1 + y_{ref} \Psi_1^T \mathbf{P} \Phi \mathbf{x} \quad (78)$$

$$\mathbf{h}_2(\mathbf{x}) \triangleq \Psi^T \mathbf{P} \Phi \mathbf{x} + \Psi_1^T \mathbf{P} \Psi_1 \quad (79)$$

$$\mathbf{H}_1 \triangleq \Psi^T \mathbf{P} \Psi. \quad (80)$$

The steady-state is

$$\mathbf{0} = (\Phi - \mathbf{I})\mathbf{x}_f + \Psi u_f + \Psi_1 y_{ref}. \quad (81)$$

Demanding $\Psi_1 = \Psi$ and $\mathbf{x}_f = \mathbf{0}$, then the results are $d = 1$ and $u_f = y_{ref}$, see Fig. 11.

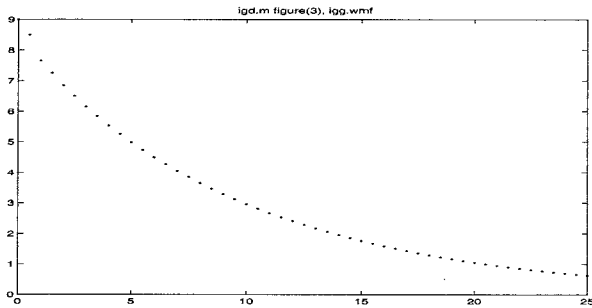


Figure 9: Lyapunov function $V(kT)$ for both signs

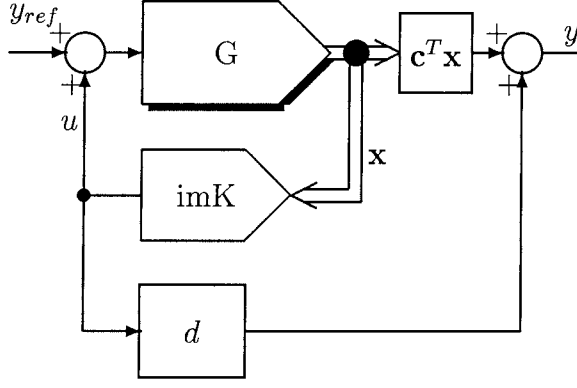


Figure 11: Iso-matrical controller with reference and $\Psi_1 = \Psi$, $d = 1$, $\mathbf{x}_f = 0$

13.2 Maximizing $-\Delta V$ under Norm Constraint of $\mathbf{u}$

Replace the Eq.(71) by

$$-(h_1 + \mathbf{u}^T \mathbf{h}_2 + 0.5 \mathbf{u}^T \mathbf{H}_1 \mathbf{u}) + \mu \frac{\mathbf{u}^T \mathbf{R} \mathbf{u}}{2} \rightarrow \max_{\mathbf{u}}. \quad (82)$$

Since ΔV has to be negative and a minimum, its negative expression is positive and should be maximized. Then

$$\mathbf{u} = -(-\mathbf{H}_1 + \mu \mathbf{R})^{-1} \mathbf{h}_2 \quad (83)$$

and from $\mathbf{u}^T \mathbf{R} \mathbf{u} = U_0^2$ the factor μ and the result $\mathbf{u}$ are obtained.

14 Robust Controller

Consider the plant characterized by the additive-type fuzzy uncertainty $\Delta \mathbf{f}(\mathbf{x}) = \gamma \mathbf{f}_1(\mathbf{x})$ with given and fixed $\mathbf{f}_1(\mathbf{x})$ and characterized by a membership function $m(\gamma)$

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}_0(\mathbf{x}) + \gamma \mathbf{f}_1(\mathbf{x}) \quad -1 \leq \gamma \leq 1. \quad (84)$$

Eq.(2) turns out as

$$\mathbf{x}^T \mathbf{P} [\mathbf{f}_0 + \gamma \mathbf{f}_1(\mathbf{x})] + \mathbf{x}^T \mathbf{P} \mathbf{g} u < -\mathbf{x}^T \mathbf{Q} \mathbf{x}, \quad (85)$$

yielding u versus γ .

The interim variable $u(\gamma)$ is converted to the robust actuating variable u_R

$$u_R = \int_{\gamma_1}^{\gamma_2} u(\gamma) m(\gamma) \frac{\partial u}{\partial \gamma} d\gamma \cdot \left[\int_{\gamma_1}^{\gamma_2} m(\gamma) \frac{\partial u}{\partial \gamma} d\gamma \right]^{-1} \quad (86)$$

where

$$u = -\frac{1}{\mathbf{x}^T \mathbf{P} \mathbf{g}} \mathbf{x}^T (\mathbf{Q} \mathbf{x} + \mathbf{P} \mathbf{f}_0 + \gamma \mathbf{P} \mathbf{f}_1) \quad (87)$$

$$\frac{\partial u}{\partial \gamma} = -\frac{1}{\mathbf{x}^T \mathbf{P} \mathbf{g}} \mathbf{x}^T \mathbf{P} \mathbf{f}_1. \quad (88)$$

Since $\frac{\partial u}{\partial \gamma}$ does not depend on γ , for $m = \text{const}$ one has the simple result

$$u_R = u(\gamma_R) \quad \text{where} \quad \gamma_R = 0.5(\gamma_1 + \gamma_2). \quad (89)$$

Based on the results of Eq.(19), right-hand side, and replacing $\mathbf{f}_0(\mathbf{x}) \triangleq \mathbf{F}_{0x} \mathbf{x}$, $\mathbf{f}_1(\mathbf{x}) \triangleq \mathbf{F}_{1x} \mathbf{x}$, the maximum gain factor $g_{\max}$ is assessed.

Depending on $\mathbf{x}$, define

$$h_{1var} \triangleq \sigma_{\max}[\mathbf{P} \mathbf{F}_{0x} + \mathbf{Q}] \cdot \|\mathbf{R}^{-1} \mathbf{G}^T \mathbf{P}\|_F \quad (90)$$

$$h_{2var} \triangleq \sigma_{\max}[\mathbf{P} \mathbf{F}_{1x}] \cdot \|\mathbf{R}^{-1} \mathbf{G}^T \mathbf{P}\|_F \quad (91)$$

$$h_{3var} \triangleq \sigma_{\min}[\mathbf{P} \mathbf{G} \mathbf{R}^{-1} \mathbf{G}^T \mathbf{P}], \quad (92)$$

then these variables determine

$$\begin{aligned} g_{\max \gamma} &= \max_{\gamma} \left[\max_{\mathbf{x}} \{ \sigma_{\max}[\mathbf{P} \mathbf{F}_{0x} + \mathbf{Q} + \gamma \mathbf{P} \mathbf{F}_{1x}] \} \right. \\ &\quad \times \|\mathbf{R}^{-1} \mathbf{G}^T \mathbf{P}\|_F \\ &\quad \times \left. [\min_{\mathbf{x}} \{ \sigma_{\min}[\mathbf{P} \mathbf{G} \mathbf{R}^{-1} \mathbf{G}^T \mathbf{P}] \}]^{-1} \right]. \end{aligned} \quad (93)$$

This result is bounded by

$$g_{\max} = \frac{\max_{\mathbf{x}} h_{1var} + \max_{\gamma} |\gamma| \cdot \max_{\mathbf{x}} h_{2var}}{\min_{\mathbf{x}} h_{3var}}. \quad (94)$$

For examples, see Appendix J.

14.1 Worst Perturbation in $\mathbf{f}$

Consider

$$\dot{V} < -\mathbf{x}^T \mathbf{Q} \mathbf{x} \quad (95)$$

$$\mathbf{x}^T \mathbf{P} \mathbf{f} + \mathbf{x}^T \mathbf{P} \mathbf{g} u + \mathbf{x}^T \mathbf{Q} \mathbf{x} < 0. \quad (96)$$

The left-hand side above should hold even for perturbed $\mathbf{f} = \mathbf{f}_0 + \mathbf{f}_d$ where $\mathbf{f}_d$ is an arbitrary perturbation $\|\mathbf{f}_d\|_F \leq F_d$. The norm of $\mathbf{f}_d$ is assumed to be a certain percentage ε of the norm of $\mathbf{f}_0$

$$F_d = \varepsilon \|\mathbf{f}_0(\mathbf{x})\|_F. \quad (97)$$

This F_d determines a hypersphere; the worst perturbation is given by $\mathbf{f}_d$ in the direction $\mathbf{P}\mathbf{x}$, then $\mathbf{x}^T \mathbf{P} \mathbf{f}_d$ is maximum in absolute value sense. For $\mathbf{f}_d$ in the direction of $\mathbf{P}\mathbf{x}$ with norm F_d , one has

$$\mathbf{f}_d = \mathbf{P}\mathbf{x} \frac{\|\mathbf{f}_d\|_F}{\|\mathbf{P}\mathbf{x}\|_F} = \mathbf{P}\mathbf{x} \frac{\varepsilon \|\mathbf{f}_0(\mathbf{x})\|_F}{\|\mathbf{P}\mathbf{x}\|_F}. \quad (98)$$

To guarantee adequate u for both signs, then for $\mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{x}^T \mathbf{P} \mathbf{f}_0 > 0$

$$\mathbf{x}^T \mathbf{P} \mathbf{f} = \mathbf{x}^T \mathbf{P} \mathbf{f}_0 - \mathbf{x}^T \mathbf{P}^2 \mathbf{x} \frac{\varepsilon \|\mathbf{f}_0(\mathbf{x})\|_F}{\|\mathbf{P}\mathbf{x}\|_F}; \quad (99)$$

otherwise the minus sign in Eq.(99) is changed to plus.

In order to weaken, or to be opposite to, $\mathbf{x}^T \mathbf{P} \mathbf{f}_0 + \mathbf{x}^T \mathbf{Q} \mathbf{x}$

$$u = -[\mathbf{x}^T (\mathbf{Q} \mathbf{x} + \mathbf{P} \mathbf{f}_0) - \text{sign}(\mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{x}^T \mathbf{P} \mathbf{f}_0) \quad (100)$$

$$\times \mathbf{x}^T \mathbf{P}^2 \mathbf{x} \varepsilon \|\mathbf{f}_0(\mathbf{x})\|_F / \|\mathbf{P}\mathbf{x}\|_F] / (\mathbf{x}^T \mathbf{P} \mathbf{g}). \quad (101)$$

Example 3. Uncertain second-order plant:

A predetermined function $\mathbf{f}_d$ is distributed referring to a constant γ ($0 < \gamma < 1$)

$$\mathbf{f}_{dg} = \begin{pmatrix} \gamma f_{d1} \\ (1 - \gamma) f_{d2} \end{pmatrix} \quad (102)$$

$$\mathbf{f}_\Delta = \varepsilon \frac{\|\mathbf{f}_0\|_F}{\|\mathbf{f}_{dg}\|_F} \mathbf{f}_{dg}. \quad (103)$$

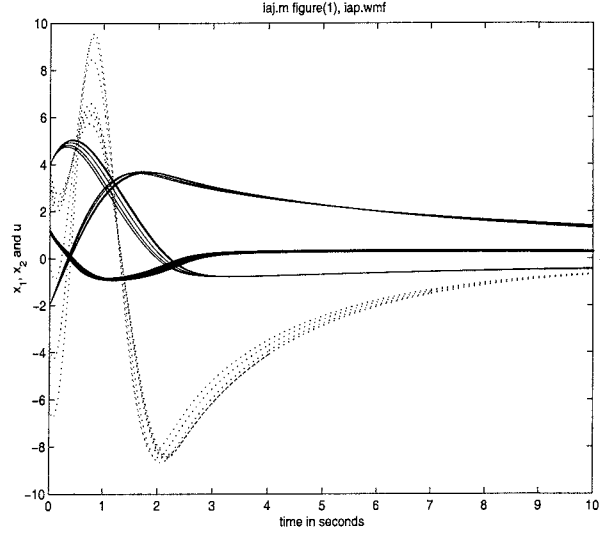


Figure 12: Robust controller and uncertain plant (in ten versions)

Thus the actual plant perturbation $\mathbf{f}_\Delta$ is always a certain percentage 100ε of the nominal $\mathbf{f}_0$ but distributed into the components of $\mathbf{f}_\Delta$ following a random parameter γ .

In Fig. 12 the state variables and the plant input are depicted for the random numbers γ representing the uncertain plant. The Lyapunov functions $V(t)$ and $\dot{V}(t)$ are depicted in Fig. 13. $\square$

14.2 Robustness and Maximum Permissible Uncertainty

Consider $\mathbf{f}$ perturbed by $\Delta \mathbf{f}$. Which $\Delta \mathbf{f}$ can be tolerated by the controller without violating $\dot{V} < -\mathbf{x}^T \mathbf{Q}_1 \mathbf{x}$? This has to hold for a given $\mathbf{Q}_1$ where $\mathbf{Q} - \mathbf{Q}_1 > 0$.

The robust controller

$$u = -\frac{\mathbf{x}^T (\mathbf{Q} \mathbf{x} + \mathbf{P} \mathbf{f})}{\mathbf{x}^T \mathbf{P} \mathbf{g}} \quad (104)$$

operates with some fixed original $\mathbf{f}$. However, the actual $\dot{V}$ contains $\Delta \mathbf{f}$

$$\dot{V} = \mathbf{x}^T \mathbf{P} (\mathbf{f} + \Delta \mathbf{f} + \mathbf{g} u) \leq -\mathbf{x}^T \mathbf{Q}_1 \mathbf{x}. \quad (105)$$

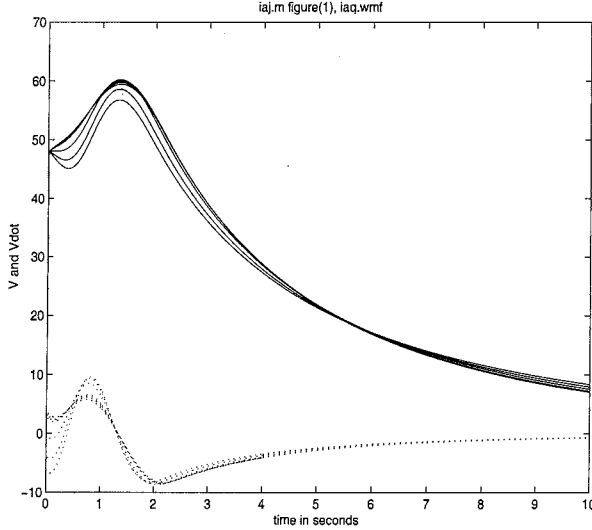


Figure 13: V and $\dot{V}$ for the robust control

From this relation, the permissible Δf is deduced

$$\mathbf{x}^T \mathbf{P} \Delta \mathbf{f} \leq \mathbf{x}^T (\mathbf{Q} - \mathbf{Q}_1) \mathbf{x} . \quad (106)$$

15 Phase-locked Loop Application

Consider the phase-locked loop of Fig. 14. Evaluating its elements, the state-space representation utilizes the phase error $e = -x_1$, the derivative $\dot{x}_1$ and phase reference zero

$$\dot{x}_1 = x_2 \quad (107)$$

$$\dot{x}_2 = -\frac{a_1}{b_1} v x_2 \cos x_1 - \frac{v}{b_1} (\sin x_1 + \beta u) .$$

An additional input u is included, resulting from the iso-matrical stabilizer of Fig. 15, interconnected with the help of a limiter $\arctan(\cdot)$ and a gain factor r .

The nominal $\beta = 1$ yields a sliding-mode and minimal steady-state phase error. At the expense of small steady-state phase error, sliding mode can be avoided by $\beta \triangleq \beta(x_2) = x_2$.

The settings are $v = 1$

$$K(s) = \frac{a_1 s + a_0}{b_1 s} , \quad \mathbf{P} = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \quad (108)$$

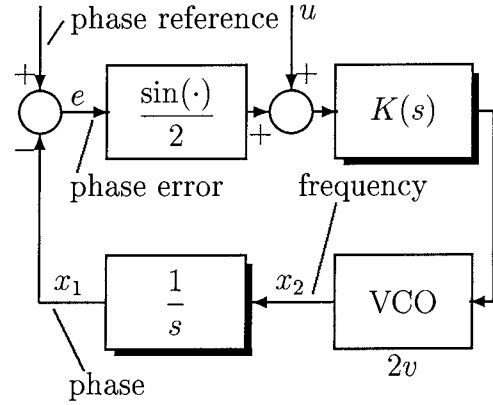


Figure 14: Phase-locked loop with measurement nonlinearity $0.5 \sin x_1$ and VCO of gain $2v$.

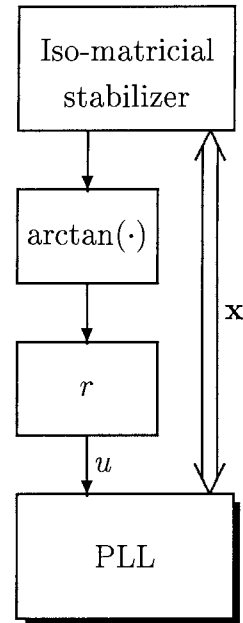


Figure 15: PLL and iso-matrical controller operating as a stabilizer

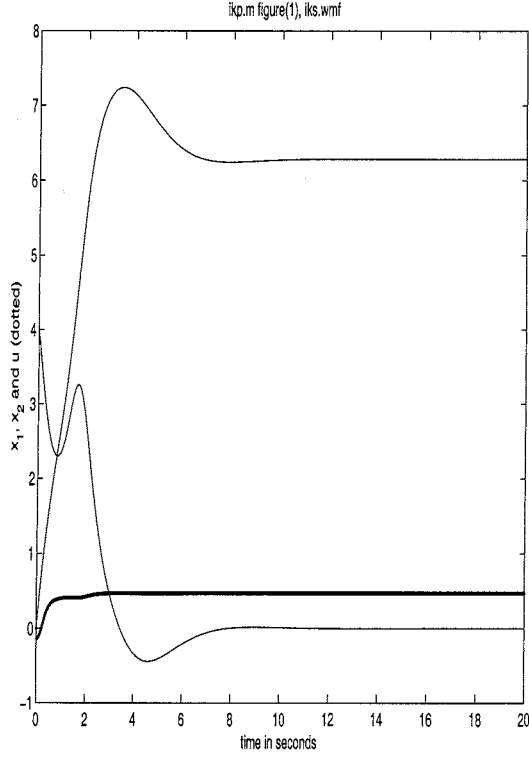


Figure 16: $\mathbf{x}(t)$ and $u(t)$ stabilized for $r = 0.3$

$$\mathbf{Q} = \mathbf{I}_2, \quad \mathbf{x}(0) = \begin{pmatrix} 2 \\ -4 \end{pmatrix}, \quad a_1 = a_0 = 1. \quad (109)$$

For $r = 0.3$ and $r = 0$, stabilization is performed or switched off, respectively.

In Figs. 16 and 17, $\mathbf{x}(t)$ and $u(t)$ are depicted, for $r = 0.3$ (stabilized) and $r = 0$ (original), respectively.

In Figs. 18 and 19, $V(t)$ and $\dot{V}(t)$ are given for $r = 0.3$ and 0, respectively.

In Fig. 20, $x_1(t)$ is shown for a set of parameters r .

Introducing the iso-Lyapunov stabilizer yields a much quicker response at the expense of a small steady-state phase error.

The need for iso-Lyapunov stabilization can be seen if a nonlinear VCO should be applied, i.e., for v replaced by $0.5v(1 + |\sin x_2|)$. In this case, the PLL is nearly unstable. With the help of the iso-Lyapunov stabilizer, good performance is achieved, see

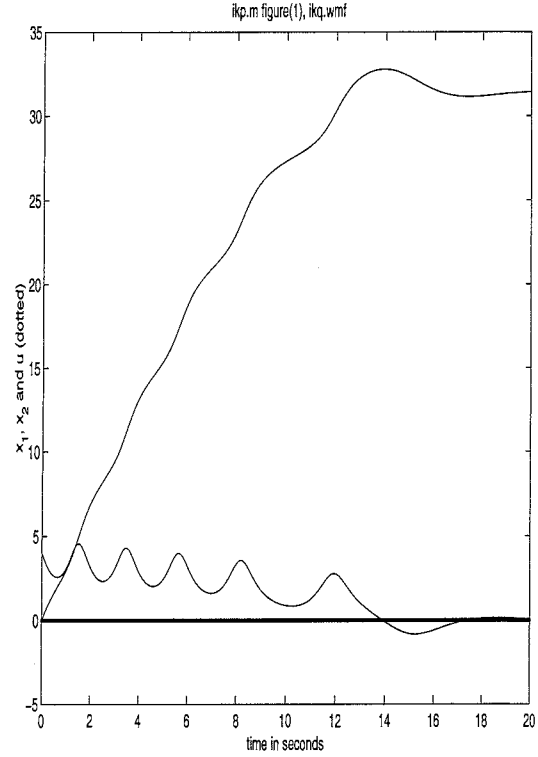


Figure 17: $\mathbf{x}(t)$ and $u(t)$ for $r = 0$

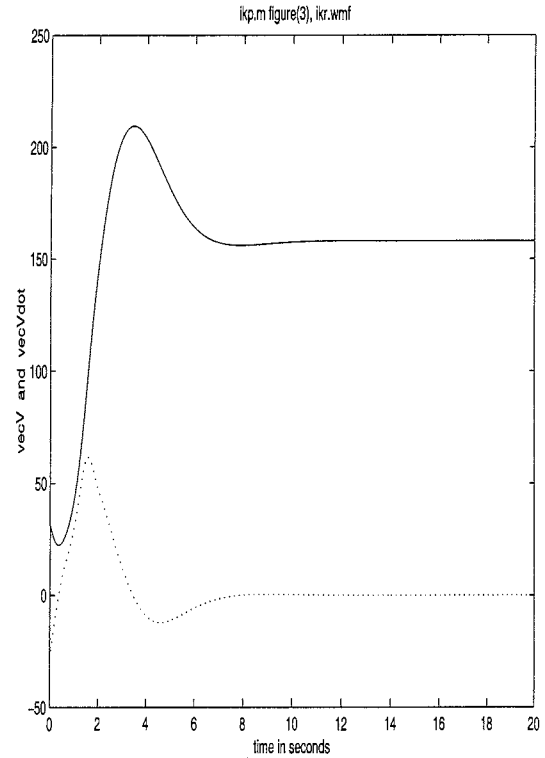


Figure 18: $V(t)$ and $\dot{V}(t)$ stabilized for $r = 0.3$

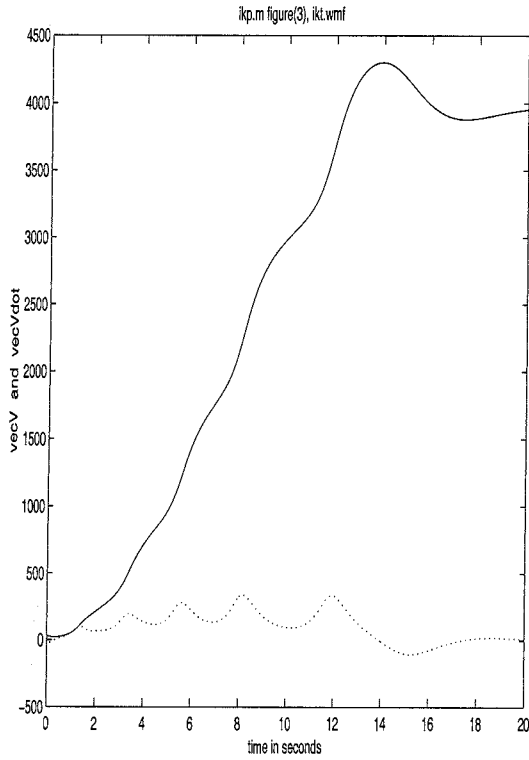


Figure 19: $V(t)$ and $\dot{V}(t)$ for $r = 0$

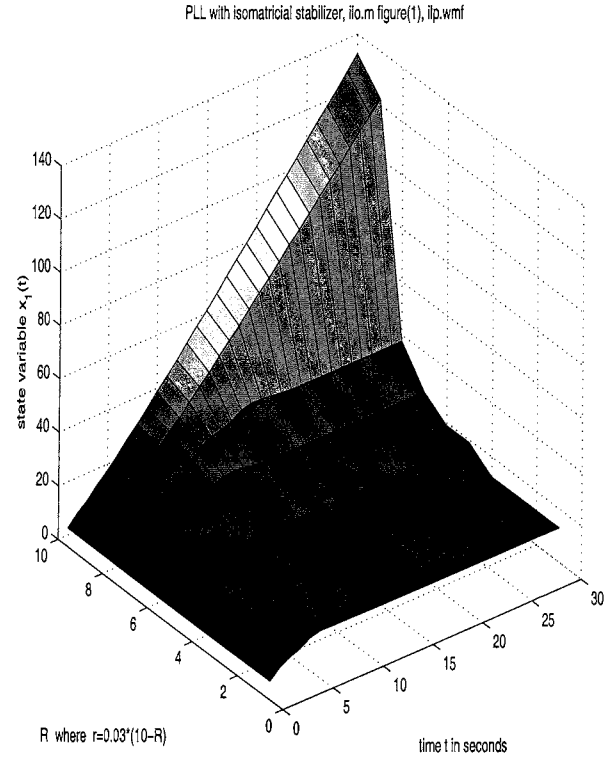


Figure 21: Stable PLL for iso-Lyapunov stabilizer included, in comparison with nearly unstable PLL

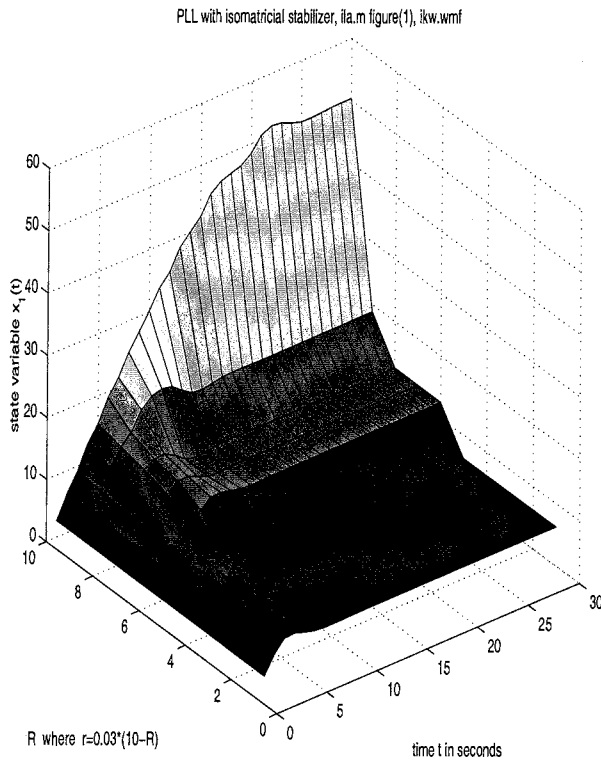


Figure 20: $x_1(t)$ with r as parameter

Fig. 21 for $\beta = 1$. $\square$

16 Conclusion

Iso-Lyapunov dynamics and iso-matrixial controllers obey to a concept of predetermined stability as a common strategy for linear and nonlinear systems. They are carried out for detailed nonlinear and linear assumptions both for SISO and MIMO systems in continuous and discrete time.

The general recommendation for applying the iso-Lyapunov concept is stabilization of nonlinear systems for mainly known nonlinearity. Application is advisable in cases where the design lacks of appropriate suggestions for stabilizing the system. By proper solution of the required matrices, robust results are easily achieved.

Maximum amplitudes and norms are assessed. Furthermore, the control system deviation and set-point assumptions are included. Moreover, the Lyapunov derivative is replaced by passivity settings.

Robustness is investigated as far as worst case perturbation is concerned.

Hard limits of activating signals facilitate the occurrence of sliding mode.

Open problems of concern are: Observation of $\mathbf{x}$ instead of direct measuring; conditions for sliding mode; dependence on initial conditions.

References

- Aeyels, D., Peuteman, J., 1998, New asymptotic stability criterion for nonlinear time-variant differential equations, *IEEE-Trans.* **43**, pp. 968-971
- Ball, A., Helton, W., Walker, L., 1993, H_∞ -Control for Nonlinear Systems with Output Feedback, *IEEE-Trans.* **AC-38**, pp. 546-559
- Bang, H., 1994, Generalized stabilizing control laws design using Lyapunov stability theory, *Proc. Int. Mechanical Eng. Congress and Exp.* pp. 49-58
- Blanchini, F., 1995, Nonquadratic Lyapunov functions for robust control, *Automatica* **31**, pp. 451-461
- Böcker, J., Hartmann, I., and Zwanzig, Ch., 1986, Nichtlineare und adaptive Regelsysteme (Springer, Berlin)
- Boyd, S., ElGhaoui, L., Feron, E., and Balakrishnan, V., 1994, Linear Matrix Inequalities in System and Control Theory (SIAM Studies in Applied Mathematics, Philadelphia)
- Collins, E.G., Sadhukhan, D., and Watson, L.T., 1999, Robust controller synthesis via non-linear matrix inequalities, *Int.J. of Control* **72**, pp. 971-980
- DeCarlo, R.A., and Drakunov, S., 1998, Unified Lyapunov setting for continuous and discrete time sliding mode control, *Proc. Int. Mechanical Eng. Congress and Exp.* pp. 547-554
- Feng, G., 1996, Improving tracking control for robots using neural networks, *Int. J. of Robotics and Automation* **11**, pp. 74-82
- Hu, B., and Michel, A.N., 2000, Stability analysis of digital feedback control systems with time-varying sampling periods, *Automatica* **36**, pp. 897-905
- Isidori, A., Astolfi, A., 1992, Disturbance Attenuation and H_∞ -Control via Measurement Feedback in Nonlinear Systems, *IEEE-Trans.* **AC-37**, pp. 1283-1293
- Jankovic, M., Sepulchre, R., and Kokotovic, P.V., 1996, Constructive Lyapunov stabilization of nonlinear cascade systems, *IEEE-Trans.* **41**, pp. 1723-1735
- Jiang, Z.P., 1999, Unified Lyapunov framework for stabilization and tracking of nonholonomic systems, *IEEE Conference on Decision and Control, Phoenix*, pp. 2088-2093
- Kosmatopoulos, E.B., Chassiakos, A., Boussalis, H., Mirmirani, M., Ioannou, P.A., 1998, Neural network control of unknown systems, *IEEE Conf. on Neural Networks*, Part 2, pp. 943-948
- Kugi, A., 2001, Nonlinear Control Based on Physical Models, *Lecture Notes in Control and Information Sciences* **260** (Springer, Heidelberg)
- Lee, J.K., Suh, B.H., and Abe, K.I., 1995, Model reference adaptive control of nonlinear system using feedback linearization, *Proc. SICE Conf.* **34**, pp. 1571-1576
- Leonessa, A., Haddad, W.M., Chellaboina, V.S., 1998, Nonlinear system stabilization via stability-based switching, *IEEE Cnf. on Decision and Control* **37**, pp. 2983-2996
- McConley, M.W., Appleby, B.D., Dahleh, M.A., Feron, E., 2000, Computationally efficient Lyapunov-based scheduling procedure for control of nonlinear systems with stability guarantees, *IEEE-Trans.* **45**, pp. 33-49
- Ortega, R., Loria, A., Nicklasson, P.J., Sira Ramirez, H., 1998, Passivity-based Control of Euler-Lagrange Systems, (Springer, London)
- Qu, Zhihua, 1998, Robust Control of Nonlinear Uncertain Systems, , pp. (John Wiley, New York)
- Robinet, R.D., Parker, G.G., Schaub, H., Junkins, J.L., 1997, Lyapunov optimal saturated control for nonlinear systems, *Journal of Guidance, Control, and Dynamics* **20**, pp. 1083-1088
- Sepulchre, R., Jankovic, M., Kokotovic, P., 1997, Constructive Nonlinear Control (Springer, London)
- Slotine, J.J.E., and Li, W., 1991, Applied Non-

- linear Control (Prentice-Hall., Englewood Cliffs)
- Su, W., de Souza, C.E., Xie, L., 1999, H_∞ control for asymptotically stable nonlinear systems, *IEEE-Trans.* **44**, pp. 989-993
- Sznaier, M., Clutier, J., Hull, R., Jacques, D., and Mracek, C., 2000, Receding horizon control Lyapunov function approach to suboptimal regulation of nonlinear systems, *Journal of Guidance, Control, and Dynamics* **23**, pp. 399-405
- Tsai, C.Y., Li, T-H.S., 1996, Design of Lyapunov function based fuzzy logic controller, *IEEE Conf. on Industrial Electronics, Control, and Instrumentation* **22**, Part 1, pp. 396-401
- Ugrinovskii, V.A., and Petersen, I.R., 2000, Guaranteed cost control of uncertain systems via Lur'e-Postnikov Lyapunov functions, *Automatica* **36**, pp. 279-285
- van der Schaft, A.J., 2000, L_2 -Gain and Passivity Techniques in Nonlinear Control (Springer, London)
- Vidal, P., 1969, Non-linear sampled-data systems (Gordon and Breach, New York)
- Vidyasagar, M., 1993, Nonlinear Systems Analysis, 2nd edn., (Prentice Hall, Englewood Cliffs New Jersey)
- Willems, J.C., 1972, Dissipative dynamical systems, Part I: General Theory, *Arch. Rational Mech. Anal.* **45**, pp. 321-351
- Yang, S.K., 1998, Lyapunov-type state feedback controller design for uncertain systems with time-varying disturbance, *Journal of Control Systems and Technology* **6**, pp. 277-282
- Yesildirek, A., and Lewis, F.L., 1994, Neural network controller for feedback linearization, *IEEE Conf. on Decision and Control* **33**, pp. 2494-2499
- Zergeroglu, E., Dixon, W.E., Dawson, D.M., and Behal, A., 1999, Lyapunov-based set point control of the acrobot, *Int. J. of Robotics and Automation* **14**, pp. 161-170

Appendices

A Lyapunov Equation

The Lyapunov equation is

$$\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} + \mathbf{Q} = \mathbf{0}. \quad (110)$$

Starting with an arbitrary $\mathbf{P} > 0$, even for some stable system matrix $\mathbf{A}$ the matrix $\mathbf{Q}$ need not be $\mathbf{Q} > 0$, it can be also $\mathbf{Q} < 0$, hence no stability answer is stated.

Example: Given $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -2 & -4 \end{pmatrix}$, assume $\mathbf{P} = \mathbf{I}$ then $\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} = \mathbf{Q} = \begin{pmatrix} 0 & -1 \\ -1 & -1 \end{pmatrix}$ and $\mathbf{Q}$ is not > 0 .

However, starting with $\mathbf{Q} > 0$ and solving the Lyapunov equation with respect to $\mathbf{P}$, then the system is stable if $\mathbf{P} > 0$. In this direction a conclusion can be drawn.

Positive definiteness $\mathbf{P} > 0$ is satisfied if $\mathbf{x}^T \mathbf{P} \mathbf{x} > 0 \quad \forall \mathbf{x} \neq \mathbf{0}$ or if $P_{11} > 0$, $P_{11} P_{22} - P_{12} P_{21} > 0, \dots \det \mathbf{P} > 0$ (all principal minors positive) or if the angle $(\mathbf{x}, \mathbf{P} \mathbf{x}) < \pi/2$.

For the extended Lyapunov equation

$$\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} + \mathbf{Q} + 2\mu \mathbf{P} = \mathbf{0}, \quad (111)$$

if $\mathbf{P} > 0$, $\mathbf{Q} > 0$ then $\Re \lambda_i[\mathbf{A}] < \mu$.

Lyapunov Functions and Storage Functions: Given $V(\mathbf{x})$ positive definite, if $\dot{V}(\mathbf{x}) = L_f V(\mathbf{x}) \leq 0$, i.e. the Lie derivative along the vector field $\mathbf{f}$ for $\mathbf{x}_f = \mathbf{0}$ is negative, then the system is stable (for ≤ 0) or asymptotically stable (for < 0) (Slotine, J.J.E., and Li, W., 1991; Vidyasagar, M., 1993).

More generally, since energy functions are appropriate Lyapunov candidates, a **dissipative system** with non-negative storage function V obeys the dissipation inequality (Ball, A., et al., 1993; Isidori, A., and Astolfi, A., 1992; Kugi, A., 2000)

$$V[\mathbf{x}(t)] - V[\mathbf{x}(t_0)] \leq \int_{t_0}^t s[\mathbf{u}(\tau), \mathbf{y}(\tau)] d\tau \quad (112)$$

where $s[\cdot]$ is the supply rate.

For the equality sign = instead of $\leq$, the system is termed lossless with respect to $s[\cdot]$. For $V(\mathbf{x})$ differentiable, the differential dissipation inequality is $\dot{V}(\mathbf{x}) \leq s[\mathbf{u}(t), \mathbf{y}(t)]$.

For $s[\mathbf{u}, \mathbf{y}] = \gamma \|\mathbf{u}\|_F^2 - \|\mathbf{y}\|^2$ the input-output stability of the system is characterized by L_2 gain less than γ .

For $\dim \mathbf{u} = \dim \mathbf{y}$ the system is passive if it is dissipative with $s = \mathbf{u}^T \mathbf{y}$.

The system is strictly input passive if it is dissipative with $s = \mathbf{u}^T \mathbf{y} - \alpha \|\mathbf{u}\|_F^2$.

A passive and lossless system is termed conservative.

An affine-input system is positive real if

$$\int_{t_0}^t \mathbf{u}^T \mathbf{y} d\tau \geq 0 \quad \text{for } \mathbf{x}(t_0) = \mathbf{0}. \quad (113)$$

A system which is reachable from $\mathbf{x}(t_0) = \mathbf{0}$ is passive iff it is positive real.

A system is positive real iff

$$L_f V(\mathbf{x}) = -\mathbf{l}^T(\mathbf{x}) \mathbf{l}(\mathbf{x}) \quad (114)$$

$$0.5[L_{g_1} V(\mathbf{x}), \dots, L_{g_m} V(\mathbf{x})] = \mathbf{h}^T(\mathbf{x}) - \mathbf{l}^T(\mathbf{x}) \mathbf{W}(\mathbf{x}) \quad (115)$$

$$\mathbf{W}^T(\mathbf{x}) \mathbf{W}(\mathbf{x}) = \mathbf{J}^T(\mathbf{x}) + \mathbf{J}(\mathbf{x}) \quad (116)$$

where $\mathbf{l}$, $\mathbf{W}$ are real-valued and $V(\mathbf{x}) > 0$.

For a linear systems with $(\mathbf{A}, \mathbf{B})$ reachable and $(\mathbf{C}, \mathbf{A})$ observable and $\mathbf{P} > 0$, the corresponding relations are

$$\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} = -\mathbf{L}^T \mathbf{L} \quad (117)$$

$$\mathbf{P} \mathbf{B} = \mathbf{C}^T - \mathbf{L}^T \mathbf{W} \quad (118)$$

$$\mathbf{W}^T \mathbf{W} = \mathbf{D}^T + \mathbf{D}. \quad (119)$$

B Passive and Dissipative Systems

The behaviour of a system is dissipative, if the energy E_1 fed to the system is bigger or equal to the difference energy E_s stored in the system in a time interval of concern; the surplus of energy has to be deleted in the system. The variable $w(\mathbf{u}, \mathbf{y})$ is a generalized function of the power (energy per time)

$$E_1 > E_s \quad (120)$$

$$E_1 = \int_{t_0}^{t_1} w(\mathbf{u}, \mathbf{y}) dt > V[\mathbf{x}(t_1)] - V[\mathbf{x}(t_2)] \quad (121)$$

where V is the energy stored in the system.

A stable LTI system is dissipative with respect to a square power function

$$w(\mathbf{u}, \mathbf{y}) = \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \end{pmatrix}^T \begin{pmatrix} \mathbf{R} & \mathbf{S} \\ \mathbf{S}^T & \mathbf{T} \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \end{pmatrix} \quad (122)$$

$\mathbf{R} = \mathbf{R}^T \in \mathcal{R}^{q \times q}$, $\mathbf{T} = \mathbf{T}^T \in \mathcal{R}^{p \times p}$, $\mathbf{S} \in \mathcal{R}^{q \times p}$ if a positive definite energy function $V(\mathbf{x}) = \mathbf{x}^T \mathbf{P} \mathbf{x}$ exists where

$$\int_0^T w(\mathbf{u}, \mathbf{y}) dt > V[\mathbf{x}(T)] - V[\mathbf{x}(0)] \quad (123)$$

is satisfied. In differential form $w \geq \dot{V}$. The relation

$$\int_0^T w(\mathbf{u}, \mathbf{y}) dt \geq 0 \quad (124)$$

is a sufficient condition.

Hyperstability $\mathbf{u}^T \mathbf{y} \geq 0$ is obtained if

$$\begin{pmatrix} \mathbf{R} & \mathbf{S} \\ \mathbf{S}^T & \mathbf{T} \end{pmatrix} \text{ is replaced by } \begin{pmatrix} \mathbf{O} & \mathbf{I} \\ \mathbf{I} & \mathbf{O} \end{pmatrix}. \quad (125)$$

An LTI System is located in a sector $[a, b]$ for $b > a$, $b < \infty$, if

$$\int_0^T (\mathbf{y} - a\mathbf{u})^T (\mathbf{y} - b\mathbf{u}) \leq 0 \quad (126)$$

$$\mathbf{u} \in L_{2e}^p, T \in [0, \infty) \quad (127)$$

$$\int_0^T (\mathbf{F}_{11}\mathbf{y} - \mathbf{F}_{12}\mathbf{u})^T (\mathbf{F}_{21}\mathbf{y} - \mathbf{F}_{22}\mathbf{u}) dt \geq 0. \quad (128)$$

Equivalences to $\mathbf{R}, \mathbf{S}$ and $\mathbf{T}$ can easily be found.

An affine-input system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u} \quad (129)$$

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) \quad (130)$$

satisfying

$$\dot{V}(t) = \mathbf{u}^T \mathbf{y} - g(t) \quad (131)$$

with V lower bounded, e.g. positive, and $g(t) \geq 0$, is termed passive. Moreover, it can be characterized by a passive mapping between input $\mathbf{u}$ and output $\mathbf{y}$.

A strictly stable linear SISO system is passive (Slotine, J.J.E., and Li, W., 1991)

$$\text{iff } \Re G(j\omega) \geq 0 \quad \forall \omega \geq 0 \quad (132)$$

$$\text{or iff } \arg G(j\omega) \leq \pi/2. \quad (133)$$

A passive system is termed dissipative if, for V lower bounded,

$$\int_0^\infty \mathbf{u}^T \mathbf{y} d\tau \neq 0 \quad \text{and} \quad \int_0^\infty g(\tau) d\tau > 0. \quad (134)$$

A passive linear differential inclusion arises if every solution $\mathbf{x}$ with $\mathbf{x}(0) = \mathbf{0}$ satisfies

$$\int_0^T \mathbf{u}^T \mathbf{y} dt \geq 0 \quad T > 0. \quad (135)$$

Dissipation η is given for (Willems, J.C., 1972; Su, W., et al., 1999; Ball, A., et al., 1993)

$$\int_0^T (\mathbf{u}^T \mathbf{y} - \eta \mathbf{u}^T \mathbf{u}) \geq 0. \quad (136)$$

The largest number η such that this relation holds is termed dissipativity.

Positive (and zero) dissipation η corresponds to passivity (Ortega, R., et al., 1998)

Passivity is given if the transfer matrix $\mathbf{G}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$ is square and positive real and if all the elements of $\mathbf{G}(s)$ are analytic in the open right-half s -plane and $\mathbf{G}(s) + \mathbf{G}^H(s) \geq 0$ holds for all $\Re s > 0$, i.e. is positive semidefinite or $\mathbf{G}(s) + \mathbf{G}^H(s)$ is positiv-real. For $\mathbf{B} \in \mathcal{R}^{n \times m}$, $\mathbf{B} = \mathbf{b}$, i.e. $m = 1$, the Nyquist plot is entirely in right half plane.

These relations are equivalent to the fact that the LMI in the variable $\mathbf{P}$ (Boyd, S., et al., 1994)

$$\begin{pmatrix} \mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} & \mathbf{P} \mathbf{B} - \mathbf{C}^T \\ \mathbf{B}^T \mathbf{P} - \mathbf{C} & -(\mathbf{D} + \mathbf{D}^T) \end{pmatrix} \leq 0 \quad (137)$$

is feasible.

Alternatively, there exists a $\mathbf{P} = \mathbf{P}^T$ satisfying the Riccati equation

$$\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} + (\mathbf{P} \mathbf{B} - \mathbf{C}^T)(\mathbf{D} + \mathbf{D}^T)^{-1}(\mathbf{P} \mathbf{B} - \mathbf{C}^T)^T = \mathbf{0}. \quad (138)$$

In special cases, nonlinear matrix inequalities are used for robust controller synthesis (Collins, E.G., et al., 1999).

Positive linear systems are characterized by positive real transfer functions, i.e.

$$\Re G(s) \geq 0 \quad \forall \Re s \geq 0 \quad (\text{right-half } s\text{-plane}) \quad (139)$$

and are strictly positive real if

$$\Re G(s - \varepsilon) \geq 0 \quad \forall \Re s \geq 0 \quad \text{and some } \varepsilon > 0. \quad (140)$$

In state-space, the linear system

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}\mathbf{u} \quad (141)$$

$$\mathbf{y} = \mathbf{c}^T \mathbf{x} \quad (142)$$

is strictly positive real iff

$$\exists \mathbf{P} > 0, \mathbf{Q} > 0 \quad (143)$$

$$\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} + \mathbf{Q} = \mathbf{0} \quad (144)$$

$$\mathbf{P} \mathbf{b} = \mathbf{c} \quad (145)$$

(Kalman Yakubovic Lemma). This also means that $G(s)$ is stable.

Given the linear autonomous system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$, for $\mathbf{A}$ strictly stable, then for

$$\dot{V} = 0.5 \mathbf{x}^T \mathbf{P} \mathbf{x} \quad (146)$$

$$\dot{V} = -0.5 \mathbf{x}^T \mathbf{Q} \mathbf{x} \quad (147)$$

$\exists \mathbf{P} > 0, \mathbf{Q} > 0$. For a non-autonomous system

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \quad \mathbf{y} = \mathbf{C}\mathbf{x} \quad (148)$$

$$\dot{V} = \mathbf{x}^T \mathbf{P} \dot{\mathbf{x}} \quad \text{from} \quad V = \mathbf{x}^T \mathbf{P} \mathbf{x} \quad (149)$$

select the additional term $\mathbf{x}^T \mathbf{P} \mathbf{B} \mathbf{u}$

$$\dot{V} \triangleq \mathbf{x}^T \mathbf{P} \mathbf{B} \mathbf{u} - 0.5 \mathbf{x}^T \mathbf{Q} \mathbf{x} . \quad (150)$$

Assuming Kalman Yakubovich

$$\mathbf{C} = \mathbf{B}^T \mathbf{P} \quad (151)$$

$$\mathbf{x}^T \mathbf{C}^T = \mathbf{y}^T \quad (152)$$

$$\mathbf{x}^T \mathbf{P} \mathbf{B} = \mathbf{y}^T \quad (153)$$

$$\mathbf{x}^T \mathbf{P} \mathbf{B} \mathbf{u} = \mathbf{y}^T \mathbf{u} = \mathbf{u}^T \mathbf{y} , \quad (154)$$

Eq.150 is a dissipative mapping between $\mathbf{u}$ and $\mathbf{y}$.
In Eq.131, $g(t)$ corresponds to $-0.5 \mathbf{x}^T \mathbf{Q} \mathbf{x}$.

The non-expansive system obeys

$$\int_0^T \mathbf{y}^T \mathbf{y} dt < \int_0^T \mathbf{u}^T \mathbf{u} dt \quad \forall \mathbf{x}(0) = \mathbf{0} \quad (155)$$

which is equivalent to $\mathbf{G}(s)$ being bounded-real, i.e.

$$\mathbf{G}^H(s) \mathbf{G}(s) \leq \mathbf{I} \quad \forall \Re s > 0 \quad (156)$$

or

$$\|\mathbf{G}\|_\infty \leq 1 \quad \forall \Re s > 0 \quad (157)$$

where

$$\|\mathbf{G}\|_\infty \triangleq \sup \{ \sigma_{\max}[\mathbf{G}] \mid \Re s > 0 \} . \quad (158)$$

For $m = 1$, this result corresponds to the Bode plot.

C Comparison with HJB Equation

Consider the affine-input system

$$\mathbf{G}_u = [\mathbf{g}_j(\mathbf{x})] \quad j = 1 \dots m \quad (159)$$

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{G}_u \mathbf{u} , \quad \mathbf{y} = \mathbf{h}(\mathbf{x}) . \quad (160)$$

Nonlinear H_2 design for a stable system at $\mathbf{x} = \mathbf{0}$ requires

$$\int_0^\infty \left(\sum_1^p y_i^2 + \sum_1^m u_i^2 \right) d\tau \rightarrow \min_{\mathbf{u}} . \quad (161)$$

Selecting

$$V[\mathbf{x}(t_0)] = \inf_{\mathbf{u}} \int_{t_0}^\infty \left(\sum_1^p y_i^2 + \sum_1^m u_i^2 \right) d\tau \quad (162)$$

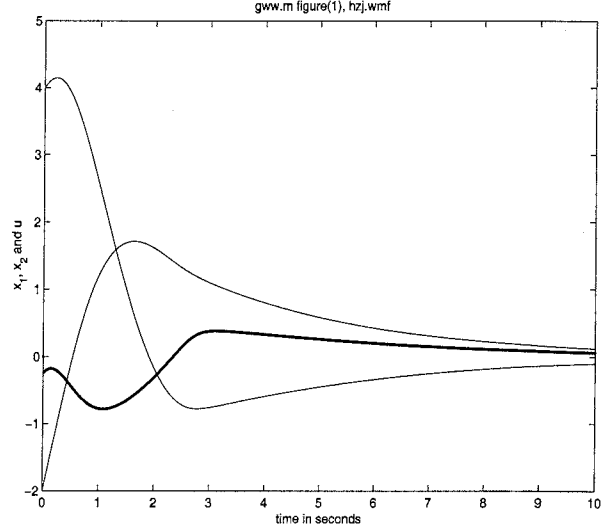


Figure 22: State variable $\mathbf{x}$ with $\mathbf{x}(0) = (-2 \ 4)^T$ and u

the result is

$$L_f V(\mathbf{x}) + \sum_1^p y_i^2 - \sum_{j=1}^m (L_{g_j} V(\mathbf{x}))^2 = 0 \quad (\leq 0) \quad (163)$$

and

$$u_j^* = -L_{g_j} V(\mathbf{x}) , \quad (164)$$

termed Hamilton-Jacobi-Bellmann equation (inequality), abbreviated HJBe(i).

For linear systems with $V = 0.5 \mathbf{x}^T \mathbf{P} \mathbf{x}$, $\mathbf{P} > 0$, HJBe(i) turns out as

$$\mathbf{x}^T (\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} + \mathbf{C}^T \mathbf{C} - \mathbf{P} \mathbf{B} \mathbf{B}^T \mathbf{P}) \mathbf{x} = 0 . \quad (165)$$

D Alternative with Absolute Denominator

An interesting alternative to Eq.(5) is

$$u = - \frac{\mathbf{x}^T (\mathbf{Q} \mathbf{x} + \mathbf{P} \mathbf{f})}{|\mathbf{x}^T \mathbf{P} \mathbf{g}|} , \quad (166)$$

see Figs. 33 through 36, e.g. for smooth transients.

Example 4. Absolute denominator: Consider Eq.(166) and the same setting as in Example 1, the results are given in Figs. 22 and 23. There is no change in sign for $\mathbf{x}^T \mathbf{P} \mathbf{g}$ (Fig. 23).

For specific settings and initial conditions, including the modulus in the denominator leads to better results although the stability condition seems to be violated. It must be guaranteed that the left-hand side $\dot{V} = \mathbf{x}^T (\mathbf{f} + \mathbf{g} u)$ is negative even if u is

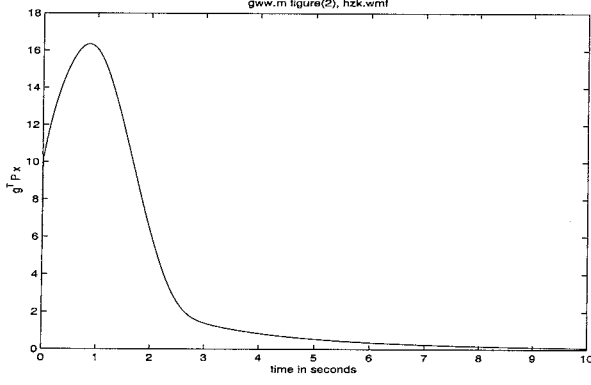


Figure 23: Denominator $\mathbf{g}^T \mathbf{P} \mathbf{x}$ versus time

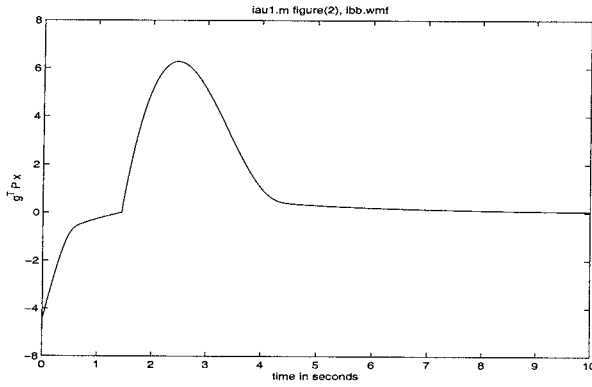


Figure 24: $\mathbf{x}^T \mathbf{P} \mathbf{g}(t)$

wrong in sign. A similar effect of Lyapunov derivative positive or negative is described in *Aeyels, D., and Peuteman, J., 1998*. $\square$

Example 5. Wrong sign: In Fig. 24 for $0 < t < 1.5$ the sign of $\mathbf{g}^T \mathbf{P} \mathbf{x}$ is wrong according to Eq.(166). The derivative $\dot{V}$ is wrong as long as $\mathbf{x}^T \mathbf{P} \mathbf{g}$ is negative, it is equal to $L = 2\mathbf{x}^T \mathbf{P} \mathbf{f} + \mathbf{x}^T \mathbf{Q} \mathbf{x}$ instead of $-\mathbf{x}^T \mathbf{Q} \mathbf{x}$. Then $\dot{V}$ (dotted characteristic in Fig. 25) corresponds to L (dotted characteristic in Fig. 26) up to 1.5 seconds, afterwards it is equal to $-\mathbf{x}^T \mathbf{Q} \mathbf{x}(t)$ ('dash dot' in Fig. 26). $\square$

E Multivariable System Examples Minimizing the Norm of $\mathbf{u}$

Example 6. Second-order system: Assuming

$$\mathbf{P} = \text{diag}[4 \ 2], \quad \mathbf{Q} = 0.5\mathbf{I}_2, \quad \mathbf{R} = 2\mathbf{I}_2 \quad (167)$$

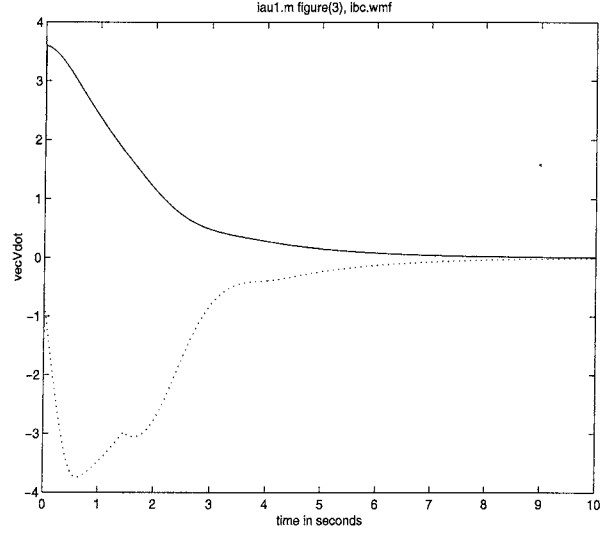


Figure 25: $V(t)$ and $\dot{V}(t)$ ('dotted')

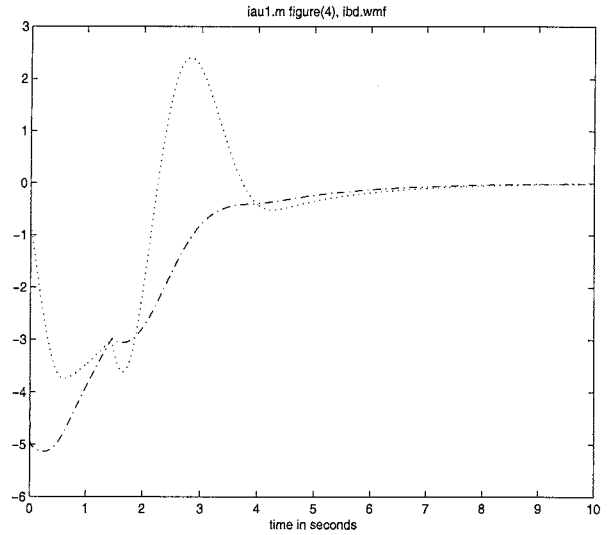


Figure 26: $L(t)$ ('dotted') and $-\mathbf{x}^T \mathbf{Q} \mathbf{x}(t)$ ('dash dot')

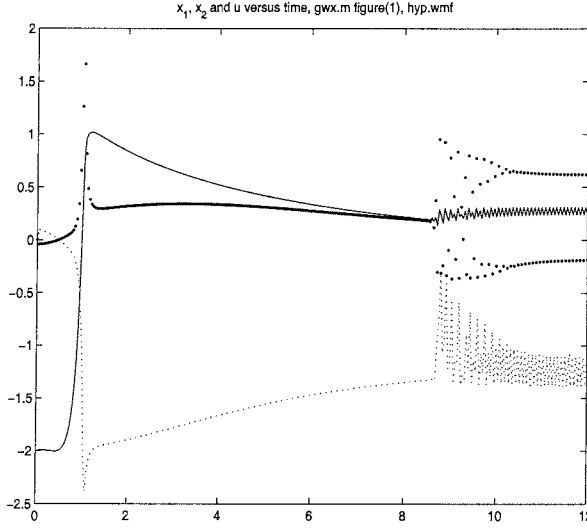


Figure 27: $x_1(t)$, $u_1(t)$, $u_2(t)$

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} x_2^2 \\ 2x_1 + 3x_2 \end{pmatrix}, \quad \mathbf{G}(\mathbf{x}) = \begin{pmatrix} 4x_1 & 4 \\ x_2 & 3 + x_1 \end{pmatrix} \quad (168)$$

in Eq.(18), the results are depicted in Fig. 27.

Interesting alternatives are obtained by changing $G_{12}(\mathbf{x}) = 4x_2$ or $0.4x_1$. $\square$

Example 7. Third-order system: For a as a parameter in $\mathbf{G}$ and Eq.(18),

$$\mathbf{P} = \text{diag}[0.4, 0.2, 1], \quad \mathbf{Q} = 0.5\mathbf{I}_3, \quad \mathbf{R} = 0.2\mathbf{I}_3 \quad (169)$$

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} x_2 \\ -2x_1 - 3x_2 \\ -x_3^2 \end{pmatrix} \quad (170)$$

$$\mathbf{G}(\mathbf{x}) = \begin{pmatrix} 40ax_1 & 4 & x_3^2 \\ x_2 & 3 + x_1 & x_3^3 \\ 0 & 0 & 2 \end{pmatrix}. \quad (171)$$

The parameter a is an element of the set $\{0.2 \ 0.4 \ 0.6 \ 0.8 \ 1\}$. In Figs. 28 through 30 the results are depicted for all the parameters a . $\square$

Example 8. Stabilization: Suppose a third-order plant with low damping quality. In Figs. 31 and 32 the result of an iso-matricial controller is shown and compared with the unforced plant dynamics without controller ('dotted'). Moreover, five uncertainty parameter values are assumed just as in the example before. (The trajectories seem to intersect, but this is only a consequence of figuring out a three-dimensional trajectory in the two-dimensional plane.) $\square$

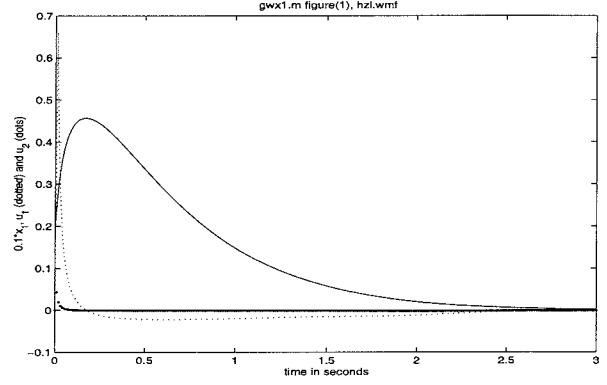


Figure 28: $x_1(t)$, $u_1(t)$ and $u_2(t)$, for $a = 5$ only

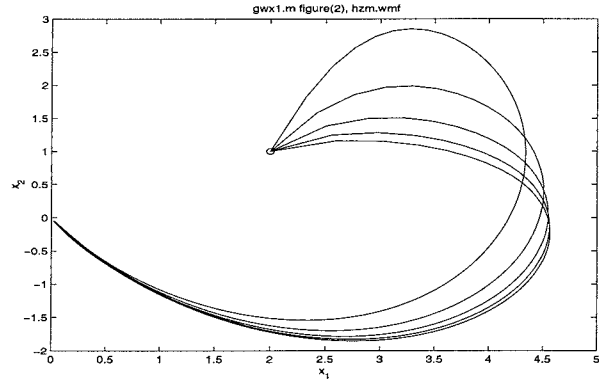


Figure 29: Phase plot x_2 versus x_1

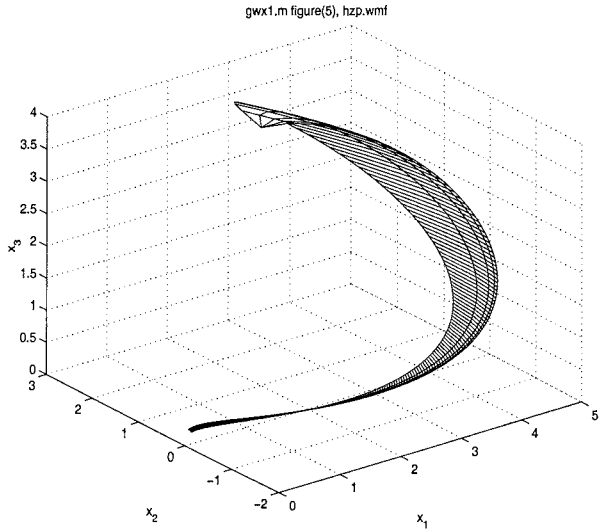


Figure 30: State space transients x_1 , x_2 , x_3

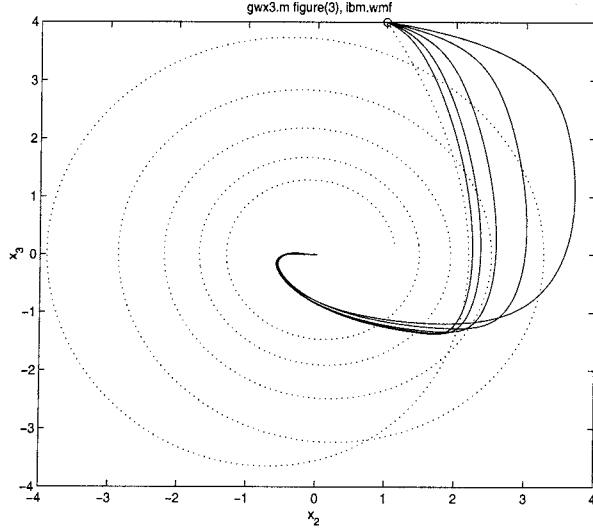


Figure 31: $x_2(t)$ and $x_3(t)$ with controller ('solid') and unforced ('dotted')

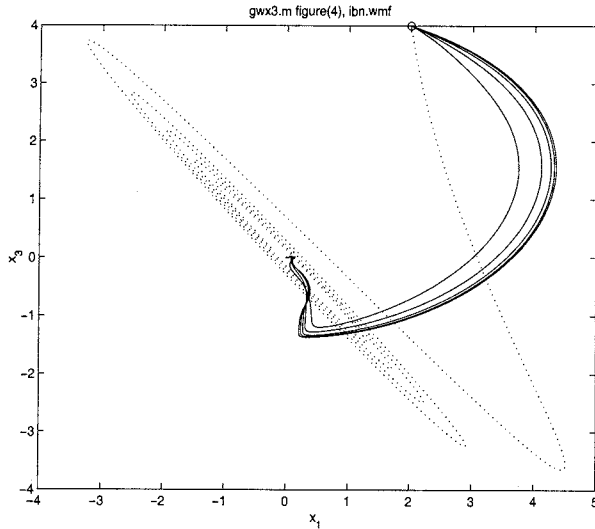


Figure 32: $x_1(t)$ and $x_3(t)$ with controller ('solid') and unforced ('dotted')

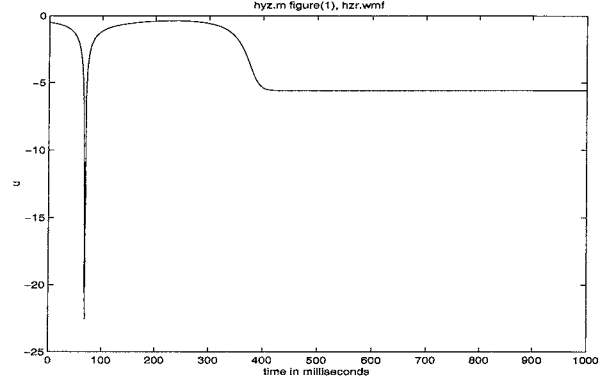


Figure 33: $u(t)$ without hard limit

F Linear SISO Examples

Example 9. Absolute value denominator:

By intentionally selecting the absolute value in the denominator of Eq.(21) one has

$$u = -\frac{\mathbf{x}^T(\mathbf{Q} + \mathbf{P}\mathbf{A})\mathbf{x}}{|\mathbf{x}^T\mathbf{P}\mathbf{b}|} = \frac{\mathbf{x}^T\mathbf{A}^T\mathbf{P}\mathbf{x}}{|\mathbf{x}^T\mathbf{P}\mathbf{b}|}. \quad (172)$$

The steady-state $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u = 0$, only for $\mathbf{x}_f \rightarrow 0$, is given by

$$\det(\mathbf{A} + \frac{\mathbf{b}\mathbf{x}^T\mathbf{A}^T\mathbf{P}}{|\mathbf{x}^T\mathbf{P}\mathbf{b}|}) = 0. \quad \square \quad (173)$$

Example 10. Linear second-order SISO system, absolute-value denominator:

$$\mathbf{Q} = 0.1\mathbf{I}_2 \quad (174)$$

$$\mathbf{A} = \begin{pmatrix} 0.15 & 1 \\ -1 & -0.3 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0.13 \\ 1 \end{pmatrix} \quad (175)$$

The result of Eq.(172) is given in Fig. 33; with the inclusion of a hard limit $|u| < 7$ in Fig. 34, with nearly no influence on $\mathbf{x}(t)$. Characteristics versus time are portrayed in Figs. 35.

In Fig. 36 the denominator $\mathbf{x}^T\mathbf{P}\mathbf{b}$ is sketched versus time. $\square$

Example 11. Linear second-order system Eq.(21): The results for true denominator are given in Figs. 37 through 43. Hard limit causes worse behaviour. $\square$

Example 12. Q depending on x: Consider settings similar to Eq.(175) but $\mathbf{Q} = 0.5 \|\mathbf{x}\|_F^{0.5e_x} \mathbf{I}_2$ for $e_x = \{1 \dots 5\}$. The phase plot of the results is given in Fig. 44. $\square$

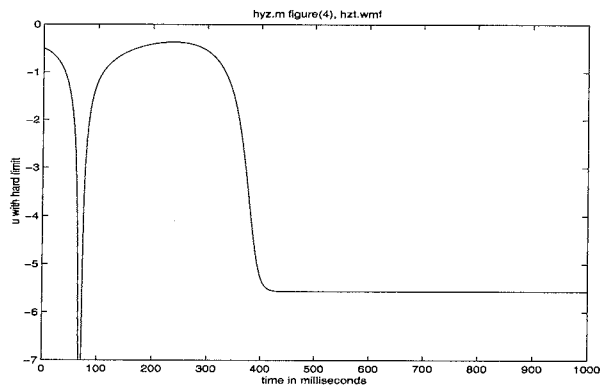


Figure 34: $u(t)$ with hard limit at 7

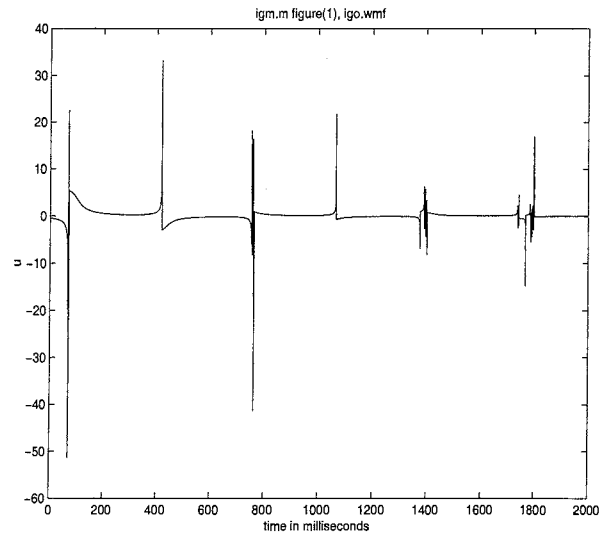


Figure 37: $u(t)$

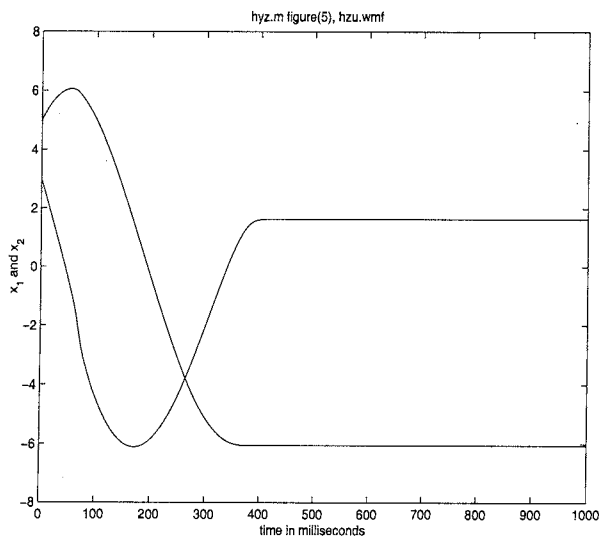


Figure 35: $\mathbf{x}(t)$ for $\mathbf{x}(0) = (5 \ 3)^T$

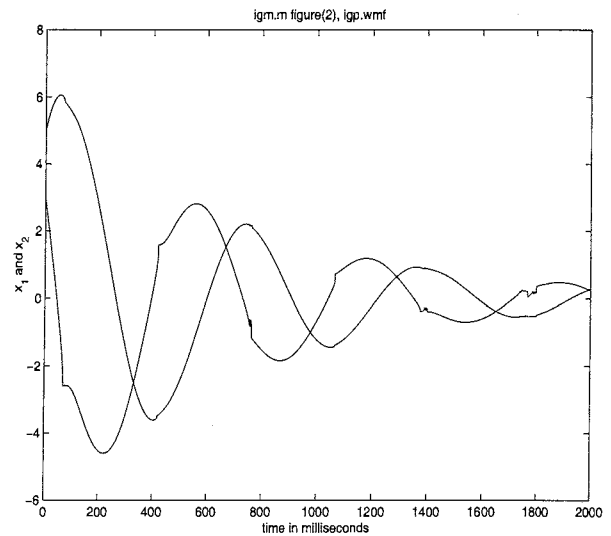


Figure 38: $\mathbf{x}(t)$

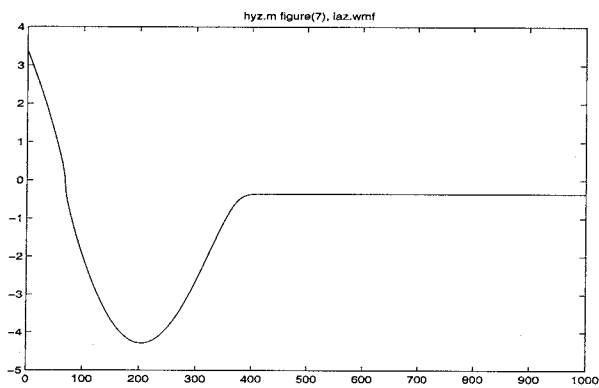


Figure 36: Denominator $\mathbf{x}^T \mathbf{P} \mathbf{b}$

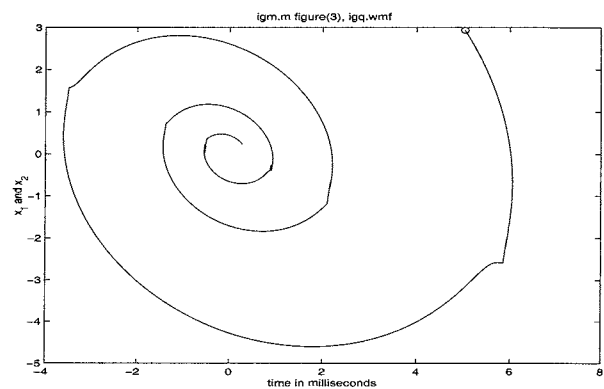


Figure 39: Phase plot $\mathbf{x}(t)$

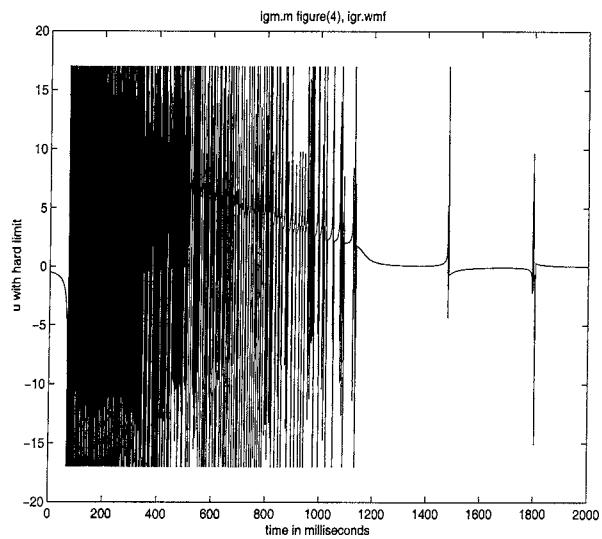


Figure 40: $u(t)$ with hardlimit

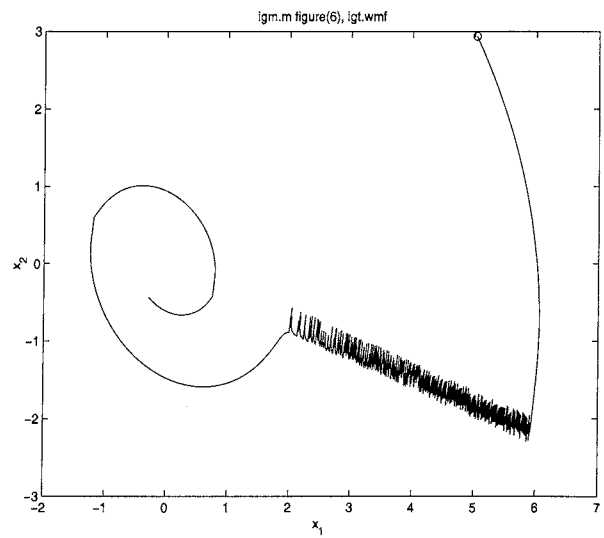


Figure 42: Phase plot x with hardlimit

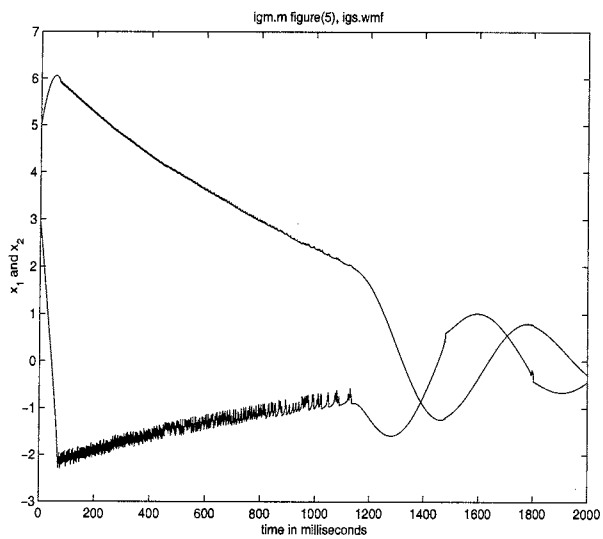


Figure 41: $x(t)$ with hardlimit

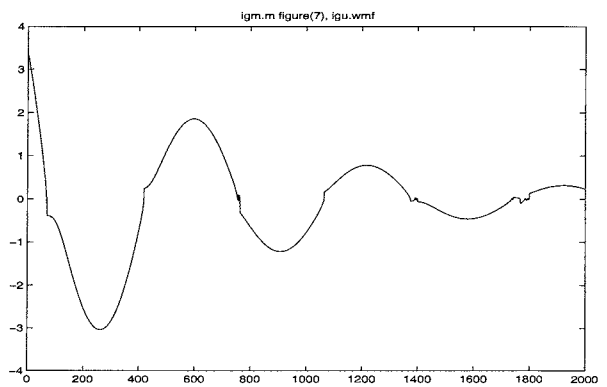


Figure 43: $x^T P b$ versus time

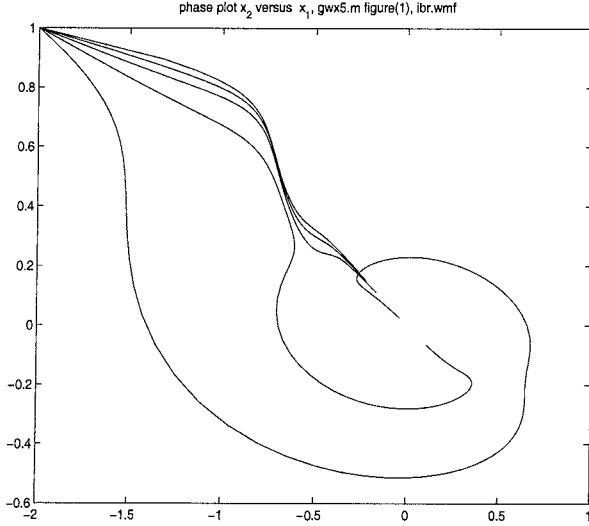


Figure 44:
Phase plot x_2 versus x_1 for Q depending
on x

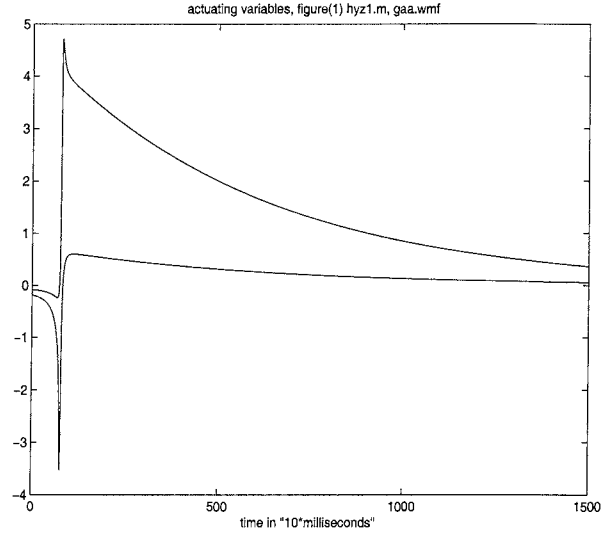


Figure 45: Actuating variables

G Linear MIMO Examples

Example 13. Linear MIMO system: For

$$A = \begin{pmatrix} 0.15 & 1 \\ -1 & -0.3 \end{pmatrix}, \quad B = \begin{pmatrix} 0.13 & 0.5 \\ 1 & 2 \end{pmatrix} \quad (176)$$

$$Q = 0.1I_2, \quad R = 0.7I_2 \text{ (or } 0.7\text{diag}[1 \ 3]) \text{,} \quad (177)$$

the results are given in Figs. 45 up to 47 from $x(0) = (5 \ 3)^T$. The Lyapunov function and its time derivative is included in Fig. 46. The gain limit is $g = 295.2$. $\square$

Example 14. Linear system: Consider a linear system with $G = B$, $f = Ax$ For the same data as before, results are discussed in Figs. 48 up to 51. Since $y \rightarrow y_{ref}$ (or $e \rightarrow 0$) in Fig. 51, the system with this structure and output Eq.(32) can be considered as a control system although x is derived from x and not from e . Moreover, from Eq.(32) $\int_0^\infty c^T x dt$ should meet $y_{ref,f}$. $\square$

Example 15. Linear System. Steady-state oscillation: Modification in u

$$u = \frac{q}{b_1^T R^{-1} b_1} R^{-1} b_1 + v \begin{pmatrix} e \\ 0 \end{pmatrix}. \quad (178)$$

There are interesting results for various assumptions in c_1 , see Table 1.

For $c_1 = (0 \ 0)^T$ in Figs. 52 up to 55 the results are presented. $\square$

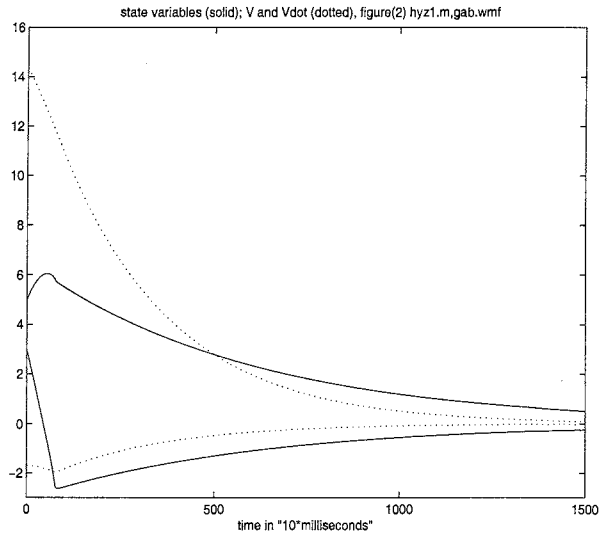


Figure 46: State variables $x(t)$, $V(t)$ and $\dot{V}(t)$

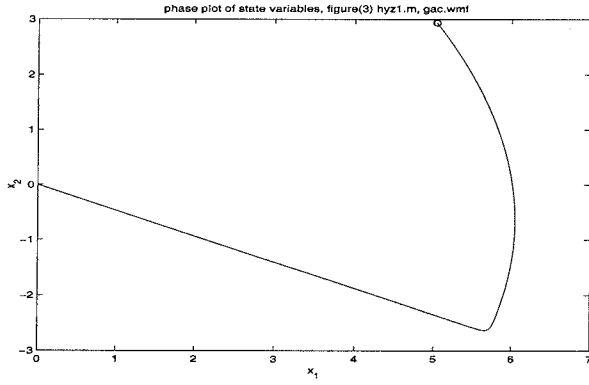


Figure 47: Phase plot

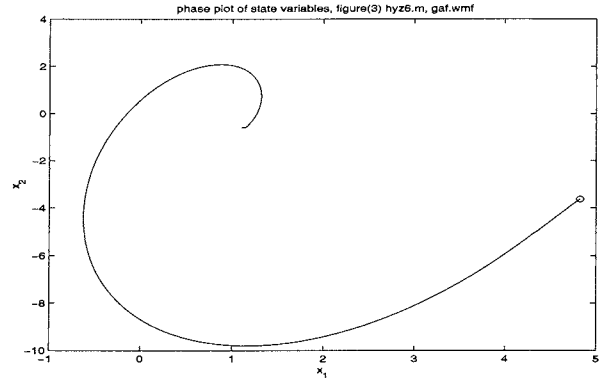


Figure 50: Phase plot

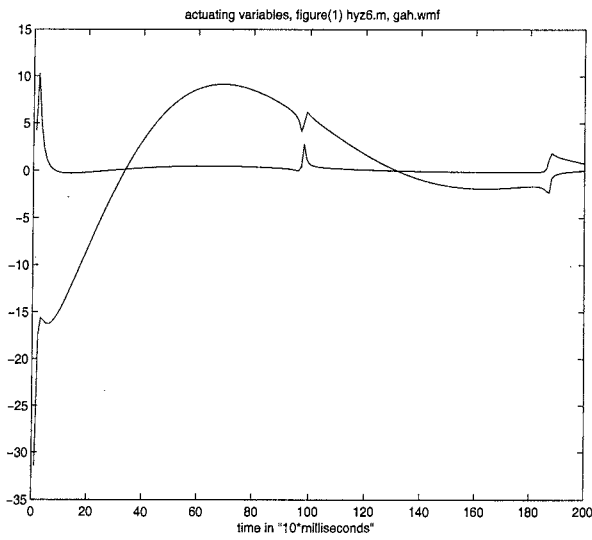


Figure 48: Actuating variables $u(t)$

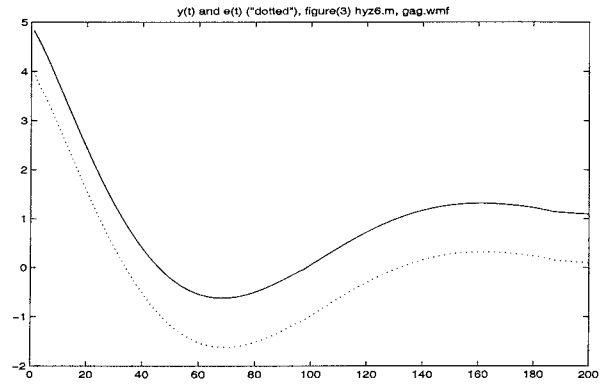


Figure 51: Output $y(t)$ and deviation $e(t)$

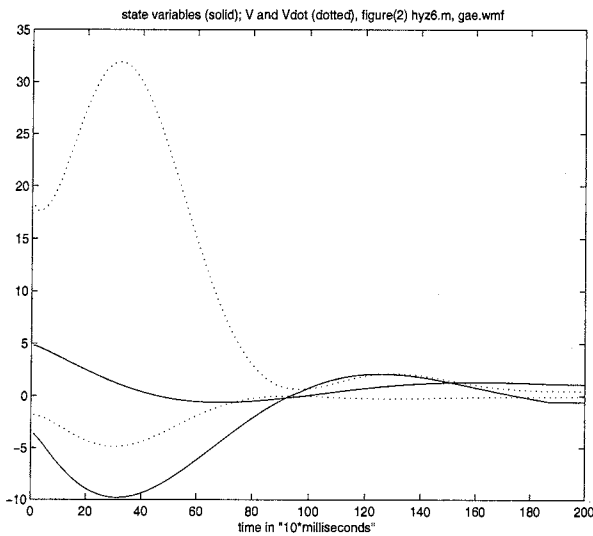


Figure 49: $x(t)$ (solid), $V(t)$ and $\dot{V}(t)$

Table 1: Results for various c_1

c_1	Result
$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$	steady-state oscillation
$\begin{pmatrix} 0.2 \\ 0 \end{pmatrix} \dots \begin{pmatrix} 1.6 \\ 0 \end{pmatrix}$	asymptotically decreasing oscillation
$\begin{pmatrix} 1.8 \\ 0 \end{pmatrix}$	unstable oscillation

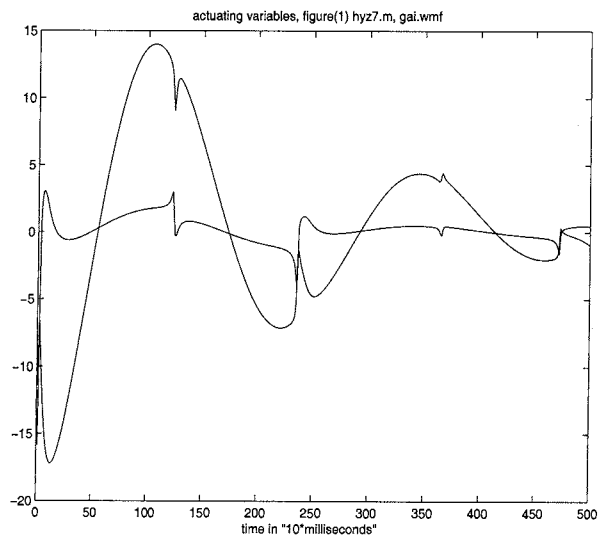


Figure 52: Actuating variables $u(t)$

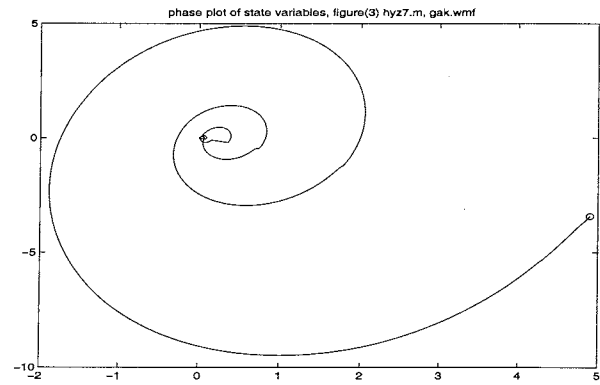


Figure 54: Phase plot with small steady-state oscillation

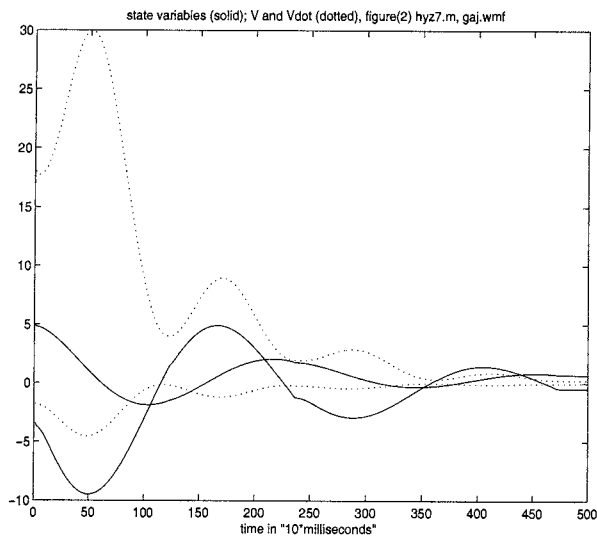


Figure 53: $x(t)$, $V(t)$ and $\dot{V}(t)$

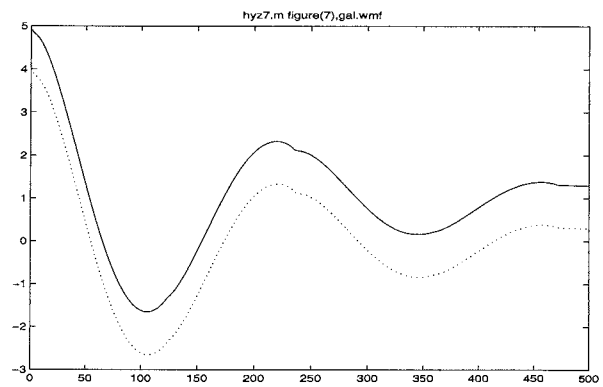


Figure 55: $y(t)$ and $e(t)$ ('dotted')

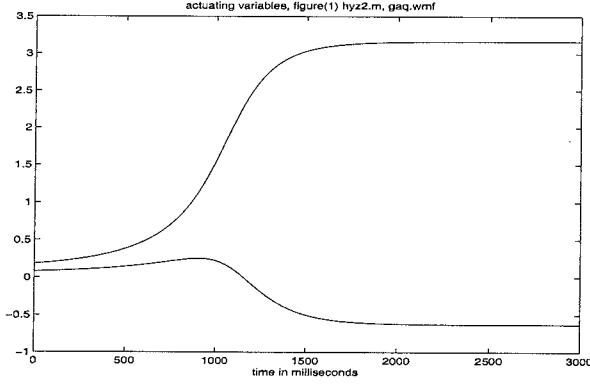


Figure 56: Actuating variable $u(t)$

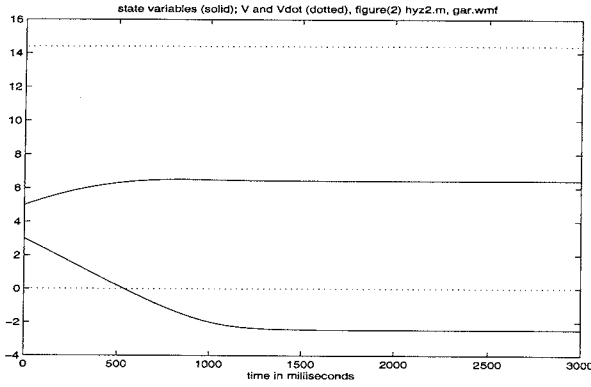


Figure 57: State variable $x(t)$, $V(t)$, and $\dot{V}(t)$

Example 16. Opposite Sign in u :

Selecting the opposite sign for u from Eq.(30) intentionally, there is an interesting but indefinite result

$$\dot{V} = x^T P \dot{x} = x^T (PA - A^T P)x \quad (179)$$

where $V = \text{konstant}$ and $\dot{V}(t) = 0$. Intermediate calculations were omitted. The matrix $PA - A^T P$ is skew-symmetric with zero main diagonal. Results are given in Figs. 56 and 57. The steady-state result $x_f = 0$, resulting from $\dot{x}_f = Ax_f + Bu_f = 0$, is traced back to

$$w_1 = \det\left(A - \frac{x_f^T A^T P x_f}{x_f^T P B R^{-1} B^T P x_f} B R^{-1} B^T P\right) = 0, \quad (180)$$

and is characterized by results very close together, see Fig. 58. $\square$

Example 17. Constrained linear system
Eq.(44): With the setting

$$A = \begin{pmatrix} 0.15 & 1 \\ -1 & -0.2 \end{pmatrix} \quad b = \begin{pmatrix} 0.13 \\ 1 \end{pmatrix} \quad (181)$$

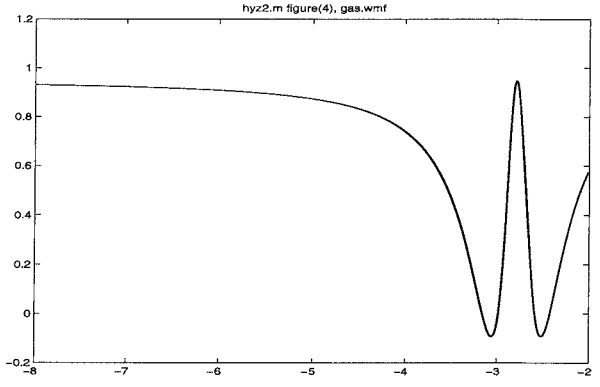


Figure 58: Eq.(180) value w_1 versus x with interrelated components

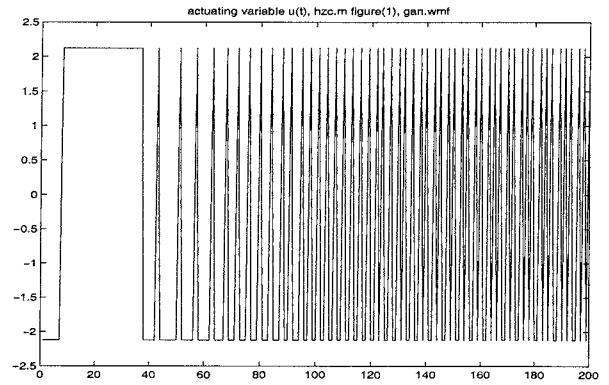


Figure 59: Actuating variable $u(t)$

$$Q = 0.1 I_2, \quad r = 2, \quad U_0 = 3, \quad x(0) = (5 \ 3)^T \quad (182)$$

the results in Figs. 59 to 61 are achieved. $\square$

Example 18. Iso-matricial controller:

For $y = c^T x$ and $v = -1/c^T A^{-1} b$ the results of a second-order example are presented in Figs. 62 to 65. Owing to actuating variable limitation $|u| \leq 1$, there are two derivatives of V : $\dot{V}$ referring to Eq.(59), and $\dot{V}_L$ actually resulting when regarding the limitation of u . $\square$

H Static and Harmonic Reference

Given f , g , c , select an admissible steady-state x_f and u_f (f final).

From $y = c^T x$, $y_{ref,f} = c^T x_f$ and the plant model

$$\dot{x} = f(x) + g(x)u + b y_{ref,f} \quad (183)$$

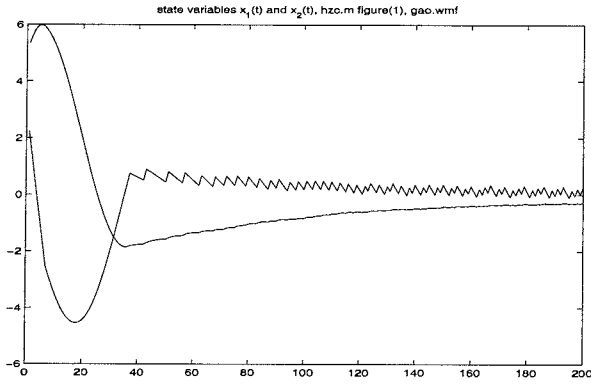


Figure 60: State variable $\mathbf{x}(t)$

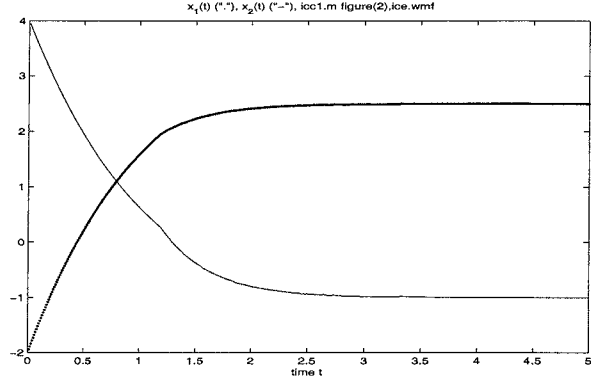


Figure 63: Iso-matricial controller $\mathbf{x}(t)$

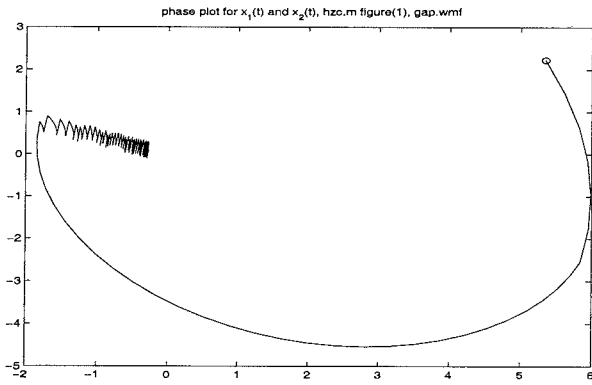


Figure 61: Phase plot

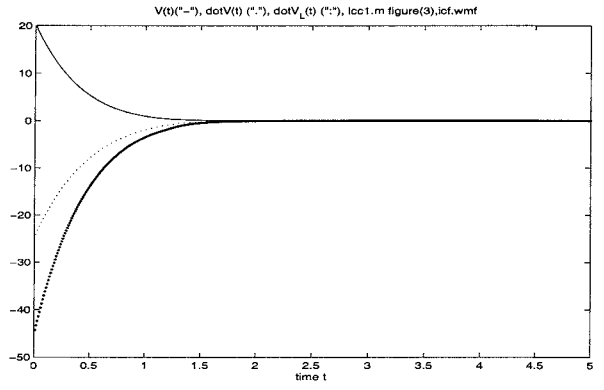


Figure 64: Iso-matricial controller Lyapunov functions $V(t)$, $\dot{V}(t)$, $\dot{V}_L(t)$

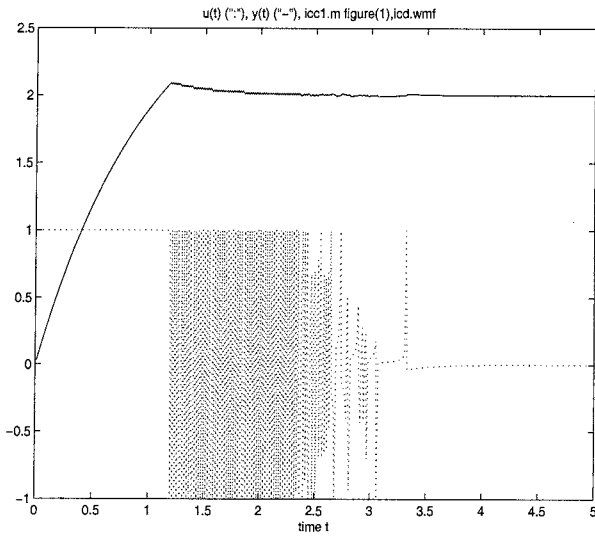


Figure 62: Iso-matricial controller $u(t)$ and $y(t)$

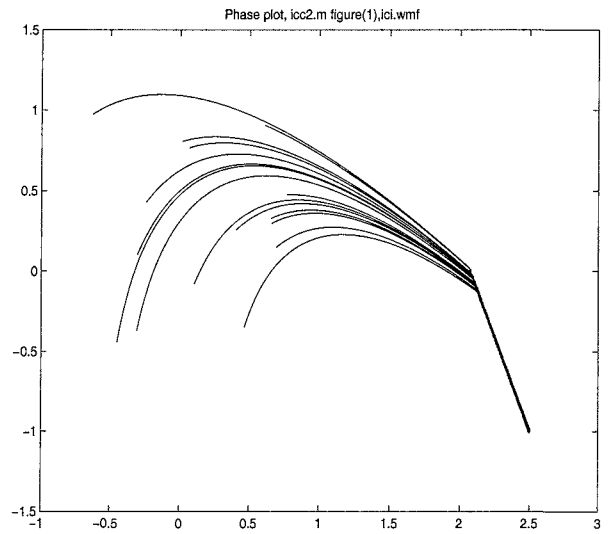


Figure 65: Phase plot $\mathbf{x}(t)$ approaching $\mathbf{x}_f = (2.5; -1)^T$ with $\|\mathbf{x}(0)\|_\infty \leq 1$

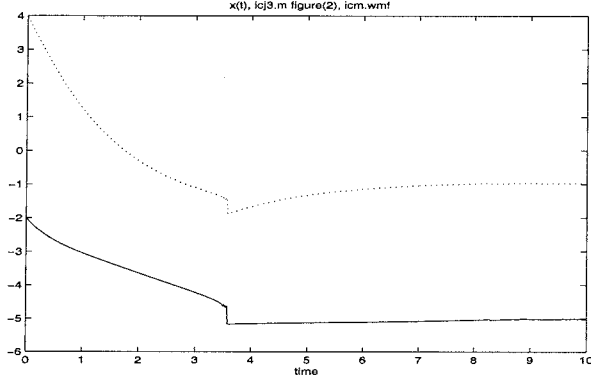


Figure 66: Step response $x(t)$

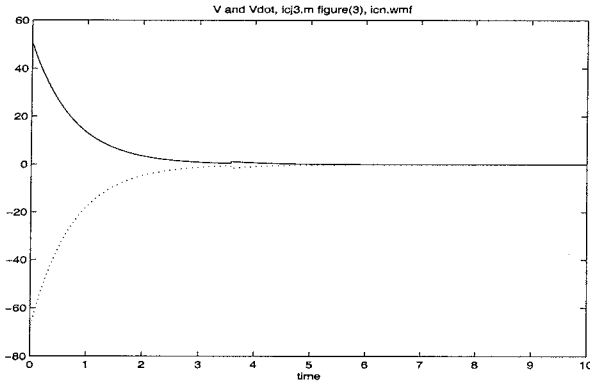


Figure 67: $V(t)$ and $\dot{V}(t)$ versus time

$$\mathbf{0} = \mathbf{f}(\mathbf{x}_f) + \mathbf{g}(\mathbf{x}_f)u_f + \mathbf{b}y_{ref,f} \quad (184)$$

$$\mathbf{b} = -\frac{\mathbf{f}(\mathbf{x}_f) + \mathbf{g}(\mathbf{x}_f)u_f}{\mathbf{c}^T \mathbf{x}_f} \quad (185)$$

Moreover,

$$\dot{V} = (\mathbf{x} - \mathbf{x}_f)^T \mathbf{P}(\mathbf{f} + \mathbf{g}u + \mathbf{b}y_{ref}) = -(\mathbf{x} - \mathbf{x}_f)^T \mathbf{q}(\mathbf{x} - \mathbf{x}_f) \quad (186)$$

$$u = -\frac{(\mathbf{x} - \mathbf{x}_f)^T [\mathbf{q}(\mathbf{x} - \mathbf{x}_f) + \mathbf{P}\mathbf{f} + \mathbf{P}\mathbf{b}y_{ref}]}{(\mathbf{x} - \mathbf{x}_f)^T \mathbf{P}\mathbf{g}} \quad (187)$$

Example 19. Static response: For a nonlinear SISO system of second order the resulting response is given in Figs. 66 and 67. $\square$

Example 20. Sine response: The results of the example above are used for a sine reference y_{ref} . For a preselected $\mathbf{x}_f$ and u_f (see Fig. 68), the static $y_{ref,f}$ is determined which would be required for obtaining the final values $\mathbf{x}_f$ and u_f . Then, $y_{ref,f}$ is replaced by a sine input. The resulting output y , which is a deteriorated sine wave, is recorded in Figs. 69 and 70 versus time and as a phase plot, respectively. $\square$

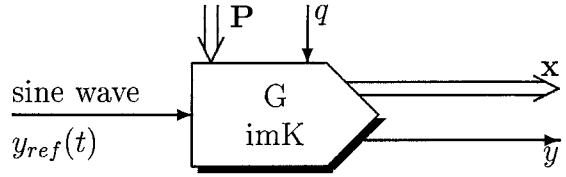
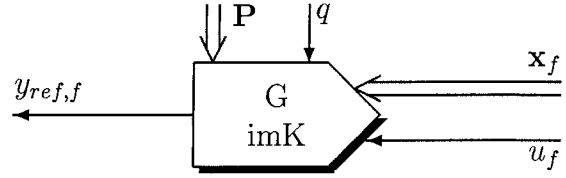
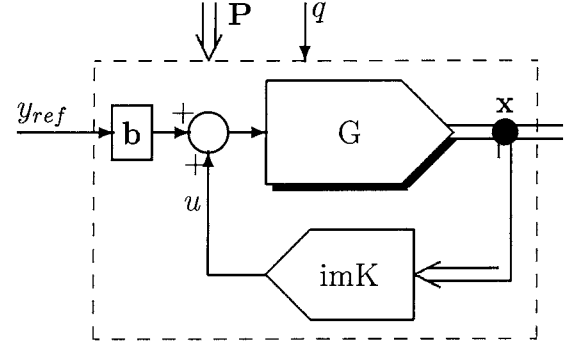


Figure 68: Design of the iso-matrical controller imK by preselecting final values $\mathbf{x}_f$ and u_f and driving the closed-loop system by a sinusoidal y_{ref}

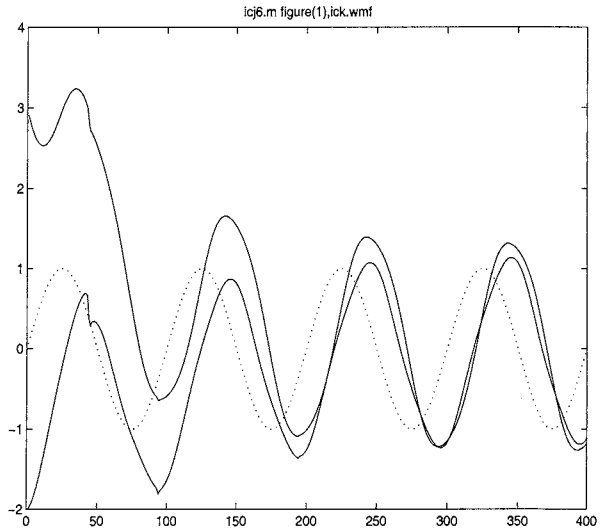


Figure 69: Sine wave y_{ref} and $x(t)$ versus time

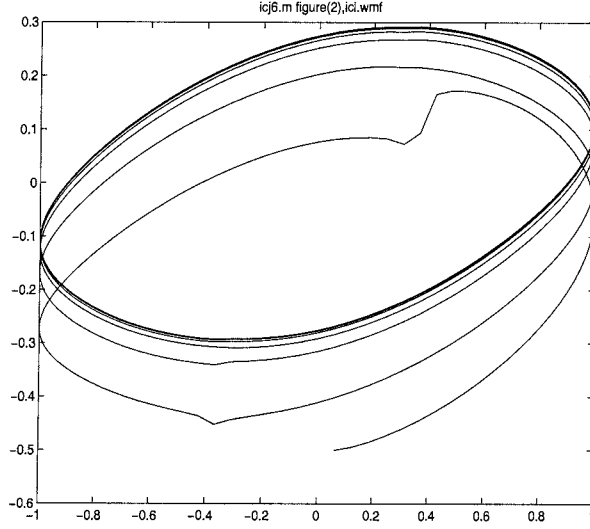


Figure 70: Phase plot y_{ref} versus y

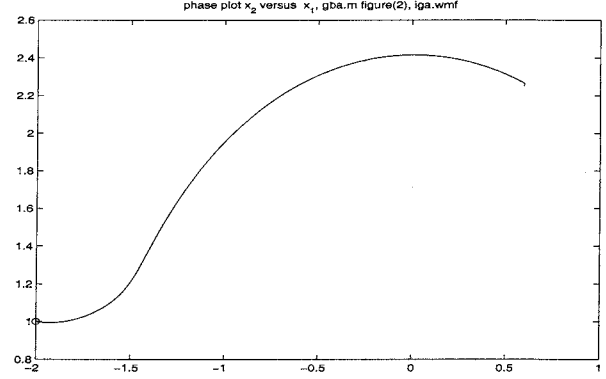


Figure 71: Phase plot $\mathbf{x}$ for semidefinite $\dot{V}$

I Passivity function

Assuming $V = \mathbf{x}^T \mathbf{P} \mathbf{x}$ and $\dot{V} \triangleq -0.5q_0(\mathbf{y}^T \mathbf{u})^2$, which replaces $-\mathbf{x}^T \mathbf{Q} \mathbf{x}$. For the output $\mathbf{y} = \mathbf{h}(\mathbf{x})$ one has

$$\mathbf{x}^T \mathbf{P} \mathbf{f} + \mathbf{x}^T \mathbf{P} \mathbf{G} \mathbf{u} = -0.5q_0(\mathbf{y}^T \mathbf{u})^2 = -0.5q_0 \mathbf{y}^T \mathbf{u} \mathbf{u}^T \mathbf{y}. \quad (188)$$

This expression serves as a condition for $0.5\mathbf{u}^T \mathbf{R} \mathbf{u}$ by using the Lagrange multiplier μ . By $\partial/\partial \mathbf{u}$

$$\mathbf{R} \mathbf{u} + \mu \mathbf{b}_1 + \mu q_0 \mathbf{y} \mathbf{y}^T \mathbf{u} = \mathbf{0} \quad (189)$$

$$\mathbf{u} = -(\mathbf{R}/\mu + q_0 \mathbf{y} \mathbf{y}^T)^{-1} \mathbf{G}^T \mathbf{P} \mathbf{x}. \quad (190)$$

Combining with Eq.(188) yields μ and the final $\mathbf{u}$. An analytically concise solution does not exist.

Example 21. Passivity function: For the settings of the example with Eqs.(167) and (168) the resulting phase plot is Fig. 71. Owing to the fact that $\dot{V}$ is only semidefinite, V tends to a constant value although $\dot{V}$ goes to zero, see Fig. 72.

Even though $\mathbf{x}$ approaches a steady-state rather quickly, the actuating signal $\mathbf{u}$ may become big and only decreases slowly. The Lagrange factor μ changes its value unexpectedly.

By adding a term $-q_1$ to $\dot{V}$, asymptotic stability is guaranteed (for $q_1 > 0$). $\square$

J Robust Controller Examples

Example 22. Maximum actuating variable:

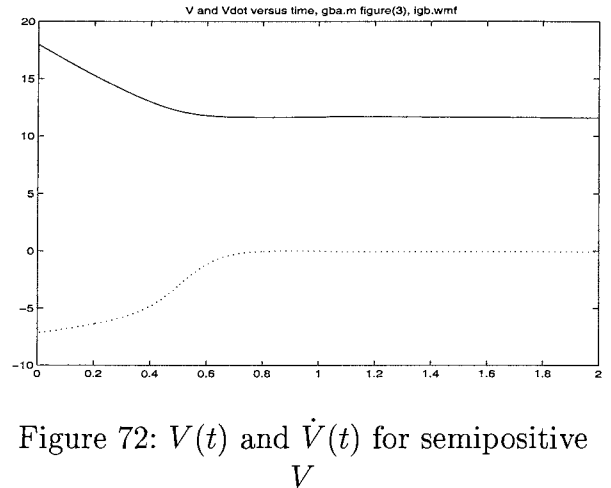


Figure 72: $V(t)$ and $\dot{V}(t)$ for semipositive V

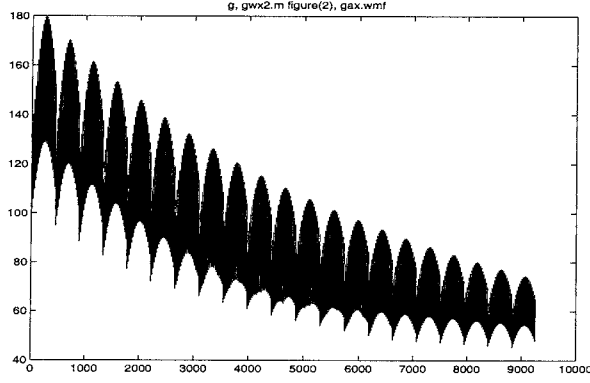


Figure 73: Factor g for variation in x_1, x_2, x_3 , for 21 assumptions each

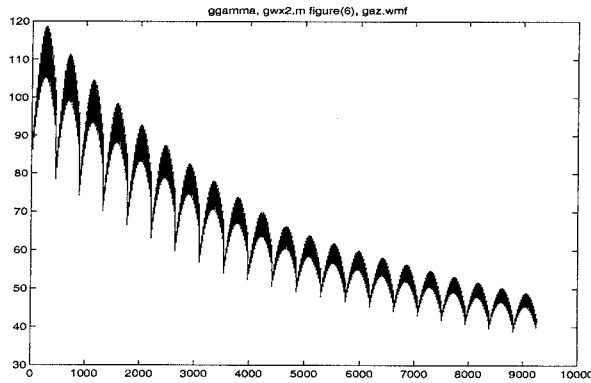


Figure 74: Factor g_γ for variation in x_1, x_2, x_3 , for 21 assumptions each

For special data, following Eqs.(90) to (92)

$$g(\mathbf{x}) = (h_{1var} + h_{2var})/h_{3var} \quad (191)$$

$$g_\gamma(\mathbf{x}) = \max_{\gamma} \sigma_{\max}[\mathbf{P}\mathbf{F}_{0x} + \mathbf{Q} + \gamma\mathbf{P}\mathbf{F}_{1x}] \times \|\mathbf{R}^{-1}\mathbf{G}^T\mathbf{P}\|_F \times \sigma_{\min}^{-1}[\mathbf{P}\mathbf{G}\mathbf{R}^{-1}\mathbf{G}^T\mathbf{P}]. \quad (192)$$

The coefficients g and g_γ are depicted in Fig. 73 and Fig. 74, respectively. There is a remarkable difference in their maxima

$$\max_{\mathbf{x}} g(\mathbf{x}) = 179.8 \quad \text{and} \quad \max_{\mathbf{x}} g_\gamma(\mathbf{x}) = 118.6. \quad \square \quad (193)$$

Example 23. Worst Case Selection:

Consider a plant with $\mathbf{f}(\mathbf{x})$ deteriorated by harmonic disturbance. The actuating variable is determined as the extreme value selected from all possible ones. This process is repeated each discrete-time computation of u . For a second-order SISO example, the results are given in Fig. 75. For comparison,

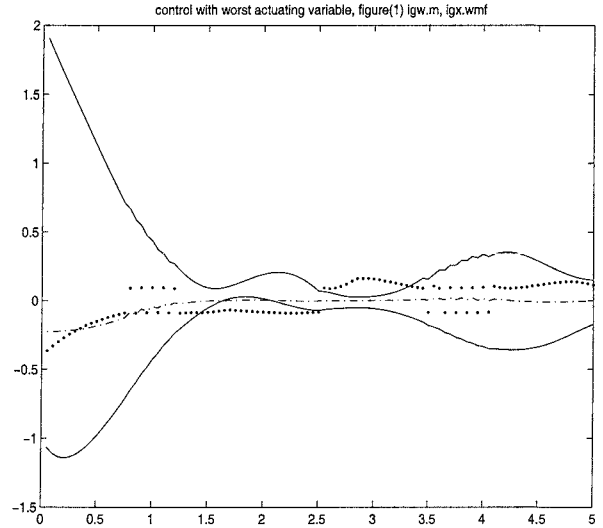


Figure 75: $x_1(t), x_2(t), u(t)$ ('dotted'), $0.02\dot{V}(t)$ ('dash dot') for the worst selection

in Fig. 76 the original u referring to Eq.(166) is depicted.

Example 24. Influence of $\mathbf{Q}$:

Varying α for $\mathbf{Q} = \alpha\mathbf{I}$, the maximum actuating variable $\max_{\mathbf{x}} u(t)$ is depicted in Fig. 77, based on Eq.(166) and on $\mathbf{f}(\mathbf{x})$ disturbed by noise in $\mathbf{x}$. $\square$

Example 25. Permissible uncertainty parameter: Consider, for simplicity, $\Delta\mathbf{f} = \alpha\mathbf{x}$ where $\alpha > 0$ is an uncertainty parameter, then from Eq. (106)

$$\alpha \leq \mathbf{x}^T(\mathbf{Q} - \mathbf{Q}_1)\mathbf{x}/(\mathbf{x}^T\mathbf{P}\mathbf{x}). \quad (194)$$

For randomly perturbed plant $\Delta\mathbf{f} = \mathbf{r}\mathbf{x}$, with random r and $0 < r < 1$, the results are depicted in Figs. 78 through 80. The region below the characteristic in Fig. 80 represents the permissible uncertainty parameter. $\square$

K Auxiliary Linear Controller

Consider $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})u$ again, with subdivided $u = u_1 + u_2$ and $u_1 = \mathbf{k}^T\mathbf{x}$, then

$$u_2 = -\frac{\mathbf{x}^T\mathbf{P}(\mathbf{f} + \mathbf{g}\mathbf{k}^T\mathbf{x}) + \mathbf{x}^T\mathbf{Q}\mathbf{x}}{\mathbf{x}^T\mathbf{P}\mathbf{g}}. \quad (195)$$

Select $\mathbf{k}$ such that for $u_2 = 0$

$$\dot{V}_1 = \mathbf{x}^T\mathbf{P}(\mathbf{f} + \mathbf{g}\mathbf{k}^T\mathbf{x}) < 0. \quad (196)$$

For the example with Eqs.(8) and (9), $\dot{V}_1$ is depicted in Fig. 81 with '•' for Eq.(196) satisfied, and 'o' not satisfied. $\square$

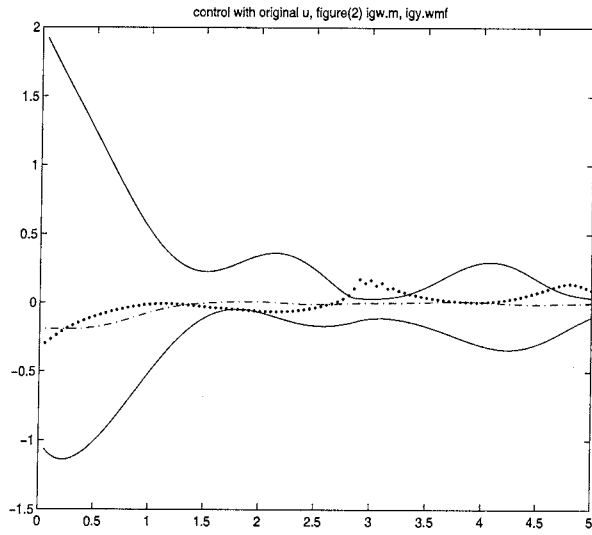


Figure 76: $x_1(t)$, $x_2(t)$, $u(t)$ ('dotted'), $0.02\dot{V}(t)$ ('dash dot')

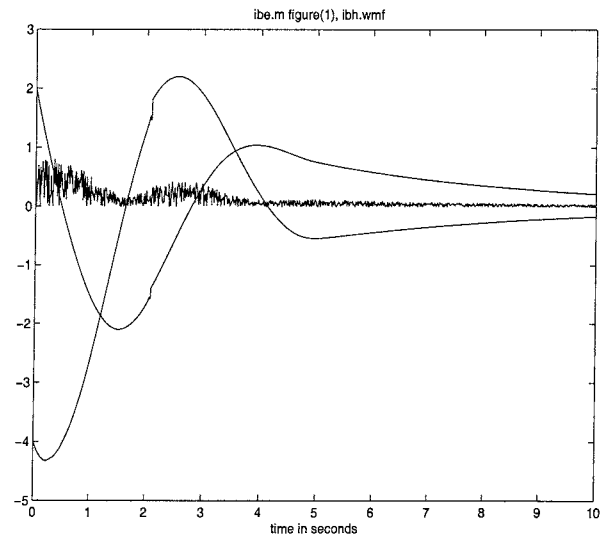


Figure 78: Perturbed $\mathbf{x}(t)$ and uncertainty signal norm $r\alpha\|\mathbf{x}\|_F(t)$

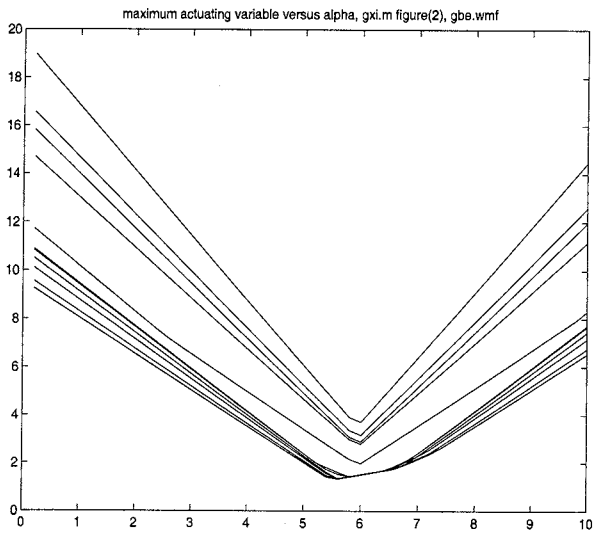


Figure 77: Maximum $\max_{\mathbf{x}} u(t)$

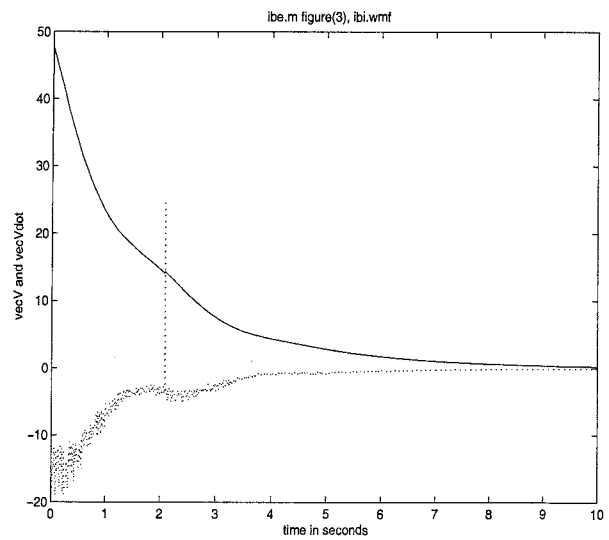


Figure 79: $V(t)$ and $\dot{V}(t)$

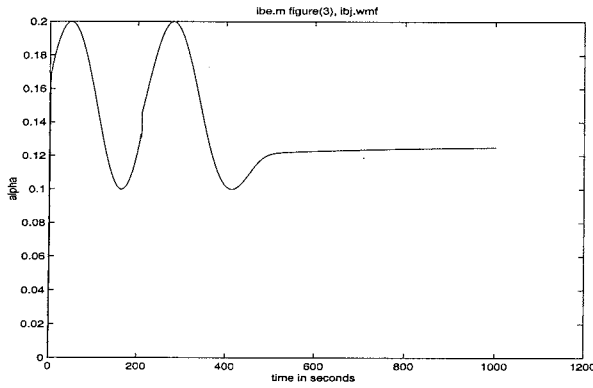


Figure 80: Permissible α as a function of time

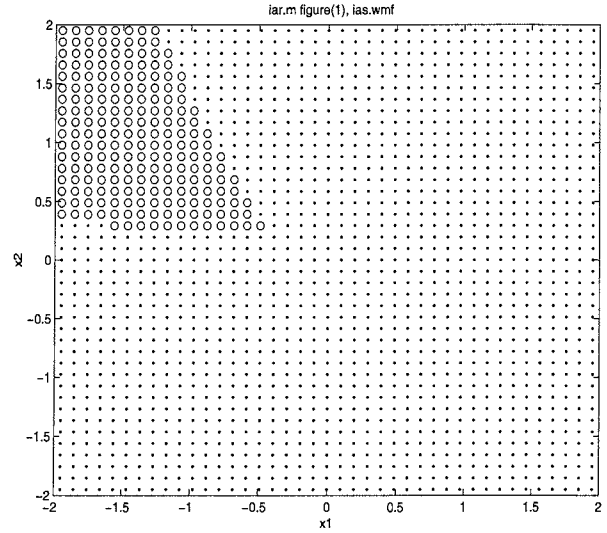


Figure 82: Feasible points for u_1

Example 26. Unstable plant:

For an unstable nonlinear plant with auxiliary controller $u_1 = \mathbf{k}^T \mathbf{x}$, in Fig. 82 the feasible points satisfying Eq.(196) are shown. In Fig. 83 the state variables $\mathbf{x}$ and u_2 are depicted, with an initial value not violating the feasible points of Fig. 82. The actuating variable enters a sliding mode at $t \doteq 5.5$.

Including absolute value in the denominator, the result is Fig. 84, with no sliding and quicker responses. $\square$

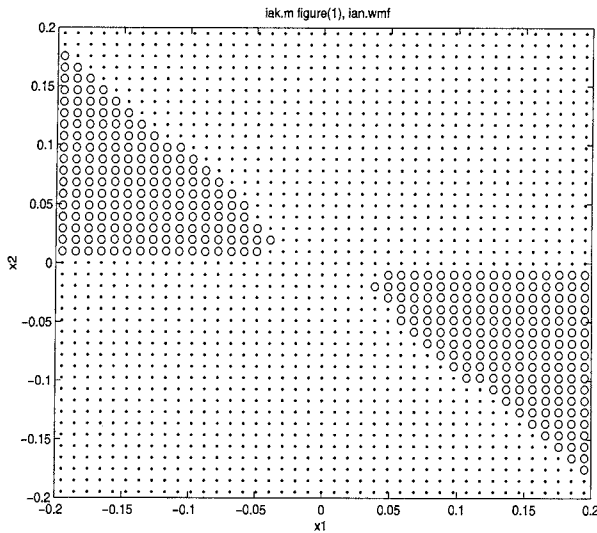


Figure 81: (x_1, x_2) -plane with regions for $\dot{V}_1 < 0$ satisfied ('●') or not ('o')

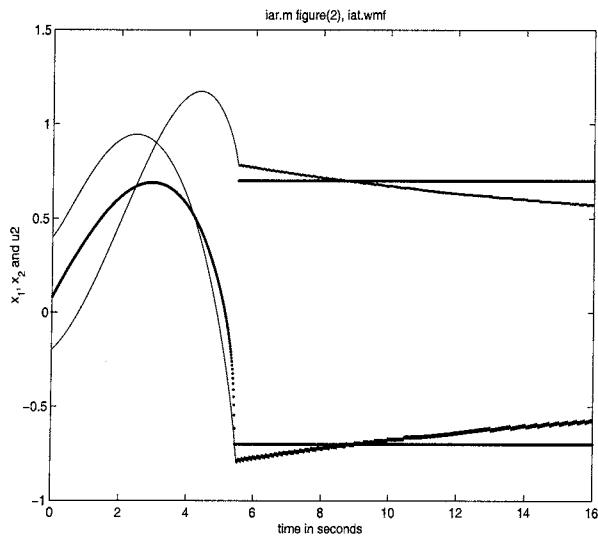


Figure 83: State variables and controlling variable u_2 for an unstable plant

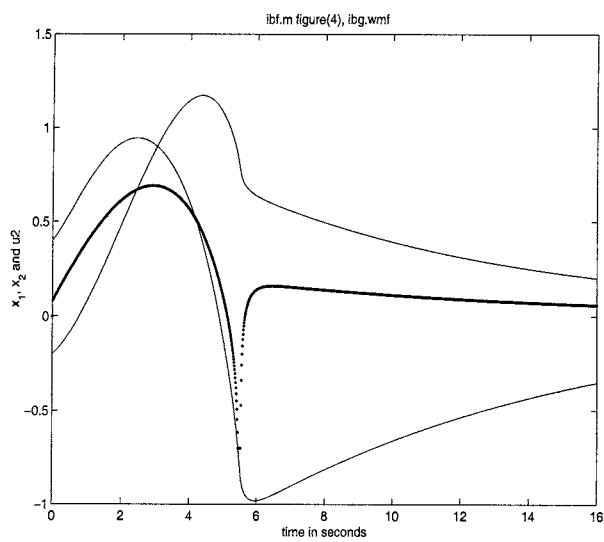


Figure 84: State variables and controlling variable u_2 for an unstable plant with absolute value denominator

Linearisierung von Lautsprechern mit AC-Verstärkern

D. Fränken, Universität Paderborn

Eine exakte Linearisierung elektrodynamischer Lautsprecher mittels statischer Zustandsrückführung erfordert in vielen Fällen den Einsatz von DC-Verstärkern. Dieser Beitrag widmet sich der Frage, wie eine linearisierende Rückführung durchzuführen ist, wenn lediglich ein AC-Verstärker zur Verfügung steht.

Einführung

Speziell beim Einsatz von Tieftonlautsprechern kleiner Membranfläche z. B. im Kfz-Bereich zeigen sich prinzipbedingte Schwächen elektrodynamischer Wandler: die zur Erzeugung ausreichender Schalldruckpegel nötigen großen Auslenkungen führen zu nicht mehr vernachlässigbaren nichtlinearen Verzerrungen. Es existiert eine Vielfalt von Ansätzen, diese Verzerrungen zu reduzieren. Ein Ansatz ist die Anwendung des bekannten Verfahrens der exakten Linearisierung [1], wobei die dynamischen Nichtlinearitäten des Lautsprechers durch geeignete Vorverzerrung des Verstärkereingangssignals kompensiert werden [2].

Um das linearisierende Signal berechnen zu können, ist die Kenntnis eines mathematischen Modells erforderlich, welches die wesentlichen dynamischen Eigenschaften des Lautsprechers hinreichend genau beschreibt. Für viele praktische Anwendungen ist der Einsatz eines Modells dritter Ordnung ausreichend, dessen Zustandsgrößen durch die Membranauslenkung, die Membrangeschwindigkeit sowie den Schwingspulenstrom gegeben sind [3]. Setzt man nun noch voraus, dass der Verstärker eine über den relevanten Frequenzbereich konstante Verstärkung aufweist, so kann das linearisierende Signal auf einfache Weise durch statische Zustandsrückführung ermittelt werden. Die dazu benötigten Momentanwerte der Zustandsgrößen werden in der hier betrachteten Anwendung mithilfe eines diskreten Zustandsschätzers näherungsweise bestimmt, um eine aufwendige Sensorik speziell auf der mechanischen Seite des Lautsprechers zu vermeiden.

In diesem Beitrag wird zur Beschreibung des Lautsprechers ein äquivalentes elektrisches Netzwerk verwendet, welches insbesondere die Passivität des Lautsprechers als inhärente Eigenschaft widerspiegelt [4, 5]. Damit ist dieses Modell selbst unmittelbar als Zustandsbeobachter einsetzbar¹.

In der Praxis zeigt sich nun aber, dass eine auf den genannten Voraussetzungen beruhende Rückführung oft nicht zu der erwarteten Verringerung nichtlinearer Verzerrungen führt. Je nach verwendetem Lautsprecher ist unter Umständen sogar eine Zunahme dieser Verzerrungen zu verzeichnen. Hauptgrund hierfür ist die Tatsache, dass handelsübliche Verstärker in der Regel als so genannte *AC-Verstärker* ausgelegt sind. Diese sind im Gegensatz zu *DC-Verstärkern* nicht in der Lage, einen Gleichanteil im linearisierenden Signal zu übertragen, wie er sich speziell als Folge unsymmetrischer Kennlinien des Lautsprechers unter der Annahme einer konstanten Verstärkung ergibt.

Im Folgenden wird dieser Tatsache durch Erweiterung der Zustandsbeschreibung um ein einfaches Modell des Verstärkers Rechnung getragen. Zwar kann für das erweiterte Modell for-

¹Eine Diskretisierung eines konkret passiven elektrischen Netzwerkes kann z. B. mithilfe der Theorie der Wellendigitalfilter stattfinden. Der interessierte Leser sei hier auf die Literatur verwiesen [6, 7, 8].

mal eine linearisierende Zustandsrückführung angegeben werden, doch wegen unzureichender Stabilitätseigenschaften des resultierenden Systems erweist sich auch dieser Ansatz als noch nicht praktikabel. Abhilfe schafft hier eine modifizierte Rückführung, die zugunsten gesicherter Stabilität auf eine vollständige Linearisierung verzichtet.

1 Netzwerkmodell des Lautsprechers und Zustandsschätzer

Bei einem dynamischen Lautsprecher ist eine Membran mit einer Schwingspule verbunden, die mittels einer flexiblen Aufhängung im Luftspalt eines Magnetsystems positioniert wird. Ein Stromfluss durch die Schwingspule bewirkt eine Beschleunigung von Schwingspule und Membran als Folge der LORENTZ-Kraft. Wesentlich verantwortlich für nichtlineare Effekte sind bei dem Lautsprecher eine ortsabhängige Eigeninduktivität der Schwingspule aufgrund einer sich ändernden Überdeckung mit dem Eisenkern, ein inhomogenes Konstantmagnetfeld sowie eine mit wachsender Auslenkung progressiv zunehmende Steifigkeit von Aufhängung und Sicke. Dagegen können der Schwingspulenwiderstand, die Masse sowie der mechanische effektive Reibungskoeffizient als konstant angenommen werden.

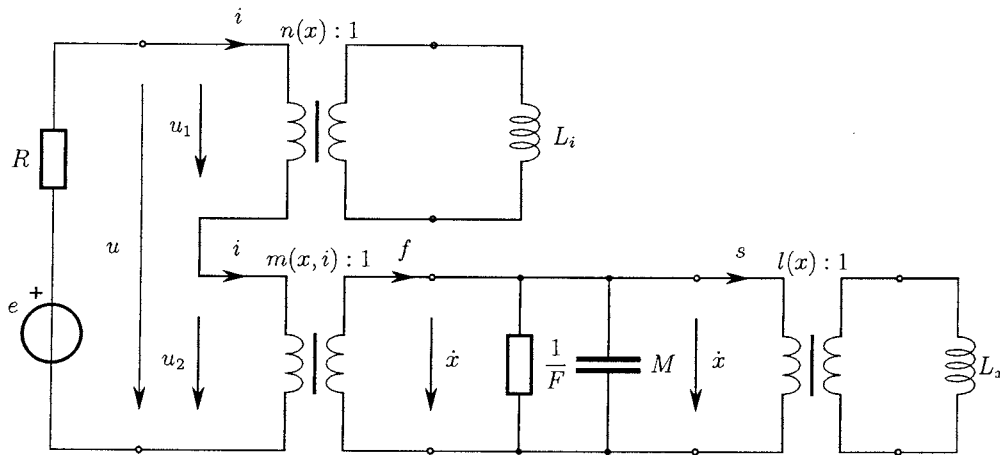


Abbildung 1: Äquivalentes Netzwerkmodell eines elektrodynamischen Lautsprechers

Abbildung 1 zeigt ein äquivalentes elektrisches Netzwerk, welches sowohl die elektrische als auch die mechanische Seite des Lautsprechers in geeigneter Weise widerspiegelt und dessen Zustandsgrößen durch die Membranauslenkung x , die Membrangeschwindigkeit $\dot{x}$ sowie den Schwingspulenstrom i gegeben sind [4, 5]. Auf der elektrischen Seite ist neben einer resistiven Quelle, die zusammenfassend Verstärkerausgang und ohmsche Verluste der Schwingspule repräsentiert, die auslenkungsabhängige positive Induktivität $L(x)$ zu finden. Die gewählte Darstellung mittels einer konstanten Induktivität $L_i > 0$ in Verbindung mit einem idealen *energieneutralen* Übertrager mit Übersetzungsverhältnis $n(x) := \sqrt{L(x)/L_i}$ mag zunächst ungewohnt erscheinen, ist aber zweckmäßig, da hieraus die *Verlustfreiheit* des Elementes Induktivität mit der gespeicherten *magnetischen Energie*

$$E_{\text{mag}} = \frac{1}{2} L(x) i^2 = \frac{1}{2} L_i [n(x) i]^2 \quad (1)$$

unmittelbar erkennbar ist. Konsequenterweise wird die Kopplung zwischen elektrischer und mechanischer Seite, die ihrerseits im Idealfall energieneutral stattfindet, durch einen idealen

Übertrager dargestellt, der die Spannung u_2 und den Strom i auf der einen Seite sowie die Geschwindigkeit v und die auf die Membran wirkende Kraft f auf der anderen Seite miteinander verbindet. Das zugehörige Übersetzungsverhältnis setzt sich gemäß

$$m(x, i) = A(x) + \frac{i}{2} \frac{dL(x)}{dx} \quad (2)$$

additiv zusammen aus dem *Kraftfaktor* $A(x)$ zwischen Spulenstrom und LORENTZ-Kraft sowie dem Faktor zwischen Strom und der durch die Auslenkungsabhängigkeit der Eigeninduktivität bedingten *Reluktanzkraft*². Der mechanische Teil des Netzwerkes wird durch das NEWTONsche Bewegungsgesetz in der Form $M\ddot{x} = f - F\dot{x} - s$ mit der Masse M , dem Reibungskoeffizienten F sowie der Rückstellkraft s beschrieben. Damit fungiert die Masse als konstante „Kapazität“, die mit der Geschwindigkeit $\dot{x}$ assoziierte *kinetische Energie* ist durch

$$E_{\text{kin}} = \frac{1}{2} M \dot{x}^2 \quad (3)$$

in gewohnter Weise gegeben. Die Rückstellkraft s wird aber durch die auslenkungsabhängige Steifigkeit $K(x) > 0$ gemäß $s(x) = K(x)x$ bestimmt. Zur Darstellung dieses nichtlinearen Zusammenhanges wird wiederum eine konstante „Induktivität“ $L_x > 0$ in Verbindung mit einem idealen Übertrager herangezogen, das zugehörige Übersetzungsverhältnis $l(x)$ ist durch die gespeicherte *potentielle Energie*

$$E_{\text{pot}} = \int_0^x s(\xi) d\xi = \int_0^x K(\xi)\xi d\xi = \frac{1}{2} L_x [l(x)s(x)]^2 = \frac{1}{2} L_x [l(x)k(x)x]^2 \quad (4)$$

festgelegt.

Vervollständigt wird das angegebene elektrische Netzwerk durch die Angabe geeigneter Kennlinien für die die Übersetzungsverhältnisse $l(x)$, $m(x, i)$ und $n(x)$ bestimmenden Größen Steifigkeit, Kraftfaktor und Induktivität. Die in der Literatur übliche Approximation mittels Polynomen kommt dabei nur eingeschränkt in Betracht, da ja eine Haupteigenschaft des Lautsprechers seine Passivität ist und somit Steifigkeit und Induktivität stets positive Werte annehmen müssen. Zwar gilt diese Einschränkung für den Kraftfaktor nicht aus Passivitätsgründen, in der Regel wird ein Lautsprecher aber so ausgelegt sein, dass ein positiver Stromfluss auch eine Membranbeschleunigung in positiver Richtung bewirkt. Daher erscheint auch hier die Forderung nach ausschließlich positiven Werten sinnvoll.

Während die tendentiell zu (dem Betrage nach) größeren Auslenkungen progressiv zunehmende Steifigkeit durch

$$K(x) = k_0 + k_1[x - x_k]^2 \quad (5)$$

mit positiven Koeffizienten k_0 und k_1 in geeigneter Weise repräsentiert werden kann, ist der entsprechend abnehmende Kraftfaktor nicht wie üblich durch eine nach unten geöffnete Parabel, sondern sinnvollerweise durch einen Ausdruck der Form

$$A(x) = \frac{a_0}{1 + a_1[x - x_a]^2} \quad (6)$$

²Der sich ergebende Ausdruck für die Reluktanzkraft wurde zuerst von CUNNINGHAM unter Ausnutzung energetischer Betrachtungen hergeleitet [9]. Daher erscheint es erstaunlich, dass zu ihrer Modellierung in der Literatur in der Regel eine gesteuerte (Strom-)Quelle herangezogen wird.

mit positiven Werten a_0 und a_1 zu approximieren. Da mit in positiver Richtung steigender Auslenkung der „wirksame“ Eisenkern immer kürzer wird, ist für die Induktivität eine streng monoton fallende Funktion anzusetzen, die der Bedingung $L(-\infty) > L(\infty) > 0$ genügt. Mit einer rein rationalen Funktion ist dieser Anforderung nicht zu genügen, ein geeigneter Ansatz ist aber durch

$$L(x) = l_0 + l_1 \left[1 - \frac{x - x_\ell}{l_2 + |x - x_\ell|} \right] \Rightarrow L'(x) = \frac{-l_1 l_2}{(l_2 + |x - x_\ell|)^2} \quad (7)$$

mit positiven Werten l_0 , l_1 und l_2 gegeben.

2 Zustandsraumbeschreibung des Lautsprechers

Ausgangspunkt der weiteren Betrachtungen ist eine Zustandsraumdarstellung des Lautsprechers, die dem angegebenen äquivalenten elektrischen Netzwerk entnommen werden kann. Mit der Festlegung $\mathbf{z} := [x, v, i]^T$ erhält man

$$\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z}, e) = \mathbf{a}(\mathbf{z}) + \mathbf{b}(\mathbf{z})e = \begin{bmatrix} a_1(\mathbf{z}) \\ a_2(\mathbf{z}) \\ a_3(\mathbf{z}) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ b_3(\mathbf{z}) \end{bmatrix} e, \quad (8)$$

wobei

$$\begin{aligned} a_1(\mathbf{z}) &= z_2, & a_2(\mathbf{z}) &= \frac{1}{M} [Az_3 + \frac{z_3^2}{2} L' - Kz_1 - Fz_2], \\ a_3(\mathbf{z}) &= -\frac{1}{L} [Rz_3 + L'z_2z_3 + Az_2], & b_3(\mathbf{z}) &= \frac{1}{L} \end{aligned} \quad (9)$$

gilt. Zur Verkürzung der Schreibweise wurde hier, wie auch im Folgenden, auf eine explizite Angabe des Argumentes bei den auslenkungsabhängigen Größen L , A und K sowie den zugehörigen Ableitungen (z. B. L') verzichtet, soweit nicht spezielle Werte gemeint sind.

Als unmittelbare Konsequenz der Tatsache, dass das in Abbildung 1 gezeigte Modell ausschließlich passive Bauelemente enthält, ist die Ruhelage $\mathbf{0}$ des verwendeten nichtlinearen Lautsprechermodells stabil. Dies lässt sich mit Hilfe der positiv definiten Energiefunktion

$$V(\mathbf{z}) = E_{\text{pot}} + E_{\text{kin}} + E_{\text{mag}} = \int_0^{z_1} K(\xi) \xi \, d\xi + \frac{1}{2} M z_2^2 + \frac{1}{2} L(z_1) z_3^2 \quad (10)$$

nachweisen. Die Berechnung von $\dot{V}$ unter Berücksichtigung der Systemgleichungen liefert nämlich mit

$$\dot{V}(\mathbf{z}, e) = e z_3 - R z_3^2 - F z_2^2 \quad (11)$$

für $e = 0$ eine negativ semidefinite Funktion. Aufgrund der für das Modell gültigen Implikation $\dot{V}(\mathbf{z}, 0) \equiv 0 \Rightarrow \mathbf{z} = \mathbf{0}$ ist die Ruhelage sogar asymptotisch stabil. Diese Aussage kann auch durch Betrachtung des um seine Ruhelage Null linearisierten freien Modells (dieses wird im Weiteren kurz als zugehöriges lineares Modell bezeichnet)

$$\dot{\mathbf{z}} = \mathbf{A} \mathbf{z} \quad \text{mit} \quad \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ -K_0/M & -F_0/M & A_0/M \\ 0 & -A_0/L_0 & -R/L_0 \end{bmatrix} \quad (12)$$

bestätigt werden, wobei hier abkürzend $K_0 = K(0)$, $L_0 = L(0)$ bzw. $F_0 = F(0)$ geschrieben wurde. Das zugehörige charakteristische Polynom,

$$\Delta(p) = p^3 + \frac{MR + F_0L_0}{ML_0}p^2 + \frac{FR + A_0^2 + K_0L_0}{ML_0}p + \frac{RK_0}{ML_0}, \quad (13)$$

ist ein HURWITZ-Polynom, da eine Kettenbruchentwicklung von

$$Z(p) := \frac{\Delta(p) - \Delta(-p)}{\Delta(p) + \Delta(-p)} \quad (14)$$

ausschließlich positive Koeffizienten aufweist.

3 Linearisierung mit DC-Verstärkern

Der Lautsprecher werde nun von einem Verstärker mit Eingangssignal u und Ausgangssignal e getrieben, der zunächst eine konstante Verstärkung G besitzen möge. Um eine vollständig linearisierende Zustandsrückführung für ein Modell der Form (8) zu gewinnen, werden üblicherweise sukzessive die zeitlichen Ableitungen eines geeignet gewählten Ausgangssignals $y = g(z)$ bestimmt. Bei dem verwendeten Lautsprechermodell erweist es sich als zweckmäßig, hierfür die Auslenkung zu verwenden. Damit erhält man

$$y = g(z) = z_1 =: g_0(z), \quad (15a)$$

$$\dot{y} = g'_0(z)f(z, e) = a_1(z) =: g_1(z), \quad (15b)$$

$$\ddot{y} = g'_1(z)f(z, e) = a_2(z) =: g_2(z), \quad (15c)$$

$$\ddot{\ddot{y}} = g'_2(z)f(z, e) =: g_3(z) + h_3(z)e, \quad (15d)$$

wobei g'_ν den (als Zeilenvektor aufzufassenden) Gradienten von g_ν bezeichnet und

$$g_3(z) = \frac{\partial g_2}{\partial z_1}g_1(z) + \frac{\partial g_2}{\partial z_2}g_2(z) + \frac{\partial g_2}{\partial z_3}a_3(z), \quad h_3(z) = \frac{\partial g_2}{\partial z_3}b_3(z) \quad (15e)$$

mit

$$\frac{\partial g_2}{\partial z_1} = \frac{1}{M}[A'z_3 + \frac{z_3^2}{2}L'' - K'z_1 - K], \quad \frac{\partial g_2}{\partial z_2} = -\frac{F}{M}, \quad \frac{\partial g_2}{\partial z_3} = \frac{1}{M}[A + z_3L'] \quad (15f)$$

gilt. Offenbar hängen hier weder das Ausgangssignal selber noch seine erste und zweite zeitliche Ableitung explizit von e ab, während die dritte Ableitung affin in e ist.

Nun wird ein Signal gesucht, welches das Gesamtsystem linearisiert, also zu einem Eingangs-/Ausgangsverhalten der Form

$$\sum_{\nu=0}^3 \lambda_\nu y^{(\nu)} = Gv \quad \text{mit} \quad \lambda_3 \neq 0 \quad (16)$$

führt. Aus den Gleichungen (15) und (16) erhält man zunächst

$$\sum_{\nu=0}^3 \lambda_\nu g_\nu(z) + \lambda_3 h_3(z)Gu = Gv \quad (17)$$

und damit, sofern $h_3(z)$ verschieden von Null ist, das linearisierende Signal u zu

$$u = \frac{v - \Gamma(z)}{\lambda_3 h_3(z)} \quad \text{mit} \quad \Gamma(z) := \frac{1}{G} \sum_{\nu=0}^3 \lambda_\nu g_\nu(z). \quad (18)$$

Bekanntermaßen stabilisiert die Rückführung (18) die Ruhelage $z = \mathbf{0}$ des freien Systems immer dann, wenn die Parameter λ_0 bis λ_3 als Koeffizienten eines HURWITZ-Polynoms gewählt werden. So kann die resultierende Zustandsraumdarstellung mit dem transformierten Zustandsvektor $\mathbf{g} = [g_0, g_1, g_2]^T$ und der Begleitmatrix $\mathbf{A}_B$ in der Form

$$\dot{\mathbf{g}} = \mathbf{A}_B \mathbf{g} \Rightarrow \dot{z} = \varphi(z) \quad \text{mit} \quad \varphi(z) := \mathbf{G}(z)^{-1} \mathbf{A}_B \mathbf{g}(z) \quad \text{und} \quad \mathbf{G}(z) := \frac{\partial \mathbf{g}(z)}{\partial z} \quad (19)$$

geschrieben werden, wobei $\mathbf{G}(z)$ die JACOBI-Matrix der Abbildung $z \rightarrow \mathbf{g}(z)$ bezeichnet. Mit

$$\frac{\partial \varphi(z)}{\partial z_\nu} = \mathbf{G}(z)^{-1} \mathbf{A}_B \frac{\partial \mathbf{g}(z)}{\partial z_\nu} - \mathbf{G}(z)^{-1} \frac{\partial \mathbf{G}(z)}{\partial z_\nu} \mathbf{G}(z)^{-1} \mathbf{A}_B \mathbf{g}(z) \quad (20)$$

liefert wegen $\mathbf{g}(\mathbf{0}) = \mathbf{0}$ eine Linearisierung dieses Modells um die Ruhelage den Ausdruck

$$\dot{z} = \mathbf{G}(\mathbf{0})^{-1} \mathbf{A}_B \mathbf{G}(\mathbf{0}) z \quad (21)$$

und damit ein asymptotisch stabiles Verhalten.

Die Wahl der Entwurfsparameter λ_ν kann also prinzipiell (sofern sie Koeffizienten eines HURWITZ-Polynoms bleiben) frei erfolgen. Oft wird man die Parameter aber so wählen, dass sich gerade das Übertragungsverhalten des linearen Modells (8) ergibt. Dies bedeutet aber, dass $\mathbf{A} = \mathbf{G}(\mathbf{0})^{-1} \mathbf{A}_B \mathbf{G}(\mathbf{0})$ gilt³. Andererseits gehört zu

$$\dot{z} = \begin{bmatrix} a_1(z) \\ a_2(z) \\ a_3(z) - b_3(z) G d(z) \end{bmatrix} \quad \text{mit} \quad d(z) := \frac{\Gamma(z)}{\lambda_3 h_3(z)}. \quad (22)$$

wegen $\Gamma(\mathbf{0}) = 0$ das lineare Modell

$$\dot{z} = \begin{bmatrix} a'_1(\mathbf{0}) \\ a'_2(\mathbf{0}) \\ a'_3(\mathbf{0}) - b_3(\mathbf{0}) G d'(\mathbf{0}) \end{bmatrix} z \quad \text{mit} \quad \begin{bmatrix} a'_1(\mathbf{0}) \\ a'_2(\mathbf{0}) \\ a'_3(\mathbf{0}) \end{bmatrix} = \mathbf{A} \quad (23)$$

sowie

$$d'(\mathbf{0}) = \frac{\frac{1}{G} \sum_{\nu=0}^3 \lambda_\nu g'_\nu(\mathbf{0})}{\lambda_3 h_3(\mathbf{0})} \quad \text{und} \quad h_3(\mathbf{0}) = \frac{A_0}{M L_0}. \quad (24)$$

Damit erhält man in dem hier betrachteten Sonderfall speziell

$$d'(\mathbf{0}) = \mathbf{0}^T. \quad (25)$$

Auf dieses Ergebnis wird später noch zurückgegriffen.

³Für die Zeilenvektoren $\mathbf{g}'_\nu(z)$ der JACOBI-Matrix $\mathbf{G}(z)$ gilt

$$\mathbf{g}'_{\nu+1}(z) = \frac{\partial \mathbf{g}'_\nu(z) \mathbf{a}(z)}{\partial z} \quad \text{mit} \quad \frac{\partial \mathbf{g}'_\nu(z) \mathbf{a}(z)}{\partial z_\mu} = \frac{\partial \mathbf{g}'_\nu(z)}{\partial z_\mu} \mathbf{a}(z) + \mathbf{g}'_\nu(z) \frac{\partial \mathbf{a}(z)}{\partial z_\mu}$$

und wegen $\mathbf{a}(\mathbf{0}) = \mathbf{0}$ insbesondere $\mathbf{g}'_{\nu+1}(\mathbf{0}) = \mathbf{g}'_\nu(\mathbf{0}) \mathbf{A}$. Damit stimmt $\mathbf{G}(\mathbf{0})$ mit der Beobachtbarkeitsmatrix $\mathbf{B}_y$ des um den Ausgangsvektor $\mathbf{c}^T = \mathbf{g}'_\nu(\mathbf{0})$ ergänzten freien Modells (12) überein, die gerade zur Transformation der Matrix $\mathbf{A}$ auf die zugehörige Begleitform herangezogen werden kann.

4 Exakte Linearisierung mit AC-Verstärkern

Im Gegensatz zur Annahme des vorigen Abschnitts soll ab jetzt ein AC-Verstärker den Lautsprecher treiben. Es hat sich gezeigt, dass für die hier im Vordergrund stehende Anwendung die folgende Modellierung des AC-Verstärkers als Hochpass erster Ordnung mit einer asymptotischen Verstärkung G und einer Zeitkonstanten τ ausreicht:

$$\tau \dot{e} + e = G\tau \dot{u} \quad \text{bzw.} \quad \tau \dot{w} = -w + u, \quad e = G[-w + u]. \quad (26)$$

Das Ziel besteht nun darin, ein linearisierendes Eingangssignal u für das Gesamtsystem zu finden, welches aus einer Kaskadierung des Verstärkers und des nichtlinearen Lautsprechers hervorgeht. Mit Gleichung (26) wird aus Gleichung (16)

$$\sum_{\nu=0}^3 \lambda_{\nu} g_{\nu}(z) + \lambda_3 h_3(z) G[u - w] = Gv, \quad (27)$$

und ein linearisierendes Eingangssignal ist formal durch

$$u = \frac{v - \Gamma(z)}{\lambda_3 h_3(z)} + w \quad (28)$$

gegeben.

An dieser Stelle erscheint es aber widersprüchlich, dass ein eventuell vorhandener Gleichanteil, der ja prinzipiell nicht durch den Verstärker übertragen werden kann, aufgrund der Linearisierung nach Gleichung (16) auf einmal doch am Ausgang des Gesamtsystems auftreten sollte. Tatsächlich ist die Übertragungsfunktion des gemäß Gleichung (28) linearisierten Modells nach wie vor von der Ordnung drei, während das Gesamtsystem insgesamt vier Zustandsgrößen besitzt. Man ist hier also mit dem Problem einer (nicht beobachtbaren) so genannten *Nullodynamik* konfrontiert. Im Falle eines für $v = 0$ identisch verschwindenden Ausgangssignals erhält man durch Auswertung der Gleichungen $g_i(z) = 0$ für $i = 0, 1, 2$ nacheinander jeweils $z_{i+1} = 0$, und die Gleichungen (26) und (28) liefern mit $\Gamma(\mathbf{0}) = 0$ die Nullodynamik

$$\tau \dot{w} = \frac{\Gamma(\mathbf{0})}{\lambda_3 h_3(\mathbf{0})} = 0. \quad (29)$$

Erwartungsgemäß tritt in der Nullodynamik ein Eigenwert Null auf, der mit der durch den Verstärker vorgegebenen Übertragungsnullstelle übereinstimmt. Damit kann auf Stabilitätsaussagen, denen eine asymptotisch stabile Nullodynamik zugrunde liegt, hier nicht zurückgegriffen werden.

Simulationsergebnisse bestätigen, dass sich die angegebene Nullodynamik für die betrachteten Modelle als kritisch erweist. Zwar findet tatsächlich über einen gewissen Zeitraum die gewünschte Eingangs-/Ausgangslinearisierung statt, doch weist das Signal u in der Regel einen linearen Trend auf, der speziell im Hinblick auf eine angestrebte Implementierung auf einem DSP und des damit verbundenen begrenzten Umsetzersignals nicht toleriert werden kann.

5 Ein näherungsweise linearisierender Ansatz

Da die angesprochene, nicht asymptotisch stabile Nullodynamik bei einer *rein statischen Zustandsrückführung* unter Einsatz von AC-Verstärkern prinzipiell auftreten wird, soll im Fol-

genden von dem Versuch Abstand genommen werden, eine *exakte* Eingangs-/Ausgangslinearisierung durchzuführen. Es wird vielmehr versucht, eine *näherungsweise* linearisierende Zustandsrückführung derart anzugeben, dass die genannten Probleme, insbesondere das Auftreten eines linearen Trendes im Signal u , vermieden werden.

Als Ansatzpunkt dient hier die dem System (26) eigene Übertragungsfunktion

$$H(p) = G \frac{p\tau}{1 + p\tau}. \quad (30)$$

Zur Bestimmung einer geeigneten Zustandsrückführung soll diese nun formal durch die Übertragungsfunktion

$$H(p) = G \frac{\alpha + p\tau}{1 + p\tau} \quad \text{mit} \quad 0 < \alpha < 1 \quad (31)$$

ersetzt werden. Diese Näherung ist so lange hinreichend genau, wie die wesentlichen Spektralanteile des Signals u deutlich oberhalb der Kreisfrequenz α/τ liegen, und lässt sich gerade dadurch erzielen, dass in der Ausgangsgleichung von (26) der Ausdruck w durch $w[1 - \alpha]$ substituiert wird. Konsequenterweise wird anstelle von (27) damit eine Zustandsrückführung der Form

$$u = \frac{v - \Gamma(z)}{\lambda_3 h_3(z)} + w[1 - \alpha] \quad (32)$$

angesetzt.

Ein Nulldynamik tritt mit diesem Ansatz nicht mehr auf. War im vorigen Abschnitt für $v = 0$ die Gleichung $\ddot{y} = 0$ in Verbindung mit Gleichung (15d) durch die Wahl der Rückführung konstruktionsbedingt automatisch erfüllt, so liefert sie nun unter Berücksichtigung des tatsächlichen Verstärkerausgangssignals $e = G(u - w)$ in Verbindung mit der Rückführung (32) die zusätzliche Bedingung

$$0 = -\alpha G \lambda_3 h_3(z) w \quad \Rightarrow \quad w = 0. \quad (33)$$

Um Stabilitätsaussagen für die Ruhelage Null zu ermöglichen, werde die Zustandsraumdarstellung des freien Modells betrachtet:

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \\ \dot{w} \end{bmatrix} = \begin{bmatrix} a_1(z) \\ a_2(z) \\ a_3(z) - b_3(z)G[d(z) + \alpha w] \\ -[d(z) + \alpha w]/\tau \end{bmatrix} \quad \text{mit} \quad d(z) := \frac{\Gamma(z)}{\lambda_3 h_3(z)}. \quad (34)$$

Das zugehörige lineare Modell lautet

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \\ \dot{w} \end{bmatrix} = \begin{bmatrix} a'_1(0) & 0 \\ a'_2(0) & 0 \\ a'_3(0) - b'_3(0)Gd'(0) & -G\alpha \\ -d'(0)/\tau & -\alpha/\tau \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ w \end{bmatrix}. \quad (35)$$

Wie zuvor seien die Parameter λ_ν als Koeffizienten eines HURWITZ-Polynoms gewählt, weiterhin gelte die Einschränkung $\alpha > 0$. Im Gegensatz zu Abschnitt 3 gewährleistet diese Vorgehensweise alleine jedoch nicht immer die asymptotische Stabilität der Ruhelage. Ob diese

gegeben ist, kann im Einzelfall anhand des charakteristischen Polynoms $\Delta(p)$ bzw. der daraus nach Gleichung (14) gebildeten Funktion $Z(p)$ überprüft werden, wobei die Beziehungen

$$\begin{aligned} g'_0(\mathbf{0}) &= [1, 0, 0], \quad g'_1(\mathbf{0}) = a'_1(\mathbf{0}), \quad g'_2(\mathbf{0}) = a'_2(\mathbf{0}), \\ g'_3(\mathbf{0}) &= \frac{1}{M^2 L_0} [F K_0 L_0, L_0 F_0^2 - M K_0 L_0 - M A_0^2, -A_0 F_0 L_0 - R M A_0] \end{aligned} \quad (36)$$

zu verwenden sind. Dennoch sollte der hier vorgestellte Ansatz über weite praktisch relevante Bereiche hin einsetzbar sein. Wählt man nämlich die Koeffizienten an dieser Stelle wieder so, wie sie durch (13) vorgegeben sind, so folgt ja gemäß Gleichung (25) die Beziehung $\mathbf{d}'(\mathbf{0}) = \mathbf{0}^T$, und die Stabilität ist für positive α offenbar stets gesichert. Die Frage, wie weit man von dieser Vorgehensweise abweichen kann, ohne die Stabilität zu gefährden, bleibt weiteren Untersuchungen vorbehalten.

6 Ein Simulationsbeispiel

Dass der dargestellte Ansatz tatsächlich zur Linearisierung elektrodynamischer Lautsprecher mit AC-Verstärkern eingesetzt werden kann, soll an dieser Stelle anhand eines Simulationsbeispiels belegt werden. Abbildung 2 zeigt die Kenngrößen des zur Simulation herangezogenen

Steifigkeit			Kraftfaktor		
k_0	1775	N/m	a_0	5,92	N/A
k_1	0,093	N/mm ³	a_1	0,04	1/mm ²
x_k	1,89	mm	x_a	-0,75	mm
Induktivität			Widerstand		
l_0	0,4	mH	R	4,137	Ω
l_1	0,58	mH			
l_2	3	mm			
x_ℓ	-0,7	mm			
Reibungskoeffizient			Masse		
F	1,3988	Ns/m	M	15	g
Verstärkung			Grenzfrequenz		
G	1		f_g	1	Hz

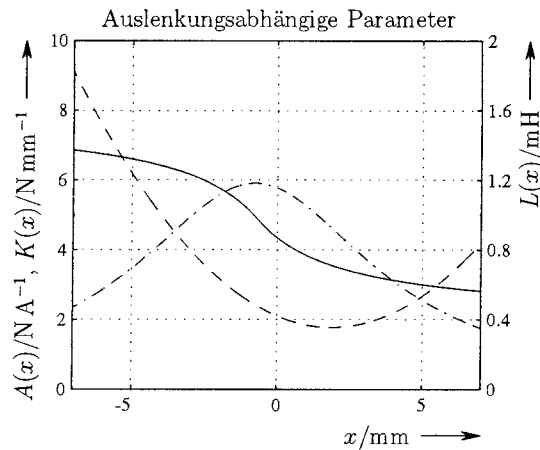


Abbildung 2: Modellparameter sowie die Verläufe von Steifigkeit (gestrichelte Linie), Kraftfaktor (Strichpunktlinie) und Induktivität (durchgezogene Linie)

Verstärkers und des Tieftonlautsprechers sowie die zugehörigen Verläufe von Steifigkeit, Kraftfaktor und Induktivität. Die ausgeprägten Unsymmetrien der Kennlinien sind bedingt durch die großen Werte x_k , x_a und x_ℓ .

Das Modell aus Verstärker und Lautsprecher wurde exemplarisch für ein aus zwei sinusförmigen Spannungen der Amplitude 8 V mit den Frequenzen $f_1 = 50$ Hz und $f_2 = 70$ Hz bestehendes Eingangssignal v und verschiedene Verstärkereingangssignale u simuliert. Abbildung 3 zeigt jeweils die Spektren der Auslenkung sowie der mit dem Schalldruck in Verbindung stehenden Membranbeschleunigung. Ohne Rückführung ergeben sich die gezeigten nichtlinearen Verzerrungen, die auch durch eine Ansteuerung nicht reduziert werden, bei der Zustandsbeobachtung und -rückführung auf der Annahme einer konstanten Verstärkung beruhen. Für eine Beobachtung und Rückführung gemäß Gleichung (32) wurde hier der Parameter

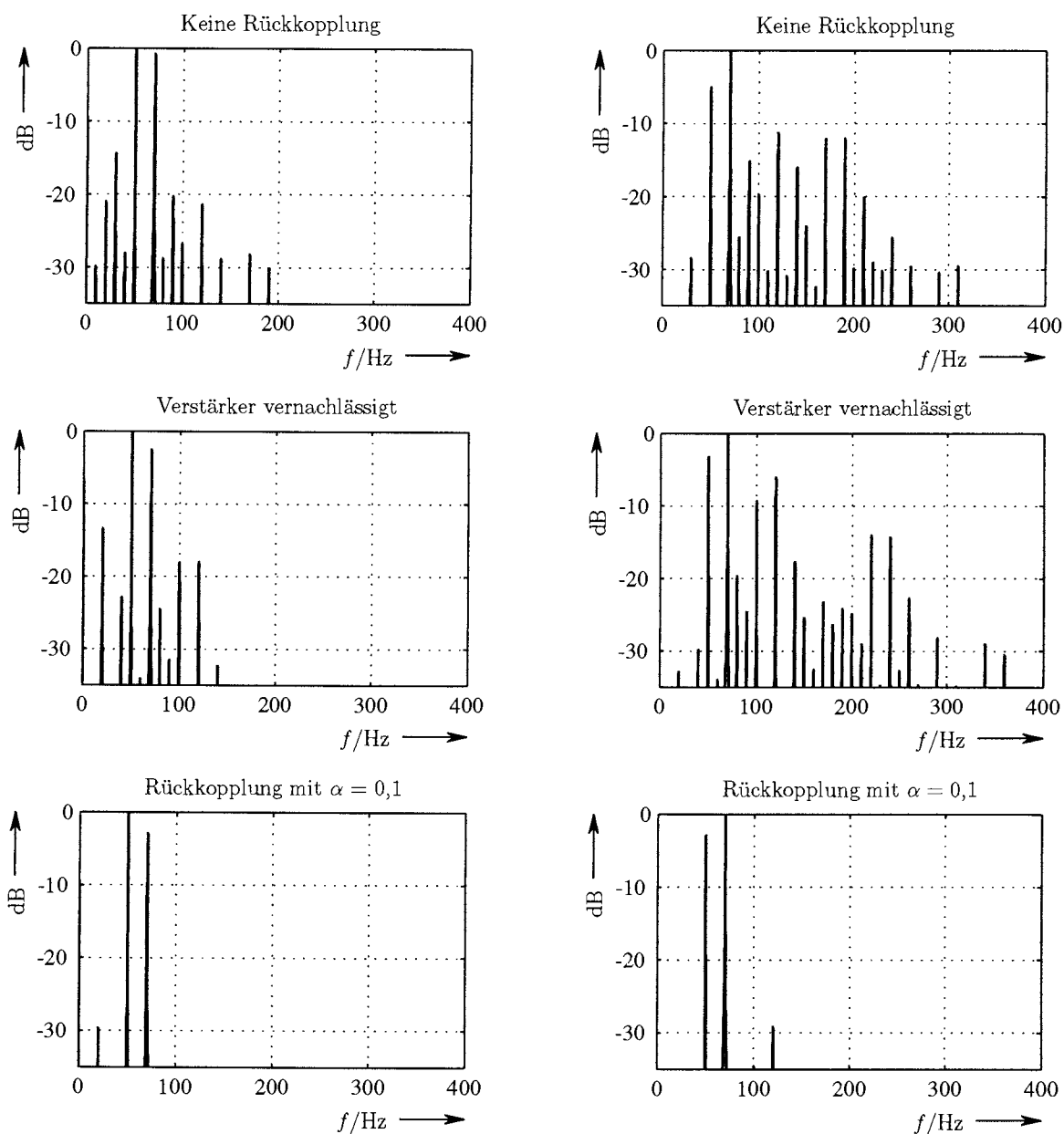


Abbildung 3: Normierte Spektren von Auslenkung (links) und Membranbeschleunigung (rechts) für eine Anregung durch zwei sinusförmige Spannungen der Amplitude 8 V mit $f_1 = 50 \text{ Hz}$ und $f_2 = 70 \text{ Hz}$

$\alpha = 0,1$ gewählt, im Vergleich zu den anderen Fällen ergeben sich deutliche Absenkungen der Verzerrungen.

7 Fazit

Ein Ansatz zur Verringerung nichtlinearer Verzerrungen von elektrodynamischen Lautsprechern unter Einsatz von AC-Verstärkern wurde vorgestellt. Aufgrund fehlender Stabilitätseigenschaften war mittels statischer Zustandsrückführung eine exakte Linearisierung nicht möglich. Mit dem eingeführten Parameter α kann, auf den jeweiligen Lautsprecher abgestimmt, ein Mittelweg zwischen Unterdrückung nichtlinearer Verzerrungen (diese ist gut für kleine α) und geringen Gleichanteil des linearisierenden Signals und damit gute Ausnutzung eines Umsetzers (dies ergibt sich für Werte nahe eins) gewählt werden.

Abschließend sei noch erwähnt, dass bei dem hier betrachteten Lautsprechermodell mit dem Wert $\alpha = 1$ ähnlich gute Ergebnisse wie zuvor erzielt werden. Mit anderen Worten, die Rückführung kann hier wie bei DC-Verstärkern erfolgen, wenn nur der AC-Verstärker im Zustandsbeobachter mit berücksichtigt wird. Allerdings ist diese letzte Aussage nicht auf alle Lautsprecher zu verallgemeinern.

Literatur

- [1] Isidori, A.: Nonlinear control systems. 3 ed. Springer, 1995.
- [2] Schurer, H.: Linearization of electroacoustic transducers. Dissertation, Universität Twente, Enschede, 1997.
- [3] Klippel, W.: Dynamic measurement and interpretation of the nonlinear parameters of electrodynamic loudspeakers. *Journal of the Audio Engineering Society* **38** (1990), 944–955.
- [4] Waßmuth, J.: Modellierung und Kompensation nichtlinearer Effekte bei elektrodynamischen Lautsprechern. Dissertation, Universität Paderborn, 2000.
- [5] Fränken, D.; Meerkötter, K.; Waßmuth, J.: Passive parametric modelling of dynamic loudspeakers. Eingereicht zur Veröffentlichung in: *IEEE Transactions on Speech and Audio Processing*.
- [6] Fettweis, A.: Wave digital filters: Theory and practice. *Proceedings of the IEEE* **74** (1986), 270–327.
- [7] Meerkötter, K.; Scholz, R.: Digital simulation of nonlinear circuits by wave digital filters. *Proceedings of the IEEE International Symposium on Circuits and Systems* (1989), 720–723.
- [8] Meerkötter, K.; Felderhoff, T.: Simulation of nonlinear transmission lines by wave digital filter principles. *Proceedings of the IEEE International Symposium on Circuits and Systems* (1992), 875–878.
- [9] Cunningham, W. J.: Non-linear distortion in dynamic loudspeakers due to magnetic effects. *Journal of the Acoustical Society of America* **23** (1949), 202–207.

DIE PHILOSOPHISCHE VORGESCHICHTE DES GESCHLOSSENEN REGELKREISES

Prof. Dr. Techn., Dimitri Kalligeropoulos,
Techn.Educ.Inst. of Piraeus, Griechenland

Die *Automaten*, **τά αὐτόματα**, die sich von selbst bewegenden Maschinen, die Maschinen die die Funktionen der Lebewesen nachahmen können, erschienen zuerst im altgriechischen Mythos.

Im Gegensatz zu den üblichen Maschinen, **τάς μηχανάς**, die sich mit Hilfe einer äußeren Energiequelle, z.B. die des Menschen, eines Tieres oder des Windes, bewegten, sollten die Automaten eine innere Energiequelle, die einen festen Bestandteil des eigenen Mechanismus ausmachte, besitzen.

Das zu erläutern und gleichzeitig zu realisieren war die parallele Aufgabe der Philosophen und Ingenieure der Antike. *“Das Studium der Pneumatik wurde von den alten Philosophen und Mechanikern sehr eifrig betrieben, indem die einen ihr Wesen theoretisch, die andern durch Vorführung von Experimenten darlegten”*, schreibt Heron in der Einführung seiner Pneumatik.

Die ersten, die eine physikalische Erläuterung des Kosmos und somit eine wissenschaftliche und philosophische Deutung über die Funktion der Automaten anstrebten waren die prosokratischen Philosophen des 6ten v.Chr. Jahrhunderts.

Die beseelten Grundelemente der Prosokratiker

Anaxagoras (500 – 428 v. Chr.) führte den Begriff des Widerspruches ein. Die Welt nach Anaxagoras bestand aus *Widersprüchen*, die sich in einem einheitlichen Ganzen beinhalteten. Dieses Ganze wurde **νοῦς**, Vernunft, Verstand genannt und hatte die Funktion einer Maschine. Durch die Drehbewegung dieser Maschine trennten sich das Warme von dem Kalten, das Trockene von dem Feuchten, das Helle vom Dunklen, das Dünne vom Dichten und bildeten die Erde und den Himmel.

Empedokles (495 – 435 v. Chr.) dagegen definierte die *vier Elemente* der Materie, Erde, Wasser, Luft und Feuer, als die Grundelemente, aus denen die Welt geschaffen wurde.

Diese primären Elemente waren als beseelt gedacht, mit Seele, Psyche, *ψυχή*, mit einer inneren Energie bestattet. Und die Auffassung über die beseelte materielle Welt entwickelte sich dann in eine technische Auffassung, ein technisches Streben der antiken Ingenieuren in ihren Konstruktionen innere Energie hinzuzufügen und sich von selbst bewegende Maschinen zu schaffen.

Das erste historische Automaton – eine fliegende Taube

Archytas aus Tarant Siziliens, Freund Platon's, Gründer der wissenschaftlichen Mechanik und bekannter Mathematiker schuf im 5. Jahrhundert v. Chr. eine solche sich von selbst bewegende Maschine, indem er als innere Energie die Ausdehnungstendenz der gepressten Luft benutzte.

“Archytas schuf das hölzerne Modell einer fliegenden Taube” berichten Favorinus und der Römer Aulus Gellius. *“Sie flog wegen eines Luftstromes aus ihrem inneren. Und sie hatte gut abgewogene Gewichte”* (Aulus Gellius, Attische Nächte, 10, 12, 9.).

Die Verwendung der “inneren Energie” des Grundelementes “Luft” führte zur Entwicklung einer neuen technischen und wissenschaftlichen Richtung: der Pneumatik.

Entsprechend wurde die “innere Energie” der anderen Grundelementen, des “Feuers”, das die Ausdehnung der erwärmten Luft oder des Dampfes hervorrief, des “Wassers”, dessen gleichförmige Bewegung eine kontrollierbare Antriebskraft erzeugte und der “Erde”, die die potenzielle Energie der Körper bewirkte, benutzt.

Das Feuer als Energiequelle

In der heronischen Pneumatik wurde das Feuer auf einem Altar dazu benutzt, um die Luft in der Altarbasis zu erwärmen und somit Wasser aus dem Boden des Altars durch ein gebogenes Rohr steigen zu lassen, bis wiederum das Feuer ausgelöscht wurde.

“Wird auf gewissen Altären ein Rauchopfer angezündet, so sollen Figuren, die daneben stehen, ein Trankopfer darbringen” (Heron, Pneumatik A, 12).

Das Feuer wurde auch dazu benutzt um den Dampfdruck in eine Drehbewegung, bei dem sogenannten heronischen Aeolusball, dem Vorfahren der Dampfmaschine, umzuwandeln. *“Über einem geheizten Kessel soll eine Kugel sich um einem Zapfen bewegen... Wird nämlich der Kessel geheizt, so ist die Folge, daß der Dampf durch ein*

Rohr in die Kugel dringt, durch die umgebogenen Röhren nach dem Deckel hin ausströmt und die Kugel zur Drehung bringt” (Heron, Pneumatik B, 40).

Das Wasser als Antriebskraft

Entsprechend wurden die Eigenschaften der Flüssigkeiten untersucht und vielfältige hydraulische Vorrichtungen für die Regelung des Wasserstandes und des Wasserflusses entdeckt. Heron beschrieb in seiner Pneumatik verschiedene hydraulische Heber (*σιφώνια*) die das Überschreiten eines gewissen Wasserstandes nicht gestatteten.

Ktesibios dagegen konstruierte am Anfang des 3ten v. Chr. Jahrhunderts seine hydraulische Uhr, die mit einem Regelmechanismus des Wasserflusses ausgerüstet war und somit die gleichförmige Bewegung eines schwimmenden Zeigers, der die Zeit auf einem drehbaren Zylinder anzeigte, ermöglichte (Beschreibung in Vitruvs Architektur, 9, 8).

Das fallende Gewicht

Was die “innere Energie” der “Erde”, der Materie, die wegen ihres Gewichtes eine potentielle Energie besaß, oder wegen ihrer Elastizität Energie anspeichern konnte, betraf, schrieb Heron in seiner Automatentechnik. *“Alle fahrenden Automaten erhalten ihren Antrieb zur Bewegung entweder durch ein Spannholz (διά ὑσπληγγος) oder durch ein Gegengewicht aus Blei (διά λείας μολιβῆς)”* (Heron, Automatentechnik, 2, 6).

Das fallende Gewicht, gebunden durch eine Schnur mit einer Antriebsachse, ermöglichte die Vorwärtsbewegung des gesamten Mechanismus und somit die Vorwärtsbewegung von sich selbst.

Das waren ein Paar Beispiele von realisierten Mechanismen der altgriechischen Technik, die die Lösung des Energieproblems der von sich selbst bewegenden Maschinen, der Automaten, in den Eigenschaften der vier Grundelemente der Natur suchten.

Die Automaten in der Philosophie des Aristotels

Diese Suche nach dem Grund, der Ursache, der Antriebskraft, die die Bewegung der Automaten ermöglichte, führte *Aristotel* (384 – 322 v. Chr.) weiter. Er versuchte damit eine philosophische Erläuterung dieser selbständigen Bewegung zu geben, und sie mit der selbständigen Bewegung der Lebewesen zu vergleichen.

“Am Anfang wundern sich alle über das Wesen der Sachen, genauso wie sich jene, die die Bewegung der Automaten mit Bewunderung anschauen, ohne die Ursache (αἰτίαν) dieser Bewegung zu begreifen, wundern” (Aristotel, *Metaphysik*, 983a, 15).

Weiter verglich der Philosoph die Bewegung eines mechanischen Automaten mit der des Samens, und unterscheidete zwischen einem inneren und einem äußeren Antrieb. *“Der Samen (σπέρμα) hat eine innere Antriebskraft (κίνησιν ἐν ἑαυτῷ), die seine eigene Bewegung bewirkt. Diese Bewegung kann jedoch von einem Samen zu einem anderen von Außen her übertragen werden. Auf diese Weise bewegen sich die Automaten der Wundermaschinen (τὰ αὐτόματα τῶν θαυμάτων).”*

Er untersuchte weiter die innere Funktion der Automaten und definierte den äußeren Antrieb bei einem Steuerungsprozeß. *“Die Elementarteile (des Samens wie auch eines Automaten) besitzen in der Ruhelage eine innere Antriebskraft (δύναμιν). Sobald etwas von Außen her den Ersten von diesen Teilen in Bewegung setzt, setzt sich gleich der Nächste in Betrieb. So funktionieren also die Automaten. Sie bewegen sich von selbst, ohne von jemandem berührt zu werden, nach einem primären jedoch externen Anstoß (ἄψάμενον)”* (Aristotel, *Genesis der Tiere*, 734b, 11).

Aristotel erlaubte sich sogar den umgekehrten Vergleich. Die Bewegungen der Tiere wurden mit denen der Automaten verglichen. *“So wie sich die Automaten durch kleine Bewegungen bewegen, indem sich ihre drehbaren Teile befreien und gegenseitig anstoßen... so bewegen sich auch die Tiere”* (Aristotel, *Bewegung der Tiere*, 701b, 2).

Die Realisierung der aristotelischen Ideen

Diese philosophische Betrachtungen, die sich auf den Automaten der klassischen Zeiten bezogen, wurden am deutlichsten in den uns überlieferten Beschreibungen der hellenistischen Automaten Herons aus Alexandria realisiert.

“Ist bei solchen Vorrichtungen zu Anfang der Automat an irgend einem Punkte aufgestellt, so wird bald nachdem wir zurückgetreten sind, der Automat nach einer bestimmten Stelle vorrücken. Wenn er dann stehen bleibt, wird der Altar vor Bacchus angezündet. Und aus seinem Thyrsus spritzt Milch oder Wasser, aus dem Becher ergießt sich Wein auf den darunter liegenden Panther, der ganze Unterbau wird an den vier Pilastern bekränzt, die Bacchantinnen ringsum umkreisen im Tanze den Tempel, Trommelwirbel und Beckenschlag wird vernehmbar... Zum Schluß fährt der Automat nach seinem Ausgangspunkte zurück. Und so wird die Vorstellung ein Ende haben” (Heron, Automatentechnik, 4, 1-3)

So beschreibt Heron die Funktion seines fahrendes Automatentheaters. Die Einzelteile dieser komplexen automatischen Maschine folgen die von Aristotel beschriebene Logik einer offenen Steuerung. So z.B. bei den automatischen Mechanismen zur Erzeugung von Tönen öffneten sich im richtigen Augenblick die Böden der Kugelbehälter und bleierne Kugeln fielen auf Pauken und Becken. So auch bei den automatischen hydraulischen Mechanismen zum regelmäßigen Fluß von Milch und Wein.

Der Begriff der Programmierung

Ein neuer Begriff findet jedoch in diesen Beschreibung seine Realisierung: Es ist der Begriff der *Programmierung*. Im fahrenden Automatentheater Herons wurde die Programmierung sämtlicher Bewegungen durch verschiedenartige Schnurwirkungen auf der Drehachse verwirklicht: Linksdrehung der antreibenden Schnur führte zur Vorwärtsbewegung des Automaten, Rechtsdrehung führte zur Rückwärtsbewegung und freie Schnurwirkung rief keine Bewegung hervor.

So entstand ein trinäres System (+1, 0, -1), das eine volle Programmierung der verschiedensten Bewegungen, eine innere wohldefinierte Logik, eine im Automaten vorhandene *“Vorahnung”*, ermöglichte. Diese *“Vorahnung”*, diese innere Programmierung die zur Regelung von automatischen Systemen geeignet ist, unterscheidet Aristotel vom äußeren *“Befehl”* einer einfachen Steuerung.

“Wenn jedes Werkzeug (ὄργανον) sein Werk vollenden konnte, entweder nach einem äußeren Befehl, von außen her gesteuert (κλειυσθέν) oder durch Vorahnung, durch eine innere Programmierung, durch eine im voraus festgelegte Funktion

(προαισθανόμενον) und wenn es automatisch funktionieren konnte, wie jene Statuen von Dädalos, die sich bewegen konnten, oder wie die Dreifüsse von Hephaistos, die der Poet Automata nannte, weil sie von selbst in die Versammlung der Götter eindringen, so, wenn automatisch das Weberschiffchen des Webstuhls weben konnte und wenn von selbst die Tasten der Gitarre Musik spielen konnten, dann bräuchten weder die Architekten einen Arbeiter, noch die Feldherren einen Sklaven” (Aristotel, Politik A 2, 4).

Diese wunderbare Beschreibung des griechischen Philosophen beschränkt sich nicht nur in einer klaren Unterscheidung zwischen den zwei Methoden der Automatisierungstechnik, sondern sieht eine breite Anwendung dieser Technik, sowohl in der Produktion als auch im Alltagsleben, voraus und verbindet gleichzeitig diese Ausbreitung der Automatisierung mit einer radikalen sozialen Umwandlung.

Jedoch sowohl das philosophische als auch das technische altgriechische Denken ging noch weiter, und entdeckte den Begriff der Rückkopplung, des geschlossenen Regelkreises.

Die philosophische Grundlage der Rückkopplung

Epimenides, ein vorsokratischer Philosoph aus Kreta des 7ten – Anfang 6ten v. Chr. Jahrhunderts, formulierte ein Paradoxon, das die reine geradlinige Logik umwarf.

Stellung A: *“Epimenides behauptet, er sei ein Kreter.*

Stellung B: *Alle Kreter, im Widerspruch zu allen anderen Griechen, lügen.”*

Wenn also Epimenides tatsächlich ein Kreter ist dann lügt er, d.h. er ist kein Kreter und umgekehrt. Diese beide Stellungen, bei denen die zweite die logische Widerlegung der ersten ist, führen den Begriff der negativen Rückkopplung, der umgekehrten Verbindung zwischen Folge (*αἰτιατόν*) und Ursache (*αἴτιον*), ein .

Heraklit (544 – 483 v. Chr.) aus Ephessos führte den dialektischen Widerspruch ein und formulierte die Sätze:

“Τά πάντα ρεῖ, Alles fließt”,

“Gott (die Welt, das Leben) ist Tag und Nacht, Winter und Sommer, Krieg und Frieden, Sättigung und Hunger” (B67)

Und weiter:

“Das was mit sich selbst im Widerspruch steht, stimmt mit sich selbst überein – Es besteht eine widersprüchliche Harmonie” (B51)

Und das, was Epimenides als ein Paradoxon feststellte und was Heraklit als eine widersprüchliche Struktur des Kosmos erklärte, formulierte Sokrates mit seinem berühmten Satz:

*“Ἐν οἶδα, ὅτι οὐδέν οἶδα, Eins weiß ich, daß ich Nichts weiß,
Eins ist Wahr, das, was seine eigene Negation akzeptiert”.*

Die Realisierung von geschlossenen Regelkreisen

Der Begriff des Kreises, der geschlossenen Schleife, der Rückkopplung und des Widerspruchs entstand also im altgriechischen philosophischen Denken und führte im sozialen Bereich zum Dialog und zur Demokratie. Die sokratische und platonische dialektische Methode, bei der die Argumente des Redners sich auf die Antworten des Mitredenden im Zwiegespräch aufbauten, sowie die direkte altgriechische Demokratie, bei der die Bürgerschaft Gesetzte über sich selbst entschloss und die Beherrschten Herrscher von sich selbst ausmachten, sind Beispiele solcher geschlossenen Schleifen.

Mathematiker wie Platon und Euklid entwarfen mathematische rekursive Algorithmen für die genaue Bestimmung des Sollwerts von irrationalen Zahlen, Algorithmen die selbst geschlossene mathematische Schleifen darstellten.

Jedoch fand der Begriff des geschlossenen Regelkreises seine konkrete Realisierung in der altgriechischen Technologie, als Ingenieure und Techniker der Automaten Regelsysteme mit Rückführung systematisch entwarfen und sorgfältig konstruierten.

Hier gewisse Beispiele solcher Konstruktionen für die Regelung des Wasserstandes aus den hellenistischen Zeiten.

1. Die Regelung von Ktesibios

Ktesibios, am Anfang der 3ten v. Chr. Jahrhunderts in Alexandria, entdeckte einen hydraulischen Mechanismus, der einen schweren Spiegel in das Geschäft seines Vaters automatisch hochheben konnte.

Er entwarf seine berühmte hydraulische Uhr, die einen Regelmechanismus des Wasserflusses besaß. Ihm wird auch der Entwurf eines einfachen Regelmechanismus für die Regelung des Wasserstandes durch einen kegelförmigen Schwimmer zugeschrieben.

Es sei ein Becher AB, der mit Wasser aus dem Behälter $\Gamma\Delta$ durch das Rohr EZH versorgt wird. In dem Becher befindet sich ein Schwimmer Π mit einem kegelförmigen Dach Θ , der in die gewölbte Öffnung H des Rohres genau paßt. Wenn man Wasser aus dem Becher abnimmt, dann fließt aus dem Behälter Wasser solange bis der Schwimmer wie ein hydraulisches Ventil funktioniert und die Öffnung H abschließt

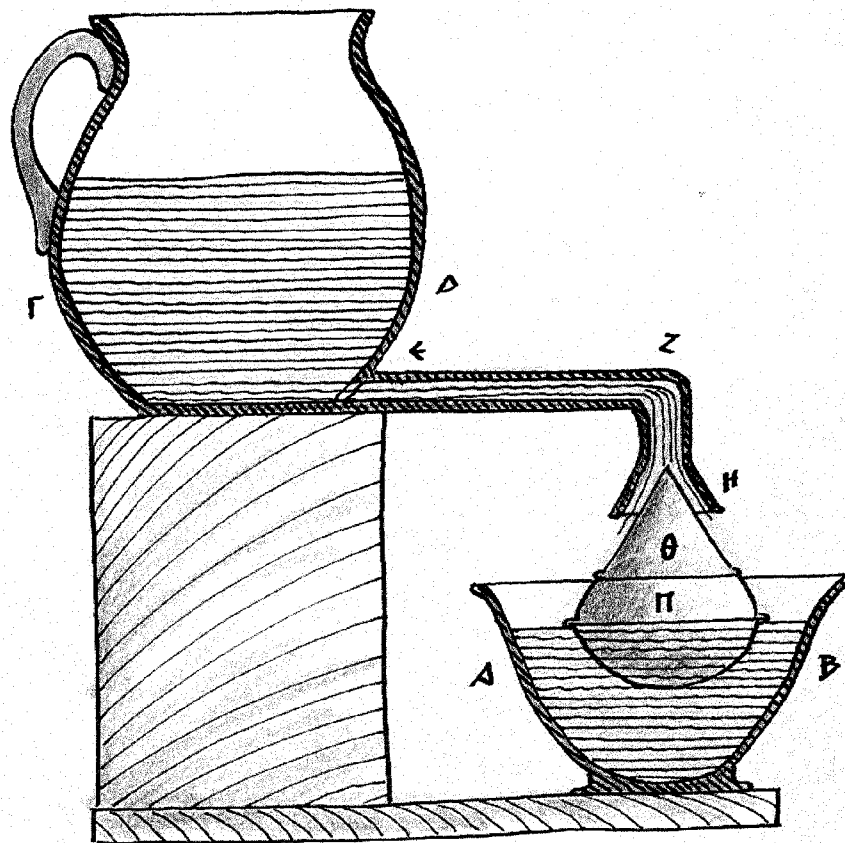


Bild 1: Ktesibios' Wasserstandsregelung durch einen kegelförmigen Schwimmer

2. Automatische Wasserstandsregelung Philons aus Byzanz

Philon, um 250 v. Chr., beschrieb in seiner Pneumatik verschiedene hydraulische Regelvorrichtungen. Eine solche Vorrichtung, die die Regelung des Weinstandes in einem Becher $\Delta\Delta$ betraff, bestand aus einer hohlen hölzernen Basis AA mit einem Hohlraum Θ , aus einer ebenfalls hohlen Säule BB mit dem Hohlraum E , aus dem Behälter $\Gamma\Gamma$, mit Wein gefüllt und in seiner Öffnung geschlossen, aus einem senkrechten Rohr K durch den gebohrten Grund des Behälters und einen Ausfuhrrohr Z .

Das System befand sich im Gleichgewicht wenn der Weinstand im Becher $\Delta\Delta$ die untere Öffnung des Rohres K abschloß und der Unterdruck im Hohlraum des Behälters $\Gamma\Gamma$ den Weinfluß aus der Öffnung Z nicht ermöglichte.

Eine Weinabnahme jedoch aus dem Becher $\Gamma\Gamma$ hatte die Enthüllung der unteren Öffnung des Rohres K und die Umwendung des Unterdrucks im Behälter zu Folge. So floß Wein aus der Ausfuhr Z solange in den Becher ein, bis das Gleichgewicht wieder erreicht wurde.

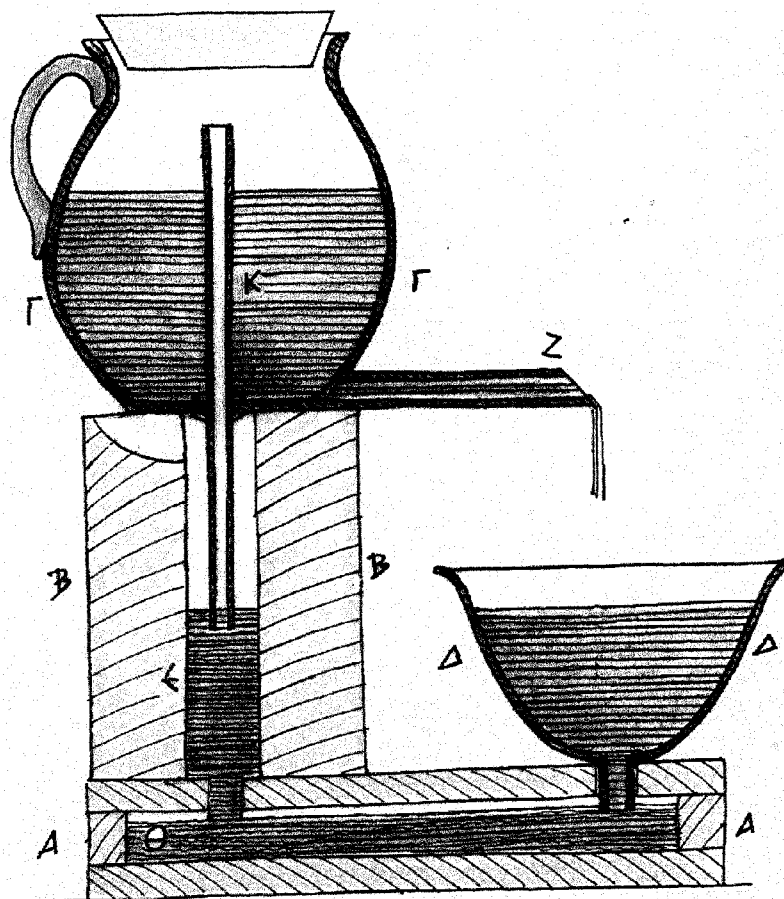


Bild 2: Automatische Regelung des Weinstandes nach Philon von Byzanz

3. Heronischer Weinautomat, mit mechanischer Rückkopplung

Heron aus Alexandria (~1tes v. Chr. Jahrhundert) beschreibt in seiner Pneumatik ein entsprechendes hydraulisches System für die Weinstandsregelung mit der Neuerung einer mechanischen Rückkopplung.

“Hat ein Gefäß mit Wein eine Ausflußröhre, unter der ein Mischkrug stellt, so soll so viel Wein, als man dem Mischkruge entnimmt, aus der Ausflußröhre zufließen.

Das Gefäß mit Wein sei AB, die Ausflußröhre ΓΔ. Es sei mit dem Scheibchen EZ, und den Stangen HΘ, KA, KO, AM, versehen. Unter der Ausflußröhre stehe ein Krug Π. Mit der Stange KO sei ein kleiner Kessel P verbunden, der sich in einem Gefäße ΣΤ befinde. Eine Röhre ΥΦ setze die Gefäße ΣΤ und Π, in Verbindung. Sind nun bei derartigen Vorrichtungen die Gefäße Π und ΣΤ leer, so liegt der Kessel P am Boden des Gefäßes ΣΤ und läßt die Ausflußröhre ΓΔ offen. Da nun die (infolgedessen ausströmende) Flüssigkeit sich in die beiden Gefäße Π und ΣΤ ergießt, so steigt der Kessel und schließt die Ausflußröhre wieder, bis man den Mischkrug von neuem ausschöpft. Dieser Vorgang wiederholt sich jedesmal, wenn man etwas herausnimmt” (Heron, Pneumatik B, 31).

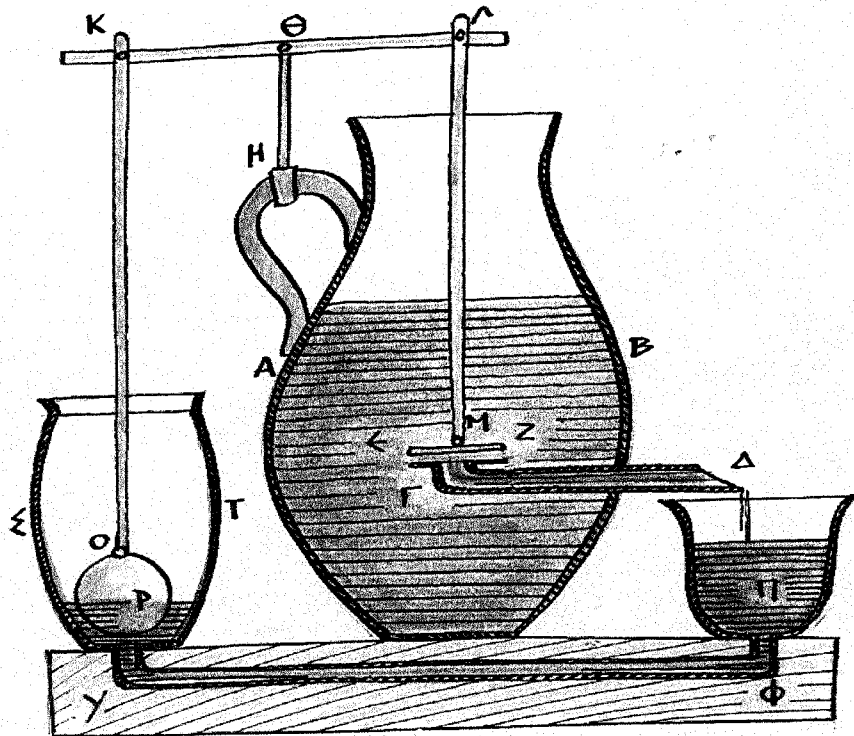


Bild 3: Heronischer Weinautomat mit Weinstandsregelung durch eine mechanische Rückkopplung

So fanden, besonders in den hellenistischen Zeiten, die philosophischen Ideen des klassischen Altertums ihre technische Realisierung.

The Platonic Dichotomy and the Euclidean Anthyphairetic Method

The first closed loop mathematical algorithms

S. Vasileiadou[†]

[†] Techn. Control Eng., Ph.D. student at Systems and Modelling Centre,
Department of Electrical, Electronic, and Information Engineering,
City University, Northampton Square, London EC1V 0HB, U.K.

Abstract

The examination of the evolution of Systems, Modelling and Control concepts through the development of the ancient Greek philosophy, science, and technology, aims at their placement in a proper historical setup. Our study passes necessarily through the fundamental steps of the philosophical and mathematical thought. In particular, the following pages concern the Platonic theoretical and the Euclidean mathematical approach to the problem of the definition of a specific element or of the determination of the exact value of an immeasurable irrational number, respectively. By means of their methods, i.e. the dichotomous method of Plato and the anthyphairetic method of Euclid, they introduce the concept of finite or infinite algorithms that can be considered as closed mathematical loops.

1. Introduction

The study of the evolution of system, modelling, and control concepts in ancient Greece reveals the close relation between the philosophical, the mathematical, and the technical thought of the Greeks. The philosophers attempt to interpret in a theoretical point of view not only the natural world, but also the mathematical and the technical issues. On the other hand, the mathematicians build their mathematical methods by taking advantage of the corresponding technical methods. For example, the ratio of two numbers, which arises from the geometrical concept of analogy, expresses the relationship between two magnitudes A and B by means of the technical process of measurement, i.e. ‘magnitude A measures magnitude B’. On the contrary, the irrational ratios, where ‘A does not measure B’, need a more complicated mathematical method for their precise computation, and demand a new philosophical approach and a new technical construction method. In the field of philosophy, it is accomplished by Plato and his dichotomous division method, whereas in the field of mathematics by Euclid and his method of anthypharesis. The common factor in these approaches is the

concept of algorithm, which in the anthyphairetic method takes a more precisely geometrical and numerical form, the form of a closed loop mathematical algorithm, and leads by unlimited approximations to the estimation of irrational numbers.

2. Ratio or Logos

Even though Greek thought and Greek mathematics up to the second century B.C. were in general completely non-arithmetised, the early Greek mathematicians seemed to find no difficulty in using freely the idea of the ratio of two numbers or of two geometrical magnitudes. They made remarkable investigations and discoveries about ratio.

The ratio or logos ρ of two magnitudes A, B represents the relationship between A and B or in general the process of transition of the one magnitude to the other $B = \rho \cdot A$. It constitutes the mathematical model or the transfer function of a simple, linear, not dynamic analogy system.

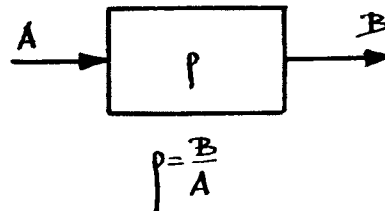


Figure 1: Ratio as the transfer function of a system

The logos of two magnitudes A and B can be: a) perfect or complete if the result is an integer, i.e. the less magnitude A measures B or A is part of B ($B = \alpha \cdot A$, where α : integer), and b) non-perfect or incomplete if the result is not an integer, i.e. magnitude A does not measure B ($B = \alpha \cdot A + v$, where the remainder $v < A$ and $v = B - \alpha \cdot A$).

The Pythagoreans were the first to consider the numbers, and in particular the integers, as the substance of all things. In the discovery of the harmonics of sounds and more precisely in the correlation between the ratios of the lengths of a vibrating string and the pitch of the produced sounds they saw a verification of their number philosophy. The finding out of ‘golden’ triangles, the sides of which were in the ratio 3:4:5, or 5:12:13, or 8:15:17, and the contemplation of them led to the great discovery of the so-called Pythagorean theorem. Therein they saw the inherent union between arithmetic and geometry, an additional confirmation of their dictum.

But the triumph was short-lived. Indeed one of the immediate consequences of the theorem was the discovery of the irrational number $\sqrt{2}$, which was made geometrically by comparing the diagonal of a square with the length of one side and finding that they cannot be expressed as the ratio of two integers. They had felt that given any two straight-line segments such as the side and the diagonal of a square there would be some integers a and b so that the ratio of the lengths of these segments would be $\frac{a}{b}$. The absence of such a ratio (*logos*) was declared by the name ‘irrational’ or ‘*alogon*’, which it is given to ratios that are not a rational number.

The phenomenon of magnitudes that couldn’t be expressed as a ratio of integer numbers disturbed the up to that point perfect concordance between things arithmetical and things geometrical. The theorem, which specified the criterion of incommensurability, comes to us from Euclid (c. 330-275 B.C.):

- If, when the less of two unequal magnitudes is continually subtracted in turn from the greater, that which is left never measures the one before it the magnitudes will be incommensurable. (*Elem. X, Prop. 2*)

The method, which is applied here, is that of reciprocal subtraction or anthypharesis, as we will examine later on.

3. The Dichotomous Division Algorithms of Plato

A century before Euclidean studies on the problem of irrational numbers by means of the anthyphairetic method that based on the concept of dichotomy, of sequence, and of mathematical algorithm, Plato introduces a particular form of his dialectic method, this one of dichotomous division. He takes advantage of the concept of dichotomy and the principle ‘the one is divided into two’ in order to compose a satisfactory definition of a specific subject. In his works, and especially in *Sophist*, Plato develops his method and quotes a number of applications, where each time he looks for the definition of a different topic. Plato firstly applies the Dichotomous Division Method on a simple and smaller thing and then tries to make it a pattern of a greater one. Thus before the overall formulation and application of his method he quotes a simple example – *angler’s definition* – through which he develops the process and the steps of Dichotomous Division Method. We choose to cite below this primary example that, on the one hand,

constitutes the pattern on which Plato is based in order to specify more serious and complicated subjects, and on the other hand, it is adequate for the platonic method to be clarified.

- Angler's definition: *Sophist*, 218d-220e

Let say that 'angler' is the parameter X that Plato intends to define. He starts with a bigger and more familiar category: Category **A**: **Man**. The next category **B** arises by dichotomy. Man is divided into: **b₁: man not having an art**, and **b₂: man having an art**. Decision: Angler is a man having an art. Category **C**: Arts are divided into: **c₁: productive or creative art**, the art of constructing or producing, and **c₂: acquisitive art**, the art of learning and cognition. Decision: The art of angle is acquisitive art. Category **D**: The acquisitive art is subdivided into: **d₁: of exchange**, that is effected by gifts, hire, purchase, and **d₂: of conquest**, which takes by force of word or deed. Decision: The art of the angler is that of the conquest. Category **E**: The conquest is subdivided into: **e₁: the open force** (fighting), and **e₂: secret force** (hunting). Decision: The art of the angler is that of hunting. Category **F**: Hunting is subdivided into: **f₁: of living prey or animal hinting**, and **f₂: of lifeless prey**. Decision: The art of the angler is that of the animal hunting. Category **G**: Living prey is subdivided into: **g₁: land-animal hunting**, and **g₂: water-animal hunting** (swimming animal). Decision: The art of the angler is that of water-animal hunting. Category **H**: Water-animal hunting is subdivided into: **h₁: capture with enclosures** (takes them when they are in nets), and **h₂: capture by striking** (takes them by a blow). Decision: The art of the angler is that of capture by striking. Category **J**: Capture by striking is subdivided into: **j₁: by the light of fire** (at night), and **j₂: by day (barbing)**. Decision: The art of the angler is of barbing.

We keep on the procedure by leaving each time the left subdivision unexamined (i.e. categories **b₁**, **c₁**, **d₁**, **e₁**, **f₁**, **g₁**, **h₁**), as it is shown in the figure below. We elaborate only the right sub-category by dividing it in smaller and smaller sub-categories to some extend where one right sub-category agrees with the initial parameter X. The definition of X results from the successive characteristics of right subdivisions (i.e. categories **b₂**, **c₂**, **d₂**, **e₂**, **f₂**, **g₂**, **h₂**).

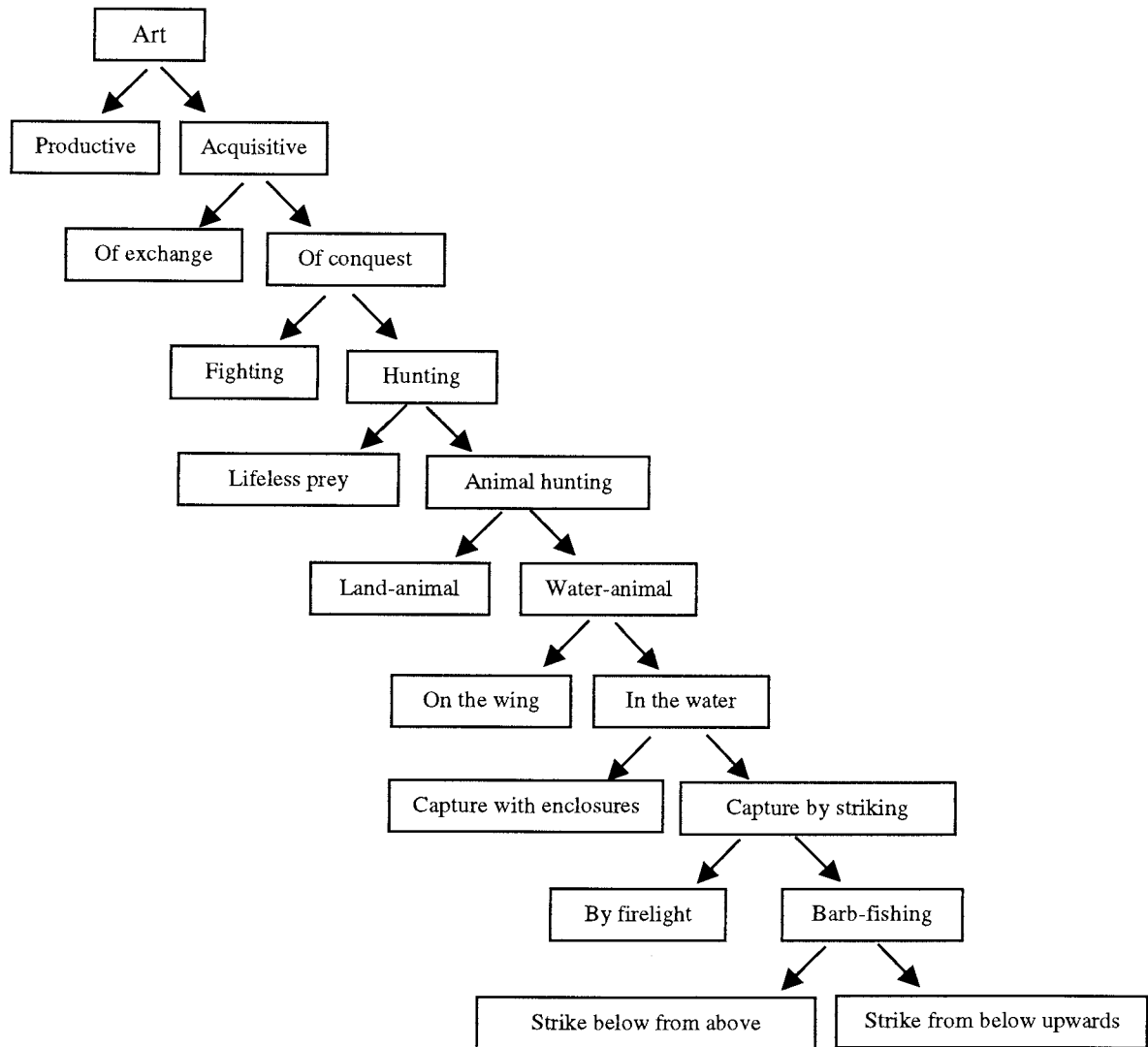


Figure 2: Dichotomous Method for the definition of angler

Definition: One half of all art was acquisitive - half of the acquisitive art was conquest or taking by force, half of this was hunting, and half of hunting was hunting animals, half of this was hunting water animals - of this again, the under half was fishing, half of fishing was striking; a part of striking was fishing with a barb, and one half of this again, being the kind which strikes with a hook and draws the fish from below upwards, is the art which we have been seeking, and which from the nature of the operation is denoted angling or drawing up.

Thus, Plato develops a theoretical method, whose main characteristic is that it consists of a number of successive and repetitive steps as in any mathematical algorithm.

4. The Anthyphairetic Algorithms of Euclid

The phenomenon of incommensurable magnitudes, of magnitudes that didn't have any common measure or couldn't be measured disturbed the pure geometrical interpretation of the world. What is $\sqrt{2}$? What is π ? How are they measured? The necessity to deal with such problems that couldn't be solved by pure geometry resulted on the one hand, in a new method of successive approximations, which in mathematics has been expressed by regressive algorithms, and on the other hand, in the distinction between finite and infinite steps of algorithms closely related to the concept of infinity.

In Book X, Euclid quotes an infinite process of successive approximations, of a 'geometrical algorithm', which allows the comparison of two unequal magnitudes:

- Two unequal magnitudes being set out, if from the greater there be subtracted a magnitude greater than its half, and from that which is left a magnitude greater than its half, and if this process be repeated continually, there will be left some magnitude which will be less than the lesser magnitude set out. (*Elem.X, Prop.I*)

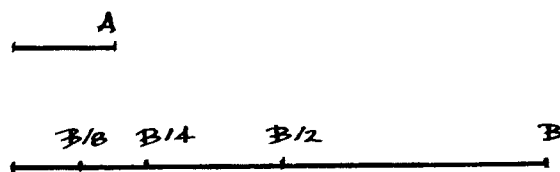


Figure 3

According to this proposition the less magnitude A can be approximated by successive bisections of the greater magnitude B.

The complete mathematical approximation of rational and irrational numbers either by the arithmetical or by the geometrical method is succeeded by the method of reciprocal subtraction or of anthypharesis. Euclid makes allusions to the method of anthypharesis or antanairesis firstly in order to define the prime numbers (*Elem. VII, Prop. I*), and secondly in order to define incommensurability (*Elem. X, Prop. II*). Both Euclidean propositions constitute a dichotomy, i.e., a distinction between the two considerable categories of numbers, the rationals and irrationals, a distinction, where the first category arises by a finite process and the second by an infinite process of anthypharesis. Let us give a more analytic description of what does it mean.

According to the anthyphairetic method: given two numbers or two lines we count a) how many times the second line can be subtracted from the first line, b) how

many times the remainder can be subtracted from second line, c) how many times the remainder of second line can be subtracted from the previous remainder, etc. and this gives a string of numbers, possibly infinite, that characterise the relationship of size between the two things, and which is called ‘anthyphairetic ratio’. Since it is based on a process of repeated subtraction of one thing from another its ancient name of *antanairesis* or *anthyphairesis* can be translated as ‘reciprocal subtraction’. The modern name of the whole process and the associated mathematical objects is ‘Euclidean algorithm’ or ‘continued fractions’, though the Euclidean algorithm is now generally construed as a division process and the continued fractions are now handled using the real numbers and a sophisticated generalization of fractions.

Let us elaborate the anthyphairetic method.

The simple anthyphairesis is the combination of the operation of subtraction and inversion, the reciprocal subtraction or the reciprocal retraction (*antanairesis*). Thus for two magnitudes A and B, where $B > A$, the anthyphairesis is: $anth(A, B) = \frac{1}{B - A}$

The reciprocal subtraction arises by the process of comparing and measuring two magnitudes. If A measures α times B, and the remainder is v , then:

$$B = \alpha \cdot A + v, \quad v < A \quad (1)$$

$$\text{Since } B > A, \text{ the ratio } \rho = \frac{B}{A} \quad (2)$$

will be greater than 1, i.e. $\rho > 1$ and by the relations (1) and (2) we have: $\rho = \alpha + \frac{1}{\rho_1}$
(3)

where $\rho_1 = \frac{A}{v} > 1$, is the comparison, the ratio of the initially less magnitude A to the first remainder v . By the relation (3) we calculate ρ_1 , which is the first anthyphairesis

$$\text{of ratio } \rho: \quad \rho_1 = \frac{1}{\rho - \alpha} \quad (4)$$

The repetition of this process has as result the following algorithm:

Step 1: For magnitudes A, B, if $B > A$ and $B = \alpha \cdot A + v$, $v < A$ then

$$\frac{B}{A} = \alpha + \frac{v}{A} \quad \text{or} \quad \rho = \alpha + \frac{1}{\rho_1} \quad \text{where} \quad \rho = \frac{B}{A} > 1, \quad \rho_1 = \frac{A}{v} > 1$$

Step 2: Comparison of the magnitudes A and v , $A > v$

$$A = \alpha_1 v + v_1, v_1 < v \quad \text{or} \quad \frac{A}{v} = \alpha_1 + \frac{v_1}{v}, \quad \rho_1 = \alpha_1 + \frac{1}{\rho_2} \quad \text{where} \quad \rho_2 = \frac{v}{v_1} > 1$$

Step 3: Comparison of the magnitudes v and v_1 ,

$$v = \alpha_2 v_1 + v_2, v_2 < v_1 \quad \text{or} \quad \frac{v}{v_1} = \alpha_2 + \frac{v_2}{v_1}, \quad \rho_2 = \alpha_2 + \frac{1}{\rho_3} \quad \text{where} \quad \rho_3 = \frac{v_1}{v_2} > 1 \quad \text{etc.}$$

The relationships of the successive anthypharesis, of the successive comparison between the less and the greater or more precisely between the less remainder and the greater preceding remainder form the next algorithm:

$$\begin{aligned} \rho &= \alpha + \frac{1}{\rho_1} \\ \rho_1 &= \alpha_1 + \frac{1}{\rho_2} & \text{Where} \\ \rho_2 &= \alpha_2 + \frac{1}{\rho_3} & \rho, \rho_1, \rho_2, \dots, \rho_v, \rho_{v+1} \text{ are ratios greater than 1} \\ & & \text{and} \\ \dots & & \alpha, \alpha_1, \dots, \alpha_v, \dots \text{ integers} \\ \rho_v &= \alpha_v + \frac{1}{\rho_{v+1}} & \text{that constitute the distributive measures of ratio } \rho \\ \dots & & \end{aligned}$$

This algorithm leads to the concise relation that determines/measures a ratio ρ (rational or irrational) by a mass of integers $(\alpha, \alpha_1, \dots, \alpha_v, \dots)$

$$\rho = \alpha + \frac{1}{\alpha_1 + \frac{1}{\alpha_2 + \dots \frac{1}{\alpha_v + \frac{1}{\alpha_{v+1} + \dots}}}}$$

This relationship, which is the result of the anthyphairetic algorithm, analyses the ratio ρ into the distributive anthyphairetic integer measures $\alpha, \alpha_1, \dots, \alpha_v, \dots$ and allows the symbolism: $\rho = [\alpha, \alpha_1, \alpha_2, \dots, \alpha_v, \dots]$

The two Euclidean proportions about anthypharesis can be formulated as follows:

1. The ratio ρ is rational if the anthyphairetic algorithm consists of v finite steps so that $\rho_{v+1} = 1$.
2. The ratio ρ is irrational if the anthyphairetic algorithm is infinite.

Let us quote some examples on the anthyphairetic algorithm:

Example 1: Numerical form of the anthyphairetic algorithm of a rational ratio

Consider the unequal magnitudes $A=3$ and $B=5$. The ratio $\rho = \frac{B}{A} = \frac{5}{3}$ is analyzed as follows:

Step 1: $\frac{5}{3} = 1 + \frac{2}{3}$, $\alpha=1$ and $v=2$

Step 2: $\frac{3}{2} = 1 + \frac{1}{2}$, $\alpha_1=1$ and $v_1=1$

Step 3: $\frac{2}{1} = 1 + 1$, $\alpha_2=1$ and $v_2=1$

Step 4: $\frac{1}{1} = 1 + 0$, $\alpha_3=1$ and $v_3=0$

So $\frac{5}{3} = [1, 1, 1, 1]$ or $\frac{5}{3} = 1 + \frac{1}{1 + \frac{1}{1 + 1}}$

Example 2: Infinite anthyphairetic algorithm of the irrational number $\sqrt{2}$

The irrational ratio $\sqrt{2}:1$ is defined as the ratio of the sides of two squares whose areas is double the one of the other, i.e. $\frac{E_\beta}{E_\alpha} = \frac{\beta^2}{\alpha^2} = 2$. Therefore $\frac{\beta}{\alpha} = \sqrt{2}$. The computation of $\sqrt{2}$ could be realized geometrically by comparing the side of a square with its diameter, or by applying relations of proportion between two squares. Consider the square $AB\Gamma\Delta$, whose sides are α and the diameter is β . We are looking for the ratio $\rho = \frac{\beta}{\alpha} = \sqrt{2}$. We construct another greater square, the $AB_1\Gamma_1\Delta_1$, whose side is $AB_1 = A\Delta_1 = \beta_1 = \alpha + \beta$. Geometrically it is shown that the diameter of it is $A\Gamma_1 = \alpha_1 = 2\alpha + \beta$.

The ratio between the diameter and the side of this square is $\rho_1 = \frac{\beta_1}{\alpha_1} = \frac{2\alpha + \beta}{\alpha + \beta}$.

Because of the analogy between the squares we have:

$$\rho_1 = \rho = \frac{\beta}{\alpha} = 1 + \frac{\alpha}{\alpha + \beta} = 1 + \frac{1}{1 + \frac{\beta}{\alpha}} = 1 + \frac{1}{1 + \rho}$$

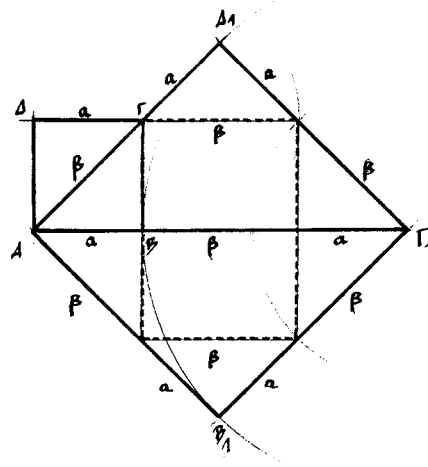


Figure 4: Geometrical interpretation of the irrational number $\sqrt{2}$

The relation $\rho = 1 + \frac{1}{1 + \rho}$ is of the form $x = f(x)$, i.e. the unknown x can be computed by a function f of its own. This kind of relation introduces an induction or a feedback, which results in an infinite process.

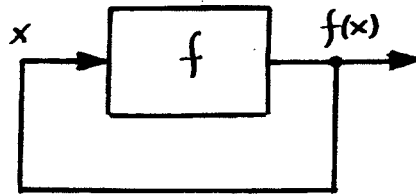


Figure 5: The infinite process of a feedback relation $x = f(x)$

The relation $x = f(x)$ entails the relation $x = f(f(x))$ and so on:
 $x = f(f(f...f(x)))$ or $x = f^x(x)$

The relation $\rho = 1 + \frac{1}{1 + \rho}$ leads to the infinite anthyphairetic relation:

$$\rho = \sqrt{2} = 1 + \frac{1}{1 + 1 + \frac{1}{1 + 1 + \frac{1}{1 + 1 + \dots}}} = 1 + \frac{1}{2 + \frac{1}{2 + \frac{1}{2 + \dots}}}, \text{ i.e. } \rho = [1, 2, 2, 2, \dots]$$

The geometrical representation of this infinite anthyphairetic process for the computation of $\sqrt{2}$ is given in the next figure.

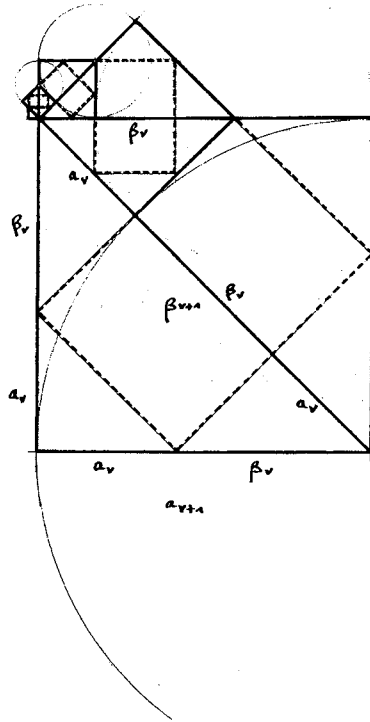


Figure 6: Infinite geometrical process of the irrational number $\sqrt{2}$

The successive approximations of the irrational number $\sqrt{2}$ are given in the following table.

$v=0$	$\rho_0 = [1] = 1$
$v=1$	$\rho_1 = [1, 2] = 1 + \frac{1}{2} = \frac{3}{2} = 1.5$
$v=2$	$\rho_2 = [1, 2, 2] = 1 + \frac{1}{2 + \frac{1}{2}} = \frac{7}{5} = 1.4$
$v=3$	$\rho_3 = \frac{17}{12} = 1.416666$
$v=4$	$\rho_4 = \frac{41}{29} = 1.4137913$
$v=5$	$\rho_5 = \frac{99}{70} = 1.4142857$
$v=6$	$\rho_6 = \frac{239}{169} = 1.4142011$

$$\begin{array}{rcl}
 v=7 & & \rho_7 = \frac{577}{408} = 1.4142156 \\
 \dots & & \dots \\
 & & \sqrt{2} = 1.4142136
 \end{array}$$

5. Summary

Plato develops the dichotomous division method based on the principle ‘the one is divided into two’ and creates a genealogical tree of all the concepts related to his subject. In other words, he tries to classify these concepts in successive levels, by means of an inductive reasoning process. He creates so a logical algorithm corresponding to this dynamical process that leads either from the general to the particular, from the result to the cause (analysis), or from the particular to the general, from the result to the cause (synthesis).

On the other hand, Euclid’s invention of the so-called anthyphairetic algorithms contributed to the exact mathematical approximation not only of commensurable but also of incommensurable quantities. These algorithms include the concept of closed loop mathematical equations and it is possible to be depicted geometrically.

References:

- [1] Fowler, D., *The Mathematics of Plato’s Academy, A New Reconstruction*, Clarendon Press, Oxford, 1999
- [2] Heath, sir Thomas, *The thirteen Books of Euclid’s Elements*, Dover Publications, Inc., New York, 1956
- [3] Hermann, W., *Philosophy of Mathematics and Natural Science*, Princeton University Press, 1949
- [4] Hutten, E., *The origins of Science*, G. Allen & Unwin Ltd., London, 1962
- [5] Klein, J., *Greek Mathematical Thought and the Origin of Algebra*, The M.I.T. Press, 1968
- [6] Kraut, R., *The Cambridge Companion to Plato*, Cambridge University Press, 1992
- [7] Stewart, J. A., *The Myths of Plato*, Centaur Press Ltd., London, 1960
- [8] Turnbull, H. W., *The Great Mathematicians*, Methuen & Co Ltd., 1951
- [9] Wedberg, A., *Plato’s Philosophy of Mathematics*, Greenwood Press, Publishers, Stockholm, 1955

Automatische Steuerung der Ultrafiltrationsrate bei Dialysepatienten

Roland Pilgram, Daniel Schneditz
Institut für Physiologie, Universität Graz, Österreich
e-mail: roland.pilgram@kfunigraz.ac.at

Abstrakt

Bei Dialysepatienten wird zur Wiederherstellung der Homöostase überschüssige Flüssigkeit aus dem Blutkreislauf entzogen. Durch das Entzugsprofil kann es zu Kreislaufinstabilitäten kommen was einen ärztlichen Eingriff erfordert. Daher werden Ultrafiltrationsprofile gefordert, die sich individuell abstimmen lassen.

Im einfachsten Fall lässt sich das dynamische Verhalten der Flüssigkeiten im vaskulären und im interstitiellen System durch ein 2-Tank System beschreiben. Wird aus dem zentralen Tank (vaskuläres System) Flüssigkeit durch Ultrafiltration (UF) entzogen, so beginnt allmählich Flüssigkeit vom peripheren Tank (interstitielles System) nachzufließen, bis sich ein neues Gleichgewicht einstellt, auch wenn der Entzug schon beendet ist. Dadurch kann es zwischenzeitlich, insbesondere gegen Ende des Entzugs, zu einem deutlich erniedrigten Flüssigkeitspegel (Blutvolumen) im zentralen Tank kommen. Da die Stabilität des Kreislaufs von diesem Blutvolumen abhängt, besteht die Möglichkeit, dass es bei zu starker Abnahme des Pegels im zentralen Tank zu Kreislaufkomplikationen kommt.

Die Parameter des 2-Tank Systems variieren von Patient zu Patient und werden durch eine Identifizierung aufgezeichneter Volumsänderungen im vaskulärem System, beruhend auf einer Sensitivitätsanalyse bestimmt. Im folgenden Schritt werden mögliche Ultrafiltrationsstrategien getestet um dann für die nächste Behandlung die für den Kreislauf am wenigsten belastende heranzuziehen. Neben Standard-Methoden, wie konstante, gepulste, und linear abnehmende UF-Rate werden mögliche weitere angeführt. Eine optimale Variante ist durch den linearen Übergang vom Ausgangs- in das Zielvolumen im zentralen Tank gegeben. Eine Möglichkeit ist die Flächenminimierung, die durch diesen und den tatsächlichen Verlauf der Blutvolumenänderung vorgegeben wird, allerdings unter bestimmten Einschränkungen: die maximale UF-Rate sowie die Dauer der UF sind bei einem fix vorgegebenen UF-Volumen begrenzt.

Die am häufigst angewendete Methode mit konstanter UFR zeigte eine Unterschreitung des Zielvolumens von ca. 4-5 %. Andere Methoden, im speziellen ein exponentiell abfallendes UF Profil weisen nur eine Unterschreitung von 1-2 % auf.

Mithilfe eines geeigneten Modells lassen sich für zukünftige Behandlungen durchaus individuell abgestimmte UFR Profile berechnen um Kreislaufinstabilitäten zu vermeiden.

Einleitung

Ein sehr wichtiges Organ in unserem Körper ist die Niere mit ihren beiden Hauptaufgaben, Entfernen von Wasser und Reinigen des Blutes von Stoffwechselendprodukten. Kommt es zu einer akuten oder terminalen Niereninsuffizienz, so ist die Folge die Einlagerung von jeglichem aufgenommenen Wasser im Gewebe und es bilden sich Ödeme. Die Konzentrationen von Harnstoff, Harnsäure und anderen Stickstoffverbindungen steigen stark an, und eine Übersäuerung des Körpers entsteht. Ohne Behandlung tritt der Tod nach 6 bis 7 Tagen ein.

Zur Zeit gibt es drei Verfahren zur Therapie des Nierenversagens.

1. Transplantation einer Spenderniere
2. Peritoneal (Bauchfell) – Dialyse
3. Hämodialyse (Blutwäsche)

Die Hämodialyse ist die zur Zeit am weitverbreiteste Therapieform des Nierenversagens. Ca. 85% aller Dialysepatienten werden mit dieser Therapie behandelt. Dem Patienten wird mittels einer arterio-venösen Kurzschlußverbindung (Gefäßzugang = operative Verbindung von Arterie und Vene) permanent Blut entzogen. Dieses wird durch einem extrakorporalen Kreislauf in den Dialysator und zurück zum Patienten geführt. Im Dialysator strömt das Blut im Gegenstrom zum Dialysat an einer Membran entlang, ein Stoffaustausch durch diese Membran findet statt. Die Membran bildet eine Sperre für Blutkörperchen, Fette, Viren und Bakterien, wo hingegen Wasser, Harnstoff und Salze ungehindert passieren können. Durch Ultrafiltration (UF) wird dem Patienten überschüssige Flüssigkeit (ca. 1-5L pro Behandlung) durch die Dialysatormembran hindurch entzogen. Die Behandlung durch Hämodialyse erfolgt meist dreimal pro Woche, zu je 3-4 Stunden.

Der Dialyseprozess stellt eine starke Kreislaufbelastung dar, was sich in der Häufigkeit kardiovaskulärer Komplikationen während der Hämodialyse widerspiegelt: die akute Blutdruckkrise ist die häufigste intradialytische Komplikation mit bis zu 25 % dar [1], die vorwiegend im letzten Drittel der Behandlung auftritt [2]. Aus der Krise kann der Patient nur durch einen sofortigen Eingriff des Pflegepersonals bzw. mittels einer Infusion befreit werden. Solche Blutdruckkrisen sind äußerst belastend für den Patienten, erhöhen die Mortalität und unterbrechen die Dialysebehandlung.

Sowohl das steigende Durchschnittsalter, als auch die ständigen Verbesserung in der Therapieführung, bewirken alljährlich einen Anstieg der Patientenzahlen [3].

Es ist abzusehen, dass in wenigen Jahren die Kosten für die nötigen Dialysebehandlungen von unserem Gesundheitssystem nicht mehr aufgebracht werden können. Dieser Umstand sowie die entstehenden Kosten für zusätzliche Betreuung bei Zwischenfällen [1] erklären das enorme Interesse, um Kreislaufkomplikationen während der Dialyse frühzeitig zu erkennen mit dem Ziel, sowohl deren Häufigkeit als auch Schweregrad herabzusetzen.

Jeder Dialysepatient reagiert individuell auf die Behandlung, deshalb sollte die Dialysebehandlung patientenadaptiv gestaltet werden! Die Modellparameter werden aus diesem Grund für jeden anders sein. In dieser Arbeit werden die Parameter eines 2-Tank Systems mittels eines Identifikationsprozesses an indirekt gemessenen Plasmavolumensänderungen beruhend auf einer Sensitivitätsanalyse [4] ermittelt. Die bestmögliche Ultrafiltrationsvariante wird dann durch Simulationsstudien gefunden, und ein optimales UF Profil könnte als Steuerung für die folgende Behandlung dienen.

Methoden

Modell: Für die Beschreibung des Flüssigkeitshaushaltes während der Ultrafiltration wurde ein Modell aus der Literatur [4,5] verwendet. Die aufgenommene Flüssigkeit, die durch ein Nierenversagen nicht mehr ausgeschieden wird, sammelt sich im Extrazellulärvolumen an. Dieses besteht im menschlichen Körper aus einem Plasmavolumen V_p (zentraler Anteil) und einem interstitiellen Volumen V_i (peripherer Anteil). Der Flüssigkeitsaustausch zwischen diesen beiden Tanksystemen wird in Abbildung 1 dargestellt.

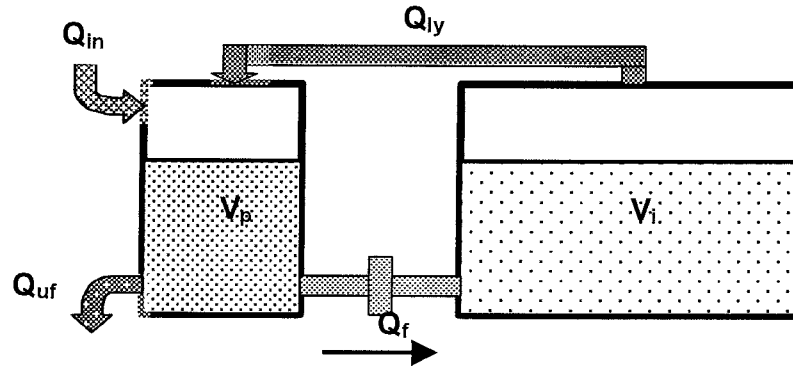


Abbildung 1: Tankmodell. Das extrazelluläre Volumen besteht aus dem Plasma Anteil (V_p , linker Tank) und einem extravaskulären Anteil (V_i , rechter Tank). Die kontinuierliche Filtration (Q_f) vom linken zum rechten Tank wird als kapillare Filtration bezeichnet. Die filtrierte Flüssigkeit wird über die Lymphbahnen (Q_{ly}) zurücktransportiert. Das Gleichgewicht des Systems kann sowohl durch Flüssigkeitsaufnahme (Q_{in}) als auch durch Flüssigkeitsentzug (Q_{uf}) gestört werden.

Im Normalfall wird Flüssigkeit vom Plasmavolumen V_p in das interstitielle Volumen V_i abfiltriert (Q_f) und gelangt über das lymphatische System Q_{ly} wieder zurück in das Plasma Kompartiment. Im Gleichgewichtszustand sind Q_f und Q_{ly} identisch. Beim Dialysepatienten ist dieser Zustand sowohl durch Flüssigkeitsaufnahme Q_{in} als auch Abgabe Q_{uf} gestört. Das Gesamtsystem lässt sich mittels folgenden Differentialgleichungen beschreiben:

$$\dot{V}_p = -Q_f + Q_{ly} - Q_{uf} + Q_{in} \quad (1)$$

$$\dot{V}_i = Q_f - Q_{ly} \quad (2)$$

Die mikrovaskuläre Filtration Q_f wird durch den Starling-Mechanismus [6] beschrieben, bestehend aus 4 Kräften: den hydrostatischen Kräften p_c und p_i und den osmotischen Kräften π_p und π_i :

$$Q_f = K_f(p_c - p_i - \pi_p + \pi_i) \quad (3)$$

Die Leitfähigkeit der Kapillarmembran ist durch den Filtrationskoeffizient K_f gegeben. Die hydrostatischen Kräfte sind nichtlineare Druck-Volumenbeziehungen, und die osmotischen Kräfte resultieren aus der Proteinkonzentration im jeweiligen Kompartiment (siehe Anhang).

Identifikation: Die Parameteridentifikation beruht auf der Sensitivitätsanalyse des nichtlinearen Modells. Es hat sich gezeigt, dass verschiedene Modellparameter zu

verschiedenen Zeitzuständen eine maximale Sensitivität haben. Die zugehörigen Messpunkte (Tabelle 1) wurden für die Identifizierung herangezogen [4].

S_N [L]	c_p [g/L]	V_p [L]	Hct [%]	c_i [g/L]	V_i [L]	K_f [L/min.mmHg]	f
Max	2.47	2.18	-0.58	-0.29	-0.29	0.14	-1.01
Zeitpunkt	Beginn UF	Beginn UF	Beginn UF	Ende der Behandlung	Ende der Behandlung	Beginn & Ende UF	Beginn UF
Mittelwert	2.28	2.04	-0.43	-0.23	-0.23	0.08	-0.66

Tabelle 1: Normalisierte Sensitivität (S_N). Es gibt drei Zeitbereiche, in denen das Modell für unterschiedliche Parameter sehr sensitiv ist: Am Beginn der Behandlung für Plasmaproteinkonzentration (c_p), Plasmavolumen (V_p), Hämatokrit (HCT), Filtrationskoeffizient (K_f) und Steigung in der Druck-Volumen Beziehung (f); am Ende der Ultrafiltration für den Filtrationskoeffizient (K_f), und nach dem Ende der Behandlung, nach dem Rebound für die Proteinkonzentration im Interstitium (c_i) und für das interstitielle Volumen (V_i), während der sogenannten Reboundphase.

Für die Identifikation wurden die Messwerte herangezogen, die durch die maximale Sensitivität der Parameter bestimmt wurden. Verwendet wurde dazu die Matlabfunktion „fminsearch“, die auf dem Verfahren von Nelder-Mead beruht [7,8].

Ultrafiltrationsprofile: Um eine vorgegebene Menge an Flüssigkeit zu entziehen wurden folgende UF Profile gewählt:

- (a) *konstante Ultrafiltrationsrate:* Es wird über den ganzen Zeitraum hinweg dieselbe Menge an Flüssigkeit entzogen.
- (b) *gepulste Ultrafiltrationsrate:* Es wird abwechselnd Flüssigkeit entzogen und dem Körper eine Erholungsphase gegönnt.
- (c) *linear abnehmende Ultrafiltrationsrate:* Zu Beginn der Behandlung wird dem Patienten am meisten entzogen, jedoch linear abnehmend mit der Behandlungsdauer.
- (d) *exponentiell abnehmende Ultrafiltrationsrate:* Zu Beginn der Behandlung wird ebenfalls am meisten Flüssigkeit entzogen, jedoch in diesem Fall exponentiell abnehmend mit der Behandlungsdauer.
- (e) *sigmoide Ultrafiltrationsrate:* Die Entzugsrate nimmt langsam zu, erreicht nach ca. 30 Minuten das Maximum und nimmt anschließend stetig ab. Mathematisch kann das UF Profil durch die Kombination zweier sigmoider Funktionen (4) beschrieben werden:

$$f(t) = \frac{1}{1 + e^{-\alpha(t-\beta)}} \quad (4)$$

- (f) UF Profil berechnet durch ein Flächenfunktional:

Wie in Abbildung 2 erkennbar, schließt $V_p(t)$ mit $V_d(t)$ eine Fläche ein. $V_d(t)$ ist der optimale Übergang vom Ausgangsvolumen ins Zielvolumen. Durch die Minimierung dieser Fläche könnte man durch Variation der UF Rate einen Angleich an die ideale Plasmavolumensänderung erzielen. Definiert wird das Funktional $f(\text{UFR})$ mit UFR, der Ultrafiltrationsrate als Veränderliche folgendermaßen (5):

$$f(UFR) = \min \int_{t=0}^{t=t_e} |V_p(t) - V_{des}(t)|^2 dt \quad (5)$$

mit der Nebenbedingung

$$0 \leq u(t) \leq UFR_{\max} \quad (6)$$

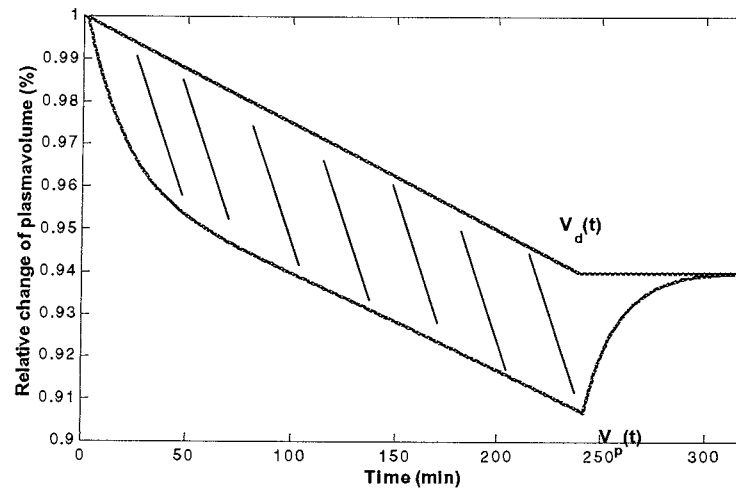


Abbildung 2: Ideale und reale Plasmavolumsänderung: Die ideale Plasmavolumsänderung $V_d(t)$, die vom Ausgangsvolumen linear abfallend zum Zielvolumen übergeht, schließt mit einer typisch gemessenen Plasmavolumsänderungen $V_p(t)$ eine Fläche (schraffierter Anteil) ein.

Für die Berechnung der Zielfunktion $f(UFR)$ wurde das Quasi-Newton-Verfahren verwendet. Zur Bestimmung einer Suchrichtung wird neben den ersten partiellen Ableitungen der Zielfunktion auch die Hesse-Matrix der zweiten Ableitungen herangezogen. Die Berechnung der Hesse-Matrix erfolgt approximativ durch Matrix-Aufdatierungsformeln, in diesem Fall mittels der BFGS-Formel.

Während die Ultrafiltrationsprofile (a) (b) (c) (d) bereits Standard Methoden sind wurde das UFR Profil (e) neu dazugenommen und das UFR Profil (f) als Ergebnis einer Optimierungsvariante verwendet.

Anhand von Modellparameter aus der Literatur (siehe Anhang) wurde der Zustand für eine Überwässerung von 4 Litern berechnet. Innerhalb von 4 Stunden sollte mit allen erwähnten UF Profilen das überschüssige Volumen entzogen werden. Im Anschluss daran werden die Parameter durch identifizierte, von einem indirekt gemessenen Plasmavolumsverlauf, ersetzt und mit Hilfe der verschiedenen UF Profile die unterschiedlichen Varianten ausgetestet.

Ergebnisse

Unter der Verwendung von plausiblen Modellparametern aus der Literatur (siehe Anhang) für das 2-Tank System, wurden Simulationsstudien mit verschiedenen Ultrafiltrationsraten durchgeführt. In Abbildung 3 sind die verschiedenen Ultrafiltrationsprofile (ufr) und die resultierenden Plasmavolumsänderungen ($V_{p,rel}$) dargestellt.

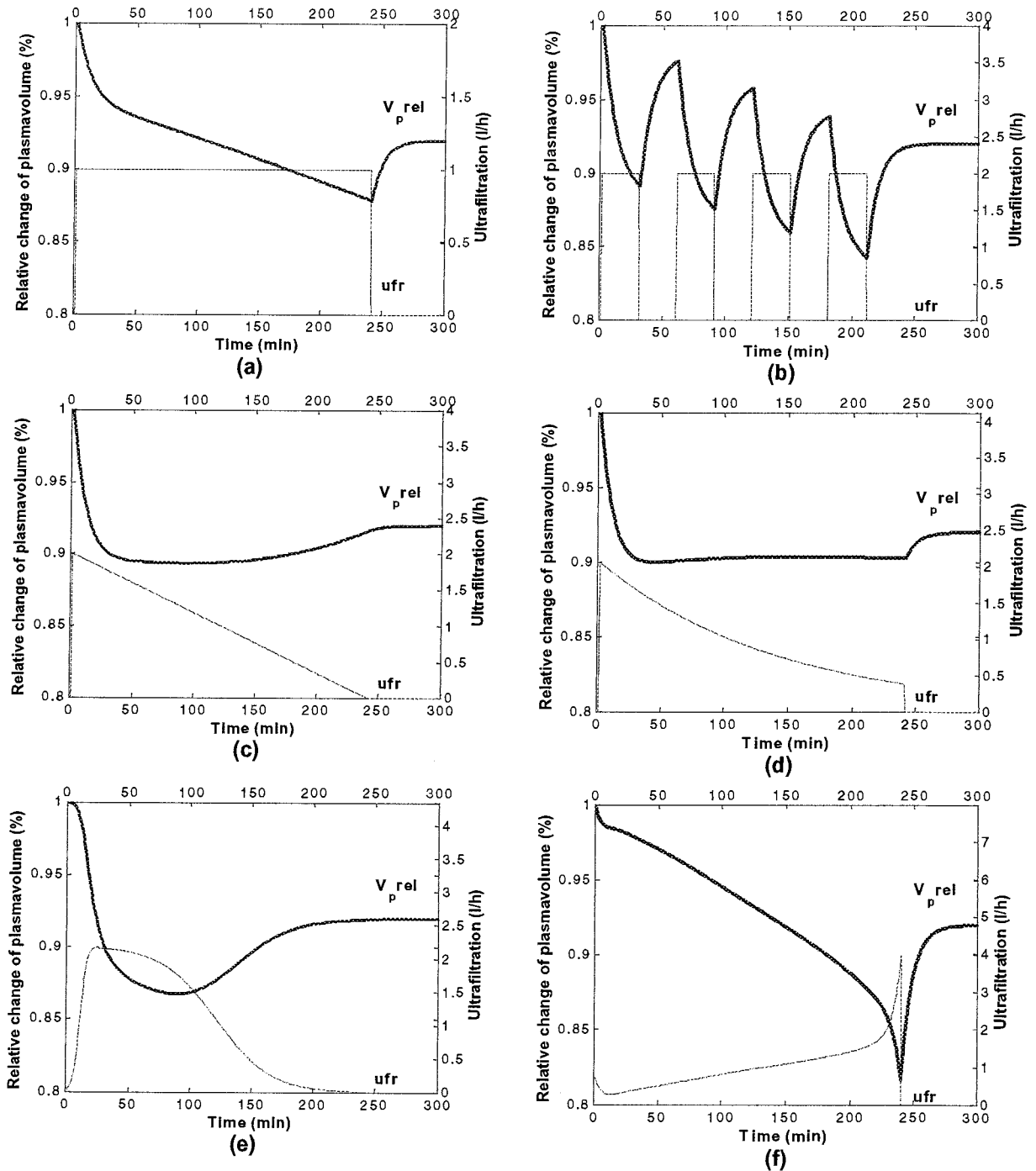


Abbildung 3: Plasmavolumsänderungen ($V_{p,rel}$) aufgrund verschiedener Ultrafiltrationsprofile (ufr). Konstante UFR (a), gepulste UFR (b), linear abnehmende UFR (c), exponentiell abnehmende UFR (d), sigmoide UFR (e) und flächenoptimierte UFR (f).

Um eine typische Dialysebehandlung nachzustellen, werden innerhalb von 4 Stunden 4 Liter Volumen entzogen. Jedes UFR Profil bewirkt also innerhalb der vorgeschriebenen Behandlungszeit den selben Volumsentzug. Durch die verschiedenen Profile ist jedoch das zwischenzeitliche Minimum unterschiedlich lang und tief ausgeprägt. Wie hoch die Unterschreitung des Zielvolumens von 92 % des Plasmaanfangsvolumens ist wird in Tabelle 2 in % angegeben.

	konstante UFR	gepulste UFR	linear abnehmende UFR	exponentiell abnehmende UFR	sigmoide UFR	flächenoptimierte UFR
[%]	-4.2	-7.7	-2.7	-2	-4.2	-11

Tabelle 2: Angabe der Unterschreitung des Zielvolumens von 92 % der verschiedenen UFR Profile.

Abbildung 4 zeigt einen typisch gemessenen Verlauf einer Plasmavolumensänderung (durchgezogene Linie) aufgrund einer gepulsten Ultrafiltration. Für die Parameteridentifizierung des 2-Tank Systems wurden die durch die Kreise hervorgehobenen Punkte verwendet. Die punktierte Linie gibt die Plasmavolumensänderung des identifizierten Modells wieder.

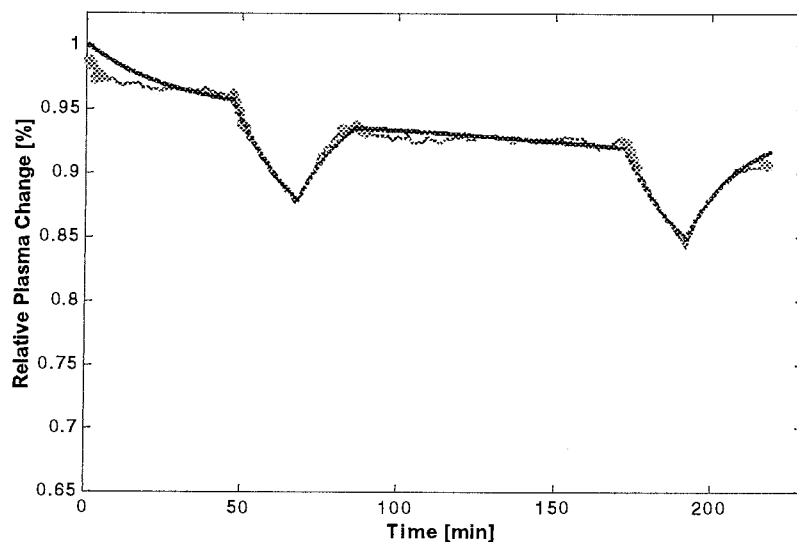


Abbildung 4: Relative Änderung des Plasmavolumens: Beispiel einer gemessenen (durchgezogene Linie) und identifizierten (punktierte Linie) Plasmavolumensänderung während einer Ultrafiltration. Die offenen Kreise beschreiben die maximale Sensitivität des Systems und werden für die Identifikation herangezogen. Das Patientenanzugsgewicht betrug 62.8 kg und Zielgewicht betrug 60.0 kg. Ein Volumen von 2.8 Liter wurde durch gepulste UFR innerhalb von 240 Minuten entzogen.

Hct(0)	$V_p(0)$	$V_i(0)$	$C_p(0)$	$C_i(0)$	K_f	f
27.51	5.11	14.50	62.18	14.38	0.0057	1.30

Tabelle 3: Identifizierte Anfangszustände und Modellparameter: Hämokrit ($Hct(0)$), Plasmavolumen ($V_p(0)$), Interstitielles Volumen ($V_i(0)$), Plasmaproteinkonzentration ($C_p(0)$), Interstitielle Proteinkonzentration ($C_i(0)$), kapillarer Filtrationskoeffizient (K_f) und die Steigung der Druck-Volumsbeziehung im vaskulären System (f).

Die identifizierten Modellparameter sind in Tabelle 3 angegeben. Mithilfe der gefundenen Parameter wird in weiterer Folge die Änderung des Plasmavolumens aufgrund unterschiedlicher UFR Profile berechnet (Abbildung 5).

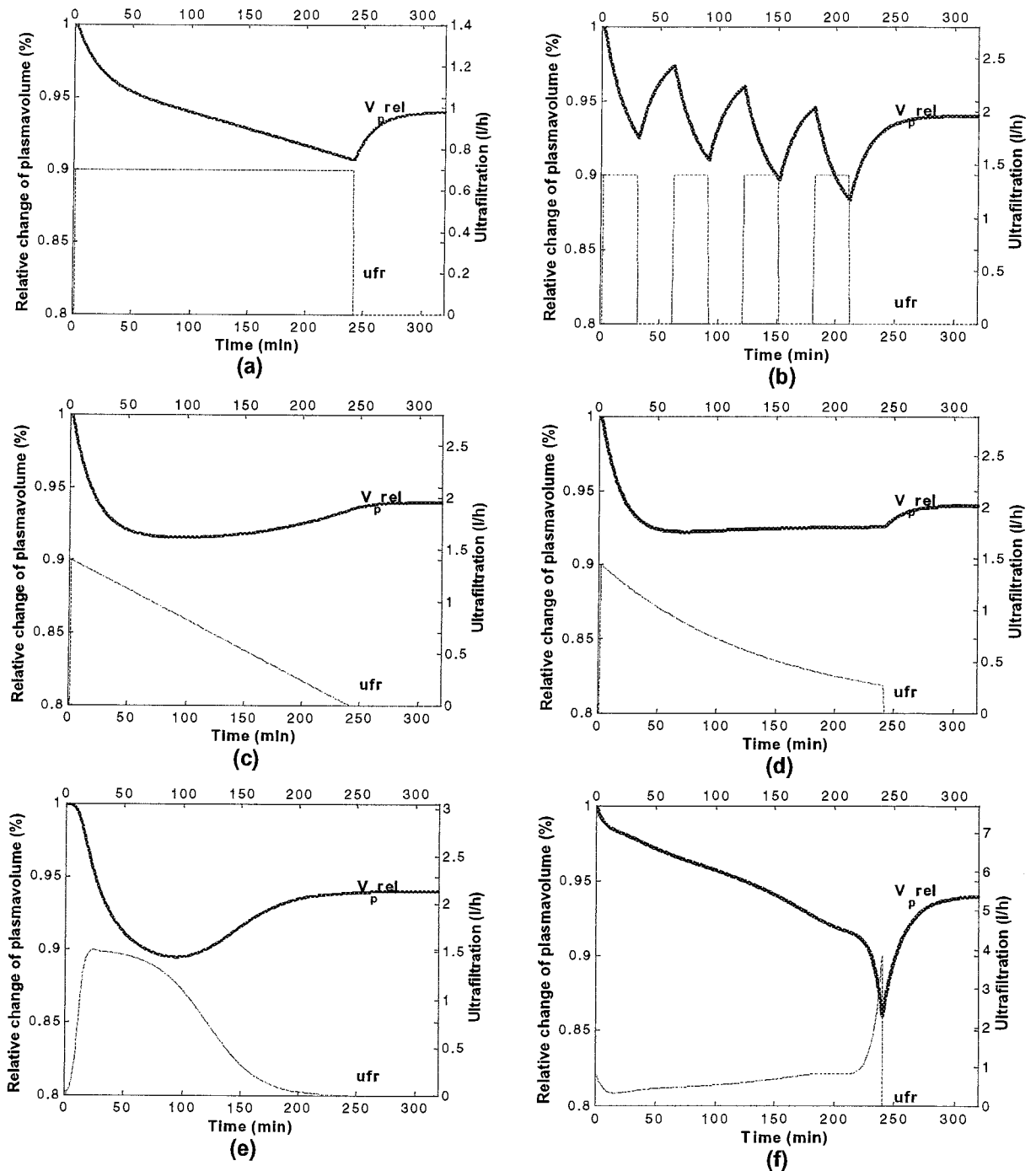


Abbildung 5: Plasmavolumensänderungen (V_{rel}), beruhend auf identifizierten Patientendaten, aufgrund verschiedener Ultrafiltrationsprofile (ufr). Konstante UFR (a), gepulste UFR (b), linear abnehmende UFR (c), exponentiell abnehmende UFR (d), sigmoide UFR (e) und flächenoptimierte UFR (f).

Auch in diesem Fall ergeben sich zwischenzeitliche Minima. Wie hoch an diesem, dem Patienten angepassten Modell, die Unterschreitung des Zielvolumens von 94 % des Plasmaanfängsvolumens ist wird in Tabelle 4 in % angegeben.

	konstante UFR	gepulste UFR	linear ab- nehmende UFR	exponentiell abnehmende UFR	sigmoide UFR	flächenoptimierte UFR
[%]	-4	-6	-3	-1.8	-4.5	-9

Tabelle 4: Angabe der Unterschreitung des Zielvolumens von 94 % der verschiedenen UFR Profile basierend auf einem Patientenangepassten 2-Tank Modells.

Diskussion

Anhand eines 2-Tank Systems, dessen Parameter einerseits durch Literaturangaben andererseits durch Identifizierung festgelegt wurden, wurde mithilfe von verschiedenen Ultrafiltrationsprofilen die Plasmavolumensänderung simuliert.

Die konstante UFR ist derzeit das meistgefahrenste Profil. Es ist ein sehr einfaches zu berechnendes Verfahren, führt aber bei 25 % der Behandlungen zu Kreislaufinstabilitäten [1]. Eine Unterschreitung des Zielvolumens von ca. 4 % ist also zu groß, um Kreislaufinstabilitäten für jedermann zu vermeiden. Werden hingegen gepulste UFR verwendet, so nimmt man zwar dem Patienten zwischenzeitlich das doppelte an Volumen im Verhältnis zur konstanten UFR, gönnt dem Kreislauf jedoch dann immer wieder Erholungsphasen. Mit ca. bis zu 7 % Unterschreitung ist diese Variante jedoch auch eine enorme, zwar nur kurze Belastung für das Kreislaufsystem. Die linear abnehmende UFR Variante eignet sich schon besser als die vorangegangenen Methoden. Eine negative Abweichung von ca. 3% vom Zielvolumen ist schon eine Verbesserung. Ein weiterer Aspekt, der für diese Variante spricht ist das nicht mehr Vorhandensein eines Reboundeffekts am Ende der Dialyse. Ersetzt man hingegen dieses Profil durch ein exponentiell abnehmendes mit einem geringen konstanten UF Anteil, so ergibt sich nur mehr eine ca. 2% Abweichung vom Zielvolumen. Dies wäre eine optimale Strategie für ein UFR Profil.

Der starke Volumsabfall zu Beginn der Dialyse kann jedoch auch in seltenen Fällen eine Kreislaufkrise auslösen. Aus diesem Grunde wurde noch eine andere Variante, ein UFR Profil, bestehend aus zwei sigmoiden Funktionen getestet. Durch das relativ langsame Anfahren der UF kommt es nicht zu dem scharfen Abfall des Volumens, jedoch kann es auch in diesem Fall zwischenzeitlich zu einer Unterschreitung des Zielvolumens von bis zu 4.5 % kommen. Dies würde in etwa dem Minima bei konstanter UFR entsprechen. Auch in diesem Fall ist kein Rebound zu beobachten. Die Idee, die Fläche, die zwischen den optimalen und gefahrenen Volumen eingeschlossen wird, durch die Einstellung der UFR zu minimieren erwies sich leider als Irrweg. Zum einen wird die meiste Flüssigkeit am Ende entzogen und zwar mit einer sehr hohen UFR, zum anderen liegt das Minimum mit -11 % Abweichung deutlich unter dem Zielvolumen.

Der vorerst nur theoretische Ansatz von Identifikation, mit nachfolgender optimaler UFR Bestimmung, könnte durchaus die Dialysebehandlung patientenschonender gestalten, müsste jedoch davor ausreichend in Probeläufen getestet werden.

Danksagung

Ein herzliches Dankeschön gilt Herrn Dr. Ring für die Ausprogrammierung des Optimierungsverfahrens sowie Herrn Prof. Hofer für die Hilfestellung. Diese Arbeit wurde vom Sonderforschungsbereich SFB, Projektnummer F 323 finanziell unterstützt.

Referenzen

- [1] Amerling R, Cu GA, Dubrow A, et al. Complications During Hemodialysis. In: Nisenson AR, Fine RN, Gentile DE (Eds.): Clinical Dialysis (3rd ed.). Prentice-Hall International, London (1995)
- [2] Santoro A, Mancini E, Spongano M et al. A haemodynamic study of hypotension during haemodialysis using electrical bioimpedance cardiography. *Nephrol Dial Transplant* 5 (Suppl 1): 147-153 (1990)
- [3] Hörl WH, de Alvaro F, Williams PF. Healthcare systems and end-stage renal disease (ESRD) therapies – an international review: access to ESRD treatments. *Nephrol Dial Transplant* 14 (Suppl 6): 10-15 (1999)
- [4] Pilgram R, Ring W, Schneditz D. Optimized Identification of Fluid Volume and Hematocrit during Ultrafiltration; Special Research Center F003. Report No. 220 – May (2001)
- [5] Chamney PW, Johnner C, Aldridge M, Krämer N, Valasco JE, Tattersall T, Au-kaidey R, Gordon and RN Greenwood. Fluid balance modelling in patients with kidney failure. *J Med Eng Technol* 23: 45-52 (1999)
- [6] Starling EH, On the absorption of fluids from the connective tissue spaces. *J Physiol (London)* 19 (1896)
- [7] Khoo MC, Physiological control systems. New York, IEEE Press. (2000).
- [8] Nelder JA, and R. Mead. A simplex method for function minimization, *Computer Journal*, Vol. 7, 308 (1965)

Anhang

Druck-Volumsbeziehungen

$$p_i(t) = aV_i(t) + \frac{b}{V_i(t) + c} \quad (7)$$

$$p_v(t) = d(V_p(t) + V_{RBC} + e)^f \quad (8)$$

$$p_c = p_v + p_{off} \quad (9)$$

Lymphfluss

$$Q_{ly} = g \tanh(hp_i) + i \quad (10)$$

Osmotischen Kräfte

$$\pi = k_1[C] + k_2[C]^2 + k_3[C]^3 \quad (11)$$

Modelparameter

Parameter	Values
a	0.05
b	-17.50
c	-6.50
d	0.80
e	-1.20
f	1.50
g	0.02
h	0.63
i	0.02
k1	0.2100
k2	0.0016
k3	9e-6
K _f [L/min*mmHg]	0.0057
V _{RBC} [L]	2

Initialisierungsparameter

Variable	Normal	Überwässert (4 Liter)
V _p [L]	3	3.25
V _i [L]	12	15.75
c _p [g/L]	70	64
c _i [g/L]	17.5	13.21
Hct [%]	40	38
p _{offset} [mmHg]	13.2	13.2
p _i [mmHg]	-2.6	-1.10
p _v [mmHg]	5.92	6.52
π _i [mmHg]	4.21	3.07
π _v [mmHg]	25.63	22.35

Servicerobotik – Umgebungserfassung und Bewegungssteuerung für einen mobilen Manipulator

Thomas Wösch und Werner Neubauer

Corporate Technology, Information & Communications
Siemens AG
Otto-Hahn-Ring 6
D-81730 München
{thomas.woesch,werner.neubauer}@mchp.siemens.de

Zusammenfassung

Im Bereich der mobilen Servicerobotik werden an die eingesetzten Robotersysteme Anforderungen gestellt, die sich von denen herkömmlicher Industrieroboter unterscheiden. Unter anderem ist der Arbeitsbereich von Servicerobotern im allgemeinen weitgehend unbekannt und kann sich über ein großes Gebiet erstrecken. Umgebungswahrnehmung und Bewegungssteuerung müssen trotz unbekannter Umgebung eine gewisse Sicherheit vor Kollisionen bieten bzw. über einen Mechanismus verfügen, der in der Lage ist, auftretende Kollisionen zu erkennen und entsprechend zu handeln. In der Servicerobotik sind Interaktionsmöglichkeiten zwischen Mensch und Maschine erwünscht, die deutlich über die derzeit verfügbaren hinausgehen. In diesem Beitrag werden ein *Kollisionsvermeidungssystem* und *Kontaktreflexsystem* vorgestellt, die zusammen mit einer „*künstlichen Haut*“ die unmittelbare Interaktion zwischen Mensch und Maschine im gleichen Arbeitsbereich ermöglichen. Die berührungslose und taktile Sensorik des Systems dient der Umgebungswahrnehmung und trägt darüber hinaus zur Sicherheit des Benutzers bei. Beispielhafte Ergebnisse werden mit Hilfe einer Simulation und anhand von Experimenten mit einem realen Roboter gezeigt.

1 Einleitung

Im Zuge der industriellen Automatisierung war man in den letzten Jahren zunehmend bestrebt, dortige Ergebnisse in den alltäglichen Bereich zu transportieren. Obwohl in einem Großteil industrieller Fertigungsanlagen bereits Roboter im Einsatz sind, die ihre Aufgabe erfolgreich lösen, geht deren Einzug in den Privathaushalt oder in das öffentliche Leben doch relativ langsam

voran. Man sieht relativ schnell, wodurch dies begründet sein mag, wenn man die unterschiedlichen Einsatzgebiete und Aufgabenbereiche betrachtet.

Für die meisten automatisierten Anlagen können im Bezug auf den Arbeitsbereich Annahmen gemacht werden, die den erfolgreichen Einsatz eines Roboters wesentlich vereinfachen. Beispielsweise ist in vielen Fällen die Einsatzumgebung bekannt und statisch. In dynamischen Umgebungen ist die Position der bewegten Objekte bzw. Roboter *a priori* zu jedem Zeitpunkt bekannt. In vielen Fällen werden Industrieroboter für Aufgaben eingesetzt, die robust, ständig wiederholend und exakt durchgeführt werden sollen. Die Bewegungsabläufe sind in den meisten Fällen immer die gleichen, eine Abweichung davon ist im allgemeinen unerwünscht und kann zu Beschädigungen des Roboters und der Umgebung führen. Die Eingabe einer Bewegungstrajektorie kann einmalig über einen Benutzer erfolgen. Wenn die Planung für den Menschen zu komplex wird, können automatisierte Planungsverfahren herangezogen werden. Mit entsprechenden Regelungen können die ausgeführten Bewegungen energetisch oder zeitlich optimiert werden. Um einen sicheren, robusten Bewegungsablauf zu ermöglichen, kann der Arbeitsbereich der Maschinen in industriellen Umgebungen in vielen Fällen verändert werden. So ist das Anbringen von Sensoren oder Landmarken durchaus üblich, damit Zustände im Arbeitsbereich sicher erkannt werden. Beispielsweise kann die Umgebung im Hinblick auf sicherheitstechnische Maßnahmen entsprechend gestaltet werden, sodaß eine Beschädigung der Roboterumgebung bzw. letztendlich eine Verletzungsgefahr für den Menschen, weitgehend ausgeschlossen werden kann. In industriellen Fertigungsanlagen spielt die „Erscheinungsform“ der Maschinen keine besondere Rolle. Nicht-mobile Roboter können entsprechend massiv gebaut werden, falls dies erforderlich ist. Die Interaktionsmöglichkeiten zwischen Mensch und Maschine gehen in der Industrierobotik nur in den wenigsten Fällen über konventionelle Möglichkeiten wie Tastatur und Bildschirm hinaus. Es ist offensichtlich, daß ein industrieller Roboter mit den genannten Eigenschaften, nicht für den Umgang mit Menschen im Alltag geeignet ist.

Serviceroboter sind, wie ihr Name schon vermuten läßt, Roboter, die einem Anwender bestimmte Dienste (engl.: Services) im alltäglichen Einsatz zur Verfügung stellen. Im Prinzip führen auch Industrieroboter Aufgaben für einen Anwender aus. Der Unterschied der Serviceroboter zu den Industrierobotern, liegt im wesentlichen in deren Einsatzgebiet. Bekannte „Arbeitszellen“ gibt es bei den Servicerobotern im Prinzip nicht. Es ist zwar möglich, daß entsprechendes Vorwissen über die Umgebung teilweise bekannt ist, trotzdem kann das Auftreten unbekannter Hindernisse hier nicht ausgeschlossen werden. Als Beispiel hierzu, betrachte man den erfolgreich eingesetzten Reinigungsroboter [1]. Der Dienst, den dieser zur Verfügung stellt, ist

in erster Linie das Reinigen von Supermärkten. Nach einer Lernphase besitzt der Roboter ein Wissen über die statische Umgebung, dh. er weiß wo Wände und Gänge sind. Im Einsatz muß der Roboter aber auch mit anderen Objekten, insbesondere Menschen, die sich in seiner Umgebung befinden, zurechtkommen. Zeitlich fest vorgeschriebene Bewegungsabläufe kann es hier nicht geben. Um Kollisionen mit der Umgebung zu verhindern, muß der Roboter Bewegungen einleiten, die nicht vorprogrammiert sind. Bei dem Einsatz von Servicerobotern soll die Umgebung möglichst unverändert bleiben. Serviceroboter müssen so ausgestattet werden, daß diese ohne spezielle Modifikation der Umwelt autonom arbeiten können. Nicht nur technische Aspekte spielen bei Servicerobotern eine Rolle. Deren Bewegungsausführung muß für den Menschen vorhersehbar und plausibel sein, um die Akzeptanz solcher Systeme zu gewährleisten. Auch deren Erscheinungsform ist von Interesse, wenn diese im Alltag operieren. Beispielsweise würden sich die wenigsten Personen wohlfühlen, wenn sie im Supermarkt neben einem massiven, bedrohlich wirkenden Roboter einkaufen müßten. Die Maschine sollte sich möglichst unauffällig in den Alltag „eingliedern“. Dazu sind aber auch fortgeschrittenere Interaktionsmöglichkeiten zwischen Mensch und Maschine, wie beispielsweise Sprach- oder Gestikerkennung, von besonderem Interesse.

In diesem Beitrag wird die Umgebungswahrnehmung und Bewegungssteuerung eines Roboters betrachtet, der imstande sein soll, in einer Alltagsumgebung von durchschnittlicher Komplexität verschiedene Aufgaben, wie beispielsweise Hol- und Bringdienste durchzuführen.

2 Motivation

Für die Bewegungssteuerung von Robotern wurden in den letzten zehn Jahren bevorzugt probabilistische Planer entwickelt [2, 3, 4]. Planer dieser Klasse zeichnen sich besonders durch die Eigenschaft aus, verhältnismäßig einfach auf höherdimensionale Problemstellungen erweiterbar zu sein. Die Beliebtheit dieser Planer ist in der steigenden Effizienz der Algorithmen im Bereich der Computergrafik zur Berechnung von Abstandsinformationen [5, 6] aus einem Umgebungsmodell und in der steigenden Rechnerleistung begründet. Trotz der ständig schneller werdenden Planer benötigt man für eine erfolgreiche Bewegungsplanung ein System, welches auf unvorhergesehene Hindernisse reagiert. Um eine Bewegungsplanung zu ermöglichen, ist ein vollständig bekanntes Umgebungsmodell erforderlich. Im Bereich der Servicerobotik kann jedoch im allgemeinen mit der eingeschränkten Sensorik die Umgebung nur unvollständig wahrgenommen werden. Die Umgebung ist, wie im einleitenden Kapitel bereits erwähnt, zudem in den meisten Fällen dynamisch,

so daß wiederholtes Neuplanen erforderlich wäre. Dabei können Kollisionen unter Umständen nicht in Echtzeit vermieden werden. Eine Möglichkeit, dieses Problem zu lösen, wird in [7] vorgeschlagen. Dabei wird ein Bewegungsplan generiert, der bei der Ausführung einen lokalen reaktiven Handlungsspielraum zuläßt. Im vorliegenden Beitrag wird ein rein reaktiver Ansatz zur Bewegungssteuerung innerhalb dieses Handlungsspielraumes vorgeschlagen. Ähnlich der Potentialfeldmethode werden dabei zielführende und kollisionsvermeidende Bewegungsvorgaben überlagert. Für sehr viele Situationen erzeugt dieses Vorgehen bereits das gewünschte Roboterverhalten. Da die Potentialfeldmethode jedoch zu lokalen Minima neigt, wird dem System zu einem späteren Zeitpunkt noch eine Planungskomponente hinzugefügt.

Zur Erzeugung der kollisionsvermeidenden Bewegungsvorgaben verwendet die vorgeschlagene Steuerung ein 3D-Modell der Umgebung, welches aus den Sensordaten eines Laserscanners berechnet wird. Sobald sich der Roboter einem Objekt im Umgebungsmodell unzulässig nähert, ermittelt ein *Kollisionsvermeidungssystem*, basierend auf den berechneten Abstandsinformationen, eine Bewegung, die den Arm in einen sicheren Abstand relativ zu den Hindernissen führt. Aufgrund der hohen Dynamik der Roboterumgebung und der zum Teil großen Abschattung des Sensorsichtbereiches durch den eigenen Roboterarm ist das Umgebungsmodell unvollständig. Deshalb ist das gesamte Robotersystem zusätzlich mit taktilen Sensoren umgeben, siehe Abbildung 1 und 4. Diese haben die Funktion einer „künstlichen Haut“¹. Sollte es zu einem oder mehreren Kontakten zwischen Hindernissen und dem Roboter kommen, leitet ein *Kontaktreflexsystem* abhängig von Ort und Anpreßdruck der Kontakte eine entsprechende Ausweichbewegung ein.

Mit der hier vorgestellten Bewegungssteuerung und der taktilen Sensorik zur Umgebungswahrnehmung, steht eine Interaktionsmöglichkeit zwischen Mensch und Robotersystem zur Verfügung, die deutlich über derzeit verwendete hinausgeht, indem es dem Menschen erlaubt ist, in direkten Kontakt mit dem Roboter zu treten und diesen entsprechend zu „führen“.

3 Experimentalsystem und Simulationsumgebung

Das Robotersystem besteht aus einem redundanten Roboterarm mit 8 Freiheitsgraden, der auf einer nicht-holonomen mobilen Plattform mit Differentialantrieb montiert ist. Am Kopf des Experimentalsystems befinden sich ein Laserscanner und eine Stereokamera. Beide Sensoren liefern die benötigten

¹Siemens-Patent, amtl. Aktz. 19959703.0

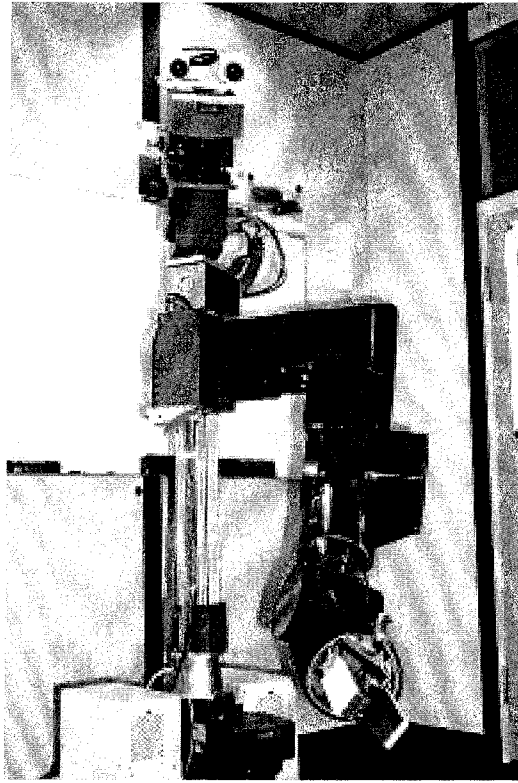


Abbildung 1: Experimentalsystem

3D-Tiefeninformationen. Der Arm des Roboters und die Plattform sind mit taktilen Sensoren ausgestattet. Für eine detailliertere Beschreibung des Experimentalaufbaus und des gesamten Steuerungsansatzes siehe [8].

Die Simulationsumgebung ermöglicht unter anderem, dem Benutzer interaktiv in eine künstlich generierte Roboterumgebung einzugreifen, beispielsweise durch das Verschieben von Objekten. Abbildung 2 zeigt einen *Screenshot* einer Simulation. Die darin abgebildeten Objekte sind hinsichtlich Geometrie maßstabsgetreu den in unserem Labor befindlichen nachgebildet. Der dargestellte Roboter ist stark vereinfacht, wie ein Vergleich mit dem realen System aus Abbildung 1 zeigt.

4 Umgebungserfassung

Für die Erstellung und Aktualisierung des Umgebungsmodells stehen drei Informationsquellen zur Verfügung. Zunächst hat der Roboter Kenntnis über die eigene Position in der Umgebung und über die Position des Roboter-

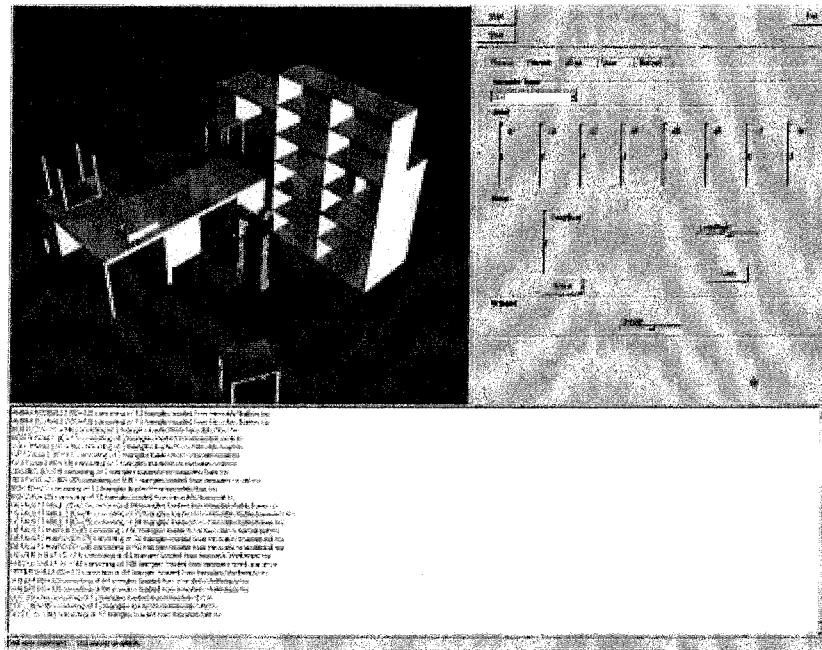


Abbildung 2: Interaktive Simulationsumgebung

armes. Mit Hilfe der bekannten Robotergeometrie kann das Kollisionsvermeidungssystem Selbstkollisionen ausschließen. Das Navigationssystem des Roboters [9] besitzt eine zweidimensionale Karte der Umgebung. Diese Karte wird entweder vorgegeben, oder der Roboter erstellt und aktualisiert diese selbst. Wände und andere Objekte dieser Karte können zusammen mit der Robotergeometrie ebenfalls in das dreidimensionale Umgebungsmodell übernommen werden. Die dritte Informationsquelle zur Aktualisierung des Umgebungsmodells sind alle Sensoren, also der Laserscanner, die Stereokamera und die taktile Haut.

Der verwendete Laserscanner (Abbildung 3) kann die Umgebung nur in einer Ebene wahrnehmen. Für eine dreidimensionale Szenenansicht wird der Scanner mit einem Servoantrieb über die Szene geschwenkt. Aus den Sensorrohdaten wird ein 3D-Oberflächenmodell der Szenenansicht generiert.

Die Stereokamera (Abbildung 3) liefert wesentlich schneller ein dreidimensionales Bild der Szene, da sie nicht geschwenkt werden muß. Die Tiefendaten der Stereokamera sind ohne den Einsatz von strukturiertem Licht jedoch dünn besetzt, falls die Objekte nicht genügend Struktur aufweisen.

Die taktile Haut (Abbildungen 4-5) dient im wesentlichen dazu, bei unvollständigem Umgebungswissen die Sicherheit für Mensch und Maschine zu gewährleisten. Jedes Flächenstück der Haut liefert dazu Position und An-

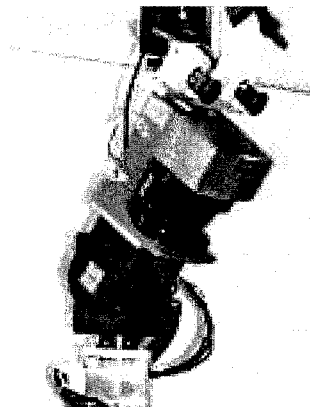


Abbildung 3: Laser-Scanner und Stereokamera

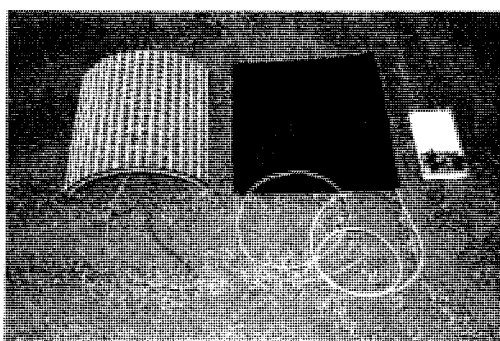


Abbildung 4: Einzelteile eines taktilen Sensors

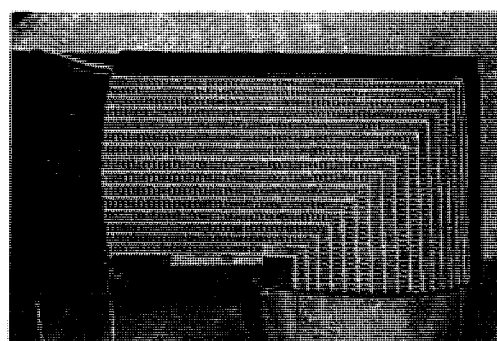


Abbildung 5: Taktile Sensoren am Roboter

preßdruck einer Kontaktstelle. Abbildung 4 zeigt die Einzelteile eines Segments der Haut, bestehend aus zwei leitfähigen Schaumstoffflächen, einer netzförmigen Isolationsschicht und einer Auswerteelektronik. Mehrere am Roboter angebrachte Hautsegmente sind in Abbildung 5 dargestellt. An dem im Vordergrund liegendem Hautstück ist die zweite Schaumstoffschicht nicht angebracht, sodaß man die netzförmige Isolationsschicht deutlich erkennen kann. Die Isolationsschicht hält im Ruhezustand die aufeinanderliegenden Schaumstoffschichten getrennt. Bei einem Kontakt werden die Schaumstoffflächen aufeinandergedrückt. Über Spannungsteiler und Übergangswiderstand beider Schichten werden der Ort des Kontakts und die Größe des Drucks ermittelt. Alle Elektroniken der Hautstücke sind durch ein Bussystem verbunden, so daß auch mehrere Kontakte gleichzeitig detektiert werden können.

5 Bewegungssteuerung

Die mit dem Experimentalsystem in Abbildung 1 zu realisierenden Manipulationsaufgaben, lassen sich in verschiedene Bewegungsklassen einteilen. Die nachfolgend aufgelisteten Bewegungsklassen leiten sich aus den unterschiedlichen Anforderungen für die Einhaltung von Position und Orientierung des Endeffektors sowie die Nutzung der Redundanz des Roboterarms während des Bewegungsablaufes ab.

- **Aufgabenorientierte Bewegung:** Die Endeffektorposition *und* -orientierung folgt einer vorgegebenen Trajektorie. Das Kollisionsvermeidungssystem kann bei Hindernissen in Aramnähe die Redundanz des Roboterarmes für Ausweichbewegungen nutzen. Ist die Ausweichbewegung nicht ohne Abweichen des Endeffektors von seiner vorgeschriebenen Trajektorie möglich, muß die Trajektorie neu geplant werden.
- **Semi-Aufgabenorientierte Bewegung:** Hier muß gegenüber der aufgabenorientierten Bewegung für die Endeffektorlage nur Position *oder* Orientierung eingehalten werden. Ein Beispiel hierfür ist der Transport eines offenen Flüssigkeitsbehälters, bei dem die Position eventuell abweichen darf, jedoch die Orientierung eingehalten werden muß.
- **Umgebungs-Aufgabenorientierte Bewegung:** Bei dieser Bewegung wird die Restriktion für die Endeffektorlage weiter gelockert. Abweichungen von der vorgegebenen Trajektorie für den Endeffektor sind bis zu einem gewissen Grad zulässig, sofern ein Kompromiß mit der Hindernisvermeidung gefunden werden kann.
- **Umgebungsorientierte Bewegung:** Hier ist keine Bewegung für den Endeffektor vorgegeben. Bewegungen werden bei dieser Klasse ausgelöst, sobald sich ein Hindernis dem Roboter unzulässig nähert.

5.1 Roboter-Segmentierung

Der Roboter wird, entsprechend seiner Kinematik, in Segmente $S_i \in \mathbb{S}$ aufgeteilt. Abbildung 6 zeigt $I = 9$ Segmente S^i ($1 \leq i \leq I$) des Roboters in unterschiedlichen Grautönen. Zu bemerken ist an dieser Stelle, daß die nicht-holonome Plattform und der Roboterarm nicht getrennt, sondern als eine kinematische Kette betrachtet werden. Um jedes Segment ist ein Sicherheitsbereich $\mathbb{Z}^i \subseteq \mathbb{R}^3$ angeordnet. Sobald der Abstand eines Robotersegment $S_i \in \mathbb{S}$ zu einem Hindernis den minimalen Sicherheitsabstand $d_0^i > 0$

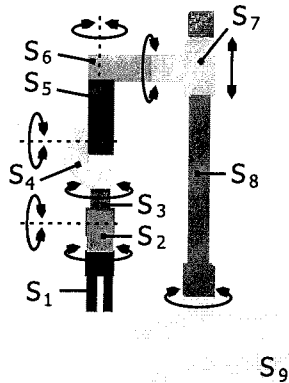


Abbildung 6: Segmente S^i des Roboters dargestellt in unterschiedlichen Grautönen.

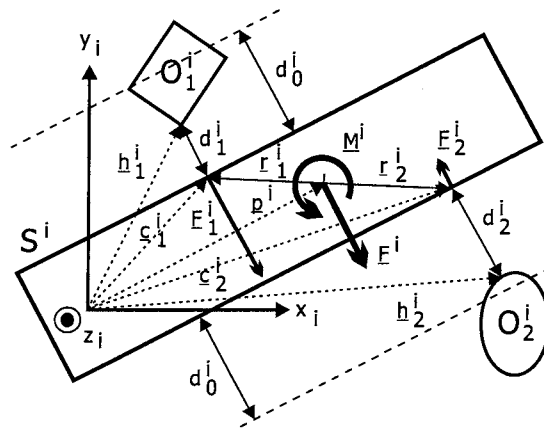


Abbildung 7: Geometrische Definitionen und virtuelle Kräfte und Momente für ein vereinfachtes Robotersegment S^i mit 2 Hindernissen in dessen Sicherheitsbereich.

unterschreitet, wird diese Sicherheitszone verletzt. Die Berechnung der daraufhin einzuleitenden Ausweichbewegung wird im folgenden Abschnitt beschrieben.

5.2 Kollisionsvermeidungsstrategie

Abbildung 7 zeigt ein vereinfachtes Segment S^i mit $L = 2$ Hindernissen innerhalb dessen Sicherheitsbereichs $\mathbb{Z}^i$. Für jedes Hindernis berechnet das Kollisionsvermeidungssystem eine virtuelle Kraft $\underline{F}_l^i \in \mathbb{R}^3$ ($1 \leq l \leq L$). Diese liegt parallel zum minimalen Abstandsvektor zwischen Hindernis und Segment

$$\underline{d}_l^i = \underline{h}_l^i - \underline{c}_l^i \quad . \quad (1)$$

Die Stärke der Kraft ist proportional zur Eindringtiefe

$$e_l^i = d_0^i - d_l^i \quad . \quad (2)$$

mit

$$d_l^i = \|\underline{h}_l^i - \underline{c}_l^i\| \quad . \quad (3)$$

Daraus ergibt sich eine virtuelle Kraft

$$\underline{F}_l^i = \begin{cases} -K_f^i e_l^i \nabla(d_l^i) & d_l^i \leq d_0^i \\ 0 & d_l^i > d_0^i \end{cases} \quad (4)$$

mit

$$\nabla(d_l^i) = \frac{h_l^i - c_l^i}{d_l^i} \quad (5)$$

Der konstante Faktor K_f^i in (4) berücksichtigt den Dimensionsunterschied zwischen Kraft und Länge. Da die Eindringweite (2) immer positiv ist, erhält man für $K_f^i > 0$ eine Kraft, die das Armsegment vom Hindernis abstößt. Im Falle mehrerer generierter Kräfte, an demselben Segment, werden diese zu einer resultierenden Kraft

$$\underline{F}^i = \sum_{l=1}^L \underline{F}_l^i \quad (6)$$

zusammengefaßt. Diese Kraft setzt dabei an einem Kontrollpunkt $\underline{p}^i$ (7) des Segments S^i an. Die Position des Kontrollpunktes errechnet sich unter Verwendung der Segmentaufpunkte $\underline{c}_n^i$ zu

$$\underline{p}^i = \frac{1}{N} \sum_{n=1}^N \underline{c}_n^i \quad (7)$$

Betrachtet man den Kontrollpunkt $\underline{p}^i$ und die einzelnen Kräfte $\underline{F}_l^i$, so ergibt sich ein Moment $\underline{M}^i \in \mathbb{R}^3$ zu

$$\underline{M}^i = K_m^i \sum_{l=1}^L \underline{r}_l^i \times \underline{F}_l^i \quad (8)$$

mit

$$\underline{r}_l^i = \underline{c}_l^i - \underline{p}^i \quad .$$

K_m^i dient dabei als Verstärkungsfaktor. Wenn sich nur ein Hindernis im Sicherheitsbereich eines Segments befindet, bildet sich kein Drehmoment um den Kontrollpunkt. Die Kraft $\underline{F}^i$ (6) und das Moment $\underline{M}^i$ (8) werden zu einem verallgemeinerten Kraft-Momenten-Vektor

$$\underline{\delta}^i = \begin{bmatrix} \underline{F}^i \\ \underline{M}^i \end{bmatrix} \quad (9)$$

zusammengefaßt, der nun alle auf Segment S^i wirkenden virtuellen Kräfte und Momente vereint.

Um aus dem Kraft-Momenten-Vektor (9) eine Ausweichbewegung zu bestimmen, wird dieser mit Hilfe der transponierten Jakobimatrix $\left(\underline{J}_p^i\right)^T$ [10] in den Gelenkraum abgebildet. Damit erhält man den resultierenden Gelenk-Momenten-Vektor

$$\underline{\tau}^i = \left(\underline{J}_p^i\right)^T \underline{\delta}^i \quad . \quad (10)$$

Betrachtet man alle Segmente, erhält man für das Gesamtsystem den Gelenk-Momenten-Vektor

$$\underline{\tau} = \sum_{i=1}^I \underline{\tau}^i \quad . \quad (11)$$

Durch die Überlagerung der Gelenk-Momenten-Vektoren (10) der Einzelsysteme zu einem gemeinsamen Vektor (11) wird das gesamte System vor möglichen Kollisionen bewahrt. Durch Multiplikation von (11) mit der Verstärkungsmatrix $\underline{K}_q(q)$, die unter anderem systembedingte Restriktionen, wie beispielsweise Gelenkwinkelbeschränkungen, maximale Geschwindigkeiten oder Beschleunigungsgrenzwerte, berücksichtigt, erhält man schließlich den Gelenkwinkelgeschwindigkeitsvektor

$$\underline{\dot{q}} = \underline{K}_q(q) \underline{\tau} \quad . \quad (12)$$

Mit obigen Geschwindigkeitsvektor (12) werden die Gelenke des Roboters angesteuert, und es kommt zu einer Ausweichbewegung.

Die taktile Sensorik des Roboters liefert bei Kontakten unmittelbar Abstandsvektoren der Länge $d_i^i = 0$. Diese Abstandsvektoren werden ohne den Umweg über das Umgebungsmodell in obigem Verfahren, zusammen mit den modellbasiert berechneten Abstandsvektoren, verwendet. Der Betrag des Kontaktdrucks wird dabei als Maß für den Abstand verwendet, der Normalenvektor zur Roboter Oberfläche als Richtungsvektor und der Ort des Kontakts als Aufpunkt am entsprechenden Segment.

6 Ergebnisse

Abbildung 8 zeigt Simulationsergebnisse, aus denen eine qualitative Aussage über die Auswirkung des Kollisionsvermeidungssystem auf das Bewegungsverhalten des Roboters abgeleitet werden kann. Der Roboter hat in diesem Szenario keine Aufgabe zu lösen, er soll lediglich Hindernissen ausweichen. Tatsächlich soll er hier nur mit seinem Arm ausweichen, Position und Orientierung der Plattform bleiben unverändert. Die Positionen der Hindernisse

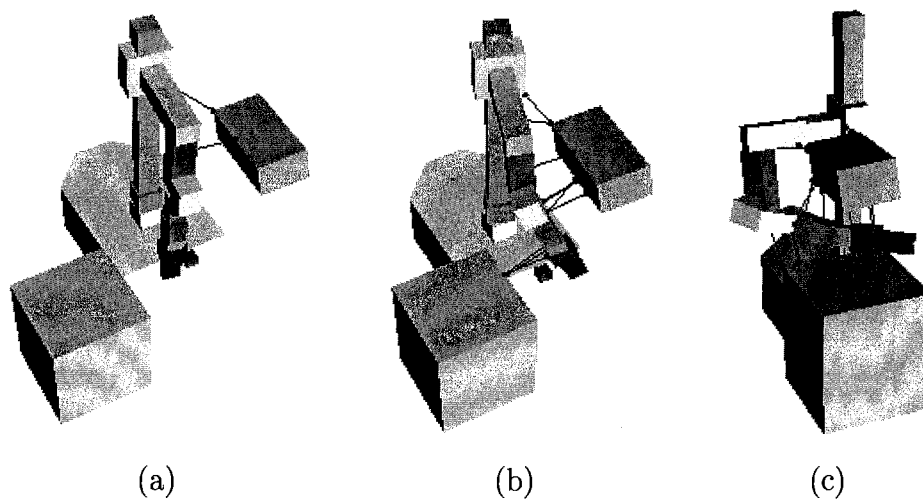


Abbildung 8: Simulationsbeispiel einer empfohlenen Armbewegung des Kollisionsvermeidungssystems, wenn zwei bekannte zum Roboter sich hin bewegende Objekte in dessen Sicherheitsbereiche gelangen.

sind dem System zu jedem Zeitpunkt bekannt. Zwei Hindernisse bewegen sich in Abbildung 8a in Richtung Roboter. Sobald ein Hindernis in einen Sicherheitsbereich des Roboters gelangt, siehe Abbildung 8a, berechnet das Kollisionsvermeidungssystem eine geeignete Ausweichbewegung, siehe Abbildungen 8b-c.

Abbildung 9 zeigt die Funktionsweise des Kontaktreflexsystems am realen Roboter. Die Steuerung des Roboters besitzt lediglich Informationen über Lage und Geometrie des Arms und der Plattform. Die Umgebung wird in diesem Anwendungsbeispiel nur über die taktile Sensorik wahrgenommen. Das Kollisionsvermeidungssystem ist deaktiviert, um unmittelbare Interaktionen mit dem Roboter zu ermöglichen. Das Kontaktreflexsystem kann somit von einem Anwender benutzt werden, um den Roboter interaktiv in eine gewünschte Position zu „führen“ (siehe Abbildungen 9a-c).

7 Zusammenfassung und Ausblick

Der vorliegende Beitrag beschreibt eine Umgebungswahrnehmung und Bewegungssteuerung für einen mobilen nicht-holonomen Manipulator, die es ermöglichen, sich in unbekannten bzw. nur teilweise bekannten Umgebungen weitgehend kollisionsfrei zu bewegen. Das Kollisionsvermeidungssystem generiert Ausweichbewegungen, die die Kollisionsfreiheit zu dynamischen Ob-

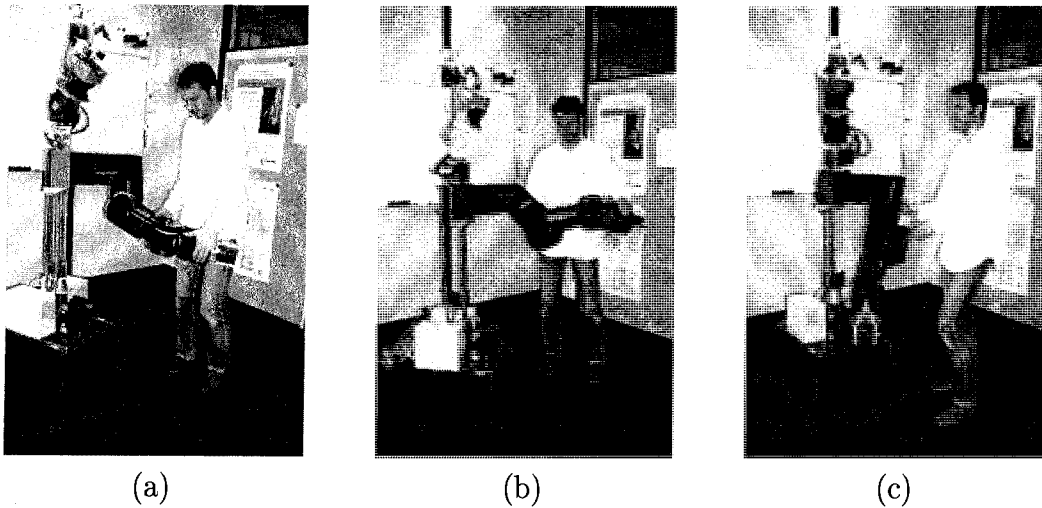


Abbildung 9: Beispiel einer Ausweichbewegung infolge eines physikalischen Kontakts am realen System.

jekten sicherstellen. Darüber hinaus wurde ein Kontaktreflexsystem vorgestellt, das über einen taktilen Sensor auf physikalische Kontakte reagiert, indem geeignete Ausweichbewegungen eingeleitet werden. Beide Ansätze beruhen auf der gleichen Strategie und können deshalb einfach zu einem Gesamtsystem integriert werden. Die resultierenden Bewegungen des Roboters erscheinen „plausibel“ und fügen sich deshalb harmonisch in das Interaktionsszenario Mensch-Maschine ein. Für die Überlagerung der Ausweichbewegungen mit den Bewegungsvorgaben für den Endeffektor gibt es verschiedene Möglichkeiten, die sich in die hier vorgeschlagenen Bewegungsklassen einteilen lassen.

Weiterführende Arbeiten beschäftigen sich mit der Erweiterung der Funktion des Sicherheitsbereichs. Neben der Eindringtiefe eines Hindernisses, können auch die Eindringgeschwindigkeit in den Sicherheitsbereich und die Verweildauer im Sicherheitsbereich berücksichtigt werden. Schließlich werden die hier vorgestellten Systeme in die bereits am Roboter implementierten planenden Komponenten integriert, um deren lokalen reaktiven Handlungsspielraum zur Kollisionsvermeidung und Interaktion auszunutzen.

Die diesem Beitrag zugrundeliegenden Arbeiten wurden teilweise mit Mitteln des Bundesministeriums für Bildung und Forschung unter den Förderkennzeichen 01IN601A2 und 01IL902DO gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Literatur

- [1] G. Lawitzky, "A Navigation System for Service Robots - From Research to Products," in *FSR 2001* (A. Halme, R. Chatila, und E. Prassler, Hrsg.), S. 15–20, Yleisjäljennös–Painopörssi, 2001.
- [2] L. E. Kavraki, P. Svetka, J. C. Latombe, und M. Overmars, "Probabilistic roadmaps for path planning in high dimensional configuration spaces," in *IEEE Transactions on Robotics and Automation* 12(4), S. 566–580, 1996.
- [3] N. M. Amato, O. B. Bayazit, L. K. Dale, C. Jones, und D. Vallejo, "Choosing good distance metrics and local planners for probabilistic roadmap methods," in *International Conference on Robotics and Automation*, (Leuven Belgium), 1998.
- [4] C. Nissoux, T. Simeon, und J. P. Laumond, "Visibility Based Probabilistic Roadmaps," in *Intelligent Robots and Systems*, 1999.
- [5] M. C. Lin und J. F. Canny, "A Fast Algorithm for Incremental Distance Calculation," in *International Conference on Robotics and Automation*, S. 1008–1014, April 1991.
- [6] E. Larsen, S. Gottschalk, M. C. Lin, und D. Manocha, "Fast Distance Queries with Rectangular Swept Shere Volumes," in *International Conference on Robotics and Automation*, (San Francisco), April 2000.
- [7] O. Brock und O. Khatib, "Elastic Strips: A Framework for Integrated Planning and Execution," in *Preprints, 6th International Symposium on Experimental Robotics (ISER'99)*, (Sydney), 1999.
- [8] G. v. Wichert, T. Wösch, S. Gutmann, und G. Lawitzky, "MobMan – Ein mobiler Manipulator für Alltagsumgebungen," in *Autonome Mobile Systeme (AMS'00)* (R. Dillmann, H. Wörn, und M. v. Ehr, Hrsg.), Informatik aktuell, S. 55–62, Springer Verlag, Heidelberg, 2000.
- [9] G. Lawitzky, "Das Navigationssystem SINAS," in *Proceedings Robotik 2000, VDI-Berichte 1582*, (Düsseldorf), S. 77–82, VDI-Verlag, June 2000.
- [10] M. W. Spong und M. Vidyasagar, *Robot Dynamics and Control*. John Wiley & Sons Inc., 1989.

Robuste Regelung eines parameterabhängigen Übertragungssystems

Stefan Doczy

Institut für Regelungstechnik

Technische Universität Graz

Inffeldgasse 16c/II, A-8010 Graz

doczy@irt.tu-graz.ac.at

Die robuste Regelung parameterabhängiger Übertragungssysteme stellt große Anforderungen an das verwendete Regelgesetz. Sind am Prozeß Messungen der veränderlichen physikalischen Parameter möglich, so können diese für den Entwurf eines parameterabhängigen Reglers verwendet werden. Mit diesem Regelgesetz können gegenüber einem parameterunabhängigen Regler wesentliche Verbesserungen erzielt werden. Gezeigt wird dies an einem Labormodell, bei dem ein meßbarer physikalischer Parameter während des Betriebes gezielt verändert werden kann. Dieses neu entwickelte Modell wird im Rahmen dieser Arbeit näher vorgestellt. Es wird nicht nur die Reglersynthese beschrieben, sondern auch auf die Implementierung des parameterabhängigen Regelgesetzes näher eingegangen.

1 Einleitung

Das Ergebnis einer analytischen Modellbildung eines realen Prozesses ist oft ein mathematisches Modell, das sich in Abhängigkeit von einigen veränderlichen physikalischen Parametern verändert. Diese können in bestimmten Fällen sogar gemessen werden, oder wenigstens mit Hilfe von oberen und unteren Schranken abgeschätzt werden. Verschiedene Methoden der Regelungstechnik erlauben es, ausgehend von einem sogenannten nominellen Modell konstante, das heißt parameterunabhängige, Regler zu entwerfen, die die veränderlichen Parameter als Modellunsicherheiten berücksichtigen. Dieser Regler ist für alle im Einsatz auftretenden Betriebszustände auf seine Funktionstüchtigkeit vorerst in einer Simulation und dann am Prozeß selbst zu testen. Erfüllt er die gestellten Anforderungen für alle betrachteten Betriebszustände, so wird er als *robuster Regler* bezeichnet.

Sind am Prozeß Messungen zur Bestimmung der veränderlichen Parameter möglich, so kann dies bereits beim Reglerentwurf berücksichtigt werden. Werden nun für verschiedene Systemkonfigurationen, oder Betriebspunkte, Regler entworfen, und diese für den gesamten Parameterbereich zu einem parameterabhängigen Regelgesetz zusammengefaßt, so ist der resultierende Regler der Klasse der sogenannten Gain Scheduling Regler zuzuordnen. In dieser Arbeit wird der Entwurf eines parameterabhängigen Reglers vorgestellt.

Dabei wird auf die Realisierung und Implementierung des Reglers, dessen Übertragungsverhalten durch ein sogenanntes lineares parametervariables (LPV) Übertragungssystem beschrieben wird, besonderer Wert gelegt.

Viele Beispiele in der Fachliteratur, in denen parameterabhängige Modelle behandelt werden, stammen aus der Luftfahrt oder dem Raketenbau. Der entworfene Regler ist dann nur in den seltensten Fällen am realen Prozeß getestet worden, was zweifellos unbefriedigend ist. Um ein reales Modell für die Beurteilung der Qualität des Reglerentwurfes zur Verfügung zu haben, wurde ein neues Labormodell mit veränderlichen Parametern aufgebaut. Zentrale Anforderung an dieses Labormodell ist es, daß Veränderungen während des Betriebes vorgenommen werden können, und daß diese Beeinflussung meßtechnisch erfaßbar ist.

2 Labormodell

Der Ausgangspunkt für den Aufbau des Labormodells ist die wohlbekannte Verladebrücke [1]. Diese kann bei der Beförderung unterschiedlicher Lasten untersucht werden, wobei keine Parameterabhängigkeit direkt im Betrieb auftritt. Wird sie mit variabler Seillänge betrieben, so wird während des Versuches Einfluß auf die Systemkonfiguration genommen. Bei einer großen Variation der Seillänge ist allerdings der Raumbedarf enorm, und auch die Messung des Seilwinkels ist schwer realisierbar. Im Laborbetrieb des Institutes für Regelungstechnik (IRT) der Technischen Universität Graz wird eine Versuchsanordnung verwendet, die modellmäßig eine große Übereinstimmung mit der Verladebrücke zeigt. Es handelt sich dabei um den sogenannten *Rotary Flexible Joint* [14]. Diese Versuchsanordnung kann zwar nicht zum Heben und Befördern von Lasten verwendet werden, das mathematische Modell zeigt allerdings eine weitreichende Übereinstimmung mit der Verladebrücke und damit ist ein Zusammenhang herstellbar. Der gegenüber der Verladebrücke wesentlich einfachere Aufbau für einen Prototyp und die leicht durchzuführenden Modifikationsmöglichkeiten am Labormodell sind wichtige Argumente für diese Realisierung.

2.1 Beschreibung des Labormodells

Ein Bild des neu entwickelten Labormodells zeigt Abbildung 1. In der schematischen Darstellung einer Seitenansicht und einer Draufsicht des Labormodells werden die wichtigsten Teile der Anordnung bezeichnet (Abbildung 2). Dabei ist trotz der einfach gehaltenen Zeichnungen wichtig, daß die auftretenden Größenverhältnisse richtig dargestellt werden. Die zentrale Anforderung an das neu entwickelte Modell ist aus regelungstechnischer Hinsicht, daß während des Betriebes willkürlich Modellparameter verändert werden können, wobei diese Veränderungen durch einen meßbaren Parameter bestimmt sind.

Der Unterteil (Gerüst) trägt die gesamte Versuchsanordnung. An ihm befestigt ist der elektrische Antrieb 1, auf dessen Getriebewelle der Oberteil des Labormodells aufgesetzt wird. Der Oberteil besteht einerseits aus dem Aufbau und andererseits aus dem Ausleger. Der Drehwinkel des Aufbaus hinsichtlich der Bezugsachse wird mit φ bezeichnet. Im Auf-

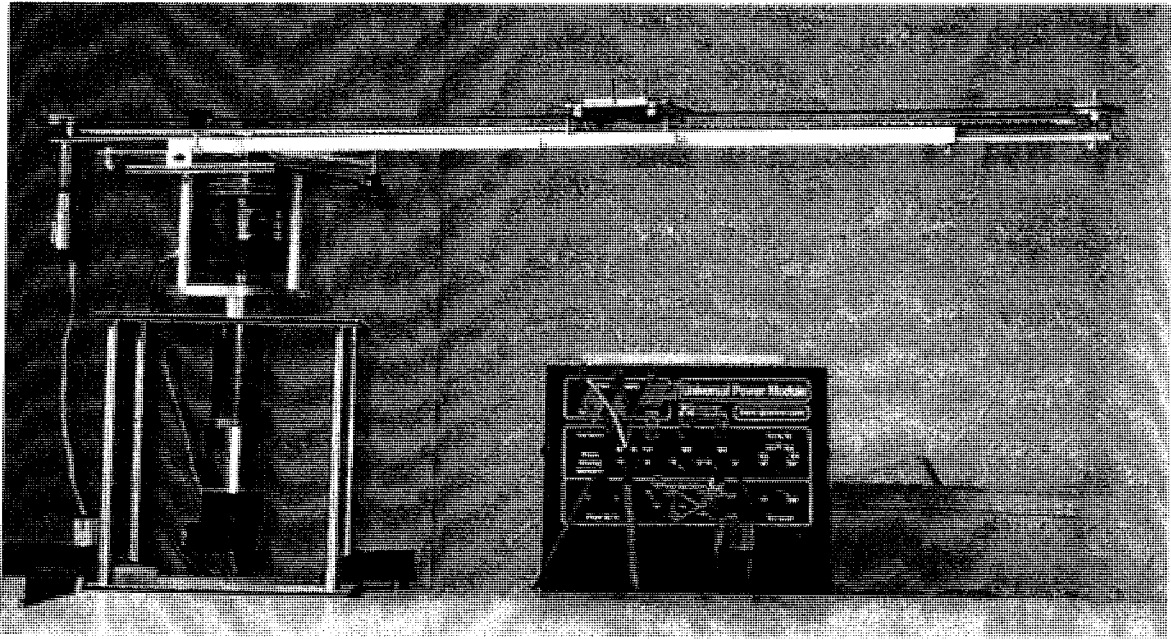


Abbildung 1: Labormodell

bau ist ein Inkrementalgeber befestigt, auf dessen Welle der Ausleger drehbar gelagert ist. Der Ausleger wird gebildet aus dem Balken und den auf ihm fest montierten Bauteilen. Dies sind der elektrische Antrieb 2 mit aufgesetzter Motorseilscheibe, die Umlenkseilscheibe und der Seilzug, die den Wagen auf dem Ausleger zu positionieren erlauben. Am Balken befestigte Schienen werden für die Führung der Wagenbewegung benötigt. Zur Positionsbestimmung des Wagens wird der Abstand l der Mitte des Wagens zur Drehachse verwendet, wobei $0 \leq l \leq 0.5 \text{ m}$ gilt. Der Ausleger ist mit dem Aufbau durch zwei Federn verbunden, wodurch er sich relativ zum Aufbau um einen Winkel α bewegen kann. Die Befestigungspunkte der Federn sowohl am Aufbau als auch am Ausleger können in bestimmten Grenzen verändert werden. Über die Prozeßanbindung wird die Verbindung zwischen dem Labormodell und einem Rechner hergestellt. Eine genaue Beschreibung des Labormodells mit Detailzeichnungen der einzelnen Bauteile wird in [7] angegeben.

2.2 Ergebnis der Modellbildung

Betrachtet man die beschriebene Anordnung als ein Übertragungssystem mit der Eingangsgröße u , das ist in diesem Fall die am elektrischen Antrieb 1 anliegende Spannung, so liefert die Anwendung des Drallsatzes auf den Aufsatz beziehungsweise den Ausleger zwei gewöhnliche lineare Differentialgleichungen zweiter Ordnung

$$\begin{aligned} \frac{d}{dt}(J_A(\dot{\alpha} + \dot{\varphi})) + k_s \alpha &= 0 \\ \frac{d}{dt}(J_A(\dot{\alpha} + \dot{\varphi}) + J_{Auf} \dot{\varphi}) &= M_A. \end{aligned}$$

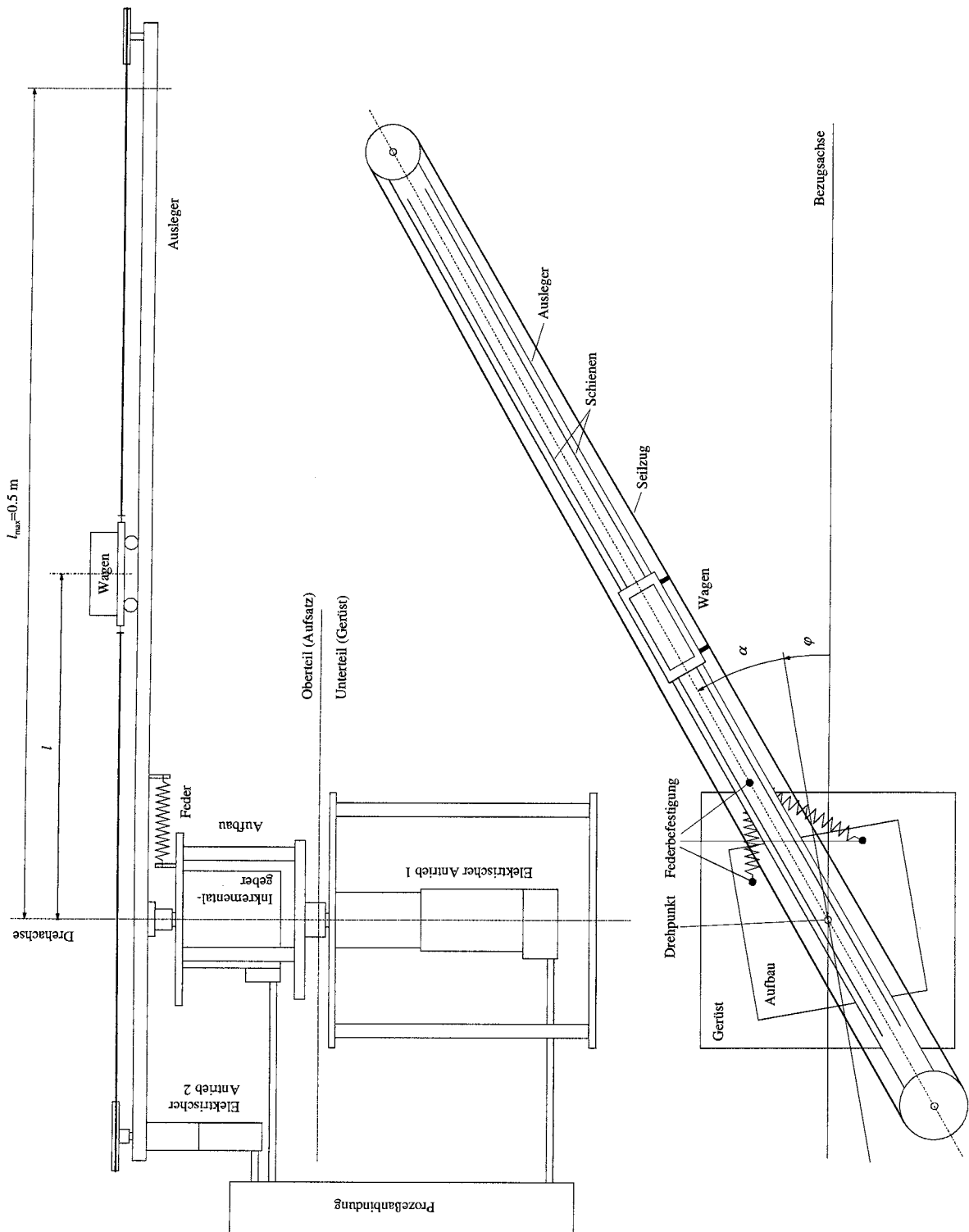


Abbildung 2: Ansichten des Labormodells

Mit dem Motordrehmoment $M_A = k_1 u - k_2 \dot{\varphi}$, hierbei sind k_1 und k_2 zwei positive Konstante, folgen daraus die Bewegungsdifferentialgleichungen

$$\begin{aligned} J_{Auf} \ddot{\varphi} &= k_s \alpha - k_2 \dot{\varphi} + k_1 u \\ J_A \ddot{\alpha} &= - \left(\frac{J_A}{J_{Auf}} + 1 \right) k_s \alpha + \left(\frac{J_A}{J_{Auf}} k_2 - \dot{J}_A \right) \dot{\varphi} - \dot{J}_A \dot{\alpha} - \frac{J_A}{J_{Auf}} k_1 u. \end{aligned}$$

Die Zahlenwerte der hier auftretenden Konstanten sind Anhang A zu entnehmen.

Das Trägheitsmoment des Auslegers J_A setzt sich nach

$$J_A = J_B + J_W \quad (1)$$

aus dem *konstanten* Trägheitsmoment J_B des Balkens und aller auf ihm fest montierten Bauelemente und dem *variablen* Trägheitsmoment J_W zusammen. Dieses wird nach

$$J_W = m_E l^2 \quad (2)$$

durch den Abstand l des Wagens von der Drehachse und der punktförmig angenommenen Ersatzmasse m_E des Wagens bestimmt. *Das durch die Position des Wagens vorgebbare Trägheitsmoment legt die Parameterabhängigkeit des mathematischen Modells fest.*

Führt man nun den Zustandsvektor

$$x := \begin{pmatrix} \varphi & \alpha & \dot{\varphi} & \dot{\alpha} \end{pmatrix}^T \quad (3)$$

ein, so ergibt sich das mathematische Modell der Versuchsanordnung in der Zustandsraumdarstellung $\dot{x} = Ax + bu$ mit der von der Wagenposition abhängigen Systemmatrix A und dem Eingangsvektor b (siehe Anhang A).

Auswirkungen der veränderlichen Wagenposition Für verschiedene aber *feste* Abstände l des Wagens von der Drehachse, das bedeutet $\dot{J}_A = 0$, werden die Auswirkungen auf die Systemeigenschaften untersucht. Die Position des Wagens auf dem Balken bestimmt nach (1) und (2) das Trägheitsmoment des Auslegers J_A . Dadurch werden die Eigenwerte der Systemmatrix A beeinflusst, die für die dynamischen Eigenschaften des Übertragungssystems prägnant sind. Ein Gesichtspunkt beim Aufbau des Modells ist ein möglichst großer Bereich für das veränderliche Trägheitsmoment J_A . Dazu ist einerseits der Verfahrensweg entsprechend groß zu realisieren und andererseits das konstante Trägheitsmoment J_B klein zu halten. In unserem Fall beträgt das Verhältnis

$$\frac{J_{A \max}}{J_{A \min}} = \frac{33.35 \cdot 10^{-3} \text{ kgm}^2}{13.65 \cdot 10^{-3} \text{ kgm}^2} = 2.44.$$

Aufgrund einer Vielzahl von Einschränkungen hinsichtlich der mechanischen Besonderheiten des Versuchsaufbaus läßt sich dieses Verhältnis nicht beliebig erhöhen. Sowohl der Verfahrensweg als auch die Masse des Wagens können aufgrund der maximal zulässigen Belastbarkeit des Inkrementalgebers nicht weiter erhöht werden.

Der in der Tabelle 1 eingeführte Parameter δ , der über die Relation

$$\delta = 4l - 1 \quad (4)$$

mit der Wagenposition l zusammenhängt, dient in weiterer Folge als beschreibender Parameter für die aktuell auftretende Systemkonfiguration. Für $\delta = 0$ befindet sich der Wagen in der Mitte des möglichen Fahrweges ($l = 0.25$ m), für $\delta = -1$ liegt der Wagen genau auf der Drehachse ($l = 0$; diese Position wird als *innen* bezeichnet) und für $\delta = 1$ steht der Wagen am Ende des Fahrweges ($l = 0.5$ m; diese Position wird als *außen* bezeichnet). Der Parameter ist nach (4) so festgelegt, daß $\delta = 0$ einem nominellen System entspricht, bei dem sowohl nach innen als auch nach außen der gleich große Fahrweg zur Verfügung steht.

Der Einfluß der Wagenposition auf die Lage der vier Eigenwerte s_i der Systemmatrix A ist aus Tabelle 1 ersichtlich. Natürlich bleibt der Eigenwert $s_1 = 0$ erhalten. Der Eigenwert

δ	Eigenwerte		
	s_1	$s_{2,3}$	s_4
-1	0	$-0.3032 \pm 16.62i$	-67.81
-0.5	0	$-0.3046 \pm 15.92i$	-67.81
0	0	$-0.3078 \pm 14.25i$	-67.80
0.5	0	$-0.3111 \pm 12.34i$	-67.79
1	0	$-0.3137 \pm 10.63i$	-67.79

Tabelle 1: Lage der Eigenwerte in Abhängigkeit von der Wagenposition

s_4 verändert sich kaum, die Auswirkung auf das Systemverhalten ist aufgrund seiner Lage bei etwa -67.8 relativ gering. Am stärksten wird das konjugiert komplexe Eigenwertpaar $s_{2,3}$ beeinflusst, das heißt, das Schwingungsverhalten des über die Federn mit dem Aufbau verbundenen Auslegers wird verändert. Eine Positionierung des Wagens von innen nach außen führt zu einer größeren Dämpfung bei kleiner werdender Schwingungskreisfrequenz. In der Abbildung 3 wird die Lage des Eigenwertes s_2 in Abhängigkeit von δ dargestellt. Zwischenwerte werden hierbei durch Punkte gekennzeichnet. Die Pfeilrichtung gilt für steigende Werte von δ , also für eine Positionsänderung von innen nach außen.

Feder und Federbefestigung Die Federn liefern das Rückstellmoment für die Bewegung des Auslegers bezüglich des Aufbaus. Dabei muß der Versuchsaufbau so konzipiert sein, daß die Belastbarkeit der Federn nicht überschritten wird. Der Ausleger wird möglichst leicht aber trotzdem stabil gewählt, um das Trägheitsmoment des Balkens niedrig zu halten, aber der Belastung durch den Wagen zu genügen. Die Beladung des Wagens und damit die Wagenmasse wird so groß gewählt, wie es die mechanische Belastbarkeit des Inkrementalgebers zuläßt, um eine möglichst große Variation des veränderlichen Trägheitsmomentes zu erzielen.

Die Befestigungspunkte der Federn sowohl am Ausleger als auch am Aufbau können in den durch den Versuchsaufbau festgelegten Grenzen frei gewählt werden. Mit der

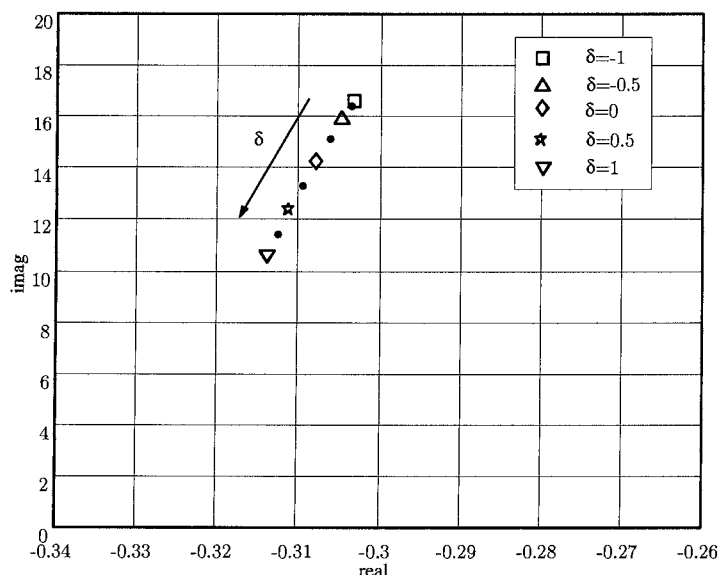


Abbildung 3: Lage des Eigenwertes s_2 in Abhängigkeit von der Wagenposition

Berücksichtigung der zulässigen Auslenkungen der Federn und dem damit verbundenen Winkel α wird dieses Gebiet sehr eingeschränkt [8]. Dieser Winkel darf aus mechanischen Gründen eine maximale Größe nicht überschreiten.

3 Reglerentwurf

3.1 Aufgabenstellung und Problemformulierung

Der Winkel φ zwischen dem Aufbau und der Bezugsachse wird während des Betriebes durch eine Führungsgröße w , die $-0.5 \text{ rad} \leq w \leq 0.5 \text{ rad}$ erfüllt, vorgegeben. Ziel der Regelung ist es, den Aufbau und damit auch den Ausleger der Führungsgröße nachzuführen. Für einen Positionswechsel von -0.5 rad auf 0.5 rad soll die Aufgabe mit geringem Überschwingen und unter Einhaltung der Stellgrößenbeschränkungen gelöst werden. Wichtigster Gesichtspunkt ist dabei, daß die Regelung *robust* gegenüber einer beliebigen Position des Wagens auf dem Ausleger sein muß. Das heißt, für jede beliebige Wagenposition ist zumindest die Stabilität des Regelkreises zu gewährleisten, man spricht in diesem Zusammenhang auch von *robuster Stabilität*.

3.1.1 Systembeschreibung Labormodell

In Abbildung 4 ist ein Blockschaltbild für die Zusammenschaltung der einzelnen Regelkreisglieder dargestellt. Auf das Labormodell G wirkt als Eingangsgröße die Stellgröße u . Die durch Messung ermittelten Ausgangsgrößen des Labormodells sind die Winkel φ und α . Als externe Eingangsgröße für die Strecke P wird die Führungsgröße w für den Winkel

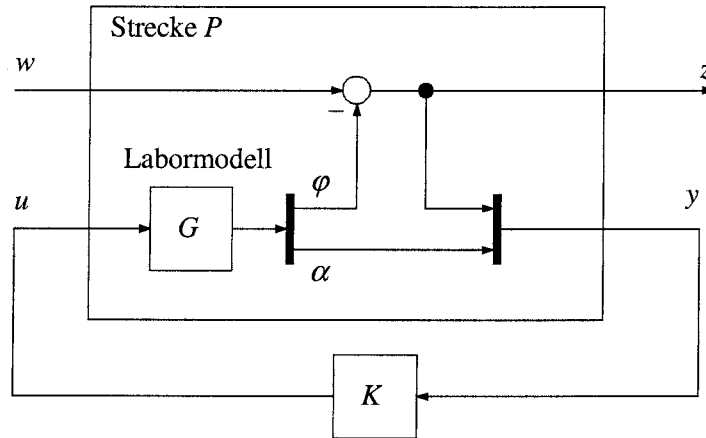


Abbildung 4: Blockschaltbild für den Regelkreis

φ vorgegeben. Mit Hilfe des Addierers wird eine Ausgangsgröße der Strecke

$$z = w - \varphi, \quad (5)$$

gebildet. Die bei der Rückkopplung verwendeten Größen werden im Vektor

$$y = \begin{pmatrix} w - \varphi \\ \alpha \end{pmatrix} \quad (6)$$

zusammengefaßt. Er dient als Eingangsgröße für den Regler. Die Ausgangsgröße des Reglers ist die Stellgröße u , die wiederum eine Eingangsgröße der Strecke P darstellt.

Das Ergebnis der Modellbildung in 2.2 ist die Zustandsraumbeschreibung des Labormodells. Nun wird angenommen, daß der Wagen eine bestimmte *nicht veränderliche Position* einnimmt. Aufgrund dieser Position ist das Trägheitsmoment des Auslegers J_A festgelegt und konstant, das heißt, für die zeitliche Ableitung gilt $\dot{J}_A = 0$. Als Ausgangsgrößen des Labormodells werden die beiden gemessenen Größen φ und α verwendet. Um die Strecke P zu beschreiben, müssen der Einfluß der externen Eingangsgröße und der Addierer $z = w - \varphi$ berücksichtigt werden. Dies führt zu dem linearen zeitinvarianten mathematischen Modell der Form (7) mit zwei Eingangsgrößen und drei Ausgangsgrößen

$$\begin{aligned} \dot{x} &= \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{k_s}{J_{Auf}} & -\frac{k_2}{J_{Auf}} & 0 \\ 0 & -\left(\frac{J_A}{J_{Auf}} + 1\right) \frac{k_s}{J_A} & \frac{k_2}{J_{Auf}} & 0 \end{pmatrix} x + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} w + \begin{pmatrix} 0 \\ 0 \\ \frac{k_1}{J_{Auf}} \\ -\frac{k_1}{J_{Auf}} \end{pmatrix} u \\ z &= \begin{pmatrix} -1 & 0 & 0 & 0 \end{pmatrix} x + w + 0u \\ y &= \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} x + \begin{pmatrix} 1 \\ 0 \end{pmatrix} w, \end{aligned}$$

das die Grundlage für den Reglerentwurf darstellt.

3.2 Entwurfsgrundlagen

Zunächst wird für ein lineares zeitinvariantes (LZI) Mehrgrößensystem ein Reglerentwurfsverfahren angeführt [16]. Die Regelstrecke P wird durch ein Zustandsmodell der Form

$$P \begin{cases} \dot{x} = Ax + B_w w + Bu \\ z = C_z x + D_{zw} w + D_z u \\ y = Cx + D_w w \end{cases} \quad (7)$$

beschrieben. Hierbei symbolisieren u die auf die Strecke wirkenden Stellgrößen, w den Vektor der externen Eingangsgrößen (wie Führungsgrößen, Störgrößen oder Meßrauschen) und y die gemessenen Ausgangsgrößen. Im Vektor z sind Systemgrößen zusammengefaßt, die zur Beurteilung des Übertragungsverhaltens herangezogen werden. Wir wollen einen dynamischen Regler K

$$K \begin{cases} \dot{\zeta} = A_K \zeta + B_K y \\ u = C_K \zeta + D_K y \end{cases} \quad (8)$$

so entwerfen, daß gewisse noch anzugebende Spezifikationen erfüllt werden. Damit erhalten wir den Regelkreis nach Abbildung 5.

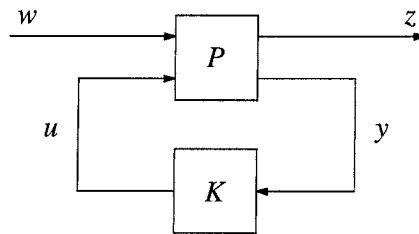


Abbildung 5: Regelkreis

Wir wollen nun die Spezifikationen näher ausführen und werden hierzu die Übertragungsmatrix $T(s)$, die das Übertragungsverhalten für den Kanal $w \rightarrow z$ beschreibt, das heißt $z(s) = T(s)w(s)$, geeignet festlegen. Folgende Anforderungen stellen wir in den Vordergrund:

- Um die Erfüllung der Robustheitsanforderungen zu gewährleisten, wird für den Kanal $w \rightarrow z$ ein optimales Übertragungsverhalten im Sinne der H_∞ Norm angestrebt.
- Es ist oft erwünscht, eine Mindest-Verfallsrate einzuhalten, oder die Dämpfung des Gesamtsystems in geeigneter Weise zu beeinflussen. Solche Anforderungen können durch Einschränkung des erlaubten Gebietes für die Lage der Pole von $T(s)$ erfüllt werden.

Die Komplexität des gestellten Problems erkennt man unmittelbar, wenn man die Übertragungsmatrix $T(s)$ des Gesamtsystems, aus dem Zustandsmodell¹ des Regelkreises nach Abbildung 5

$$\begin{aligned}\dot{x}_{cl} &= \mathcal{A}x_{cl} + \mathcal{B}w \\ z &= \mathcal{C}x_{cl} + \mathcal{D}w\end{aligned}$$

berechnet. Dafür wird die in der Literatur übliche Notation

$$T(s) = \mathcal{D} + \mathcal{C}(sE - \mathcal{A})^{-1}\mathcal{B} = \left(\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array} \right)$$

verwendet. Damit erhält man:

$$\left(\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array} \right) = \left(\begin{array}{cc|c} A + BD_KC & BC_K & B_w + BD_KD_w \\ B_KC & A_K & B_KD_w \\ \hline C_z + D_zD_KC & D_zC_K & D_{zw} + D_zD_KD_w \end{array} \right). \quad (9)$$

3.2.1 Formulierung als lineare Matrixungleichung

Lineare Matrixungleichungen (linear matrix inequalities - LMIs) sind in den letzten Jahren in vielen Bereichen der Regelungstechnik eingesetzt worden [3, 9]. Sowohl in der Analyse als auch in der Synthese von Regelkreisen werden zu diesem Thema immer mehr Hilfsmittel (Software) zur Verfügung gestellt. Vor allem aufgrund der Fortschritte der verschiedenen Lösungsalgorithmen, die kommerziell angeboten werden, erfreut sich dieses Werkzeug heute einer großen Beliebtheit [12].

Es soll zunächst die Definition einer LMI angegeben werden. Ausgehend von einem Variablenvektor $\xi^T = (\xi_1 \ \cdots \ \xi_m)$ und gegebenen symmetrischen (n, n) Matrizen F_i ($i = 0, \dots, m$) bilden wir die Matrix $F(\xi)$ nach der Vorschrift

$$F(\xi) := F_0 + \sum_{i=1}^m F_i \xi_i. \quad (10)$$

Das heißt, $F(\xi)$ ist eine symmetrische (n, n) Matrix, deren Elemente affin in ξ sind. Verlangen wir, daß $F(\xi)$ positiv beziehungsweise negativ (semi-)definit ist, so entstehen Ungleichungen der Form

$$F(\xi) > 0, \quad (F(\xi) \geq 0) \quad \text{beziehungsweise} \quad F(\xi) < 0, \quad (F(\xi) \leq 0).$$

Eine Ungleichung dieser Form wird LMI genannt.

Für LMIs gibt es einige Möglichkeiten zur Umformung, die für die Reglersynthese oftmals vorteilhaft sind. Sei T eine reguläre $n \times n$ Matrix, so gilt (sogenannte *Kongruenztransformation*):

$$T^T F(\xi) T \geq 0 \iff F(\xi) \geq 0. \quad (11)$$

¹Für die Indizierung der Variablen wird *cl* für *closed loop* verwendet.

Für eine Matrix F , die aus Matrizen $Q = Q^T$, $R = R^T$ und S anhand der Vorschrift

$$F = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}$$

gebildet wird, sind folgende drei Aussagen äquivalent (sogenanntes *Schur Komplement*):

$$\begin{aligned} F &> 0 \\ Q &> 0 \text{ und } R - S^T Q^{-1} S > 0 \\ R &> 0 \text{ und } Q - S R^{-1} S^T > 0 \end{aligned} \quad (12)$$

Für $S = 0$ schließt man daraus, daß eine blockdiagonale Matrix genau dann positiv definit ist, wenn das für ihre Blöcke zutrifft. Hieraus resultiert, daß das System von Matrixungleichungen $F_1(\xi) > 0, \dots, F_k(\xi) > 0$ äquivalent ist zu *einer* LMI mit

$$F(\xi) := \text{diag}(F_1(\xi), \dots, F_k(\xi)).$$

Es wird nun gezeigt, wie die verschiedenen Anforderungen an den Regelkreis mit Hilfe von LMIs formuliert und gelöst werden. Die interne Stabilität des Regelkreises wird mit Hilfe der zu ermittelnden quadratischen Ljapunovfunktion

$$V(x_{cl}) = x_{cl}^T P x_{cl}$$

garantiert. Hierzu muß die positiv definite Matrix P die Ungleichung

$$A^T P + P A < 0$$

erfüllen. Jede weitere Anforderung wird dann als zusätzliche Bedingung für die Ljapunovfunktion hinzugenommen und in Form einer weiteren Ungleichung angegeben!

H_∞ Norm Die Forderungen nach interner Stabilität und der Beschränktheit der Norm $\|T\|_\infty$ können unter Verwendung des sogenannten Bounded Real Lemmas elegant als LMI formuliert werden [3]. Denn genau dann, wenn eine positiv definite Matrix P und eine Zahl $\gamma > 0$ existieren, sodaß die Ungleichung

$$\begin{pmatrix} A^T P + P A & P B & C^T \\ B^T P & -\gamma E & D^T \\ C & D & -\gamma E \end{pmatrix} < 0 \quad (13)$$

erfüllt ist, gilt $\|T\|_\infty < \gamma$.

Vorgabe eines Polgebietes Eine Platzierung der Eigenwerte der Systemmatrix A und damit der Pole des geschlossenen Regelkreises in einem Gebiet der komplexen Ebene kann auch in der Form von LMI Anforderungen an die Ljapunovmatrix P formuliert werden. Ein sogenanntes LMI Gebiet ist dabei jedes Gebiet $\mathcal{R}$ in der komplexen Ebene, welches als

$$\mathcal{R} = \{z \in \mathbb{C} | L + zM + \bar{z}M^T < 0\}$$

mit geeigneten Matrizen $L = L^T$ und M angegeben werden kann. Anzumerken ist, daß $\mathcal{R}$ ein *konvexes Gebiet* ist. Das Polgebiet kann dabei in Form von Kreisen, Ellipsen, vertikalen beziehungsweise horizontalen Streifen oder konischen Sektoren angegeben werden, beziehungsweise aus dem Durchschnitt einzelner Gebiete gebildet werden.

Bei unseren Betrachtungen ist das vorgebbare Gebiet durch vier Parameter (p_1 , p_2 , p_3 und p_4) bestimmt. Deren Festlegung ist aus Abbildung 6 ersichtlich.

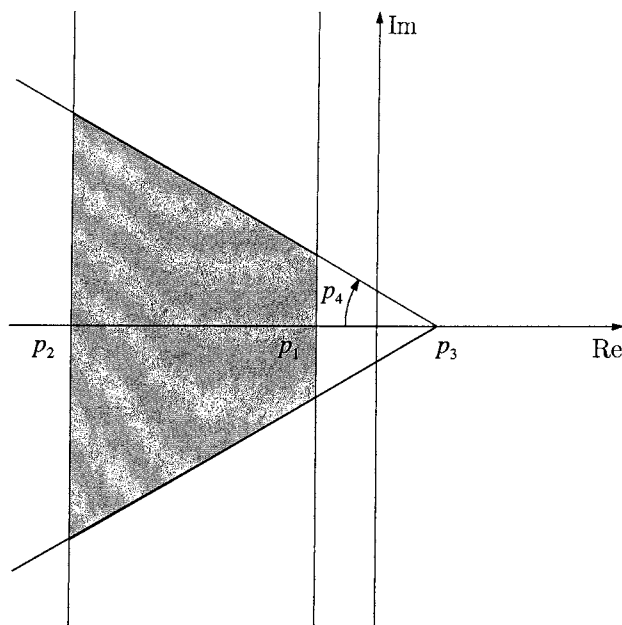


Abbildung 6: Parameter zur Festlegung des Polgebietes

- Die Schranke p_1 legt das Abklingverhalten der „langsamsten“ Eigenbewegungen fest.
- Durch die Schranke p_2 wird sichergestellt, daß die Realteile der Pole einen bestimmten Minimalwert nicht unterschreiten, was im Hinblick auf die Stellgrößenbeschränkung sinnvoll ist.
- Die Vorgabe des Sektors mit dem beschreibenden Parameterpaar (p_3, p_4) dient zur Einhaltung bestimmter Dämpfungen.

Aus der Vorgabe des LMI Gebietes resultiert dann auf einfache Weise eine LMI in der Variablen P und der Systemmatrix $\mathcal{A}$ [5, 6].

3.2.2 Reglersynthese für lineare zeitinvariante Übertragungssysteme

Die Zusammenfassung der Anforderungen an das Verhalten des Regelkreises als LMI Beschränkungen für die Systemmatrizen und die Ljapunovmatrix des Regelkreises führt nun zum Syntheseproblem. Dessen rechnerunterstützte Lösung liefert den gesuchten Regler. Die bei der Reglersynthese festzulegenden Größen sind:

- die Reglerparameter A_K , B_K , C_K und D_K ,
- die Ljapunovmatrix P und
- zusätzliche Hilfsvariablen wie etwa γ (Schranke für die H_∞ Norm).

Die Erfüllung aller Spezifikationen mit Hilfe *einer einzigen Ljapunovmatrix* $P > 0$ kann zu konservativen Lösungen führen.

Allerdings besteht jetzt das Problem darin, daß zunächst alle Bedingungen *nichtlinear* in den Reglerparametern A_K , B_K , C_K , D_K und der Ljapunovmatrix P sind. Mit der in [16] vorgestellten Methode kann diese fundamentale Schwierigkeit jedoch elegant beseitigt werden. Dazu wird eine nichtlineare Transformation der Reglerparameter eingeführt, und die Matrizenungleichungen werden geeigneten Kongruenztransformationen unterworfen. Die erforderlichen Schritte werden in Anhang B näher erläutert.

3.3 Entwurf eines robusten LZI Reglers

Nach dem in 3.2 beschriebenen Verfahren wird nun ein System von LMIs aufgestellt. Dafür werden in dem in MATLAB entwickelten Reglerentwurfsprogramm die in der LMI-Toolbox zur Verfügung gestellten Funktionen verwendet [10]. Das Ergebnis ist ein linearer zeitinvarianter Regler K der Form (8) mit der Eingangsgröße y nach (6) und der Ausgangsgröße u . Die Reglerordnung $n = 4$ entspricht der Systemordnung des Labormodells. Die Entwurfsparameter zur Festlegung des gewünschten Polgebietes lauten:

$$p_1 = -10, \quad p_2 = -50, \quad p_3 = 10 \quad \text{und} \quad p_4 = \frac{\pi}{4}. \quad (14)$$

Für den ersten Entwurf wird nun angenommen, daß sich der Wagen bei $\delta = 0$ ($l = 0.25$ m) *in der Mitte* des möglichen Auslenkungsbereiches befindet. Das Zustandsraummodell des berechneten LZI Reglers wird im Abschnitt 4.2 angeführt.

3.3.1 Stabilitätsanalyse

Laut den Anforderungen muß die asymptotische Stabilität des Regelkreises für alle möglichen Wagenpositionen erfüllt werden. Hierzu sind in Abbildung 7 die Eigenwerte der Systemmatrix $\mathcal{A}$ des geschlossenen Regelkreises in Abhängigkeit von der Wagenposition dargestellt, wobei das durch die Entwurfsparameter festgelegte Polgebiet eingezeichnet ist.

Aus der Abbildung können nun folgende Ergebnisse abgelesen werden:

- Der ermittelte Regler erfüllt für $\delta = 0$ die Vorgabe, daß alle Eigenwerte der Systemmatrix $\mathcal{A}$ im definierten Polgebiet liegen.
- Für die Extrempositionen $\delta = -1$ und $\delta = 1$ befinden sich einige Eigenwerte außerhalb des definierten Polgebietes, jedoch in der linken offenen s -Halbebene.

Der Regler erfüllt also die Robustheitsanforderung hinsichtlich der Stabilität jedoch nicht jene bezüglich des angestrebten Polgebietes.

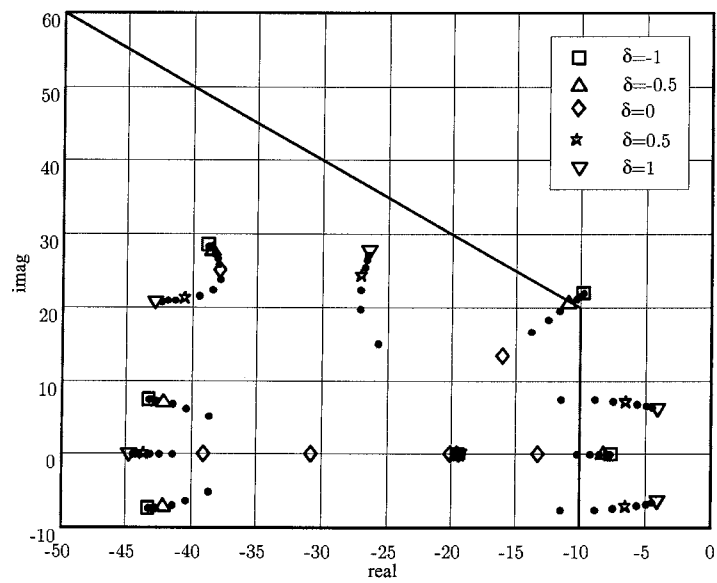


Abbildung 7: Lage der Eigenwerte in Abhängigkeit von der Wagenposition

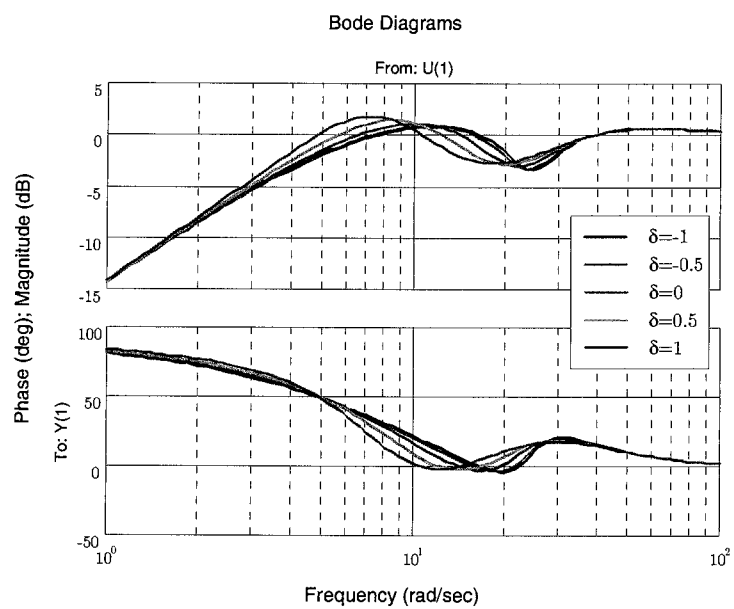


Abbildung 8: Bodediagramme der Übertragungsfunktionen T für verschiedene Wagenpositionen

3.3.2 Werte der H_∞ Norm von T für verschiedene Wagenpositionen

Der aufgrund der Optimierung erreichte Minimalwert in der H_∞ Norm von T wird um etwa 8.6 % erhöht, falls der Wagen die Extremposition $\delta = 1$ einnimmt ($\|T\|_{\infty, \delta=0} \leq 1.139$). Die Bodediagramme der Übertragungsfunktionen sind Abbildung 8 zu entnehmen. Daraus ist auch die Robustheit des Entwurfes ersichtlich.

3.3.3 Messungen am Labormodell

Für drei Experimente mit den Wagenpositionen $\delta = -1$, $\delta = 0$ und $\delta = 1$ werden die Ergebnisse in der linken Hälfte der Abbildung 9 dargestellt. Um den Einfluß der Wagenposition deutlich zu machen, werden für einen Positionswechsel von -0.5 rad auf 0.5 rad der Winkel zwischen Aufbau und Bezugsachse φ , der Winkel zwischen Ausleger und Aufbau α und die Stellgröße u dargestellt.

Die Abhängigkeit des Überschwingens von der Wagenposition ist sehr leicht ersichtlich. Für $\delta = -1$ tritt kein Überschwingen auf, hingegen für $\delta = 1$ ist das Überschwingen sehr deutlich erkennbar. Die Regelung ist zwar robust, aber die erzielbaren Ergebnisse sind stark abhängig vom veränderlichen Parameter des Modells.

Durch einen weiteren Entwurf soll nun geklärt werden, inwieweit sich durch das Einbeziehen der gemessenen Wagenposition in das Regelgesetz Verbesserungen erzielen lassen.

3.4 Entwurf eines parameterabhängigen Regelgesetzes

3.4.1 Gain Scheduling Verfahren

Eine Möglichkeit parameterabhängige Regelgesetze zu entwerfen, stellen sogenannte *Gain Scheduling Verfahren* dar [2, 11, 15]. Dabei werden für verschiedene im Betrieb vorkommende Parameterkombinationen, oder verschiedene *Betriebspunkte* der Regelstrecke, Regler entworfen, und im Betrieb mit Hilfe geeigneter Methoden zwischen diesen Reglern umgeschaltet. In unserem Fall ist die momentane Position des Wagens eine meßbare Größe. Sie legt die Parameterabhängigkeit des Systems fest, beinhaltet also wesentliche Informationen zur aktuellen Systembeschreibung, und sie wird die Parameterabhängigkeit des Regelgesetzes bestimmen! Nach folgender Strategie wollen wir nun vorgehen:

- Ausgehend von einem bestimmten Betriebspunkt, der durch den Entwurfparameter δ charakterisiert ist, und dem zugehörigen LZI Modell der Regelstrecke, wird mit dem in 3.2 vorgestellten Verfahren ein linearer und konstanter Regler ermittelt.
- Dieser Entwurf wird für eine festgelegte Anzahl von ausgewählten Betriebspunkten durchgeführt. Diese Betriebspunkte sind repräsentativ für den gesamten Parameterbereich. Damit erhalten wir zu jedem Betriebspunkt je einen zugehörigen Regler.
- Die nun vom jeweiligen Wert des Entwurfparameters δ abhängigen Regler werden zu einem linearen parametervariablen (LPV) Regler zusammengefaßt, das heißt, die

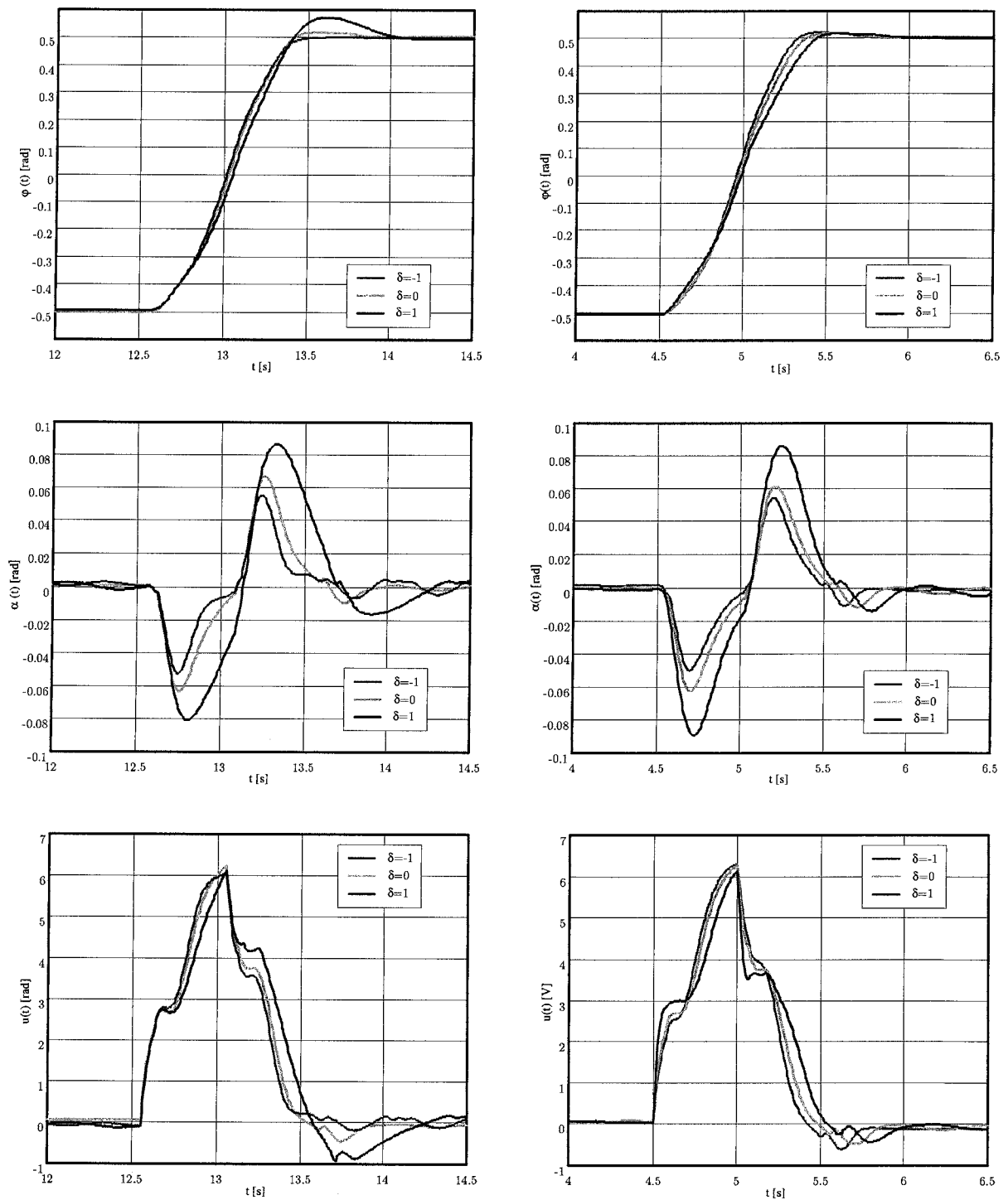


Abbildung 9: Vergleich der Meßergebnisse LZI Regler (links) und LPV Regler (rechts)

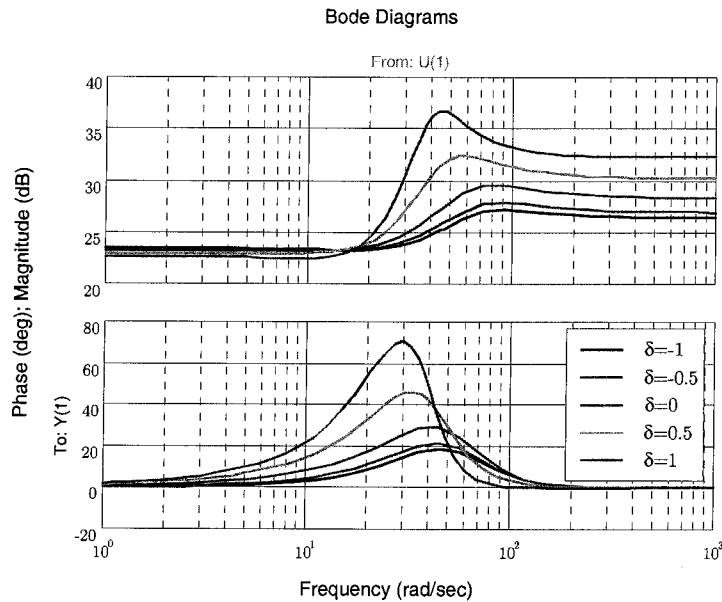


Abbildung 10: Bodediagramme der Reglerübertragungsfunktionen K_1

Parameterabhängigkeit des Reglers von den Betriebspunkten wird durch ein LPV Übertragungssystem² beschrieben.

Im Betrieb selbst sind dann keine weiteren Berechnungen nötig, der Regler ändert nur in Abhängigkeit vom gemessenen Parameter sein Übertragungsverhalten.

Das Reglerübertragungsverhalten wird durch zwei Übertragungsfunktionen vierter Ordnung $K_1(\delta, s)$ und $K_2(\delta, s)$ charakterisiert. Mit der für alle Entwürfe gleichen Vorgabe des gewünschten Polgebietes nach (14) ergeben sich in Abhängigkeit von der Wagenposition die einzelnen Regler. In Abbildung 10 sind zur Verdeutlichung der Parameterabhängigkeit die Bodediagramme der Übertragungsfunktionen $K_1(\delta, s)$ für das Übertragungsverhalten von der Regelabweichung z nach u zu sehen.

Beide Übertragungsfunktionen werden in unserem Fall durch je neun Koeffizienten charakterisiert, die vom Entwurfparameter δ abhängig sind. Diese Abhängigkeit kann mit Hilfe von Ausgleichspolynomen des geeignet zu wählenden Grades m mathematisch erfaßt werden. Die Implementierung der Übertragungsfunktionen $K_1(\delta, s)$ und $K_2(\delta, s)$ sowohl für die Simulation als auch für die Verwendung am Labormodell, wird in Kapitel 4 beschrieben. Die resultierenden Reglerdaten werden im Abschnitt 4.2 angeführt.

²Bei einem LPV Übertragungssystem der Form

$$\begin{aligned}\dot{x} &= A(\rho(t))x + B(\rho(t))u \\ y &= C(\rho(t))x + D(\rho(t))u\end{aligned}$$

sind Elemente der Systemmatrizen abhängig von einem zeitvariablen Parameter $\rho(t)$. In unserem Fall ist die Abhängigkeit in der Systembeschreibung des Labormodells und in weiterer Folge in der Systembeschreibung des entworfenen Reglers durch den gemessenen Parameter δ gegeben.

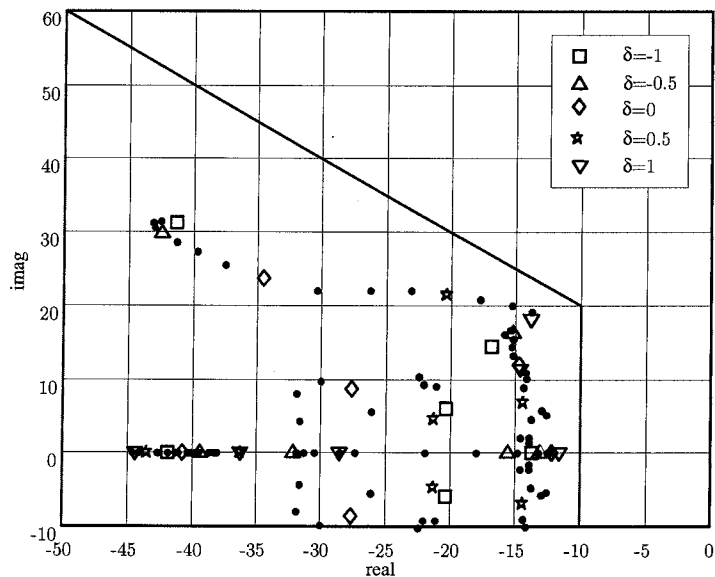


Abbildung 11: Lage der Eigenwerte in Abhängigkeit von der Wagenposition bei Verwendung des LPV Reglers

Wichtig ist, daß zu bestimmten Wagenpositionen ein Regler entworfen wird, der eingesetzte Regler aber mit Koeffizienten realisiert wird, die durch eine Approximation der Abhängigkeit der errechneten Koeffizienten vom Entwurfsparameter δ gebildet werden.

3.4.2 Stabilitätsanalyse

Die Lage der Eigenwerte von $\mathcal{A}$ in Abhängigkeit vom Parameter δ ist in Abbildung 11 dargestellt. Somit ist gezeigt, daß für jede beliebige Wagenposition nicht nur alle Eigenwerte der Systemmatrix in der linken offenen s -Halbebene liegen, sondern auch in dem für den Entwurf vorgegebenen Polgebiet. Es ist zu beachten, daß dies für den Stabilitätsbeweis für ein *zeitvariables* Übertragungssystem aber nicht ausreicht [4]. Eine Möglichkeit die robuste Stabilität des Übertragungssystems zu zeigen, eröffnet die in [13] angegebene Vorgehensweise. Dafür sind sowohl für die Strecke als auch für den Regler mathematische Modelle einer speziellen Form zu ermitteln, die vom beschreibenden Parameter δ abhängen. Der Stabilitätsnachweis wird dann unter der Voraussetzung $-1 \leq \delta \leq 1$ gezeigt. Dabei geht jedoch der prinzipielle physikalische Zusammenhang zwischen dem Labormodell und dem beschreibendem System verloren. Dieser umfassende Stabilitätsbeweis wird an dieser Stelle nicht weiter verfolgt.

3.4.3 Werte der H_∞ Norm von T für verschiedene Wagenpositionen

Der aufgrund der Optimierung erreichte Minimalwert in der H_∞ Norm von T wird um etwa 4.0 % erhöht, falls der Wagen die Extremposition $\delta = -1$ einnimmt ($\|T\|_{\infty, \delta=0} \leq 1.142$). Die Bodediagramme der Übertragungsfunktionen sind Abbildung 12 zu entnehmen. Daraus ist auch die Robustheit des Entwurfes ersichtlich.

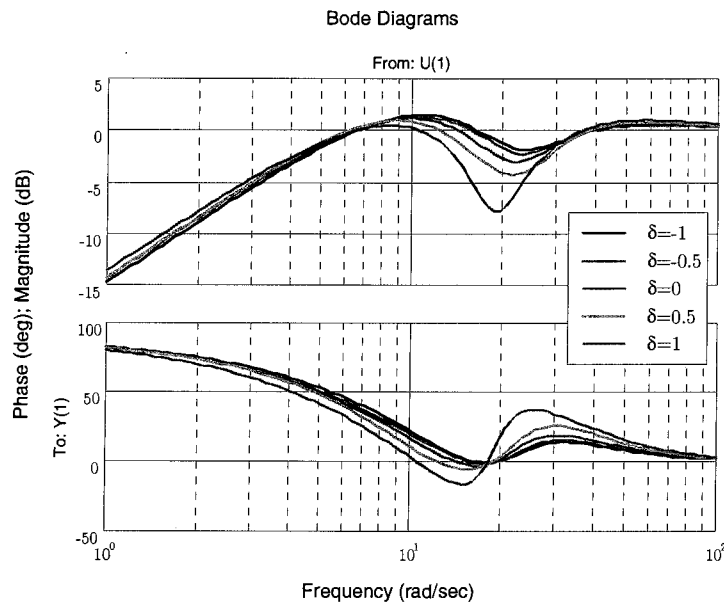


Abbildung 12: Bodediagramme der Übertragungsfunktionen T für verschiedene Wagenpositionen

Vergleicht man die beiden maximal auftretenden Werte der H_∞ Norm der Übertragungsfunktion T für den LZI Regler mit $\|T\|_{\infty, \max} = 1.237$ und für den LPV Regler mit $\|T\|_{\infty, \max} = 1.188$, so liegt die durch das parameterabhängige Regelgesetz erzielte Verbesserung in der H_∞ Norm bei 4.12 %.

3.4.4 Messungen am Labormodell

Es werden die Ergebnisse aus drei Experimenten mit den Wagenpositionen $\delta = -1$, $\delta = 0$ und $\delta = 1$ in der rechten Hälfte der Abbildung 9 miteinander verglichen. Das Überschwingen ist nun praktisch unabhängig von der Wagenposition und außerdem für $\delta = 1$ am geringsten. Dabei ist noch auffällig, daß am Beginn des Positionswechsels von -0.5 rad auf 0.5 rad die schnellste Bewegung für $\delta = 1$ auftritt. Dies ist auch aus dem Stellgrößenverlauf u ersichtlich, da in diesem Fall anfänglich die größte Stellgröße ausgegeben wird. Es kommt zu einer großen Auslenkung α des Auslegers bezüglich des Aufbaus. Diese Tatsache ist aufgrund des maximalen Trägheitsmomentes auch verständlich. Weiters nimmt der Einfluß von Reibungseffekten auf das Schwingungsverhalten bei zunehmendem Trägheitsmoment ab.

Aus der Gegenüberstellung der Ergebnisse für den LZI Regler und den LPV Regler in Abbildung 9 sind die erzielbaren Verbesserungen durch den Einsatz eines parameterabhängigen Regelgesetzes eindeutig ersichtlich.

4 Reglerimplementierung

Die ermittelten Regler wurden vorerst in der Simulation getestet und dann am Labormodell eingesetzt. Dazu müssen die Regler in Simulink implementiert werden. Dies ist für die Simulation einfacher, da hier Rechenzeit beziehungsweise Echtzeit der Regelung keine Rolle spielen. Verschiedene Möglichkeiten der Reglerrealisierung werden in [8] vorgestellt. Hier wird die direkte Implementierung gezeigt, die dann auch für die Prozeßregelung verwendet wird.

4.1 Direkte Implementierung

Durch die in Abschnitt 3.4 eingeführten Übertragungsfunktionen $K_1(\delta, s)$ und $K_2(\delta, s)$ wird das Übertragungsverhalten des Reglers mit einer minimalen Anzahl von Koeffizienten beschrieben. Für die Zustandsraumdarstellung ist es wichtig, daß die Anzahl der in den Systemmatrizen verwendeten Koeffizienten aufgrund der zu implementierenden Parameterabhängigkeit möglichst klein ist. Die beiden Übertragungsfunktionen werden nun durch ein Zustandsraummodell in Beobachternormalform realisiert, welches die vorhin genannte Bedingung erfüllt und aus dem die Koeffizienten der beiden Reglerübertragungsfunktionen

$$\begin{aligned} K_1(\delta, s) &= d_1(\delta) + \frac{\tilde{b}_{13}(\delta)s^3 + \tilde{b}_{12}(\delta)s^2 + \tilde{b}_{11}(\delta)s + \tilde{b}_{10}(\delta)}{s^4 + a_3(\delta)s^3 + a_2(\delta)s^2 + a_1(\delta)s + a_0(\delta)} \\ K_2(\delta, s) &= d_2(\delta) + \frac{\tilde{b}_{23}(\delta)s^3 + \tilde{b}_{22}(\delta)s^2 + \tilde{b}_{21}(\delta)s + \tilde{b}_{20}(\delta)}{s^4 + a_3(\delta)s^3 + a_2(\delta)s^2 + a_1(\delta)s + a_0(\delta)} \end{aligned}$$

direkt ablesbar sind:

$$\begin{aligned} \dot{\lambda} &= \begin{pmatrix} 0 & 0 & 0 & -a_0 \\ 1 & 0 & 0 & -a_1 \\ 0 & 1 & 0 & -a_2 \\ 0 & 0 & 1 & -a_3 \end{pmatrix} \lambda + \begin{pmatrix} \tilde{b}_{10} & \tilde{b}_{20} \\ \tilde{b}_{11} & \tilde{b}_{21} \\ \tilde{b}_{12} & \tilde{b}_{22} \\ \tilde{b}_{13} & \tilde{b}_{23} \end{pmatrix} y \\ u &= (0 \ 0 \ 0 \ 1) \lambda + (d_1 \ d_2) y. \end{aligned} \quad (15)$$

Die Abhängigkeit der Koeffizienten vom Entwurfsparameter δ wird durch Ausgleichspolynome des Grades m erfaßt. Es gilt dabei für $i = \{0, 1, 2, 3\}$

$$\begin{aligned} \tilde{b}_{1i} &= \tilde{b}_{1i,0} + \tilde{b}_{1i,1}\delta + \tilde{b}_{1i,2}\delta^2 + \dots + \tilde{b}_{1i,m}\delta^m \\ \tilde{b}_{2i} &= \tilde{b}_{2i,0} + \tilde{b}_{2i,1}\delta + \tilde{b}_{2i,2}\delta^2 + \dots + \tilde{b}_{2i,m}\delta^m \\ a_i &= a_{i,0} + a_{i,1}\delta + a_{i,2}\delta^2 + \dots + a_{i,m}\delta^m \\ d_1 &= d_{1,0} + d_{1,1}\delta + d_{1,2}\delta^2 + \dots + d_{1,m}\delta^m \\ d_2 &= d_{2,0} + d_{2,1}\delta + d_{2,2}\delta^2 + \dots + d_{2,m}\delta^m. \end{aligned} \quad (16)$$

Bei der nun vorgestellten Methode werden die Zustandsdifferentialgleichungen und die Ausgangsgleichung des Übertragungssystems direkt realisiert. Es wird also ein Übertragungssystem mit den beiden Eingangsgrößen z und α und der Ausgangsgröße u gebildet.

Die Zustandsdifferentialgleichungen abgeleitet aus der Beobachternormalform nach (15) lauten dabei

$$\begin{aligned}\dot{\lambda}_1 &= -a_0\lambda_4 + \tilde{b}_{10}z + \tilde{b}_{20}\alpha \\ \dot{\lambda}_2 &= \lambda_1 - a_1\lambda_4 + \tilde{b}_{11}z + \tilde{b}_{21}\alpha \\ \dot{\lambda}_3 &= \lambda_2 - a_2\lambda_4 + \tilde{b}_{12}z + \tilde{b}_{22}\alpha \\ \dot{\lambda}_4 &= \lambda_3 - a_3\lambda_4 + \tilde{b}_{13}z + \tilde{b}_{23}\alpha\end{aligned}\tag{17}$$

und die Ausgangsgleichung

$$u = \lambda_4 + d_1z + d_2\alpha.$$

Ein Simulink Koppelplan für diesen Regler mit Ausgleichspolynomen dritten Grades ist in Abbildung 13 dargestellt. Die Ermittlung der für (17) benötigten parameterabhängigen Koeffizienten geschieht auf folgende Weise:

Der mit K_A bezeichnete Simulinkblock ist ein Matrix Multiplizierer (*Matrix Gain*), durch den die Matrix

$$K_A = \begin{pmatrix} a_{0,3} & a_{0,2} & a_{0,1} & a_{0,0} \\ a_{1,3} & a_{1,2} & a_{1,1} & a_{1,0} \\ a_{2,3} & a_{2,2} & a_{2,1} & a_{2,0} \\ a_{3,3} & a_{3,2} & a_{3,1} & a_{3,0} \end{pmatrix}$$

mit den Elementen, die nach (16) die Koeffizienten der Ausgleichspolynome festlegen, mit dem Vektor

$$\begin{pmatrix} \delta^3 & \delta^2 & \delta^1 & \delta^0 \end{pmatrix}^T$$

multipliziert wird. Am Ausgang des Blockes wird somit der Datenvektor

$$\begin{pmatrix} a_0 & a_1 & a_2 & a_3 \end{pmatrix}^T$$

gebildet, der einen Teil der parameterabhängigen Koeffizienten für (17) repräsentiert.

Die Berechnung der restlichen parameterabhängigen Koeffizienten wird mit den in K_{B1} , K_{B2} , K_{D1} und K_{D2} übergebenen Koeffizienten der Ausgleichspolynome durchgeführt. Die Daten der einzelnen Matrizen werden in Abschnitt 4.2 angeführt.

Dieser Regler der Ordnung $n = 4$ mit $l = 2$ Eingangsgrößen, für den bei der Verwendung von Ausgleichspolynomen dritten Grades ($m = 3$) 56 Parameter zur Beschreibung notwendig sind, kann neben der Simulation auch am Labormodell verwendet werden. Die verwendete Software schafft es problemlos, für den derart implementierten Regler lauffähige Programme zu generieren.

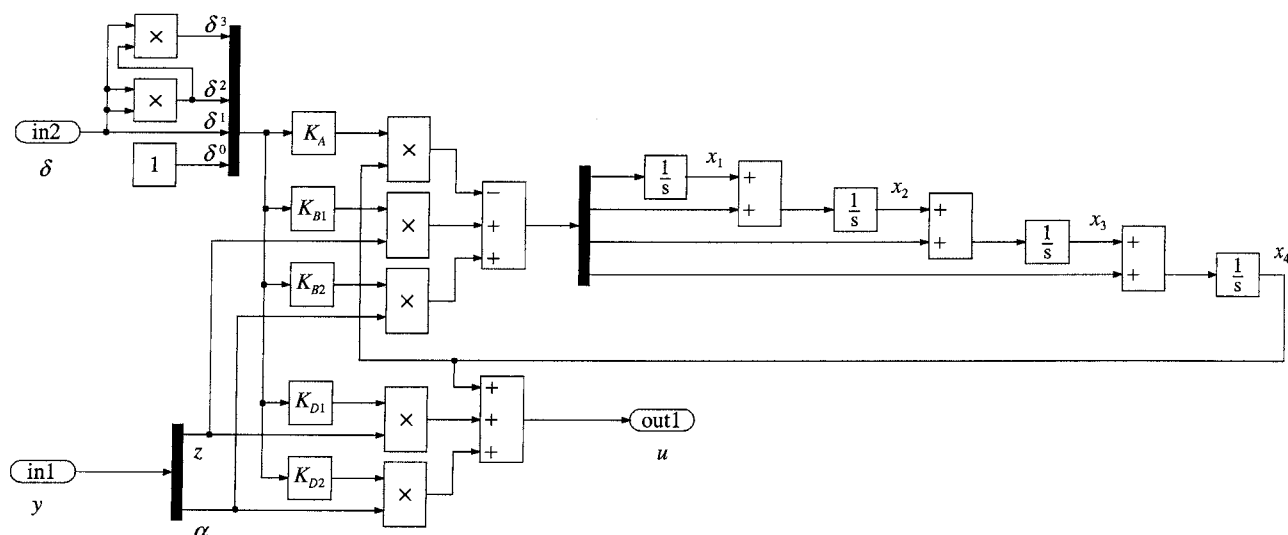


Abbildung 13: Simulink Koppelplan für die direkte Implementierung

4.2 Reglerdaten

Das Zustandsraummodell des in Abschnitt 3.3 berechneten LZI Reglers lautet:

$$\dot{\lambda} = \begin{pmatrix} 0 & 0 & 0 & -4.037 \cdot 10^6 \\ 1 & 0 & 0 & -3.505 \cdot 10^5 \\ 0 & 1 & 0 & -9.971 \cdot 10^3 \\ 0 & 0 & 1 & -1.425 \cdot 10^2 \end{pmatrix} \lambda + \begin{pmatrix} -4.686 \cdot 10^7 & 2.290 \cdot 10^8 \\ -3.333 \cdot 10^6 & 1.648 \cdot 10^7 \\ -4.415 \cdot 10^4 & 2.579 \cdot 10^5 \\ 1.467 \cdot 10^1 & 4.715 \cdot 10^2 \end{pmatrix} y \quad (18)$$

$$u = (0 \ 0 \ 0 \ 1) \lambda + (2.603 \cdot 10^1 \ -3.168 \cdot 10^1) y$$

Mit diesem Regler sind die Experimente, deren Ergebnisse in der linken Hälfte der Abbildung 9 zu sehen sind, durchgeführt worden.

Das in Abschnitt 3.4 beschriebene Entwurfsverfahren liefert bei der Verwendung von Ausgleichspolynomen des Grades $m = 3$ die in Tabelle 2 angeführten Zahlenwerte. Die Experimente am Labormodell, deren Ergebnisse in der rechten Hälfte der Abbildung 9 gezeigt werden, sind mit dem durch diese Daten festgelegten Regler durchgeführt worden.

5 Zusammenfassung

Bei dem neu aufgebauten Labormodell wird durch Positionierung des Wagens auf dem Ausleger das Trägheitsmoment verändert. Die dem Parameter δ zugeordnete Wagenposition kann gemessen werden.

Das in Abschnitt 3.2 vorgestellte Reglerentwurfsverfahren liefert für ein lineares zeitinvariantes Übertragungssystem einen LZI Regler. Dieses Verfahren wird in Abschnitt 3.3 verwendet, um einen robusten Regler zu entwerfen. In Abschnitt 3.4 werden für verschiedene Wagenpositionen, die durch den Parameter δ eindeutig festgelegt sind, Regler

K_A	i	$a_{i,3}$	$a_{i,2}$	$a_{i,1}$	$a_{i,0}$
	0	$3.0050 \cdot 10^5$	$-7.3996 \cdot 10^5$	$-7.0106 \cdot 10^5$	$3.7261 \cdot 10^6$
	1	$4.5204 \cdot 10^4$	$-6.8268 \cdot 10^4$	$-1.1841 \cdot 10^5$	$3.2237 \cdot 10^5$
	2	$1.1658 \cdot 10^3$	$-1.7040 \cdot 10^3$	$-3.6232 \cdot 10^3$	$9.3747 \cdot 10^3$
	3	$8.8455 \cdot 10^0$	$-1.1610 \cdot 10^1$	$-2.6153 \cdot 10^1$	$1.3839 \cdot 10^2$

K_{B1}	i	$\tilde{b}_{1i,3}$	$\tilde{b}_{1i,2}$	$\tilde{b}_{1i,1}$	$\tilde{b}_{1i,0}$
	0	$3.0128 \cdot 10^6$	$-1.7084 \cdot 10^6$	$-2.8752 \cdot 10^7$	$-4.3238 \cdot 10^7$
	1	$8.5155 \cdot 10^4$	$2.6733 \cdot 10^5$	$-1.1415 \cdot 10^6$	$-3.0335 \cdot 10^6$
	2	$6.0852 \cdot 10^3$	$1.0936 \cdot 10^4$	$-8.5091 \cdot 10^3$	$-4.1084 \cdot 10^4$
	3	$2.1457 \cdot 10^1$	$8.1126 \cdot 10^0$	$-2.1439 \cdot 10^1$	$1.2830 \cdot 10^1$

K_{B2}	i	$\tilde{b}_{2i,3}$	$\tilde{b}_{2i,2}$	$\tilde{b}_{2i,1}$	$\tilde{b}_{2i,0}$
	0	$2.9648 \cdot 10^7$	$-9.0732 \cdot 10^7$	$-8.3502 \cdot 10^7$	$1.9259 \cdot 10^8$
	1	$4.1184 \cdot 10^6$	$-6.7640 \cdot 10^6$	$-1.0846 \cdot 10^7$	$1.3645 \cdot 10^7$
	2	$1.0429 \cdot 10^5$	$-1.2999 \cdot 10^5$	$-3.1121 \cdot 10^5$	$2.1041 \cdot 10^5$
	3	$6.6919 \cdot 10^2$	$-6.2589 \cdot 10^2$	$-2.2775 \cdot 10^3$	$3.0905 \cdot 10^2$

	j	$b_{j4,3}$	$b_{j4,2}$	$b_{j4,1}$	$b_{j4,0}$
K_{D1}	1	$-3.4051 \cdot 10^{-1}$	$4.7121 \cdot 10^0$	$1.0227 \cdot 10^1$	$2.6068 \cdot 10^1$
K_{D2}	2	$-1.6730 \cdot 10^0$	$1.6116 \cdot 10^1$	$7.6845 \cdot 10^0$	$-2.6156 \cdot 10^1$

Tabelle 2: Koeffizienten der Ausgleichspolynome

entworfen, wodurch ein vom Parameter δ abhängiger Regler bestimmt wird, der mit einem linearen parametervariablen Übertragungssystem beschrieben werden kann. Durch Verwendung dieses Regelgesetzes wird eine wesentliche Verbesserung im Regelungsverhalten gegenüber dem ursprünglichen Regler erreicht. Die Position des Wagens fließt in das Regelgesetz ein und erlaubt somit eine beliebige Positionierung am Ausleger ohne negative Auswirkungen auf das Systemverhalten.

A Technische Daten

Zusammenfassung der Systemparameter Für den Reglerentwurf in 3.3 und 3.4 werden die Systemmatrix A und der Eingangsvektor b

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{k_s}{J_{Auf}} & -\frac{k_2}{J_{Auf}} & 0 \\ 0 & -\left(\frac{J_A}{J_{Auf}} + 1\right) \frac{k_s}{J_A} & \frac{k_2}{J_{Auf}} - \frac{\dot{J}_A}{J_A} & -\frac{\dot{J}_A}{J_A} \end{pmatrix} \quad b = \begin{pmatrix} 0 \\ 0 \\ \frac{k_1}{J_{Auf}} \\ -\frac{k_1}{J_{Auf}} \end{pmatrix}$$

des zu dem Labormodell ermittelten mathematischen Modells für eine bestimmte feste Position, das heißt $\dot{J}_A = 0$, verwendet. Die Werte aller angeführten Terme sind:

$$\begin{aligned} \frac{k_s}{J_{Auf}} &= 43.571 \frac{\text{Nm}}{\text{rad} \cdot \text{kgm}^2} \\ -\left(\frac{J_A}{J_{Auf}} + 1\right) \frac{k_s}{J_A} &= [-317.4, -155.66] \frac{\text{Nm}}{\text{rad} \cdot \text{kgm}^2} \\ \frac{k_2}{J_{Auf}} &= 68.416 \frac{(\text{Nm})^2}{\text{V} \cdot \text{kgm}^2} \\ \frac{k_1}{J_{Auf}} &= 24.394 \frac{\text{Nm}}{\text{V} \cdot \text{kgm}^2}. \end{aligned}$$

B Reglerentwurfsverfahren

Die wesentlichen Schritte des in [16] angeführten Reglerentwurfsverfahrens werden hier knapp zusammengefaßt.

Festlegung der Entwurfsvariablen Mit der Anzahl n der Zustandsvariablen der Regelstrecke und der Ordnung k des Reglers werden die Matrizen P und P^{-1} zerlegt in

$$P = \begin{pmatrix} Y & N \\ N^T & \star \end{pmatrix} \quad \text{und} \quad P^{-1} = \begin{pmatrix} X & M \\ M^T & \star \end{pmatrix} \quad (19)$$

mit den symmetrischen $n \times n$ Matrizen X und Y . Wegen $PP^{-1} = E$ gilt

$$P \begin{pmatrix} X \\ M^T \end{pmatrix} = \begin{pmatrix} E \\ 0 \end{pmatrix}$$

und es können zwei Matrizen

$$\Pi_1 := \begin{pmatrix} X & E \\ M^T & 0 \end{pmatrix} \quad \text{und} \quad \Pi_2 := \begin{pmatrix} E & Y \\ 0 & N^T \end{pmatrix} \quad (20)$$

definiert werden, sodaß $P\Pi_1 = \Pi_2$ erfüllt wird. Nun werden die neuen Reglervariablen wie folgt festgelegt:

$$\hat{A} := NA_K M^T + NB_K CX + YBC_K M^T + Y(A + BD_K C)X \quad (21)$$

$$\hat{B} := NB_K + YBD_K \quad (22)$$

$$\hat{C} := C_K M^T + D_K CX \quad (23)$$

$$\hat{D} := D_K. \quad (24)$$

Die neuen Variablen $\hat{A}$, $\hat{B}$ und $\hat{C}$ sind dabei $n \times n$, $n \times n_u$ beziehungsweise $n_y \times n$ Matrizen. Wenn M und N vollen Zeilenrang haben und die Matrizen $\hat{A}$, $\hat{B}$, $\hat{C}$, $\hat{D}$, X und Y gegeben sind, so können die Reglermatrizen A_K , B_K , C_K und D_K immer bestimmt werden. Die Lösung ist dabei eindeutig, wenn M und N quadratische ($k = n$) und invertierbare Matrizen sind. Die Motivation für diese Transformation liegt in den folgenden sich aus (20) bis (24) ergebenden Identitäten:

$$\begin{aligned}\Pi_1^T P \mathcal{A} \Pi_1 &= \Pi_2^T \mathcal{A} \Pi_1 = \begin{pmatrix} AX + B\hat{C} & A + B\hat{D}C \\ \hat{A} & YA + \hat{B}C \end{pmatrix} \\ \Pi_1^T P \mathcal{B} &= \Pi_2^T \mathcal{B} = \begin{pmatrix} B_w + B\hat{D}D_w \\ YB_w + \hat{B}D_w \end{pmatrix} \\ C\Pi_1 &= (C_z X + D_z \hat{C} \quad C_z + D_z \hat{D}C) \\ \Pi_1^T P \Pi_1 &= \Pi_1^T \Pi_2 = \begin{pmatrix} X & E \\ E & Y \end{pmatrix}.\end{aligned}\tag{25}$$

Sollen nun die Spezifikationen in eine LMI für die Synthese umgewandelt werden, so ist eine Kongruenztransformation durchzuführen. Das Ergebnis kann mit den Identitäten aus (25) in den Entwurfsvariablen dargestellt werden.

Der vollständige Katalog für die jeweilige Kongruenztransformation einer Anforderung ist [16] zu entnehmen. Diese Ungleichungen sind affin in $\hat{A}$, $\hat{B}$, $\hat{C}$, $\hat{D}$, X und Y . Damit ist die Lösbarkeit dieser LMI hinreichend für die Existenz eines Reglers mit dem gewünschten Übertragungsverhalten.

Konstruktion des Reglers Jedenfalls ist

$$\begin{pmatrix} X & E \\ E & Y \end{pmatrix} > 0\tag{26}$$

eine immer zu erfüllende Ungleichung. Aus (26) folgt mit dem Schurkomplement (12)

$$Y > 0 \quad \text{und} \quad X - Y^{-1} > 0,$$

sodaß $E - XY$ nicht singulär ist. Damit können immer quadratische und nicht singuläre Matrizen M und N gefunden werden, die

$$MN^T = E - XY$$

erfüllen³. Weiters sind damit auch die Matrizen Π_1 und Π_2 nicht singulär, und es können die Reglerparameter anhand von

$$D_K : = \hat{D}\tag{27}$$

$$C_K : = (\hat{C} - D_K C X) M^{-T}\tag{28}$$

$$B_K : = N^{-1}(\hat{B} - Y B D_K)\tag{29}$$

$$A_K : = N^{-1}(\hat{A} - N B_K C X - Y B C_K M^T - Y(A + B D_K C)X) M^{-T}.\tag{30}$$

ermittelt werden.

³Aus der Partitionierung von P und P^{-1} (19) folgt wegen $PP^{-1} = E$ unmittelbar für die links oben gebildete Blockmatrix: $YX + NM^T = E$.

Literatur

- [1] ACKERMANN J.: Robuste Regelung, Springer-Verlag, 1993
- [2] ASTROEM K., WITTENMARK B.: Adaptive Control, Addison-Wesley Publishing Company, 1995
- [3] BOYD S., EL GHAOU L., FERON E., BALAKRISHNAN V.: Linear Matrix Inequalities in System and Control Theory, SIAM, Philadelphia, 1994
- [4] CHI-TSONG CHEN: Introduction to Linear System Theory, Holt, Rinehart and Winston, 1970
- [5] CHILALI M., GAHINET P.: H_∞ Design with Pole Placement Constraints: an LMI Approach, IEEE Trans. Autom. Control, 41 (1996), S. 358-367
- [6] CHILALI M., GAHINET P., APKARIAN P.: Robust Pole Placement in LMI Regions, IEEE Trans. Autom. Control, 44 (1999), S. 2257-2270
- [7] DOCZY S.: Labormodell mit variablen Systemparametern, Interner Bericht, Institut für Regelungstechnik der Technischen Universität Graz, 1999
- [8] DOCZY S.: Regelung parameterabhängiger Übertragungssysteme, Dissertation, Institut für Regelungstechnik der Technischen Universität Graz, 2000
- [9] EL GHAOU L., NICULESCU S.: Advances in Linear Matrix Inequality Methods in Control, SIAM, Philadelphia, 2000
- [10] GAHINET P., NEMIROVSKI A., LAUB A., CHILALI M.: LMI Control Toolbox for Use with MATLAB, The MathWorks, Inc., 1995
- [11] HELMERSSON A.: Methods for Robust Gain Scheduling, PhD Thesis, Linköping University, Linköping, Sweden, 1995
- [12] NESTEROV Y., NEMIROVSKII A.: Interior-Point Polynomial Algorithms in Convex Programming, SIAM, Philadelphia, 1994
- [13] PACKARD A.: Gain scheduling via linear fractional transformations, System and Control Letters, 22 (1994), S. 79-92
- [14] QUANSER CONSULTING: A comprehensive and modular Laboratory for Control Systems Design and Implementation, 1995
- [15] RUGH W., SHAMMA J.: Research on gain scheduling, Automatica, 36 (2000), S. 1401-1425
- [16] SCHERER C., GAHINET P., CHILALI M.: Multi-Objective Output-Feedback Control via LMI Optimization, IEEE Trans. Autom. Control, 42 (1997), S. 896-911