SSRP 2011

17. Steirisches Seminar über Regelungstechnik und Prozessautomatisierung

S. Moschik, N. Dourdoumas (Hrsg.)

5. - 8. 9. 2011
Schloss Retzhof, Leibnitz, Österreich

ISBN: 978-3-901439-09-4

http://www.irt.tugraz.at

TU Graz
Graz University of Technology

Vorwort

Das Steirische Seminar über Regelungstechnik und Prozessautomatisierung findet in diesem Jahr zum 17. Mal im Schloss Retzhof, dem Bildungshaus des Landes Steiermark, statt. Die Veranstaltung wird in zweijährigem Rhythmus vom Institut für Regelungs- und Automatisierungstechnik der Technischen Universität Graz organisiert. Sie hat zum Ziel, aktuelle Arbeiten in Bereichen der Regelungs- und Prozessautomatisierungstechnik sowie der Regelungstheorie in universitärer und industrieller Forschung zur Diskussion zu stellen. Die Beiträge gehören also einem breiten Spektrum von Problemstellungen an. In dem vorliegenden Tagungsband sind die eingelangten Manuskripte zusammengefasst. Den Autoren sei an dieser Stelle für die Sorgfalt bei der Erstellung ihrer Beiträge herzlich gedankt.

Graz, im September 2011

Inhaltsverzeichnis

Vorsteuerungs-Entwurf im Frequenzbereich

Michael Zeitz
Universität Stuttgart, Institut für Systemdynamik
Pfaffenwaldring 9, D-70569 Stuttgart
zeitz@isys.uni-stuttgart.de

Kurzfassung

Modellbasierte Vorsteuerungen werden in vielen industriellen Anwendungen zur gezielten Beeinflussung des Führungsverhaltens einer Folgeregelung eingesetzt. Beim Vorsteuerungs-Entwurf muss unterschieden werden, ob die Solltrajektorien – wie bei einem Arbeitspunktwechsel – vorab geplant oder extern in Realzeit vorgegeben werden. Im ersten Fall erfolgt der Vorsteuerungs-Entwurf *offline* und die entworfenen Trajektorien werden z. B. in einer Look-up-Tabelle gespeichert. Der zweite Fall, der häufig in Servosystemen vorkommt, erfordert, dass die Steuertrajektorien *online* berechnet und aufgeschaltet werden. Die Flachheits-Methodik eröffnet einen einfachen Zugang zum Offline-Entwurf von Vorsteuerungen. In dem Beitrag wird für lineare zeitinvariante Systeme erläutert, wie der flachheitsbasierte Vorsteuerungs-Entwurf im Frequenzbereich durchgeführt wird. Der Online-Entwurf verwendet ein Sollwert-Filter, um bei minimalphasigen Systemen eine hinreichend glatte Solltrajektorie für die Bestimmung der Steuertrajektorie zu erzeugen. Für nicht-minimalphasige Systeme wird die Steuertrajektorie in einem Modellregelkreis aus den vorgegebenen Sollwerten simulativ bestimmt und in Realzeit aufgeschaltet. Schließlich sind im Frequenzbereich relativ einfache Aussagen über die Robustheit einer Vorsteuerung gegenüber frequenzabhängigen Modellfehlern möglich.

1 Einleitung

In vielen industriellen Anwendungen wird der klassische Regelkreis (closed-loop) aus Strecke und Regler – wie in Abb. 1 dargestellt – durch eine dynamische Sollwertführung und modellbasierte Vorsteuerung (open-loop) erweitert. Die Struktur aus Vorsteuerung (feedforward control) und Regler (feedback control) besitzt zwei Freiheitsgrade (two-degree-of-freedom control scheme), um das Führungs- und das Störverhalten unabhängig voneinander zu entwerfen [1]-[5].

Der Entwurf einer modellbasierten Vorsteuerung ist eigentlich im Zeitbereich beheimatet. Der Vorsteuerungs-Entwurf im Frequenzbereich bietet sich an, wenn das Ein-/Ausgangsmodell als Frequenzgang identifiziert wird und als Übertragungsfunktion vorliegt. Außerdem ist der Frequenzbereichs-Entwurf naheliegend, wenn der Regler im Frequenzbereich entworfen wird.

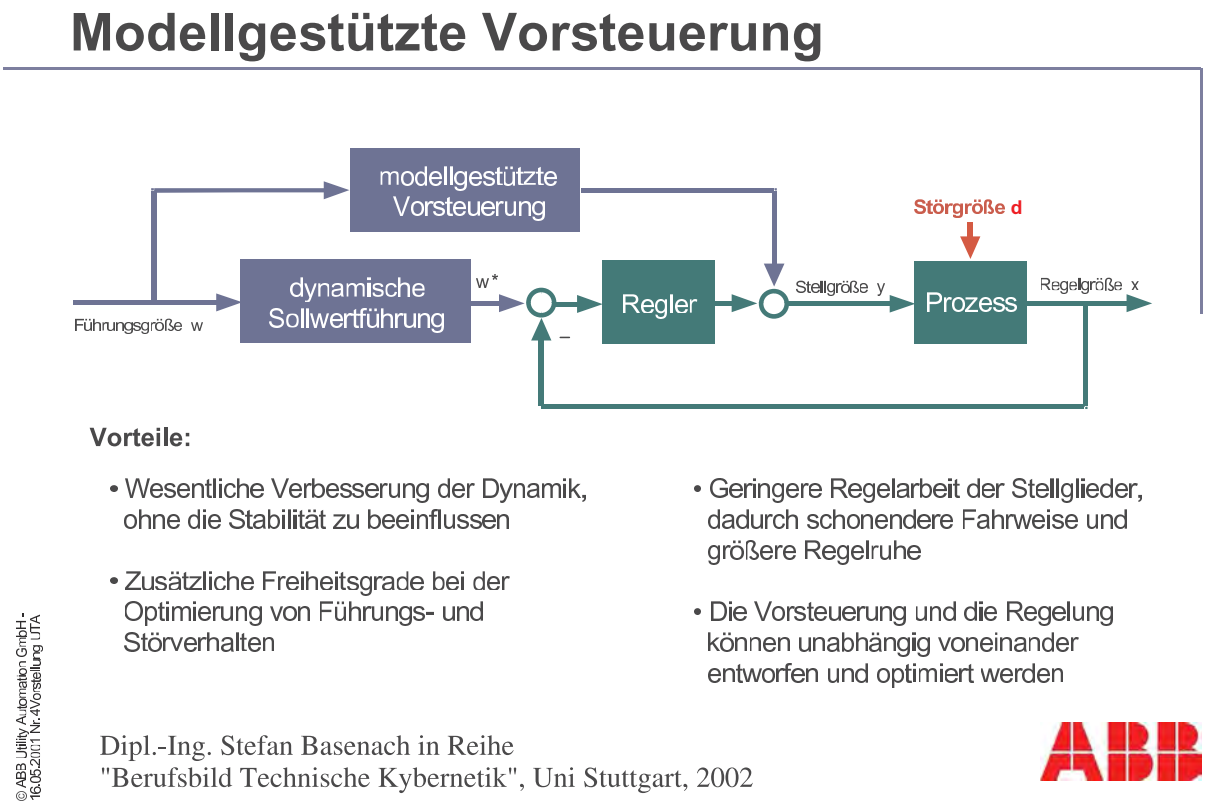


Abbildung 1: Zwei-Freiheitsgrad-Regelkreisstrukur mit einer modellgestützten Vorsteuerung, einer dynamischen Sollwertführung und einem meist klassischen PI-, PD- oder PID-Regler; Quelle: S.Basenach in "Berufsbild Technische Kybernetik", Universität Stuttgart 2002, http://www.kyb-alumni.de/kybalumni/public/events.shtml .

Bei dem Entwurf einer Vorsteuerung muss unterschieden werden, ob die Solltrajektorien – wie bei einem Arbeitspunktwechsel – vorab geplant oder extern z.B. durch einen Joystick vorgegeben werden [4]. Im ersten Fall erfolgt der Vorsteuerungs-Entwurf *offline* und die entworfenen Trajektorien werden z. B. in einer Look-up-Tabelle gespeichert und in Realzeit aufgeschaltet. Der zweite Fall, der häufig in Servosystemen vorkommt, erfordert, dass die Steuertrajektorien *online* bestimmt und aufgeschaltet werden.

Wie anläßlich des Steirischen Seminars 2009 erläutert, eröffnet die *Flachheits*-Methodik einen einfachen Zugang zum Offline-Entwurf von Vorsteuerungen [6], [7]. Im Mittelpunkt stehen dabei die Bestimmung des inversen Streckenmodells und die geeignete Planung der Solltrajektorien z.B. zur Verbindung von zwei Arbeitspunkten. In dem vorliegenden Beitrag wird für lineare zeitinvariante SISO-Systeme erläutert, wie der flachheitsbasierte Vorsteuerungs-Entwurf im Frequenzbereich durchgeführt wird, siehe auch [8].

Der Online-Entwurf einer Vorsteuerung wird maßgeblich durch die Voraussetzungen bezüglich der Differenzierbarkeit der Solltrajektorie und der Minimalphasigkeit des Systems beeinflusst. Die notwendige Differenzierbarkeit kann durch ein Sollwert-Filter erreicht werden [1]. Falls das System minimalphasig ist, wird die Steuertrajektorie mit Hilfe des inversen Systemmodells realisiert. Für nicht-minmalphasige SISO-Systeme kann ein Vorschlag von *G.Roppenecker* benutzt werden, um die Steuer- und Referenztrajektorien

mit einem Modellregelkreis aus den vorgegebenen Sollwerten simulativ zu bestimmen und in Realzeit aufzuschalten [9].

Bei modellbasierten Entwurfsverfahren spielen die Modellgenauigkeit bzw. die Modellfehler eine wichtige Rolle. Für die Anwendung einer modellbasierten Vorsteuerung stellt sich die Frage, bis zu welcher Modellfehlergrenze das Folgeverhalten einer Regelung durch eine Vorsteuerung verbessert wird. Im Frequenzbereich sind relativ einfache Aussagen über die Robustheit einer Regelung mit und ohne Vorsteuerung gegenüber frequenzabhängigen Modellfehlern möglich [11].

In den nachfolgenden Abschnitten wird die Wirkung einer Vorsteuerung in einem Zwei-Freiheitsgrad-Regelkreis erläutert. Außerdem werden die Bedingungen für den modellbasierten Vorsteuerungs-Entwurf untersucht und gezeigt, in welcher Weise diese beim *Offline*- und beim *Online*-Entwurf berücksichtigt werden. Abschließend werden die Robustheitseigenschaften einer Vorsteuerung gegenüber Modellfehlern betrachtet.

2 Regelkreis mit zwei Freiheitsgraden

Die Bedingungen für den modellbasierten Vorsteuerungs-Entwurf sollen am Beispiel eines linearen zeitinvarianten SISO-Systems mit Eingang $u(t)$ und Ausgang $y(t)$ erläutert werden. Dabei wird angenommen, dass das Ein–/Ausgangsverhalten durch die teilerfremde Übertragungsfunktion

$$G(s) = \frac{Y(s)}{U(s)} = \frac{P(s)}{Q(s)} = \frac{b_0 + b_1 s + \ldots + b_m s^m}{a_0 + a_1 s + \ldots + a_n s^n} \quad (b_m \neq 0,\ n \geq m \geq 0) \tag{1}$$

im Bildbereich der Laplace-Transformation[1] beschrieben wird; dabei sind die Polynome $P(s)$ und $Q(s)$ meist so normiert, dass $a_n = 1$ gilt. Für den Vorsteuerungs-Entwurf spielt der Differenzgrad $r = n - m \geq 0$ eine wichtige Rolle. Die Zahl r wird auch als *relativer Grad* des Systems bezeichnet und definiert die Anzahl $n - m$ der Nullstellen im Unendlichen. Ein anderes Kriterium betrifft die Lage der Nullstellen im Endlichen. Falls alle Nullstellen einen negativen Realteil besitzen, ist das System *minimalphasig* und der Entwurf einer Vorsteuerung ist – wie nachfolgend erläutert – relativ einfach. Dagegen erfordert der Vorsteuerungs-Entwurf für *nicht-minimalphasige* Systeme mit mindestens einer instabilen Nullstelle spezielle Maßnahmen.

Die Wirkung einer Vorsteuerung $G_V(s)$ in der Zwei-Freiheitsgrad-Regelkreisstruktur nach Abb. 2 ergibt sich aus der Gleichung für die Regelgröße

$$Y(s) = \underbrace{\frac{G(s)\,(G_V(s) + G_R(s))}{1 + G(s)\,G_R(s)}}_{G_{Y^*}(s)}\, Y^*(s) + \underbrace{\frac{1}{1 + G(s)\,G_R(s)}}_{G_D(s)}\, D(s) \tag{2}$$

in Abhängigkeit von dem Sollwert $Y^*(s)$ und der Störung $D(s)$. Die Übertragungsfunktionen $G_{Y^*}(s)$ und $G_D(s)$ beschreiben das Führungs- bzw. das Störverhalten des Regelkreises. Aus Gleichung (2) folgt, dass das Regelkreisverhalten sowohl durch den Regler

[1] Die Bezeichnungen im Bildbereich orientieren sich an [5] mit der komplexen Bildvariablen $s = \sigma + j\omega$ und den Großbuchstaben für die transformierten Größen.

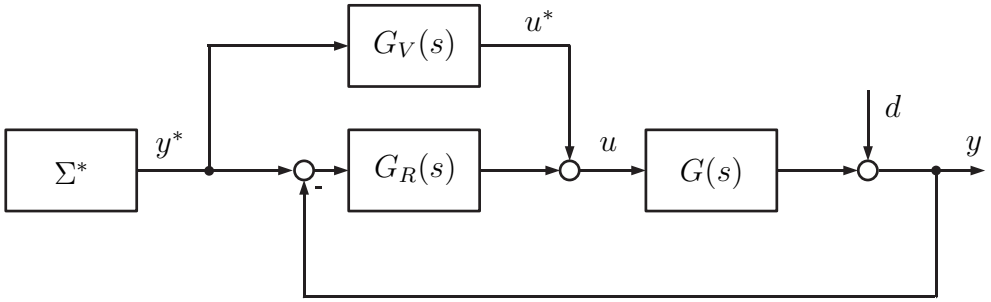


Abbildung 2: Linearer Zwei-Freiheitsgrad-Regelkreis mit Regelstrecke $G(s)$, Regler $G_R(s)$, Vorsteuerung $G_V(s)$ und Sollwertgeber Σ^* für eine Folgeregelung $y(t) \to y^*(t)$.

$G_R(s)$ als auch durch die Vorsteuerung $G_V(s)$ beeinflusst werden kann. Dabei orientiert sich der Regler-Entwurf an dem Stabilitäts- und Störverhalten, während die Vorsteuerung sich nur auf das Führungsverhalten auswirkt. Interessanterweise gibt es im Unterschied zum Regler-Entwurf für den Vorsteuerungs-Entwurf nur wenige systematische Verfahren und erst neuerdings entsprechende Erläuterungen in Lehrbüchern [5].

Das Entwurfsziel für eine Vorsteuerung $G_V(s)$ lässt sich am besten anhand der Führungs-Übertragungsfunktion $G_{Y^*}(s)$ in (2) erklären, zumal die Vorsteuerung praktisch immer in Verbindung mit einer Regelung eingesetzt wird. Die Gleichung (2) gilt natürlich auch für den klassischen Regelkreis ohne Vorsteuerung ($G_V(s) = 0$) mit dem bekannten Zusammenhang

$$G_{Y^*}(s) + G_D(s) = 1$$

zwischen den Führungs- und Stör-Übertragungsfunktionen. In diesem Fall bildet der Regler-Entwurf den einzigen Freiheitsgrad und basiert auf einem Kompromiss zwischen dem Führungs- und dem Störverhalten des Regelkreises.

Im Idealfall wird durch eine modellbasierte Vorsteuerung $G_V(s)$ erreicht, dass der Ausgang $y(t)$ einem vorgegebenen (zulässigen) Sollverlauf $y^*(t)$ identisch folgt: $y(t) = y^*(t)$. Dies setzt voraus, dass das Modell $G(s)$ der Strecke exakt bekannt ist und dass keine Störung vorkommt: $d(t) = 0$. Für diese Annahmen folgen aus (1) und (2) die Gleichungen

$$G_V(s) = G^{-1}(s) = \frac{Q(s)}{P(s)} \quad \Rightarrow \quad P(s)\,U^*(s) = Q(s)\,Y^*(s) \tag{3}$$

für eine exakte (nominelle) Vorsteuerung. Diese Gleichungen beschreiben das *inverse System* und bilden die Grundlage für die Planung einer geeigneten Solltrajektorie $y^*(t)$ und die Berechnung der Steuertrajektorie $u^*(t)$. Im Zeitbereich gehört zu (3) die lineare zeitinvariante Differenzialgleichung (Dgl.) m-ter Ordnung

$$b_m \frac{d^m u^*}{dt^m} + \ \ldots\ + b_1 \frac{du^*}{dt} + b_0\, u^* = a_0\, y^* + a_1 \frac{dy^*}{dt} + \ \ldots\ + a_n \frac{d^n y^*}{dt^n}\,, \quad m \le n \tag{4}$$

für die Bestimmung von $u^*(t)$ in Abhängigkeit von der Solltrajektorie $y^*(t)$. Die Dgl. (4) ist geeignet, um die Bedingungen für die Existenz und die Realisierung einer Vorsteuerung zu erläutern.

Die Existenz-Bedingung betrifft die Abhängigkeit der rechten Seite der Dgl. (4) von den Zeitableitungen der Solltrajektorie $y^*(t)$. Die genaue Abhängigkeit der Lösung $u^*(t)$ von den y^*-Ableitungen erkennt man, wenn die Dgl. (4) m-mal über die Zeit integriert wird:

$$\begin{aligned} u^*(t) &= \frac{1}{b_m}\,[\,a_n\frac{d^r y^*}{dt^r} + \ldots + a_{m+1}\frac{dy^*}{dt} + a_m y^* + \int^t (a_{m-1}y^* - b_{m-1}u^* + \ldots \qquad (5) \\ &\quad + \int^t (a_1 y^* - b_1 u^* + \int^t (a_0 y^* - b_0 u^*)\,dt)\,dt)\,\ldots)\,dt\,], \quad r = n - m\,. \end{aligned}$$

Damit die Steuertrajektorie $u^*(t)$ stetig ist, muss die Trajektorie $y^*(t)$ mindestens $(n-m)$-mal differenzierbar sein, was dem relativen Grad r des Systems (1) entspricht:

$$u^*(t) \in \mathcal{C}^0 \quad \Longleftrightarrow \quad y^*(t) \in \mathcal{C}^r, \quad r = n - m\,. \qquad (6)$$

Dies bedingt für den Offline-Entwurf der Vorsteuerung, dass die Solltrajektorie $y^*(t)$ entsprechend geplant werden muss. Falls die Solltrajektorie $y^*(t)$ extern vorgegeben wird, bedeutet die Differenzierbarkeits-Bedingung (6) eine wesentliche Einschränkung für den Vorsteuerungs-Entwurf und das erreichbare Folgeverhalten.

Die Realisierung der Vorsteuerung erfordert die Bestimmung der Steuertrajektorie $u^*(t)$ als Lösung der Dgl. (4) und deren Aufschaltung in Realzeit am Streckeneingang. Die erste Realisierbarkeits-Bedingung hängt mit der Stabilität der Dgl. (4) zusammen, wenn diese durch eine numerische Integration gelöst wird. Aus dem Vergleich mit der Übertragungsfunktion $G(s)$ in (1) entnimmt man, dass das Zählerpolynom $P(s)$ das charakteristische Polynom der Dgl. (4) darstellt und dass die Stabilität der Dgl. durch die Lage der Nullstellen von $G(s)$ bzw. $P(s)$ bestimmt wird. Dies bedeutet, dass Nullstellen in der rechten Halbebene zu exponentiell wachsenden numerischen Fehlern führen, wenn die Steuertrajektorie $u^*(t)$ mittels einer numerischen Intergration der Dgl. (4) berechnet wird. Daher kann die Vorsteuerung für nicht-minimalphasige Systeme nicht unmittelbar mit den Gleichungen (3) bzw. (4) des inversen Systems entworfen werden.

Die zweite Realisierbarkeits-Bedingung betrifft die Kenntnis der Zeitableitungen $d^i y^*/dt^i,\ i = 1, \ldots, r$ in (5). Bei der Offline-Planung der Solltrajektorie $y^*(t)$ stehen auch deren Ableitungen zur Verfügung und können bei der Berechnung der Steuertrajektorie $u^*(t)$ verwendet werden. Bei einer externen Vorgabe von $y^*(t)$ sind deren Zeitableitungen meist nicht bekannt bzw. existieren auch nicht; daher können in diesem Fall die Gleichungen (4) bzw. (5) nur in dem Sonderfall $r = 0$ bzw. $m = n$ für die Bestimmung der Steuertrajektorie $u^*(t)$ herangezogen werden.

Schließlich muss die entworfene Vorsteuerung in Realzeit realisiert werden; dies erfordert, dass die numerische Integration der Dgl. (4) zumindest genauso schnell wie die Systemdynamik sein muss. Aufgrund der angenommenen Linearität und Zeitinvarianz der Dgl. kann diese offline in eine exakte zeitdiskrete Übertragungsfunktion überführt und dann online effizient gelöst werden.

In den folgenden Abschnitten wird erläutert, in welcher Weise die Bedingungen bezüglich der Differenzierbarkeit der Solltrajektorie und der Minimalphasigkeit des Systems beim *Offline*- und beim *Online*-Entwurf einer Vorsteuerung berücksichtigt werden.

3 Offline-Entwurf der Vorsteuerung

Ausgangspunkt für den modellbasierten Vorsteuerungs-Entwurf linearer SISO-Systeme sind die Gleichungen (3)-(5) des inversen Systems. Falls die Aufgabenstellung für die Vorsteuerung wie z. B. bei einem Arbeitspunktwechsel vorab bekannt ist, kann bei dem *Offline*-Entwurf die Differenzierbarkeits-Bedingung (6) durch die geeignete Planung einer r-mal differenzierbaren Solltrajektorie $y^*(t)$ berücksichtigt werden; dies schließt ein, dass die Ableitungen $d^i y^*/dt^i,\ i = 1, \ldots, r$ für die Bestimmung der Steuertrajektorie $u^*(t)$ zur Verfügung stehen.

Für den maximalen relativen Grad $r = n$ bzw. $m = 0$ vereinfacht sich die Dgl. (4) zu einer algebraischen Gleichung für die Steuertrajektorie

$$u^*(t) = \frac{1}{b_0}\,[\,a_0\,y^*(t) + a_1\frac{dy^*(t)}{dt} + \,\ldots\, + a_n\frac{d^n y^*(t)}{dt^n}\,] \tag{7}$$

in Abhängigkeit von der Solltrajektorie $y^*(t) \in \mathcal{C}^n$ und deren n ersten Ableitungen. Im Bildbereich ist die Vorsteuerung $G_V(s) = Q(s)/b_0$ als Quotient des Nennerpolynoms $Q(s)$ und der Verstärkung b_0 der Übertragungsfunktion (1) definiert.

Der maximale relative Grad $r = n$ ist typisch für einfache mechanische Modelle und führt – wie in Abb. 3 beispielhaft für die Bewegung $M\ddot{y} = u$ einer Masse M dargestellt – zu einer doppelt differenzierenden Vorsteuerung $G_V(s) = Ms^2$. Die Stetigkeit der Vorsteuerung $u^*(t) \in \mathcal{C}^0$ setzt – wie ebenfalls in Abb. 3 dargestellt – die Planung einer zweimal stetig differenzierbaren Solltrajektorie $y^*(t) \in \mathcal{C}^2$ voraus.

Aus dem Beispiel entnimmt man, dass die Vorsteuerung für ein SISO-System mit dem relativen Grad $r = n$ mehr oder weniger intuitiv entworfen werden kann. Demgegenüber erfordert die Bestimmung der Steuertrajektorie $u^*(t)$ für $r < n$ die Lösung der Dgl. (4), wobei deren Stabilität vorausgesetzt werden muss. Um die Integration der Dgl. (4) – insbesondere wenn diese instabil ist – zu umgehen, kann man eine neue Größe

$$Z(s) = \frac{U(s)}{Q(s)} = \frac{Y(s)}{P(s)} \tag{8}$$

einführen, siehe auch [8]. Mittels $Z(s)$ bekommt man im Bildbereich die differenzial-algebraischen Parametrierungen

$$U(s) \;=\; (a_0 + a_1\,s + \,\ldots\, + a_n\,s^n)\,Z(s)\,, \tag{9}$$

$$Y(s) \;=\; (b_0 + b_1\,s + \,\ldots\, + b_m\,s^m)\,Z(s) \tag{10}$$

für den Eingang und Ausgang[2]. Diese Gleichungen sind geeignet, die Steuertraktorie $u^*(t)$ und die Ausgangstrajektorie $y^*(t)$ in Abhängigkeit von einer n-mal differenzierbaren Trajektorie $z^*(t) \in \mathcal{C}^n$ und deren Ableitungen zu berechnen. Dies verlangt, dass die Steuerungsaufgabe in den z-Koordinaten vorliegt und die zugehörige Solltrajetorie $z^*(t)$ geplant wird.

[2] Aus den Gleichungen (9) und (10) kann die Größe $Z(s)$ mit Hilfe der *Bezout*-Identität in Abhängigkeit von $U(s)$ und $Y(s)$ berechnet werden [8].

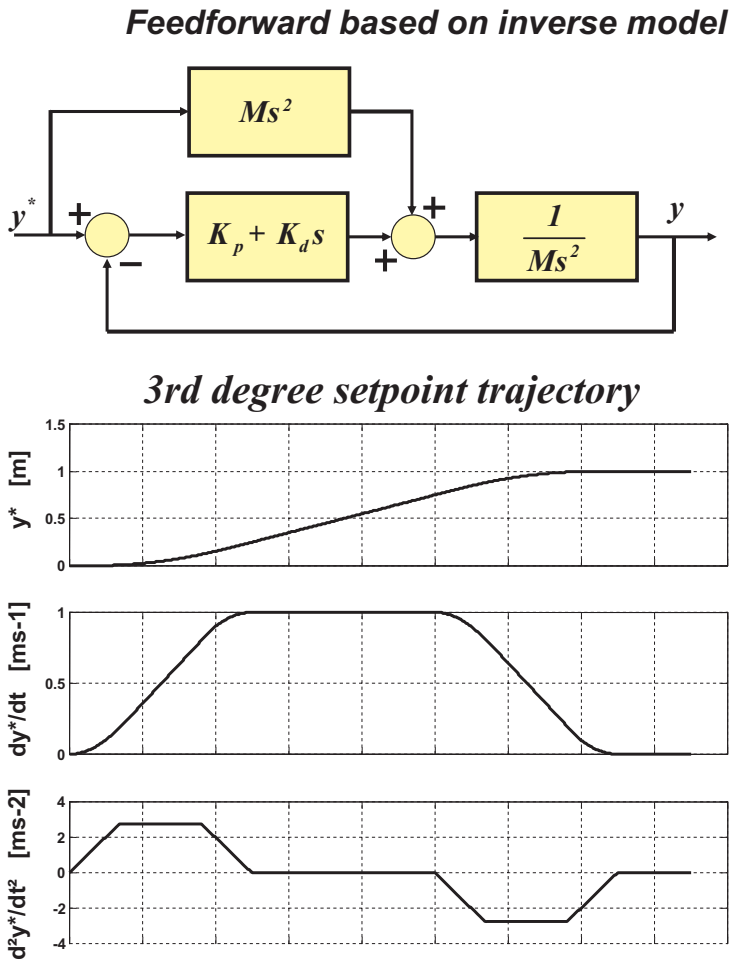


Abbildung 3: Zwei-Freiheitsgrad-Regelkreis für die Folgeregelung einer Masse M mit der Vorsteuerung Ms^2 und einem PD-Regler $K_p + K_d s$ sowie der Planung einer zweimal differenzierbaren Solltrajektorie $y^*(t) \in \mathcal{C}^2$; Quelle: M.Steinbuch, Control Systems Technology, Department of Mechanical Engineering Eindhoven University of Technology 2003.

Die Planung der Solltrajektorie $z^*(t)$ ist für einen Arbeitspunktwechsel relativ einfach. Im ersten Schritt müssen die gegebenen Arbeitspunkte

$$y(0) = y_0^* \quad \longrightarrow \quad y(T) = y_T^*\,, \quad \left.\frac{d^i y}{dt^i}\right|_{t=0,T} = 0\,, \;\; i > 0 \tag{11}$$

mittels (8) in die z^*-Koordinaten umgerechnet werden:

$$z^*(0) = z_0^* = \frac{y_0^*}{b_0} \quad \longrightarrow \quad z^*(T) = z_T^* = \frac{y_T^*}{b_0}\,, \quad \left.\frac{d^i z^*}{dt^i}\right|_{t=0,T} = 0\,, \;\; i > 0\,. \tag{12}$$

Im zweiten Schritt werden die beiden Randpunkte – wie in Abb. 4 dargestellt – beispielsweise durch ein n-mal differenzierbares Polynom vom Grad $2n+1$ verbunden:

$$z^*(t) = z_0^* + (z_T^* - z_0^*) \sum_{i=n+1}^{2n+1} p_i \left(\frac{t}{T}\right)^i \in \mathcal{C}^n, \;\; t \in [0\,, T]\,. \tag{13}$$

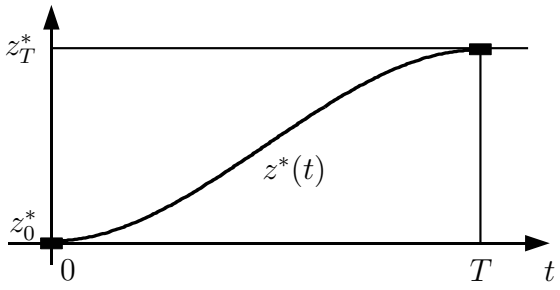


Abbildung 4: Solltrajektorie $z^*(t) \in \mathcal{C}^n$ für einen Arbeitspunktwechsel $z^*(0) = z_0^* \rightarrow z^*(T) = z_T^*$ in dem Zeitintervall $t \in [0\,, T]$.

Die Koeffizienten $p_i,\ i = n+1, \ldots, 2n+1$ sind im Anhang für die Systemordnungen $n = 1$ bis $n = 7$ tabelliert. Wenn man die Solltrajektorie $z^*(t)$ in die Eingangs-Parametrierung (9) einsetzt, ergibt sich eine stetige Steuertrajektorie $u^*(t) \in \mathcal{C}^0$. Anhand des Verlaufs von $u^*(t)$ kann man überprüfen, ob eventuelle Stellgrößen-Beschränkungen eingehalten werden. Gegebenenfalls muss die Transitionszeit T verlängert werden.

Die Referenztrajektorie $y^*(t)$ für die Regelung des Ausgangs bekommt man durch Einsetzen der Solltrajektorie $z^*(t)$ in (10). Aufgrund der Eingangs- und Ausgangs-Parametrierungen mit derselben Basisgröße $z^*(t)$ ist gewährleistet, dass die beiden Trajektorien $u^*(t)$ und $y^*(t)$ konsistent sind. Auf diese Weise sind alle Komponenten zur Realisierung der Zwei-Freiheitsgrad-Folgeregelung nach Abb. 1 entworfen. In Abb. 5 ist das konkrete Entwurfsergebnis für die modellbasierte Vorsteuerung und das Sollwert-Filter zusammen mit einem PID-Regler dargestellt. Daraus folgt, dass das Sollwert-Filter und die Vorsteuerung durch die Zähler- und Nennerpolynome $P(s)$ und $Q(s)$ der Übertragungsfunktion $G(s)$ festgelegt sind. Die Abb. 5 enthält mit $P(s) = 1$ und $Q(s) = Ms^2$ auch die in Abb. 3 dargestellte Folgeregelung einer Masse, wenn der PID-Regler durch einen PD-Regler ersetzt wird.

Wie bereits erwähnt, werden die offline entworfenen Trajektorien $u^*(t)$ und $y^*(t)$ z. B. in Look-up-Tabellen gespeichert und in Realzeit aufgeschaltet. In dem Blockschaltbild für den linearen Zwei-Freiheitsgrad-Regelkreis in Abb. 5 bedeutet dies, dass der Sollwertgeber Σ^*, das Sollwert-Filter $P(s)$ und die Vorsteuerung $Q(s)$ das Entwurfsergebnis beschreiben, aber selbst nicht realisiert werden.

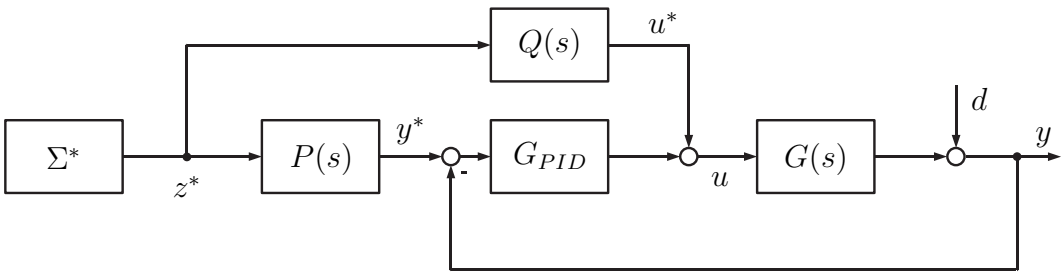


Abbildung 5: Linearer Zwei-Freiheitsgrad-Regelkreis mit Regelstrecke $G(s)$, PID-Regler, Sollwertgeber Σ^*, Vorsteuerung $Q(s)$ und Sollwert-Filter $P(s)$.

Die wesentliche Aufgabe beim Offline-Entwurf der Vorsteuerung betrifft die Planung einer differenzierbaren Solltrajektorie $z^*(t)$ für die in (8) eingeführte Größe $z(t)$. Falls bei einem Arbeitspunktwechsel (11) Beschränkungen der Ein- und Ausgangsgrößen gegeben sind, können diese beim Offline-Entwurf der Trajektorien $y^*(t)$ und $u^*(t)$ mit dem in [4], [12] beschriebenen Verfahren berücksichtigt werden; dabei wird der Arbeitspunktwechsel (11) bzw. (12) als Randwertaufgabe mit den entsprechenden Ein- und Ausgangs-Beschränkungen formuliert und der Steuerungsentwurf auf die numerische Lösung eines nichtlinearen algebraischen Gleichungssystems zurückgeführt.

Abschließend noch eine Anmerkung zu der beim Vorsteuerungs-Entwurf benutzten Größe $Z(s)$ bzw. $z(t)$: Wenn man diese Größe einem fiktiven Ausgang zuordnet, besitzt das zugehörige Ein-/Ausgangsverhalten (9) den relativen Grad $r = n$; dies entspricht dem Sonderfall (7), für den die beiden Ausgänge identisch sind: $z = y$. Ein realer oder fiktiver Ausgang eines SISO-Systems mit dem relativen Grad $r = n$ wird als *flacher Ausgang* im Sinne einer mathematischen Basisgröße bezeichnet [13]-[15]. Für die Existenz eines flachen Ausgangs eines linearen SISO-Systems (1) muss lediglich dessen Steuerbarkeit vorausgesetzt werden, was bei einer teilerfremden Übertragungsfunktion $G(s)$ gegeben ist. Aus Sicht der Flachheits-Methodik stellen die Gleichungen (9) und (10) die differenziellen Ein- und Ausgangs-Parametrierungen im Frequenzbereich dar [2], [8]. Diese Parametrierungen bilden die Grundlage für den flachheitsbasierten Offline-Entwurf einer modellbasierten Vorsteuerung für steuerbare lineare SISO-Systeme (1).

4 Online-Entwurf der Vorsteuerung

Der *Online*-Entwurf der Vorsteuerung kommt in Frage, wenn die Solltrajektorie $y^*(t)$ für den Ausgang vorab nicht bekannt ist. Dabei muss im Unterschied zum *Offline*-Entwurf außerdem davon ausgegangen werden, dass in den meisten Fällen die extern vorgegebene Trajektorie $y^*(t)$ nur stetig ist und dass deren Zeitableitungen nicht existieren: $y^*(t) \in \mathcal{C}^0$. Wie bereits erwähnt, folgt aus der Integraldarstellung (5), dass ohne Kenntnis der Zeitableitungen $d^i y^*/dt^i,\ i = 1, \ldots, r$ die Steuertrajektorie $u^*(t)$ nur für den Sonderfall $r = 0$ bzw. $m = n$ berechnet werden kann. Dabei muss natürlich die Stabilität der Dgl. (4) bzw. die Minimalphasigkeit des Systems (1) vorausgesetzt werden. Nachfolgend wird der Online-Entwurf einer Vorsteuerungs für *minimalphasige* und anschließend für *nicht-minimalphasige* SISO-Systeme behandelt.

4.1 Minimalphasige SISO-Systeme

Der Vorsteuerungs-Entwurf für *minimalphasige* Systeme (1) mit $r > 0$ bzw. $m < n$ verlangt, dass die Solltrajektorie r-mal differenzierbar ist [1]. Dazu wird die Vorsteuerung – wie in Abb. 6 dargestellt – als Reihenschaltung aus einem Sollwert-Filter $G_F(s)$ und der inversen Übertragungsfunktion $G^{-1}(s)$ realisiert:

$$G_V^*(s) = G^{-1}(s)\, G_F(s)\,,\ \ G_F(s) = \frac{1}{Q_F(s)}\,,\ \ Q_F(s) = \alpha_0 + \alpha_1 s + \ \ldots\ + \alpha_r s^r\,. \qquad (14)$$

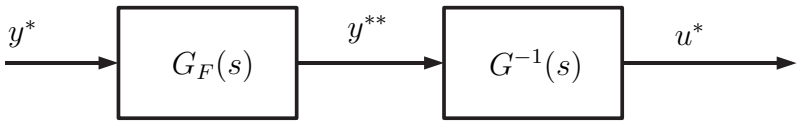


Abbildung 6: Vorsteuerung $G_V^*(s)$ aus Sollwert-Filter $G_F(s)$ und inversem System $G^{-1}(s)$.

Das Sollwert-Filter $G_F(s)$ ist ein Tiefpass mit dem Nennerpolynom $Q_F(s)$; die Tiefpass-Ordnung r wird durch den relativen Grad $r = n - m$ des Systems (1) festgelegt. Durch das Sollwert-Filter $G_F(s)$ wird die r-mal differenzierbare Sollgröße $y^{**} \in \mathcal{C}^r$ realisiert, für welche die inverse Übertragungsfunktion $G^{-1}(s)$ realisierbar ist.

Durch die Einführung des Sollwert-Filters $G_F(s)$ ergibt sich anstelle von (3) das neue inverse System:

$$P^*(s)\,U^*(s) = Q(s)\,Y^*(s) \quad \text{mit} \quad P^*(s) = Q_F(s)\,P(s) = \beta_0 + \beta_1 s + \ldots + \beta_n s^n\,. \tag{15}$$

Im Zeitbereich gehört zu (15) die lineare zeitinvariante Dgl. n-ter Ordnung

$$\beta_n \frac{d^n u^*}{dt^n} + \ldots + \beta_1 \frac{du^*}{dt} + \beta_0\, u^* = a_0\, y^* + a_1 \frac{dy^*}{dt} + \ldots + a_n \frac{d^n y^*}{dt^n} \tag{16}$$

für die Bestimmung der Solltrajektorie $u^*(t)$ in Abhängigkeit von $y^*(t)$. Die Stabilität der Dgl. (16) verlangt, dass die Nullstellen der Polynome $P(s)$ und $Q_F(s)$ einen negativen Realteil besitzen. Dies entspricht der Minimalphasigkeit des Systems (1) und der Stabilität des Tiefpassfilters (14).

Aus den Gleichungen (14)-(16) folgt, dass das Tiefpassfilter $G_F(s)$ bei dem Vorsteuerungs-Entwurf eine doppelte Funktion besitzt: Einerseits definiert $G_F(s)$ die neue Solltrajektorie $y^{**}(t)$ bzw. das dynamische Sollwert-Filter in Abb. 1. Andererseits wird durch $G_F(s)$ die Realisierbarkeit der Vorsteuerung $G_V^*(s)$ bzw. des inversen Systems (15) sicher gestellt. Während die Mindestordnung r des Filters festliegt, kann mittels der Filterparameter $\alpha_i, i = 0, \ldots, r$ das Folgeverhalten der Vorsteuerung bezüglich der Solltrajektorie $y^*(t)$ beeinflusst werden. Die Wahl der Parameter richtet sich nach der Bandbreite von Nutz- und Störanteilen in der Solltrajektorie $y^*(t)$ und nach den Zeitkonstanten des Systems (1).

Als Beispiel für die Anwendung eines Sollwert-Filters in Verbindung mit einer modellbasierten Vorsteuerung wird die in Abb. 7 dargestellte Zwei-Freiheitsgrad-Folgeregelung eines piezo-elektrischen Antriebs in einem Riesenteleskop zur stellaren Interferometrie betrachtet. Der Antrieb, welcher auf einer elastischen Plattform montiert ist, dient zur Vibrationskompensation zwischen zwei optischen Primärspiegeln und soll die Spiegelpositionen einem externen Signal nachführen. Das Antriebs-Modell 26. Ordnung wurde mittels PRBS-Anregung[3] im Frequenzbereich identifiziert und bildet die Grundlage für den Entwurf des Sollwert-Filters und der Vorsteuerung. Als Sollwert-Filter wird ein zeitdiskretes FIR-Filter[4] verwendet, um das externe Referenzsignal geeignet zu glätten. Die Filterdynamik wird durch die in Abb. 7 dargestellte Sprungantwort beschrieben. Das erzielte Folgeverhalten ist ebenfalls für eine Sprungantwort dargestellt, wobei in dem Zwei-Freiheitsgrad-Regelkreis ein PI-Regler verwendet wird.

[3]Pseudo Random Binary Signal
[4]Finite Impulse Response

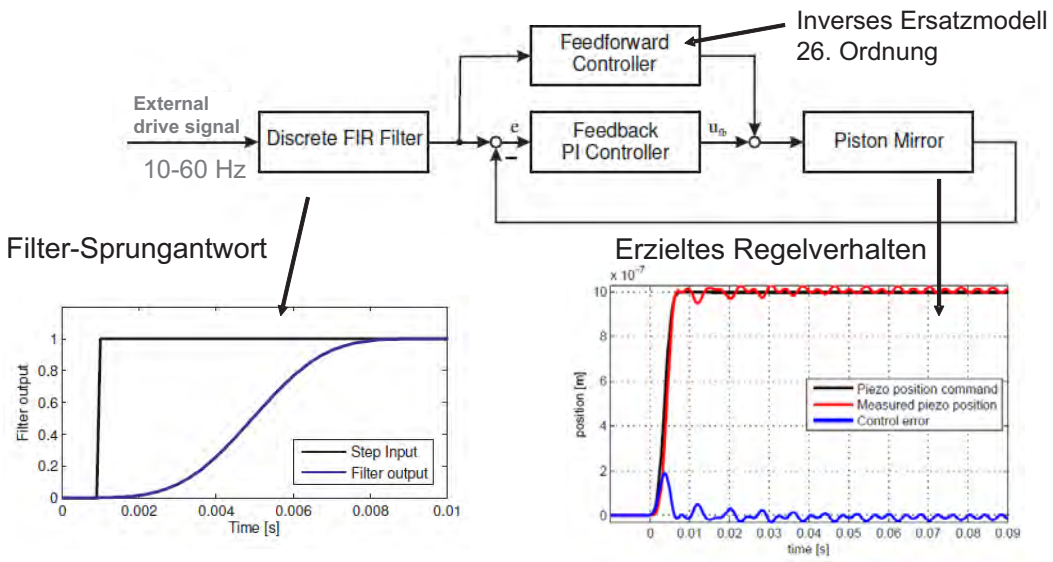


Abbildung 7: Zwei-Freiheitsgrad-Folgeregelkreis mit Sollwert-Filter für einen piezo-elektrischen Antrieb; Quelle: T.Ruppel, Modellierung, Identifikation und Regelung eines piezo-elektrischen Antriebs auf einer elastischen Trägerplattform, Institut für Systemdynamik, Universität Stuttgart 2011.

4.2 Nicht-minimalphasige SISO-Systeme

Die numerischen Ansätze zur Inversion *nicht-minimalphasiger* Systeme [3], [16] gehen von einer *Offline*-Realisierung aus und kommen daher für einen Vorsteuerungs-Entwurf nicht in Frage, wenn die Solltrajektorie in Realzeit vorgegeben wird. Die Systeminversion kann umgangen werden, wenn man – wie in [9] vorgeschlagen – die Steuertrajektorie in einem Modellregelkreis aus der vorgegebenen Solltrajektorie simulativ bestimmt und in Realzeit am Eingang der Regelstrecke aufschaltet. Diese Vorgehensweise, die in [9] im Zeitbereich und im Zustandsraum entwickelt wird, ist in Abb. 8 als Zwei-Freiheitsgrad-Struktur für

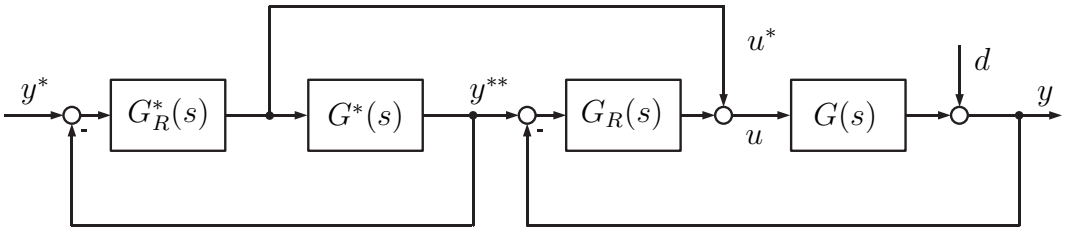


Abbildung 8: Zwei-Freiheitsgrad-Struktur mit einem Modellregelkreis zur Bestimmung der Steuer- und Referenztraktorien $u^*(t)$ bzw. $y^{**}(t)$.

eine Ausgangsregelung im Frequenzbereich dargestellt.

Das Folgeverhalten $y^*(t) \rightarrow y^{**}(t)$ des ungestörten Modellregelkreises in Abb. 8 wird durch den Entwurf des Reglers $G_R^*(s)$ festgelegt, wobei beliebige Entwurfsverfahren verwendet werden können. Die wesentliche Idee des Modellregelkreises ist, die simulierte Steuertrajektorie $u^*(t)$ am realen Streckeneingang aufzuschalten und die simulierte Ausgangstrajektorie $y^{**}(t)$ als Referenztrajektorie für den Regler $G_R(s)$ zu verwenden. Die beiden Größen werden durch die Gleichungen

$$U^*(s) = \frac{G_R^*(s)}{1 + G_R^*(s)\,G^*(s)}\,Y^*(s)\,, \quad Y^{**}(s) = \underbrace{\frac{G_R^*(s)\,G^*(s)}{1 + G_R^*(s)\,G^*(s)}}_{G_{Y^{**}}(s)}\,Y^*(s) \tag{17}$$

in Abhängigkeit von der Sollgröße $Y^*(s)$ definiert. Daraus folgt, dass die beiden Größen

$$U^*(s) = G^{*-1}(s)\,Y^{**}(s) \tag{18}$$

über die inverse Übertragungsfunktion $G^{*-1}(s)$ verknüpft sind, ohne dass $G^*(s)$ unmittelbar invertiert werden muss. Das heißt, der Zusammenhang (18) gilt auch für *nichtminimalphasige* Systeme.

Für die Zwei-Freiheitsgrad-Regelung nach Abb. 8 bekommt man für die Ausgangsgröße $Y(s)$ die Gleichung

$$Y(s) = \underbrace{G_{Y^{**}}(s)\,Y^*(s)}_{Y^{**}(s)} + \underbrace{\frac{1}{1 + G_R(s)\,G(s)}}_{G_D(s)}\,D(s) \tag{19}$$

in Abhängigkeit von der Sollgröße $Y^*(s)$ und der Störgröße $D(s)$, wenn die Übertragungsfunktionen $G^*(s)$ und $G(s)$ in den beiden Regelkreisen übereinstimmen: $G^*(s) = G(s)$. In dem Sonderfall $G_R^*(s) = G_R(s)$ beschreibt die Gleichung (19) das Übertragungsverhalten eines Regelkreises ohne Vorsteuerung. In (19) gibt es mit den beiden Reglern $G_R^*(s)$ und $G_R(s)$ zwei Freiheitsgrade zum unabhängigen Entwurf der Führungs- und der Stör-Übertragungsfunktionen $G_{Y^{**}}(s)$ bzw. $G_D(s)$.

Das Vorsteuerungs-Konzept [9] kann auch zur Kompensation einer gemessenen Störung verwendet werden [10]. Hierzu wird ein Modellregelkreis mit der gemessenen Störung realisiert, um – wie in Abb. 8 – die simulativ bestimmten Steuer- und Referenztrajektorien in dem Regelkreis aufzuschalten. Der eigentliche Regler dient dann zur Stabilisierung und zur Robustifizierung. Im Fall von linearen Systemen können Vorsteuerung, Störgrößenaufschaltung und Regelung unabhängig voneinander entworfen werden.

5 Robustheit einer Regelung mit Vorsteuerung

Beim Einsatz einer modellbasierten Vorsteuerung stellt sich die Frage nach dem Einfluss von Modellfehlern. Diese Frage ist eng verknüpft mit der Robustheit einer Regelung mit und ohne Vorsteuerung gegenüber Modellfehlern. Ausgangspunkt der Robustheitsanalyse

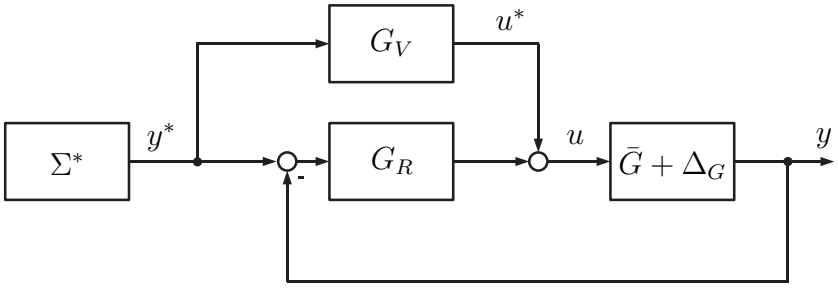


Abbildung 9: Zwei-Freiheitsgrad-Regelkreis mit nominellem Streckenmodell $\bar{G}(j\omega)$, Modellfehler $\triangle_G(j\omega)$, Regler $G_R(j\omega)$, nomineller Vorsteuerung $G_V(j\omega) = \bar{G}^{-1}(j\omega)$ und Sollwertgeber Σ^*.

ist der in Abb. 9 dargestellte Zwei-Freiheitsgrad-Regelkreis mit einem frequenzabhängigen Modellfehler $\triangle_G(j\omega)$. Dabei wird angenommen, dass der Regler $G_R(j\omega)$ und die Vorsteuerung $G_V(j\omega) = \bar{G}^{-1}(j\omega)$ mit dem nominellen Streckenmodell $\bar{G}(j\omega)$ entworfen werden[5]. In Anlehnung an [11] wird nachfolgend der Einfluss des frequenzabhängigen Modellfehlers $\triangle_G(j\omega)$ auf das Führungsverhalten der Regelung ohne und mit einer Vorsteuerung untersucht. Dabei wird der Einfluss von Störungen nicht berücksichtigt.

Die Gleichung für das frequenzabhängige Führungsverhalten der Regelung ohne Vorsteuerung ($G_V(j\omega) = 0$) lautet

$$Y_R(j\omega) = \frac{G(j\omega)\, G_R(j\omega)}{1 + G(j\omega)\, G_R(j\omega)}\, Y^*(j\omega)\,, \quad G(j\omega) = \bar{G}(j\omega) + \triangle_G(j\omega) \tag{20}$$

in Abhängigkeit von dem Modellfehler $\triangle_G(j\omega)$. Dabei wird vorausgesetzt, dass die Stabilität der Regelung durch den Modellfehler nicht verändert wird. Zu (20) gehört der auf den Sollwert $Y^*(j\omega)$ bezogenene Regelfehler

$$\delta_R(j\omega) = \frac{Y_R(j\omega) - Y^*(j\omega)}{Y^*(j\omega)} = \frac{-1}{1 + G(j\omega)\, G_R(j\omega)}\,. \tag{21}$$

In dieser Gleichung erscheint der Modellfehler $\triangle_G(j\omega)$ im Nenner.

Dieselbe Fehleranalyse kann man für die Regelung mit einer nominellen Vorsteuerung $G_V(j\omega) = \bar{G}^{-1}(j\omega)$ durchführen. Dann lautet die Gleichung für das Führungsverhalten

$$Y_{R+V}(j\omega) = \frac{G(j\omega)\, (\bar{G}^{-1}(j\omega) + G_R(j\omega))}{1 + G(j\omega)\, G_R(j\omega)}\, Y^*(j\omega)\,. \tag{22}$$

Im nominellen Fall $G(j\omega) = \bar{G}(j\omega)$ bzw. $\triangle_G(j\omega) = 0$ ergibt sich das exakte Führungsverhalten $Y_{R+V}(j\omega) = Y^*(j\omega)$. Im realen Fall $\triangle_G(j\omega) \neq 0$ bekommt man entsprechend (21) den relativen Regelfehler

$$\delta_{R+V}(j\omega) = \frac{Y_{R+V}(j\omega) - Y^*(j\omega)}{Y^*(j\omega)} = \frac{\triangle_G(j\omega)/\bar{G}(j\omega)}{1 + G(j\omega)\, G_R(j\omega)} \tag{23}$$

[5]Wegen der Annahme eines frequenzabhängigen Modellfehlers $\triangle_G(j\omega)$ werden in diesem Kapitel anstelle der Übertragungsfunktionen die Frequenzgänge der einzelnen Komponenten betrachtet.

für die Regelung mit Vorsteuerung. Im Unterschied zu (21) wirkt sich der Modellfehler $\triangle_G(j\omega)$ in (23) auch auf den Zähler aus. Daraus folgt der frequenzabhängige *Break Even Point*

$$| \delta_{R+V}(j\omega) | \leq | \delta_R(j\omega) | \quad \text{für} \quad | \triangle_G(j\omega)/\bar{G}(j\omega) | \leq 1 \tag{24}$$

für die Verwendung einer Vorsteuerung. Mit anderen Worten ausgedrückt kann man sagen: Bis zu einem 100-prozentigen Modellfehler besitzt die Regelung mit Vorsteuerung im Frequenzbereich ein günstigeres Führungsverhalten als eine Regelung ohne Vorsteuerung. Das Ergebnis (24) ist ein Beleg für den vorteilhaften Einsatz einer modellbasierten Vorsteuerung, wenn das Streckenmodell hinreichend genau ist.

6 Zusammenfassung

Der Entwurf von modellbasierten Vorsteuerungen ist wegen der Planung oder Vorgabe von zeitabhängigen Solltrajektien eigentlich im Zeitbereich beheimatet. Wie in dem vorliegenden Beitrag gezeigt, können bei linearen zeitinvarianten Systemen der *Online*- und der *Offline*-Entwurf einer Vorsteuerung auch im Frequenzbereich durchgeführt werden. Der Frequenzbereichs-Entwurf ist naheliegend, wenn das Systemmodell als Übertragungsfunktion vorliegt. Darüber hinaus bietet der Frequenzbereich Vorteile beim Entwurf einer Vorsteuerung mit Sollwert-Filter, wenn für die Solltrajektorien frequenzabhängige Bandbreiten der Nutz- und Störanteile bekannt sind. Schließlich sind im Frequenzbereich einfache Aussagen über die Robustheit einer Regelung mit und ohne Vorsteuerung gegenüber frequenzabhängigen Modellfehlern möglich. Im Hinblick auf die Behandlung der Vorsteuerung in der Lehre lässt sich feststellen, dass der modellbasierte Entwurf genauso wie der Regler-Entwurf für lineare zeitinvariante Systeme im Frequenzbereich vermittelt werden kann und wegen der praktischen Relevanz auch sollte.

Danksagung

Das Beitragsthema resultiert aus einer Diskussion mit Prof. Dr.-Ing. Wolfgang Krämer an der Fachochschule Ingolstadt über die Vermittlung des Vorsteuerungs-Entwurfs im Zeit- und im Frequenzbereich. Im Zusammenhang mit der Darstellung des Online-Entwurfs einer Vorsteuerung in Kapitel 4.2 war eine Mail-Korrespondenz mit Prof. Dr.-Ing. Günter Roppenecker an der Universität Erlangen-Nürnberg über die beiden Publikationen [6] und [9] sehr hilfreich. Schließlich sind in den Beitrag verschiedene praktische Erfahrungen beim Einsatz von Vorsteuerungen am Institut für Systemdynamik (Direktor: Prof. Dr.-Ing. Oliver Sawodny) eingeflossen.

Anhang

Die folgende Tabelle enthält die Koeffizienten $p_i,\ i = n+1, \ldots, 2n+1$ der polynomialen Solltrajektorie $z^*(t)$ in (13) für die Systemordnungen $n = 1$ bis $n = 7$. Wegen $z^*(T) = z_T$ gilt die leicht nachprüfbare Bedingung $\sum_{i=n+1}^{2n+1} p_i = 1$.

n	p_{n+1}	p_{n+2}	p_{n+3}	p_{n+4}	p_{n+5}	p_{n+6}	p_{n+7}	p_{n+8}
1	3	-2						
2	10	-15	6					
3	35	-84	70	-20				
4	126	-420	540	-315	70			
5	462	-1980	3465	-3080	1386	-252		
6	1716	-9009	20020	-24024	16380	-6006	924	
7	6435	-40040	108108	-163800	150150	-83160	25740	-3432

Tabelle 1: Koeffizienten für die polynomiale Solltrajektorie $z^*(t)$ in(13)

Literatur

[1] G.Kreisselmeier: Struktur mit zwei Freiheitsgraden. Automatisierungstechnik 47 (1999), 266–269.

[2] V.Hagenmeyer, M.Zeitz: Flachheitsbasierter Entwurf von linearen und nichtlinearen Vorsteuerungen. Automatisierungstechnik 52(2004), 3–12.

[3] K.Graichen, V.Hagenmeyer, M.Zeitz: A new approach to inversion-based feedforward control design for nonlinear systems. Automatica 41(2005), 2033–2041.

[4] K.Graichen: Feedforward Control Design for Finite-Time Transition Problems of Nonlinear Systems with Input and Output Constraints. Shaker Verlag, Aachen 2006.

[5] J.Lunze: Regelungstechnik 1. Springer, Heidelberg-Dordrecht-London-New York 2010.

[6] M.Zeitz: Flachheit – Eine nützliche Methodik auch für lineare SISO-Systeme. In: M.Horn, M.Hofbaur, N.Dourdoumas (Hrsg.): 16. Steirisches Seminar über Regelungstechnik und Prozessautomatisierung 2009, Schloss Retzhof – Leibnitz/Österreich, Institut für Regelungs- und Automatisierungstechnik, Technische Universität Graz, 3–15.

[7] M.Zeitz: Differenzielle Flachheit: Eine nützliche Methodik auch für lineare SISO-Systeme. Automatisierungstechnik 58(2010), 5–13.

[8] M.Fliess, R.Marquez: Continuous–time linear predictive control and flatness: A modul–theoretic setting with examples. International Journal of Control 73(2000), 606–623.

[9] G.Roppenecker: Zustandsregelung linearer Systeme – Eine Neubetrachtung. Automatisierungstechnik 57(2009), 491–498.

[10] C.Wurmthaler, A.Kühnlein: Modellgestützte Vorsteuerung für messbare Störungen. Automatisierungstechnik 57(2009), 328–331.

[11] S.Devasia: Should Model-Based Inverse Inputs Be Used as Feedforward under Plant Uncertainty? IEEE Transactions on Automatic Control 47(2002), 1865–1871.

[12] K.Graichen, M.Zeitz: Inversionsbasierter Vorsteuerungsentwurf mit Ein- und Ausgangsbeschränkungen. Automatisierungstechnik 54(2006), 187-199.

[13] M.Fliess, J.Lévine, P.Martin, P.Rouchon: Flatness and defect of non–linear systems: Introductory theory and examples. International Journal of Control 61(1995), 1327–1361.

[14] R.Rothfuß: Anwendung der flachheitsbasierten Analyse und Regelung nichtlinearer Mehrgrößensysteme. VDI–Fortschritt–Berichte Nr.8/664, VDI-Verlag, Düsseldorf, 1997.

[15] R.Rothfuß, J.Rudolph, M.Zeitz: Flachheit: Ein neuer Zugang zur Steuerung und Regelung nichtlinearer Systeme. Automatisierungstechnik 45(1997), 517–525.

[16] D.Chen, B.Paden: Stable inversion of nonlinear non-minimum phase systems. International Journal of Control 64(1996), 81-97.

Verfahren zur Ordnungsreduktion für Modellprädiktive Regelung

Johannes Unger, Martin Kozek, Stefan Jakubek
Vienna University of Technology,
Institute for Mechanics and Mechatronics,
A-1040 Vienna, Austria
unger@impa.tuwien.ac.at,
kozek@impa.tuwien.ac.at,
jakubek@impa.tuwien.ac.at*

Kurzfassung

Diese Arbeit stellt ein Verfahren zur Reduktion der Ordnung von komplexen modellprädiktiven Regelern vor, welches auf einer Parametrierung der Eingangstrajektorien basiert. Diese Parametrierung wird aus der Analyse von Messungen bestimmt, welche durch Messungen im tatschlichen Regelkreis oder aus Simulationen gewonnen werden. Es werden dabei repräsentative Snapshots aller Stellgrößen im Zeitbereich durch die Karhunen-Loeve Transformation auf eine orthonormale Basis projiziert, welche in Folge durch basierend auf einem analytischen Kriterium reduziert werden. In der Arbeit wird dargestellt, wie systematisch bei der Erzeugung passender Snapshots vorgegangen werden kann. Zusätzlich erfolgt eine Analyse der Stabilität des Regelkreises mit reduziertem MPC. Es wird anhand eines industriellen Trocknungsprozesses gezeigt, dass eine signifikante Ordnungsreduktion der Reglerkomplexität bei annähernd gleicher Reglerperformance möglich ist.

1 Einleitung

Modellprädiktive Regelung (MPC) ist eine moderne Methode der Prozessregelung mit einer breiten Palette von technologisch und ökonomisch vorteilhaften Anwendungen, siehe [13, 1]. Viele verschiedene Formulierungen von MPC wurden während der letzten Jahrzehnte entwickelt; dennoch besteht die wesentliche Struktur des Regelschemas immer aus einem expliziten oder impliziten Modell des Prozesses sowie aus einer online Optimierung der Stellgrößen über einen vordefinierten Regelhorizont für einen gegebenen Prädiktionshorizont der Regelgrößen, siehe [14].

*Korrespondenz bitte an diese Adresse

Die Anwendbarkeit und Leistungsfähigkeit einer MPC-Anwendung hängt besonders von der Qualität des Modells und der Umsetzbarkeit der online Optimierung ab. Aus diesem Grund werden oft Modelle reduzierter Ordnung in MPC Anwendungen verwendet, z.B. [8]. Dieser Zugang reduziert aber nicht die Anzahl der Entscheidungsvariablen [17], welche im Fall der MPC-Optimierung die Anzahl der Stellgrößen multipliziert mit der Länge des Regelhorizonts ist. Vor allem bei verteilt-parametrischen Systemen (Distributed Parameter Systems, DPS) kommt es oft zu numerischen Problemen mit Standard-Algoithmen für den Reglerentwurf [24].

Eine leistungsfähige Methode für die Modellreduktion von DPS ist die Proper Orthogonal Decomposition (POD). Eingeführt von [12], wurde die POD durch die Verwendung von sogenannten *Data Snapshots* weiter ausgebaut, [10]. Ein aktueller Überblick enthält Informationen über die Anwendung von MPC für DPS, siehe [16]. Im Gegensatz zu den Methoden, welche im vorliegenden Aufsatz vorgestellt werden, zielen die dort vorgestellten modalen Zerlegungen auf endlich dimensionale Modelle ab und reduzieren damit die Komplexität des Reglerentwurfs [3, 4, 21, 2, 28]. In [6] werden Signale im Ortsbereich in eine kleine Anzahl von Skalierungs- und Wavelet-Koeffizienten transformiert. Ein anderer Zugang wird in [7] präsentiert, wo charakteristische Eigenschaften einer Störung mittels Karhunen- Loeve Transformation (KLT) gewonnen werden.

Als Alternative zu den oben erwähnten Modellreduktionsmethoden zielen viele Veröffentlichungen darauf ab, lediglich die Anzahl der Entscheidungsvariablen zu reduzieren. In [23] werden die Entscheidungsvariablen durch geeignete Eingangstrajektorien parametriert, um die Dimension der Optimierung zu reduzieren.

Die vielseitige Methode der Singularwertzerlegung (Singular Value Decomposition, SVD) wurde oft angewendet, um hochdimensionale dynamische Prozessmessungen auf einen niedrigdimensionalen Unterraum zu projizieren. Verschiedene Zugänge für die Reduktion einer hoch-dimensionalen Informationsreihe ohne signifikanten Performance-Verlust sind z.B. dargestellt in [5] und [27]. In [27] wird die Verwendung einer positiv linearen Zerlegung für Mustererkennung präsentiert.

Der Schwerpunkt dieses Beitrags besteht aus der Präsentation und Diskussion einer Methode, um die Dimension des MPC Optimierungsproblems signifikant zu reduzieren. Dabei wird eine geeignete Parametrierung der Eingangstrajektorien vorgeschlagen, welche durch eine Analyse und Zerlegung von tatsächlichen Stellgrößen gewonnen wird.

Erstens wird eine reale MPC Implementierung üblicherweise durch vorhergehende Simulationsstudien abgesichert, wobei ein genaues Prozessmodell mit einem MPC-Regler voller Ordnung getestet wird. In diesem Fall kann die vorgeschlagene Ordnungsreduktion die Stellgrößen aus der Simulation als geeignete Eingangstrajektorien verwenden. Zweitens stellt eine MPC Implementierung oft den Ersatz für eine konventionelle Automatisierung dar. Die Stellgrößen des konventionellen Schemas können zumindest als erster Entwurf für geeignete Eingangstrajektorien verwendet werden.

Abbildung 1 zeigt diese beiden Möglichkeiten um Stellgrößen-Signale für die Analyse und Zerlegung zu gewinnen. Es ist notwendig, dass die Signale ${}^{i}\boldsymbol{W}$ alle wichtigen Prozess-Charakteristiken umfassen, so z.B. typische Störungen und Sollwert-Änderungen oder auch Kombinationen von aktiven Beschränkungen. Zu diesem Zweck wird das Konzept der *Principal Control Moves* (PCM) eingeführt: Die PCM bezeichnen Schnappschüsse

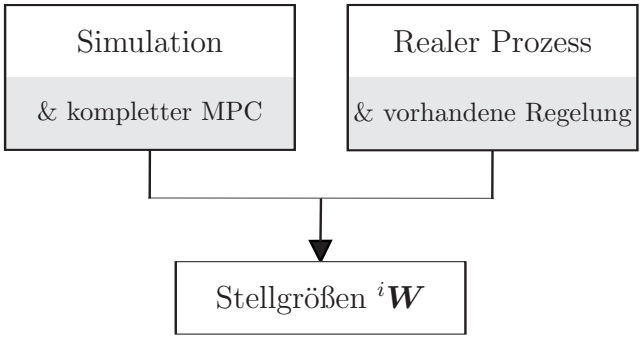


Abbildung 1: Zwei grundsätzliche Zugänge zur Gewinnung von geeigneten Schnappschüssen von Stellgrößen-Signalen ${}^{i}\boldsymbol{W}$.

(*Snapshots*) der Stellgrößen, welche während repräsentativer Regelmanöver aufgezeichnet worden sind.

Im Folgenden werden diese PCM mittels der Karhunen-Loeve Transformation zerlegt. Durch Anwendung einer Singularwertzerlegung auf die Menge aller identifizierten PCM wird eine orthonormale Basis zusammen mit einem Proper Orthonormal Value (POV) berechnet. Dieser POV wird in der Folge verwendet, um nur jene Komponenten aus der Basis auszuwählen, welche signifikant zur Performance der Regelung beitragen. Die optimale Stellgrößenfolge über den Regelhorizont wird schließlich aus einer gewichteten Überlagerung dieser ausgewählten Komponenten gebildet. Die neuen Entscheidungsvariablen sind folglich stark in ihrer Anzahl reduziert.

2 Modellprädiktive Regelung

Für die Verwendung in einem online MPC-Schema wird das Streckenmodell üblicherweise durch ein lineares zeitdiskretes Zustandsraum-Modell beschrieben [15]

$$\begin{aligned} \boldsymbol{x}(k+1) &= \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}\Delta\boldsymbol{u}(k) \\ \boldsymbol{y}(k) &= \boldsymbol{C}\boldsymbol{x}(k), \end{aligned} \tag{1}$$

wobei $\boldsymbol{x} \in I\!R^n$ den Zustandsvektor, $\Delta\boldsymbol{u} \in I\!R^{n_u}$ den Vektor der Stellgrößen und $\boldsymbol{y} \in I\!R^{n_y}$ den Ausgangsvektor beschreiben.

Die hier verwendete MPC-Formulierung basiert auf einer iterativen Optimierung mit endlichem Horizont für das Modell (1). Zu jedem Zeitschritt k wird eine zukünftige optimale Reihe von Stellinkrementen $\Delta\boldsymbol{u}(k)$ für einen gleitenden Regelhorizont N_c bestimmt. Basierend auf dem momentanen Zustandsvektor $\boldsymbol{x}(k)$ muss das folgende Kriterium minimiert werden:

$$\begin{aligned} J \;=\; \sum_{i=0}^{N_c-1} \Big\{ \boldsymbol{x}(k+i+1)^T \boldsymbol{Q}\boldsymbol{x}(k+i+1) + \Delta\boldsymbol{u}^T(k+i)\boldsymbol{R}\Delta\boldsymbol{u}(k+i) \Big\} \\ + \boldsymbol{x}(k+N_c)^T \boldsymbol{P}\boldsymbol{x}(k+N_c). \end{aligned} \tag{2}$$

Die Gewichtsmatrizen $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$, $\boldsymbol{R} \in \mathbb{R}^{n_u\times n_u}$ sind symmetrisch und positiv definit während die Endgewichtsmatrix $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ die positiv definite Lösung der algebraischen Riccati Gleichung darstellt

$$\boldsymbol{P} = \boldsymbol{A}^T\boldsymbol{P}\boldsymbol{A} - \boldsymbol{K}^T\tilde{\boldsymbol{R}}\boldsymbol{K} + \boldsymbol{Q} \tag{3}$$

mit $\tilde{\boldsymbol{R}} = \boldsymbol{R} + \boldsymbol{B}^T\boldsymbol{P}\boldsymbol{B}$ und $\boldsymbol{K} = \tilde{\boldsymbol{R}}^{-1}\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{A}$. Die Folge der zukünftigen Stellinkremente $\Delta\boldsymbol{U}(k) \in \mathbb{R}^{N_c\cdot n_u\times 1}$ ist die Entscheidungsvariable, gegeben durch

$$\Delta\boldsymbol{U}(k) = [\,\Delta\boldsymbol{u}(k+1)^T \;\ldots\; \Delta\boldsymbol{u}(k+N_c)^T\,]^T. \tag{4}$$

Beschränkungen werden üblicherweise durch Ungleichungen zusätzlich zum ursprünglichen MPC-Problem formuliert. In kompakter Schreibweise ergibt sich, vgl. [25]:

$$\boldsymbol{M}\Delta\boldsymbol{U} \le \boldsymbol{N}. \tag{5}$$

Eine formale Definition des beschränkten Regelproblems ist daher gegeben durch

$$\mathscr{P}(\boldsymbol{x}(k)): \quad \Delta\boldsymbol{U}^*(\boldsymbol{x}(k)) = \arg\min J(\boldsymbol{x}(k), \Delta\boldsymbol{U}) \;\text{ mit }\; \boldsymbol{M}\Delta\boldsymbol{U} \le \boldsymbol{N}. \tag{6}$$

In Matrix-Schreibweise kann die Kostenfunktion (2) ausgedrückt werden durch

$$J = \boldsymbol{x}^T\boldsymbol{F}^T\bar{\boldsymbol{Q}}\boldsymbol{F}\boldsymbol{x} + 2\Delta\boldsymbol{U}^T\boldsymbol{\Phi}^T\bar{\boldsymbol{Q}}\boldsymbol{F}\boldsymbol{x} + \Delta\boldsymbol{U}^T(\boldsymbol{\Phi}^T\bar{\boldsymbol{Q}}\boldsymbol{\Phi} + \bar{\boldsymbol{R}})\Delta\boldsymbol{U}. \tag{7}$$

wobei $\boldsymbol{x}$ eine kurze Schreibweise für $\boldsymbol{x}(k)$ ist, sowie $\bar{\boldsymbol{Q}} = \text{diag}\,[\,\boldsymbol{Q},\ldots,\boldsymbol{Q},\boldsymbol{P}\,]$ und $\bar{\boldsymbol{R}} = \text{diag}\,[\,\boldsymbol{R},\ldots,\boldsymbol{R}\,]$. Die Matrizen $\boldsymbol{F} \in \mathbb{R}^{N_c\cdot n\times n}$ und $\boldsymbol{\Phi} \in \mathbb{R}^{N_c\cdot n\times N_c\cdot n_u}$ prädizieren den Zustandsvektor der Strecke basierend auf $\boldsymbol{x}(k)$ und $\Delta\boldsymbol{U}(k)$. Die *unbeschränkte* optimale Sequenz der Stellinkremente ist gegeben durch

$$\Delta\boldsymbol{U}^*(\boldsymbol{x}(k)) = -(\boldsymbol{\Phi}^T\bar{\boldsymbol{Q}}\boldsymbol{\Phi} + \bar{\boldsymbol{R}})^{-1}\boldsymbol{\Phi}^T\bar{\boldsymbol{Q}}\boldsymbol{F}\boldsymbol{x}. \tag{8}$$

3 Definition und Parametrisierung der PCMs

Bei der Minimierung von (7) unter der Nebenbedingung (5) ist die Anzahl der Entscheidungsvariablen in einem quadratischen Program $N_c \cdot n_u$ festgelegt durch die Länge aller zukünftigen Stellinkremente $\Delta\boldsymbol{U}$, wie in (4) definiert. Eine mögliche Lösung (siehe z.B. [23, 18]) ist die Parametrierung von $\Delta\boldsymbol{U}$ entsprechend

$$\Delta\boldsymbol{U} = \boldsymbol{\Omega}\boldsymbol{p}, \tag{9}$$

wobei $\boldsymbol{\Omega} \in \mathbb{R}^{N_c\cdot n_u\times\nu_p}$ eine Matrix ist, deren Spalten eine Basis für $\Delta\boldsymbol{U}$ bilden, und $\boldsymbol{p} \in \mathbb{R}^{\nu_p\times 1}$ ist die neue Entscheidungsvariable mit reduzierter Dimension ($\nu_p \le N_c\cdot n_u$). Da die Spalten von $\boldsymbol{\Omega}$ den Verlauf der Stellinkremente für jede Stellgröße bis zum Regelhorizont enthalten werden sie hier *principal control moves* PCM genannt. Die Anzahl der Spalten ν_p wird deswegen als Ordnung der PCM bezeichnet.

Das beschränkte Optimierungsproblem (6) wird gelöst indem $\Delta\boldsymbol{U}$ durch (9) ersetzt wird. Die Formulierung der Kostenfunktion nach (7) wird damit zu

$$J_p = \boldsymbol{x}^T\boldsymbol{F}^T\bar{\boldsymbol{Q}}\boldsymbol{F}\boldsymbol{x} + 2\boldsymbol{p}^T\boldsymbol{\Omega}^T\boldsymbol{\Phi}^T\bar{\boldsymbol{Q}}\boldsymbol{F}\boldsymbol{x} + \boldsymbol{p}^T\boldsymbol{\Omega}^T(\boldsymbol{\Phi}^T\bar{\boldsymbol{Q}}\boldsymbol{\Phi} + \bar{\boldsymbol{R}})\boldsymbol{\Omega}\boldsymbol{p}, \tag{10}$$

und das Optimierungsproblem (6) mit $\boldsymbol{p}$ als Entscheidungsvariable ist gegeben durch

$$\mathscr{P}_p(\boldsymbol{x}(k)): \quad \boldsymbol{p}^*(\boldsymbol{x}(k)) = \arg\min J_p(\boldsymbol{x}(k), \boldsymbol{p}) \;\text{ s.t. }\; \boldsymbol{M}\boldsymbol{\Omega}\boldsymbol{p} \le \boldsymbol{N}. \tag{11}$$

4 Methodik der Ordnungsreduktion

4.1 Principal Control Manoeuvres

Die Hauptfrage besteht nun darin, die rrichtigenPPCM für die Basis $\mathbf{\Omega}$ zu finden. Das Ziel ist dabei, die Ordnung der PCM ν_p zu minimieren und gleichzeitig Stabilität sowie eine Mindestperformance zu garantieren. Dazu wird das Konzept der *Principal Control Manoeuvres* eingeführt. Diese sind transiente Vorgänge im geschlossenen Regelkreis, welche charakteristische Stellgrößen erfordern. Formale Anforderungen und Methoden zur Identifikation von geeigneten Manövern sind in [26, 9] angegeben. Diese Manöver sind ausgezeichnet durch

- typische Sollwertänderungen im geschlossenen Regelkreis,
- typische Störungen, welche kompensiert werden sollen,
- verschiedene Amplituden von Sollwertänderungen oder Störungen, welche eine zunehmende Anzahl von Beschränkungen bei Stellgrößen und Zuständen aktivieren.

Es ist zu beachten, dass die Anzahl der aktiven Beschränkungen vor allem von den Amplituden der betroffenen Variablen abhängen. Es entsteht damit ein nichtlineares Optimierungsproblem, welches auch die Aufgabe geeignete PCM zu finden nichtlinear macht. Eine geschlossene Lösung ist jedenfalls schwer anzugeben, dennoch kann ein strukturierter Zugang für die Erfassung hinreichender Manöver angegeben werden:

1. Von jedem Manöver werden die Hauptkomponenten der Entscheidungsvariablen mit Hilfe der KLT extrahiert.

2. Die zusammengefassten Hauptkomponenten von allen Manövern werden einer zweiten Singularwertzerlung zugeführt, um mögliche Redundanzen zu entfernen.

Beide Schritte können iterativ durchgeführt werden, um eine effiziente Identifikation zu ermöglichen.

Schritt 1: Extraktion einzelner Control Moves mittels Karhunen-Loeve Transformation (KLT)

Jedes Manöver (gekennzeichnet durch den Index $i = 1, \ldots, q$) wird über eine Zeitspanne von ${}^{i}N$ Samples aufgezeichnet. Für jeden Zeitpunkt k wird die vollständige Reihe der tatsächlichen Stellinkremente $\Delta\boldsymbol{U}(k)$ (4) aufgezeichnet. Mit der Matrix ${}^{i}\boldsymbol{W} \in I\!R^{N_c \cdot n_u \times {}^{i}N}$, welche alle ${}^{i}N$ Aufzeichnungen von $\Delta\boldsymbol{U}(k)$ für das i-te Manöver enthält

$$ {}^{i}\boldsymbol{W} = [\, \Delta\boldsymbol{U}(k) \quad \Delta\boldsymbol{U}(k+1) \quad \ldots \quad \Delta\boldsymbol{U}(k + {}^{i}N - 1)\,], \tag{12} $$

können die Hauptkomponenten aller $\Delta\boldsymbol{U}$ mit der Singularwertzerlegung von ${}^{i}\boldsymbol{W}$ berechnet werden [22]:

$$ {}^{i}\boldsymbol{W} = {}^{i}\boldsymbol{U}\, {}^{i}\boldsymbol{S}\, {}^{i}\boldsymbol{V}^{T}. \tag{13} $$

Die orthonormale Matrix ${}^{i}\boldsymbol{U} \in I\!R^{N_c \cdot n_u \times N_c \cdot n_u}$ stellt eine Basis für alle möglichen Control Moves dar. Insbesondere ist jede Spalte von ${}^{i}\boldsymbol{U}$ eine Hauptkomponente der Spalten von ${}^{i}\boldsymbol{W}$. Die Singularwerte ${}^{i}\sigma_j$, welche die Diagonale von ${}^{i}\boldsymbol{S}$ bilden, bewerten die relative

Bedeutung der Hauptkomponenten; die ${}^i\sigma_j$ werden daher auch als *Proper Orthogonal Values (POV)* bezeichnet ([11]).

Die Singularwerte erfüllen zwei wichtige Aufgaben:

1. Sie zeigen qualitativ die geeignete Ordnung an, welche benötigt wird, damit das jeweilige Manöver ausgeführt werden kann und sind daher wichtig für die Wahl von ν_p.

2. Sie sind hilfreich bei der Entscheidung, wie lange Stellgrößen für ein bestimmtes Manöver aufgezeichnet werden müssen. Wenn die Singularwertzerlegung für jede Aufzeichnung wiederholt ausgeführt wird, so werden die Singularwerte der anwachsenden Matrix ${}^i\boldsymbol{W}$ zuerst stark variieren. Sobald die Singularwerte nur noch geringe Änderungen aufweisen, kann die Aufzeichnung beispielsweise gemäß dem folgenden Kriterium gestoppt werden:

$$\max_j \frac{\operatorname{abs}\left({}^i\sigma_{j,{}^iN+1} - {}^i\sigma_{j,{}^iN}\right)}{{}^i\sigma_{j,{}^iN}} < {}^i\epsilon. \tag{14}$$

   Dabei ist ${}^i\sigma_{j,{}^iN}$ der j-te Singularwert von ${}^i\boldsymbol{W}$ mit ${}^iN$ Messungen und ${}^i\epsilon$ ist ein frei zu wählender Schwellwert.

Für jedes der q Manöver werden die Control Moves auf die folgende Weise reduziert: Die durch ${}^i\boldsymbol{U}$ repräsentierte Basis wird auf die ersten m_i Spalten beschränkt ($m_i < n_u \cdot N_c$):

$${}^i\boldsymbol{U}_{red} = \left[\boldsymbol{u}_{i,1} \ldots \boldsymbol{u}_{i,m_i}\right]. \tag{15}$$

Schritt 2: Kombination der PCM

Aus der Menge der q Basen ${}^i\boldsymbol{U}_{red}$ muss die Matrix der PCM $\boldsymbol{\Omega}$ in (9) festgelegt werden. Zu diesem Zweck werden die Basen ${}^i\boldsymbol{U}_{red}$ zu einer Matrix $\boldsymbol{W}_\Sigma \in I\!R^{n_u \cdot N_c \times \sum_{i=1}^q m_i}$ vereinigt, welche wiederum einer Singularwertzerlegung unterworfen wird:

$$\boldsymbol{W}_\Sigma = \left[{}^1\boldsymbol{U}_{red} \quad {}^2\boldsymbol{U}_{red} \ldots {}^q\boldsymbol{U}_{red}\right] = \boldsymbol{U}_\Sigma \boldsymbol{S}_\Sigma \boldsymbol{V}_\Sigma^T. \tag{16}$$

Die Links-Singularvektoren $\boldsymbol{U}_\Sigma$ spannen eine Basis für $\boldsymbol{\Omega}$ auf, wobei die Singularwerte $\sigma_{\Sigma,j}$ wieder als Kriterium für die Reduktion auf eine bestimmte Ordnung $\nu_p \leq \sum_{i=1}^q m_i$ verwendet werden können:

$$\boldsymbol{\Omega} = \boldsymbol{U}_{\Sigma,red} = \left[\boldsymbol{u}_{\Sigma,1} \ldots \boldsymbol{u}_{\Sigma,\nu_p}\right]. \tag{17}$$

Es sei an dieser Stelle nochmals betont, dass *beide* Schritte wesentlich der Ordnungsreduktion dienen: Da aufgrund der Beschränkungen im Allgemeinen ein nichtlineares Optimierungsproblem zu lösen ist, wird Schritt 1 auf alle Control Moves für ein bestimmtes Manöver angewendet (siehe Abschnitt 4.1), während Schritt 2 auf die Vereinigung aller PCM angewendet wird (4.1).

4.2 Stabilitätsanalyse für den reduzierten MPC

Dieser Abschnitt beschreibt eine grundlegende Stabilitätsanalyse für den reduzierten modellprädiktiven Regler basierend auf der Stabilitätstheorie von Lyapunov. Diese baut auf dem in [18] präsentierten Ansatz auf und bedient sich der Lösung der Ricattigleichung (3) als Kandidat für eine Lyapunovfunktion. Basierend auf dieser werden Bedingungen für exponentielle Stabilität angegeben, welche auf einer minimalen Ordnung ν_p basieren.

Als Voraussetzung werde die Rückführverstärkungen für den originalen und den reduzierten MPC angegeben. Für den originalen MPC voller Ordnung erhält man bei Minimierung von (7) die Verstärkung $\bar{\boldsymbol{K}}$ zu

$$\bar{\boldsymbol{K}} = \boldsymbol{H}^{-1}\bar{\boldsymbol{F}} \tag{18}$$

wobei $\boldsymbol{H} = \boldsymbol{\Phi}^T\bar{\boldsymbol{Q}}\boldsymbol{\Phi} + \bar{\boldsymbol{R}}$ die Hessematrix bezeichnet und $\bar{\boldsymbol{F}} = \boldsymbol{\Phi}^T\bar{\boldsymbol{Q}}\boldsymbol{F}$ bedeutet. Analog erhält man die Rückführverstärkung des reduzierten MPC gem. (10) zu

$$\boldsymbol{K}_{\nu_p} = \boldsymbol{\Omega}(\boldsymbol{\Omega}^T\boldsymbol{H}\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}^T\bar{\boldsymbol{F}}. \tag{19}$$

$$\boldsymbol{E}_{\nu_p} = \bar{\boldsymbol{K}} - \boldsymbol{K}_{\nu_p}. \tag{20}$$

Definition 1. *Für eine feste Ordnung ν_p werden die folgenden Matrizen definiert:*

$$\bar{\boldsymbol{E}}_{\nu_p} = \boldsymbol{D}\boldsymbol{E}_{\nu_p}$$
$$\bar{\boldsymbol{K}}_{\nu_p} = \boldsymbol{D}\boldsymbol{K}_{\nu_p}$$

mit

$$\boldsymbol{D} = [\, \boldsymbol{I}_{N_u \times n_u} \quad \boldsymbol{0}_{n_u \times (N-1)n_u} \,]$$

und $\boldsymbol{E}_{\nu_p}$, $\boldsymbol{K}_{\nu_p}$ aus (20) bzw. (19).

Theorem 1 (Stabilität des reduzierten MPC)**.** *Für eine feste PCM-Ordnung ν_p wird die Matrix $\boldsymbol{Q}_{\nu_p}$ definiert:*

$$\boldsymbol{Q}_{\nu_p} = \boldsymbol{Q} - \bar{\boldsymbol{E}}_{\nu_p}^T\tilde{\boldsymbol{R}}\bar{\boldsymbol{E}}_{\nu_p}. \tag{21}$$

Der Regelkreis ist exponentiell stabil, wenn

$$\delta_{\nu_p} = \frac{\boldsymbol{x}_{max}^T\boldsymbol{Q}_{\nu_p}\boldsymbol{x}_{max}}{\boldsymbol{x}_{max}^T\boldsymbol{P}\boldsymbol{x}_{max}} < 1$$

gilt, wobei $\boldsymbol{P}$ die Lösung der Ricattigleichung (3) darstellt und $\boldsymbol{x}_{max}$ der Eigenvektor zum größten Eigenwert der Matrizen $\boldsymbol{P}, \boldsymbol{Q}_{\nu_p}$ ist.

Beweis:

Es wird die Dynamik des geschlossenen Regelkreises betrachtet:

$$\boldsymbol{x}(k+1) = (\boldsymbol{A} - \boldsymbol{B}\bar{\boldsymbol{K}}_{\nu_p})\boldsymbol{x}(k) = (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{K})\boldsymbol{x}(k) + \boldsymbol{B}\bar{\boldsymbol{E}}_{\nu_p}\boldsymbol{x}(k), \tag{22}$$

mit $\boldsymbol{K}$ gem. Gl. (3). Als Kandidat für eine Lyapunovfunktion wird $\mathscr{V}(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}$ gewählt, wobei $\boldsymbol{P}$ die Lösung der Ricattigleichung (3) darstellt. Für den geschlossenen Regelkreis (22), welcher mit dem reduzierten MPC gebildet wird, verlangt man

$$\triangle\mathscr{V}(\boldsymbol{x}(k)) = \mathscr{V}(\boldsymbol{x}(k+1)) - \mathscr{V}(\boldsymbol{x}(k)) \leq -\delta\mathscr{V}(\boldsymbol{x}(k)).$$

Nach Einsetzen von $\boldsymbol{x}(k+1)$ und Umformung erhält man

$$\triangle\mathscr{V}(\boldsymbol{x}(k)) = -\boldsymbol{x}(k)^T\left\{\boldsymbol{Q} + (\boldsymbol{K} - \bar{\boldsymbol{E}}_{\nu_p})^T\boldsymbol{R}(\boldsymbol{K} - \bar{\boldsymbol{E}}_{\nu_p}) - \bar{\boldsymbol{E}}_{\nu_p}^T\tilde{\boldsymbol{R}}\bar{\boldsymbol{E}}_{\nu_p}\right\}\boldsymbol{x}(k),$$

$$\triangle\mathscr{V}(\boldsymbol{x}(k)) \leq -\boldsymbol{x}(k)^T\left\{\boldsymbol{Q} - \bar{\boldsymbol{E}}_{\nu_p}^T\tilde{\boldsymbol{R}}\bar{\boldsymbol{E}}_{\nu_p}\right\}\boldsymbol{x}(k).$$

Mit der Abkürzung $\boldsymbol{Q}_{\nu_p} = \boldsymbol{Q} - \bar{\boldsymbol{E}}_{\nu_p}^T\tilde{\boldsymbol{R}}\bar{\boldsymbol{E}}_{\nu_p}$ entsteht die Forderung

$$\boldsymbol{Q}_{\nu_p} \succeq \delta\boldsymbol{P}. \tag{23}$$

Dies bedeutet, dass für beliebige Vektorn $\boldsymbol{x}$ ein maximales δ derart zu finden ist, dass

$$\boldsymbol{x}^T\boldsymbol{Q}_{\nu_p}\boldsymbol{x} \geq \boldsymbol{x}^T\delta\boldsymbol{P}\boldsymbol{x} \tag{24}$$

gilt. Diese Aufgabe kann als beschränktes Optimierungsproblem formuliert werden:

$$\boldsymbol{x}^T\delta\boldsymbol{P}\boldsymbol{x} = \max_{\boldsymbol{x},\delta}! \qquad \text{mit} \quad \boldsymbol{x}^T\boldsymbol{Q}_{\nu_p}\boldsymbol{x} = c, \quad c \in I\!R^+$$

Die Lagrangefunktion für dieses Problem ist

$$\mathscr{L}(\boldsymbol{x},\delta) = \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} - \frac{1}{\delta}(\boldsymbol{x}^T\boldsymbol{Q}_{\nu_p}\boldsymbol{x} - c). \tag{25}$$

Die Lösung erhält man aus $\partial\mathscr{L}/\partial\boldsymbol{x} = 0$:

$$(\boldsymbol{P} - \lambda'\boldsymbol{Q}_{\nu_p})\boldsymbol{x} = 0 \quad \text{mit} \quad \lambda' = \frac{1}{\delta}.$$

Aus dieser Gleichung geht $\boldsymbol{x}_{max}$ als Eigenvektor zum größten Eigenwert λ' des Matrizenpaars $\boldsymbol{P}, \boldsymbol{Q}_{\nu_p}$. Wenn nun auch die Beschränkung $\boldsymbol{x}^T\boldsymbol{Q}_{\nu_p}\boldsymbol{x} = c$ erfüllt wird, ist der Beweis vollständig:

$$\delta_{\nu_p} = \frac{\boldsymbol{x}_{max}^T\boldsymbol{Q}_{\nu_p}\boldsymbol{x}_{max}}{\boldsymbol{x}_{max}^T\boldsymbol{P}\boldsymbol{x}_{max}}. \tag{26}$$

Der größte Eigenwert λ' entspricht jener Abklingrate δ, für die (24) gerade noch für *alle* $\boldsymbol{x}$ erfüllt ist. Eine höhere Abklingrate tritt nur noch in Teilbereichen des Zustandsraums auf.

Theorem 2 (Wahl der reduzierten Reglerordnung)**.** *Für ein bestimmtes ν_p wird $\delta_{\nu_p} \in (0;1]$ aus Gl. (26) betrachtet. Der geschlossene Regelkreis ist dann für alle $\nu_{p,i} \geq \nu_{p1} \geq \nu_p$ stabil, wenn*

$$||\boldsymbol{E}_{\nu_{p1}}||_2 < \sqrt{\frac{||\boldsymbol{Q} - \delta_{\nu_p}\boldsymbol{P}||_2}{||\tilde{\boldsymbol{R}}||_2}}$$

gilt.

Beweis:

Für eine bestimmt Ordnung ν_p seien $\boldsymbol{x}_{max}$ und δ_{ν_p} Lösungen von (26). Dann ergibt sich aus (24)

$$0 \leq \boldsymbol{x}^T \bar{\boldsymbol{E}}_{\nu_p}^T \tilde{\boldsymbol{R}} \bar{\boldsymbol{E}}_{\nu_p} \boldsymbol{x} \leq \boldsymbol{x}^T (\boldsymbol{Q} - \delta_{\nu_p} \boldsymbol{P}) \boldsymbol{x} \quad \forall \boldsymbol{x}. \tag{27}$$

Ersetzt man die rechte Seite der Gleichung durch ihre obere Schranke, so erhält man

$$\boldsymbol{x}^T \bar{\boldsymbol{E}}_{\nu_p}^T \tilde{\boldsymbol{R}} \bar{\boldsymbol{E}}_{\nu_p} \boldsymbol{x} \leq ||\boldsymbol{Q} - \delta_{\nu_p} \boldsymbol{P}||_2 \quad \forall \boldsymbol{x},\ ||\boldsymbol{x}||_2 = 1. \tag{28}$$

Anwendung der Chauchy-Schwarz Ungleichung auf die linke Seite von (28) liefert

$$||\boldsymbol{x}^T \bar{\boldsymbol{E}}_{\nu_p}^T \tilde{\boldsymbol{R}} \bar{\boldsymbol{E}}_{\nu_p} \boldsymbol{x}||_2 \leq ||\bar{\boldsymbol{E}}_{\nu_p}||_2^2 ||\tilde{\boldsymbol{R}}||_2. \tag{29}$$

Aus (29) kann nun eine untere Schranke $\nu_{p1} \geq \nu_p$ für eine minimale reduzierte Ordnung bei gefordertem δ_{ν_p} gewonnen werden:

$$||\bar{\boldsymbol{E}}_{\nu_{p1}}||_2 \leq ||\boldsymbol{D}||_2 ||\boldsymbol{E}_{\nu_{p1}}||_2 < \sqrt{\frac{||\boldsymbol{Q} - \delta_{\nu_p} \boldsymbol{P}||_2}{||\tilde{\boldsymbol{R}}||_2}}. \tag{30}$$

5 Anwendung auf einen industriellen Trocknungsprozess

In diesem Kapitel sollen die Ergebnisse der Ordnungsreduktionsprinzipien der PCM-Methode anhand eines linearisierten industriellen Trocknungsprozesses demonstriert werden. Das Ziel des Prozesses ist es, Faservlies mit gewünschter Restfeuchte zu erzeugen (Ausgangsgröße y, $n_y = 1$). Das Modell des Faser-Trockners wird in [20] vorgestellt, wobei das lineare Modell fünf Zustände umfasst ($n = 5$). Die Auslegung des beschränkten MPC wird in [19] beschrieben. In Abbildung 2 ist das Trocknermodell schematisch dargestellt.

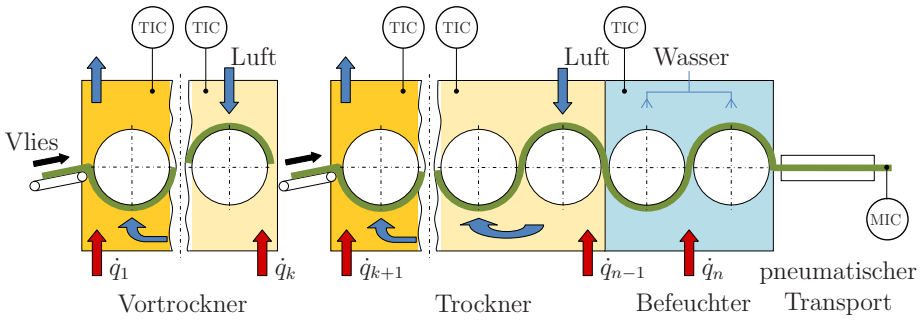


Abbildung 2: Schematische Darstellung des Faservlies-Trockners. Der Transportpfad des Faservlieses und die Flussrichtungen sind eingezeichnet. Die Pfeile mit $\dot{q}_i$ repräsentieren die in das System eingebrachte Energy.

Ein Band aus Faservlies wird im Trockner über rotierende Siebtrommeln transportiert, durch die beheizte Luft gesaugt wird. Der Trockner ist dabei in acht Beheizungszonen aufgeteilt, für welche die Temperaturen die Stellgrößen darstellen. Eine zusätzliche Stellgröße ist die Menge an Wasser die am Ende des Trockners eingesprüht wird, um eine Homogenisierung der Feuchtigkeit zu erzielen. Dadurch ergeben sich neun Stellgrößen für den Prozess ($n_u = 9$). Die messbare Störgröße z des Prozesses stellt die Abweichung der Dicke des Faservlieses von der Nominalstärke von $40mm$ dar. Alle Stellgrößen sind sowohl in Absolutwert als auch in Änderungsrate beschränkt.

5.1 PCM Identifikation

Um ${}^{i}\boldsymbol{W}$ gemäß (12) aufzeichnen zu können, wurden vier Regelmanöver ($q = 4$) betrachtet, wobei für jedes Manöver verschiedene Störungssprünge mit unterschiedlichen Amplituden aufgebracht wurden. Die unterschiedlichen Störamplituden $\Delta z = [-10, -20, -30, -40]$ wurden dabei mit dem Nominalwert von $40mm$ überlagert. Dadurch wurde erreicht, dass bei der Regelung schrittweise die unterschiedlichen Beschränkungen aktiv wurden. In Abbildung 3 ist die Systemantwort bei einer Störung von $\Delta z = -40$ dargestellt.

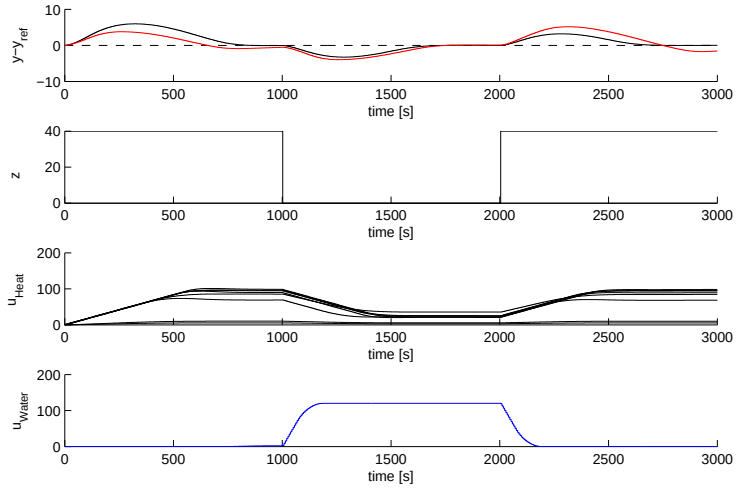


Abbildung 3: Systemantwort auf einen Störungssprung von $\Delta z = -40$ (Regelmanöver $i = 4$. Rot: LQR, Schwarz: MPC)

Der Kontrolhorizont des MPC beträgt $N_c = 10$ Abtastwerte, wodurch die Anzahl an Entscheidungsvariablen $N_c \cdot n_u = 90$ ergibt. Weiters wurde ein LQR-Regler als Vertreter der konventionellen Regelstrukturen ausgewählt und für Vergleichszwecke entworfen.

Die Ordnungsreduktion m_i ($i = 1, \ldots, 4$) gemäß (15) wurde mit $m_i = 25$ ($m_i = 70$ für die LQR-basierte Analyse) gewählt, wodurch sich vier Basen ${}^{i}\boldsymbol{U}_{red}$ mit je 25 (bzw. 70) orthonormalen Basisvektoren ergeben. Die Zerlegung der resultierenden $\boldsymbol{W}_{\Sigma}$-Matrix (16)

ergibt normierte Singularwerte, die in Abbildung 4 dargestellt sind. Errechnet wurden die Singularwerte mit (13) und (14) aus Snapshot-Daten des MPC.

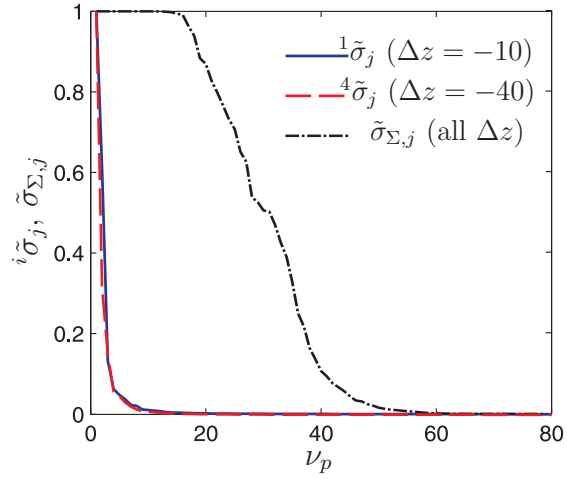


Abbildung 4: Normierte Singularwerte für verschiedene Störungssprünge ${}^i\sigma_j$ und die kombinierten Principal Control Moves $\sigma_{\Sigma,j}$, aufgetragen über die PCM-Ordnung ν_p (Snapshot-Daten von der MPC-trainierten Variante).

Ersichtlich ist, dass nur eine geringe Anzahl an Singularwerten dominant sind, wodurch gezeigt wird, dass eine Anzahl von $\nu_p \approx 10$ Freiheitsgraden für die Principal Control Moves ausreichend sind.

5.2 Resultate

Die Validierung wurde mit gemessenen Störsignalen z des realen Prozesses durchgeführt. Die Referenz stellt dabei die Simulation eines konventionellen MPC-Regler mit voller Ordnung dar, welcher in [19] präsentiert wurde.

Die Leistungsfähigkeit des konventionellen MPC-Reglers ist in Abbildung 5 dargestellt, wobei erkennbar ist, dass die Anregung durch die Störung starke Störungen der Dynamik des geschlossenen Regelkreises provoziert sowie die Beschränkungen des MPC erreicht werden (z.B. Eingespritztes Wasser u_{Water} zwischen $t = 3200\ s$ und $4000\ s$.)

Um auch eine quantitative Größe für die Leistungsfähigkeit zu geben, wurde der quadratische Regelfehler (MSE) für die Faserfeuchtigkeit für reduzierte MPC Varianten mit unterschiedlichen Freiheitsgraden in Tabelle 1 zusammengefasst.

Aus Tabelle 1 ist eindeutig ersichtlich, dass erst ab einer PCM-Ordnung von $\nu_p < 10$ eine signifikante Minderung des MSE für die MPC-trainierte Variante auftritt (Bei $\nu_p = 10$ ist die Abweichung unter 1%). Im Fall der LQR-trainierten Variante tritt eine Abweichung des MSE bereits vorher auf, wobei für $\nu_p = 45$ akzeptable Ergebnisse erzielt werden, wie in Abbildung 6 dargestellt ist. Es konnte somit im MPC-trainierten Fall mit $\nu_p = 10$, eine Reduktion von 90% erreicht werden.

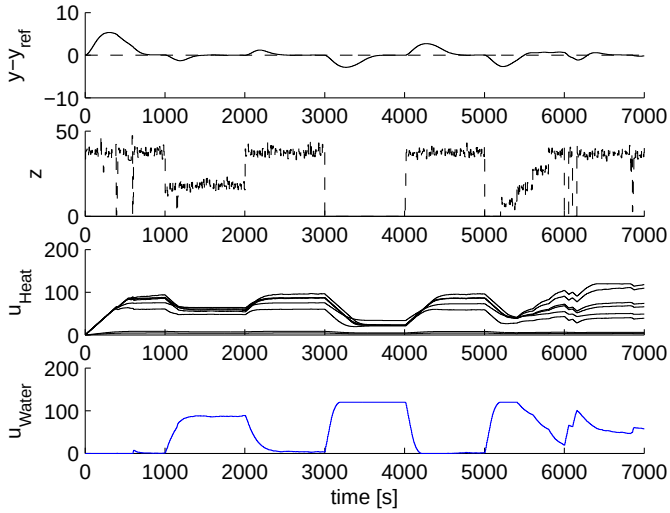


Abbildung 5: Leistungsfähigkeit des MPC mit voller Ordnung

PCM-Ordnung ν_p max. # an Entscheidungsvariablen	MSE MPC-trainiert	MSE LQR-trainiert
1	0.975	-
3	0.958	-
5	0.832	-
7	0.785	-
10	0.382	-
15	0.378	-
25	0.3780	0.8283
30	0.3788	0.5435
40	0.3776	0.4909
90 (konv. MPC)	0.378	0.3776

Tabelle 1: MSE Werte für verschiedene PCM-Ordnungen und unterschiedlichen Snapshot-Quellen.

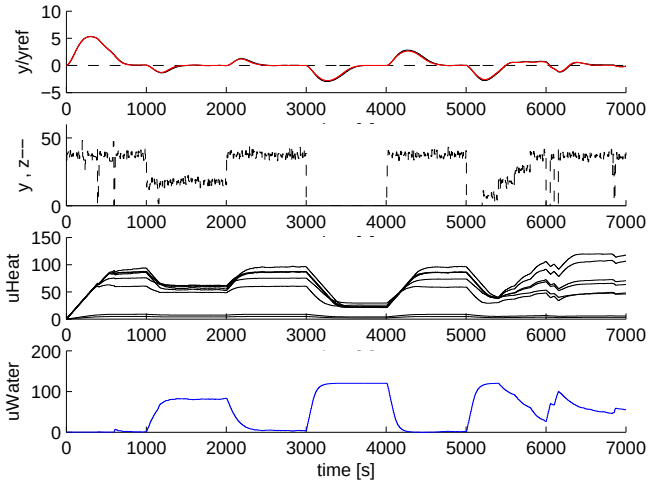


Abbildung 6: Leistungsfähigkeit des reduzierten MPC mit verschiedenen Snapshots. Schwarz: MPC Snapshots $\nu_p = 10$, Rot: LQR Snapshots $\nu_p = 45$.

6 Schlussfolgerungen

Es wurde eine Methode für die Reduktion der Komplexität der Online-Optimierung des MPC-Regelkonzeptes beschrieben. Die Hauptidee besteht darin, Snapshot-Daten der Stellgrößen zu verwenden und diese mit Hilfe der Karhunen-Loeve Transformation (KLT) auf eine orthonormale Basis zu projizieren. Nur eine Teilmenge der sogenannten Principle Control Moves (PCM) wird dabei benötigt, um die Online MPC Optimierung zu parametrieren.

Identifikationsdaten für PCM, die repräsentierende Regelmanöver darstellen, können aus Simulationen oder auch von realen Anlagen stammen. Die Reglerstruktur muss dabei keine MPC-Regelung sein, muss jedoch jedes Manöver regeln können. Für die Identifizierung sowie die Wahl der Anzahl an Regelmanövern werden Prozesswissen oder extensive Simulationen benötigt. Es ist ein analytisches Kriterium gegeben, um iterativ Regelmanöver zu sammeln bis ausreichend Merkmale aufgezeichnet sind. Die Ordnungsreduktion ist in zwei Schritten durchgeführt: Zuerst wird jede PCM auf ein Minimum an Komponenten reduziert. Danach werden die verbleibenden Komponenten vereint und ein zweiter Ordnungsreduktionsschritt analog zum ersten Reduzierungsschritt durchgeführt. Durch die Zwei-Schritt-Reduktion wird sichergestellt, dass eine ausreichende Basis trotz möglicher auftretender Nichtlinearitäten, wie z.B. Sättigung oder weitere Beschränkungen, ausgewählt wird. Eine Stabilitätsanalyse wurde dargestellt, die exponentielle Stabilität für den reduzierten MPC zeigt, wobei sich diese vorerst auf den unbeschränkten Fall konzentriert.

Die Leistungsfähigkeit der beschriebenen Methode ist an einem industriellen Trocknungsprozesses demonstriert worden, an welchem eine Reduktion der Entscheidungsvariablen (von 90 auf 10) ohne signifikante Leistungsfähigkeitsminderung durchgeführt wurde.

Literatur

[1] E.F. Camacho and C. Bordons. *Model Predictive Control*. Springer, 2nd edition, 2004.

[2] D. Ding and X. Gu. Predictive control based on wavelets of second-order linear parameter-constant distributed parameter systems. In *Proceedings of the 5th world congress on intelligent control and automation*, pages 645–649, 2004.

[3] S. Dubljevic and P. D. Christofides. Predictive control of parabolic PDEs with boundary control actuation. *Chemical Engineering Science*, 61(18):6239–6248, 2006.

[4] S. Dubljevic, N.H. El-Farra, P. Mhaskar, and P.D. Christofides. Predictive control of parabolic PDEs with state and control constraints. *International Journal of Robust and Nonlinear Control*, 16:749–772, 2006.

[5] S. Esakkirajan, T. Veerakumar, and P. Navaneethan. Best Basis Selection using Singular Value Decomposition. *Seventh International Conference on Advances in Pattern Recognition*, 2009.

[6] J. Fan and G.A. Dumont. A novel model reduction method for sheet forming processes using wavelet packets. In *Decision and Control, Proceedings of the 40th IEEE Conference on*, volume 5, pages 4820–4825. IEEE, 2001.

[7] Baris Haznedar and Yaman Arkun. Single and multiple property CD control of sheet forming processes via reduced order infinite horizon MPC algorithm. *Journal of Process Control*, 12(1):175 – 192, 2002.

[8] S. Hovland, K. Willcox, and J.T. Gravdahl. MPC for large-scale systems via model reduction and multiparametric quadratic programming. In *Decision and Control, 2006 45th IEEE Conference on*, pages 3418 –3423, 13-15 2006.

[9] K. Kunisch and S. Volkwein. Proper orthogonal decomposition for optimality systems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 42:1–23, 2008.

[10] L. Sirovich. Turbulence and the dynamics of coherent structures, part I-III. *Quaterly of Applied Mathematics*, 45:561–590, 1987.

[11] Wu Zhong Lin, Yao Jiang Zhang, and Er Ping Li. Proper orthogonal decomposition in the generation of reduced order models for interconnects. *IEEE Transactions on Advanced Packaging*, 31(3):627 –636, aug. 2008.

[12] J.L. Lumley. *The structure of inhomogeneous turbulence*, pages 166–178. Nauka, Moscow, 1967.

[13] J.M. Maciejowski. *Predictive Control with Constraints*. Pearson, 1st edition, 2001.

[14] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 36(6):789 – 814, 2000.

[15] K. Ogata. *Discrete-Time Control Systems*. Prentice Hall, 2nd edition, 1995.

[16] Radhakant Padhi and Sk. Faruque Ali. An account of chronological developments in control of distributed parameter systems. *Annual Reviews in Control*, 33(1):59–68, April 2009.

[17] S.S. Rao. *Engineering Optimization*. Wiley, 4th edition, 2009.

[18] Osvaldo J. Rojas, Graham C. Goodwin, Maria M. Seron, and Arie Feuer. An SVD based strategy for receding horizon control of input constrained linear systems. *Int. J. Robust Nonlinear Control*, 14:1207–1226, 2004.

[19] A. Schuster and M. Kozek. Constrained model predictive control implementation in an industrial drying process. In *IEEE International Conference on Computational Cybernetics, 2009. ICCC 2009*, pages 191–196, 2009.

[20] Alexander Schuster, M. Kozek, B. Voglauer, and A. Voigt. Grey-Box Modelling of a Viscose-Fibre Drying Process. *Mathematical and Computer Modelling of Dynamical Systems*, page (accepted), 2011.

[21] H. Shang, J.F. Forbes, and M. Guay. Characteristics-based model predictive control of distributed parameter systems. In *Proceedings of the American Control Conference*, pages 4383–4388, 2002.

[22] G. Strang. *Linear Algebra and its Applications*. Thomson Brooks/Cole, 3rd edition, 1998.

[23] Petter Tondel and Tor A. Johansen. Complexity reduction in explicit linear model predictive control. In *15th Triennial World Congress of the International Federation of Automatic Control*, 2002.

[24] Jeremy G. VanAntwerp, Andrew P. Featherstone, Richard D. Braatz, and Babatunde A. Ogunnaike. Cross-directional control of sheet and film processes. *Automatica*, 43(2):191 – 211, 2007.

[25] L. Wang. *Model Predictive Control System Design and Implementation Using* MATLAB. Springer, 1st edition, 2009.

[26] K. Willcox and J. Peraire. Balanced model reduction via the proper orthogonal decomposition. *AIAA Journal*, 40:2323–2330, 2002.

[27] Oh Sang Woo, Alvis Salenieks, and Michael Mavrovouniotis. A positive linear decomposition for identifying patterns in dynamic process measurements. *at-Automatisierungstechnik*, pages 486–494, 10/2006 2006.

[28] D. Zheng, K.A. Hoo, and M.J. Piovoso. Finite dimensional modeling and control of distributed parameter systems. In *Proceedings of the American Control Conference*, pages 4377–4382, 2002.

Ein nichtlineares System zum Lösen von Sattelpunktproblemen und Linearen Programmen

H.-B. Dürr, S. Zeng und C. Ebenbauer
Universität Stuttgart, IST, Pfaffenwaldring 9, 70550 Stuttgart
{hans-bernd.duerr*, shen.zeng, ce}@ist.uni-stuttgart.de

Zusammenfassung

Die Arbeit stellt ein nichtlineares System vor, das zum Lösen von konvexen Optimierungsproblemen konstruiert wurde. Unter Ausnutzung der Sattelpunkteigenschaft der Lagrange-Funktion kann globale Konvergenz unter schwachen Voraussetzungen gezeigt werden. Der Fokus der Arbeit liegt in der Anwendung der Methode auf lineare Programme, die eine wichtige Klasse innerhalb der Optimierungsprobleme bilden und bei vergleichbaren Methoden häufig Konvergenzprobleme hervorrufen.

1 Einleitung

In dieser Arbeit werden konvexe Optimierungsprobleme der Form

$$\begin{aligned} &\inf_x f(x) \\ &\text{s.t. } g_i(x) \leq 0, i = 1, \ldots, m \end{aligned} \tag{1}$$

betrachtet. Hierbei sind $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$ konvexe Funktionen. Dem Optimierungsproblem (1) wird die Lagrange-Funktion

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \tag{2}$$

mit $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m_+$ zugeordnet. Aus der Optimierungstheorie ist bekannt, dass ein Sattelpunkt $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m_+$ von (2) eine Lösung $x^* \in \mathbb{R}^n$ des Optimierungsproblems (1) liefert. Hierbei heißt ein Punkt $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m_+$ Sattelpunkt von L, wenn für alle $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m_+$ gilt:

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*).$$

*Korrespondenz bitte an diese Adresse

Unter der Voraussetzung (Slater Bedingung), dass ein Vektor $\tilde{x} \in \mathbb{R}^n$ mit der Eigenschaft $g_i(\tilde{x}) < 0$ existiert, gilt sogar die Umkehrung, d.h. für jede Lösung von (1) existiert ein nichtnegativer Lagrange-Multiplikator $\lambda^* \in \mathbb{R}^m_+$, so dass (x^*, λ^*) ein Sattelpunkt von (2) ist.

Ziel dieser Arbeit ist eine neue Methode vorzustellen, welche Optimierungsprobleme der Form (1) löst. Die Grundidee basiert darauf ein nichtlineares System so zu konstruieren, dass jede Trajektorie des Systems zu einem Sattelpunkt von (2) und somit zu einer Lösung von (1) konvergiert. Insbesondere soll in dieser Arbeit die Konvergenz für den wichtigen Speziallfall der linearen Programmierung, d.h. wenn f, g_i lineare (affine) Funktionen sind, untersucht werden.

Im folgenden Absatz wird der in dieser Arbeit vorgeschlagene Ansatz motiviert und einige verwandte Ergebnisse aus der Literatur diskutiert. Ein Gebiet auf dem Sattelpunktmethoden häufig angewendet werden, ist die Optimierung von Netzwerkprotokollen. Oft werden diese Methoden eingesetzt wenn es darum geht Netzwerkressourcen großer verteilter Kommunikationsnetzwerke optimal aufzuteilen. Solche Netzwerke, wie sie das Internet darstellt, sind häufig durch lokale Zustandsinformationen und große Zeitverzögerungen bei der Informationsübertragung charakterisiert. Die Optimierungsprobleme können in diesem Fall so formuliert werden, dass die Lagrange-Multiplikatoren als Kosten, die bei der Nutzung einer Ressource anfallen, interpretiert werden können. Aus der Berechnung der Lagrange-Multiplikatoren, die nur bei Sattelpunktmethoden explizit anfallen, ergeben sich praktische Methoden um die Schwierigkeiten zu bewältigen, die diese Randbedingungen mit sich bringen. Für eine nähere Erläuterung sei auf [9, 14] verwiesen. Die Idee Differentialgleichungen zum Lösen von Sattelpunktproblemen einzusetzen geht zurück auf K. J. Arrow, L. Hurwicz and H. Uzawa sowie G. W. Brown and J. von Neumann, siehe [1, 6, 11]. Das von Arrow, Hurwicz, Uzawa vorgeschlagene System (AHU-System) lautet:

$$\dot{x} = -\left(\frac{\partial L(x,\lambda)}{\partial x}\right)^\top = -\nabla f(x) - \sum_{i=1}^{m} \lambda_i \nabla g_i(x) \tag{3a}$$

$$\dot{\lambda}_i = \mathcal{P}(\lambda_i, g_i(x)), i = 1, \ldots, m, \tag{3b}$$

mit dem Projektionsoperator $\mathcal{P}(\lambda_i, g_i(x)) := 0$ wenn $\lambda_i = 0$ und $g_i(x) < 0$, und $\mathcal{P}(\lambda_i, g_i(x)) := g_i(x)$ sonst. Um zu einem Sattelpunkt der Lagrange-Funktion zu konvergieren, wird mit Hilfe des negativen Gradienten bezüglich der „primalen“ Variablen x die Lagrange-Funktion minimiert wohingegen sie mit dem positiven Gradienten bezüglich der „dualen“ Variablen λ maximiert wird. Der Projektionsoperator schaltet den Einfluss des Gradienten aus, sobald die Lagrange-Multiplikatoren drohen negativ zu werden.

Ein ähnliche Herangehensweise ist in [2, 10] zu finden. Das nichtlineare System ist von der Form $\dot{x} \in BR_1(y) - x, \dot{y} \in BR_2(x) - y$, wobei $BR_i(\cdot)$ die *best response* bezüglich einer Taktik des Gegenspielers bezeichnet. $BR_i(\cdot)$ liefert die Lösung eines Optimierungsproblems, die zu jedem Zeitpunkt ermittelt werden muss. Das nichtlineare System, das das Gegenspiel von x und y ausnutzt und somit zu einem Sattelpunkt eines Nullsummenspiels konvergiert wurde für spieltheoretische Untersuchungen entwickelt und hat nur einen eingeschränkten Anwendungsbereich bezüglich allgemeinen Optimierungsproblemen.

In vielen Anwendungen kommt das zuvor genannte AHU-System und dessen Varianten zum Einsatz, siehe [9, 12]. In dieser Arbeit soll eine neue, alternative Methode vorgestellt werden. Ebenso wie die zuvor genannten Methoden, nutzt das System die Sattelpunkteigenschaft der Lagrange-Funktion aus, hat jedoch einige Vorteile gegenüber den zuvor genannten Methoden.

Eine der Voraussetzungen für die garantierte Konvergenz des AHU-System ist eine *strikt* konvexe Lagrange-Funktion in x. Diese Annahme wird von linearen Programmen nicht erfüllt, weshalb das AHU-System nicht direkt auf diese Problemklasse angewendet werden kann. Im Gegensatz dazu, ist die hier vorgestellte Methode auch für lineare Programme geeignet. Desweiteren enthält das nichtlineare System im Vergleich zum *best response*- und dem AHU-System, weder einen Projektionsoperator noch den $BR_i(\cdot)$-Operator, sondern setzt sich nur aus den Funktionen f, g_i und deren Ableitungen zusammen. Deshalb ist das System einfach zu implementieren und kann beispielsweise auf semi-definite Programme erweitert werden.

Wie in der Netzwerkoptimierung als auch in anderen Anwendungsgebieten, kommen Methoden in Form von *zeitkontinuierlichen* Systemen zum Einsatz. Obwohl eine Implementierung auf einem Rechner in diskreter Form realisiert wird, ergeben sich aus mathematischer Sicht viele Vorteile bei einer zeitkontinuierlichen Form. Beispielsweise können Hilfsmittel aus der Systemtheorie und Regelungstechnik zur Untersuchung der Stabilität und Robustheit verwendet werden. Darüber hinaus ist ein zeitkontinuierliches System unabhängig von der Wahl der Diskretisierung. Bei einem zeitdiskreten System der Form $x_{k+1} = x_k + \alpha f(x_k)$ muss dagegen die Schrittweite in die Analyse mit einbezogen werden.

Die Arbeit ist wie folgt gegliedert. In Abschnitt 2 werden die notwendigen mathematischen Begriffe und Grundlagen aufgeführt, auf die die Inhalte der folgenden Abschnitte aufbauen. In Abschnitt 3 werden zwei nichtlineare Systeme für strikt konvexe Optimierungsprobleme eingeführt. Dieser Abschnitt beruht weitestgehend auf [7]. In Abschnitt 4 wird die Anwendung eines der Systeme auf lineare Programme behandelt, sowie erstmalig ein Beweis der globalen Stabilität geliefert. Abschnitt 5 enthält zur Verdeutlichung ein numerisches Beispiel. In Abschnitt 6 wird eine Zusammenfassung und ein Ausblick gegeben.

2 Mathematische Begriffe und Grundlagen

In diesem Abschnitt sollen einige Bezeichnungen und Sachverhalte aus der konvexen Optimierung dargestellt werden.

Die euklidische Norm eines Vektors v wird mit $||v|| = \sqrt{\sum_{i=1}^{n} v_i^2}$ bezeichnet. Eine Matrix A ist positiv semi-definit (definit), kurz $A \geq 0$ ($A > 0$), wenn sie symmetrisch $A = A^\top$ und alle ihre Eigenwerte nichtnegativ (positiv) sind. Eine Matrix A heißt Hurwitz, wenn alle ihre Eigenwerte in der komplexen offenen linken Halbebene liegen. Die Menge der nichtnegativen (strikt positiven), reellen Vektoren in $\mathbb{R}^n$ wird mit $\mathbb{R}^n_+$ ($\mathbb{R}^n_{++}$) bezeichnet. Der Operator $\text{diag}(v) : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ ordnet einem Vektor v eine Diagonalmatrix zu, wobei die Komponenten v_i des Vektors den Hauptdiagonalelementen entsprechen.

Sei $f : \mathbb{R}^n \to \mathbb{R}$, $L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ eine C^1- bzw. C^2-Funktion, so bezeichnen

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x}\right)^\top = \left[\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n}\right]^\top, \tag{4}$$

$$\nabla L(x,\lambda) = \left(\frac{\partial L(x,\lambda)}{\partial x}\right)^\top = \left[\frac{\partial L(x,\lambda)}{\partial x_1}, \ldots, \frac{\partial L(x,\lambda)}{\partial x_n}\right]^\top, \tag{5}$$

$$\nabla^2 L(x,\lambda) = \frac{\partial^2 L(x,\lambda)}{\partial x^2}. \tag{6}$$

Eine Funktion $f \in C^2$ ist genau dann konvex, wenn $\nabla f(x)^\top (y - x) \leq f(y) - f(x)$ oder $\nabla^2 f(x) \geq 0$ gilt. Ferner ist f strikt konvex genau dann wenn für alle $x \neq y$, $\nabla f(x)^\top (y - x) < f(y) - f(x)$, und $\nabla^2 f(x) > 0$ impliziert für alle x ebenfalls dass f eine strikt konvexe Funktion ist.

Ein bekanntes Ergebnis aus der konvexen Optimierung besagt, dass ein Sattelpunkt von (2) immer auch eine Lösung des Problems (1) ist, siehe [4, 8, 13].

Theorem 1. *Sei (x^*, λ^*) ein Sattelpunkt von (2). Dann ist x^* Lösung von (1).*

Die Umkehrung gilt im Allgemeinen nur wenn die Slater Bedingung erfüllt ist.

Theorem 2. *Sei die Slater Bedingung erfüllt und seien f, g_i, $i = 1, \ldots, m$ konvex. Dann gilt: x^* ist eine Lösung von (1) genau dann wenn ein λ^* existiert, so dass (x^*, λ^*) ein Sattelpunkt von (2) in $\mathbb{R}^n \times \mathbb{R}^m_+$ ist. D.h. $(x^*, \lambda^*) = \arg\min_{x \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^m_+} L(x, \lambda)$.*

Der Satz von Kuhn, Karush und Tucker liefert notwendige und hinreichende Optimalitätsbedingungen.

Theorem 3. *Sei die Slater Bedingung erfüllt, und seien $f, g_i \in C^1$, $i = 1, \ldots, m$ und konvex. Dann gilt: x^* ist eine Lösung von (1) genau dann wenn ein $\lambda^* \in \mathbb{R}^m_+$ existiert, so dass die folgenden drei Bedingungen (KKT-Bedingungen) erfüllt sind:*

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0, \tag{7a}$$

$$\lambda_i^* g_i(x^*) = 0, \ i = 1, \ldots, m, \tag{7b}$$

$$g_i(x^*) \leq 0, \ i = 1, \ldots, m. \tag{7c}$$

Ein weiteres zentrales Konzept der konvexen Optimierung ist die Dualität. Dem Optimierungsproblem (1) wird das duale Problem der Form

$$\sup_{\lambda \in \mathbb{R}^m_+} g(\lambda) \tag{8}$$

zugeordnet, wobei

$$g(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda) \tag{9}$$

als die duale Funktion bezeichnet wird. Als Dualitätslücke wird die Differenz zwischen dem optimalen Wert des primalen Problems (1) und des dualen Problems (8) bezeichnet.

Diese Dualitätslücke ist immer nichtnegativ und verschwindet, wenn die Slater Bedingung erfüllt ist. Im Falle von linearen Programmen

$$\begin{aligned} &\inf_x c^\top x \\ &\text{s.t. } a_i^\top x \leq b_i, i = 1, \ldots, m \end{aligned} \tag{10}$$

mit $c, a_i \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, ist die Existenz eines Sattelpunktes immer garantiert, sofern es eine Lösung des Problems gibt. Die Lagrange-Funktion zu (10) ist gegeben durch:

$$L(x, \lambda) = c^\top x + \sum_{i=1}^{m} \lambda_i (a_i^\top x - b_i) \tag{11}$$

und das duale Problem kann explizit angegeben werden:

$$\begin{aligned} &\sup_\lambda \; -b^\top \lambda \\ &\text{s.t. } \sum_{i=1}^{m} \lambda_i a_i = -c, \\ &\qquad \lambda \geq 0. \end{aligned} \tag{12}$$

Bei linearen Programmen verschwindet die Dualitätslücke immer, sofern das primale oder das duale Problem eine Lösung besitzt. In diesem Fall gilt

$$c^\top x^* = -b^\top \lambda^*. \tag{13}$$

Ist der Sattelpunkt der Lagrange-Funktion für ein lineares Programm eindeutig, so ist dies auch einzige Lösung des primalen bzw. dualen Optimierungsproblems, siehe [5], S. 190 und [15], S. 161.

Lemma 1. *Sei (x^*, λ^*) der einzige Sattelpunkt von (11). Dann gilt: x^* ist die Lösung von (10) und λ^* ist die Lösung von (12). Ferner gilt die strikte Komplementarität, d.h. $\lambda_i^* = 0 \Leftrightarrow a_i^\top x^* - b_i \neq 0$ und $a_i^\top x^* - b_i = 0 \Leftrightarrow \lambda_i^* \neq 0$.*

Die KKT-Bedingungen (7a) – (7c) für lineare Programme sind in folgender Form gegeben:

$$c + \sum_{i=1}^{m} \lambda_i^* a_i = 0, \tag{14a}$$

$$\lambda_i^* (a_i^\top x^* - b_i) = 0, \; i = 1, \ldots, m, \tag{14b}$$

$$a_i^\top x^* - b_i \leq 0, \; i = 1, \ldots, m. \tag{14c}$$

Weiterhin wird folgendes Lemma gebraucht.

Lemma 2. *Sei $A \in \mathbb{R}^{n \times n}$ und $B \in \mathbb{R}^{n \times m}$, dann ist*

$$C = \begin{bmatrix} A & B \\ -B^\top & 0 \end{bmatrix} \tag{15}$$

Hurwitz wenn $A < 0$ und $\text{Rang}(B) = m$.

Der Beweis für Lemma 2 ist in [3] auf S. 232 zu finden.

3 Konvexe Programme

In diesem Abschnitt werden zwei nichtlineare System vorgestellt, die speziell auf strikt konvexe Optimierungsprobleme angewendet werden können. Das erste System gleicht dem AHU-System bis auf den Projektionsoperator in der λ-Dynamik (3b). An dessen Stelle werden die Nebenbedingungen g_i mit λ_i multipliziert. Wenn λ in $\mathbb{R}^m_{++}$ initialisiert wurde, kann leicht festgestellt werden, dass auf diese Weise die Lagrange-Multiplikatoren immer nichtnegativ bleiben. Das zweite System beinhaltet zusätzlich noch einen Kompensationsterm, der vor allem bei linearen Programmen eine wichtige Rolle spielt.

Das erste System hat die folgende Form:

$$\dot{x} = -\nabla L(x,\lambda) = -\nabla f(x) - \sum_{i=1}^{m} \lambda_i \nabla g_i(x) \tag{16a}$$

$$\dot{\lambda}_i = \lambda_i g_i(x), i = 1,\dots,m. \tag{16b}$$

Es ist leicht festzustellen, dass alle Ruhelagen des Systems (16) die KKT-Bedingungen (7a) und (7b) erfüllen. Unter folgenden Annahmen kann Konvergenz zur x-Komponente des Sattelpunktes von L und damit zur Lösung von (1) garantiert werden:

(A1) Die Lagrange-Funktion $L \in C^1$ ist strikt konvex in x.

(A2) Es existiert mindestens ein Sattelpunkt (x^*,λ^*) von (2).

Theorem 4. *Seien die Annahmen (A1) und (A2) erfüllt. Dann konvergiert jede in $\mathbb{R}^n \times \mathbb{R}^m_{++}$ initialisierte Lösung $(x(t),\lambda(t))$ des Systems* (16) *zu einem Sattelpunkt (x^*,λ^*) der Lagrange-Funktion* (2)*. Ferner ist x^* eindeutig und jeder Sattelpunkt ist stabil.*

Beweis. Der Beweis ist in zwei Schritte unterteilt. Im ersten Schritt wird zunächst gezeigt, dass mit (A1) und (A2) für zwei Sattelpunkte (x_a^*,λ_a^*) und (x_b^*,λ_b^*) gilt: $x_a^* = x_b^*$. Dies kann durch einen Widerspruch gezeigt werden: Sei $x_a^* \neq x_b^*$, so muss aufgrund der strikten Konvexität von L bezüglich x gelten:

$$L(x_a^*,\lambda_b^*) \leq L(x_a^*,\lambda_a^*) < L(x_b^*,\lambda_a^*) \tag{17}$$

und

$$L(x_b^*,\lambda_a^*) \leq L(x_b^*,\lambda_b^*) < L(x_a^*,\lambda_b^*). \tag{18}$$

Also muss $L(x_b^*,\lambda_a^*) < L(x_b^*,\lambda_a^*)$ sein. Das ist offensichtlich ein Widerspruch und daraus folgt $x_a^* = x_b^*$.

Für den zweiten Schritt wird die Lyapunov-Funktion betrachtet:

$$V(x,\lambda) = \frac{1}{2}(x-x^*)^\top(x-x^*) + \sum_{i\in A_c} \lambda_i - \lambda_i^* - \lambda_i^* \ln\frac{\lambda_i}{\lambda_i^*} + \sum_{i\in I_c} \lambda_i \tag{19}$$

mit $A_c = \{i : \lambda_i^* \neq 0\}$ und $I_c = \{i : \lambda_i^* = 0\}$. Es kann gezeigt werden, dass die Lyapunov-Funktion positiv definit und radial unbeschränkt auf $\mathbb{R}^n \times \mathbb{R}^m_{++}\backslash(x^*,\lambda^*)$ ist.

Die Lie-Ableitung von V entlang der Trajektorien von (16) ist

$$\begin{aligned}\dot{V} &= -(x-x^*)^\top \nabla L(x,\lambda) + \sum_{i=1}^{m}(\lambda_i - \lambda_i^*)g_i(x) \\ &= -(x-x^*)^\top \nabla L(x,\lambda) + \frac{\partial L(x,\lambda)}{\partial \lambda}(\lambda - \lambda^*).\end{aligned} \tag{20}$$

Mit (A1) ist L strikt konvex in x und konkav in λ. Somit gilt

$$\begin{aligned}\dot{V} &\leq L(x^*,\lambda) - L(x,\lambda) + L(x,\lambda) - L(x,\lambda^*) \\ &= L(x^*,\lambda) - L(x,\lambda^*).\end{aligned} \tag{21}$$

Da es sich bei (x^*,λ^*) um einen Sattelpunkt von L handelt, gilt $L(x^*,\lambda) \leq L(x,\lambda^*)$, somit ist $\dot{V} \leq 0$ und damit ist jeder Sattelpunkt (x^*,λ^*) stabil. Für $x \neq x^*$ gilt aufgrund der strikten Konvexität von L bezüglich x, dass $L(x^*,\lambda) < L(x,\lambda^*)$.

Betrachtet wird nun die Menge

$$M = \left\{(x,\lambda) \in \mathbb{R}^n \times \mathbb{R}_+^m : x = x^*; \lambda_i g_i(x^*) = 0, i = 1,\ldots,m\right\}. \tag{22}$$

Mit $\dot{V} = 0$ folgt nun, dass $L(x^*,\lambda) = L(x,\lambda^*)$ und damit

$$f(x^*) + \sum_{i=1}^{m}\lambda_i g_i(x^*) = f(x) + \sum_{i=1}^{m}\lambda_i^* g_i(x). \tag{23}$$

Da L strikt konvex in x und x^* eindeutig ist, folgt aus (23), dass $x = x^*$. Desweiteren folgt aus (23), dass $\sum_{i=1}^{m}\lambda_i g_i(x^*) = 0$ und da $g_i(x^*) \leq 0$ ist, muss damit auch gelten: $\lambda_i g_i(x^*) = 0, i = 1,\ldots,m$. Also konvergieren $(x(t),\lambda(t))$ zu M. Da alle Elemente in M die KKT-Bedingungen (7a) – (7c) erfüllen, enthält M genau die Sattelpunkte von L. □

Die Systeme (3) und (16) können nicht direkt auf lineare Programme angewendet werden, da im Allgemeinen keine Konvergenz garantiert werden kann. Die Linearisierung der Systeme, die im Falle des AHU-Systems (3) nur unter der Annahme, dass $\lambda^* \in \mathbb{R}_{++}^m$ möglich ist, gibt keine Aussage über die Stabilität des Sattelpunktes (x^*,λ^*). Es wird nun ein System vorgestellt, das auch für lineare Programme anwendbar ist.

Das zweite System hat die folgende Form:

$$\dot{x} = -\nabla L(x,\lambda) - (\nabla^2 L(x,\lambda))^{-1}\left(\sum_{i=1}^{m}\lambda_i g_i(x)\nabla g_i(x)\right) \tag{24a}$$

$$\dot{\lambda}_i = \lambda_i g_i(x), \quad i = 1,\ldots,m. \tag{24b}$$

Genauso wie beim ersten System (16) erfüllen alle Ruhelagen des Systems (24) die KKT-Bedingungen (7a) und (7b). Im Folgenden wird die Grundidee der Funktionsweise erläutert. Die Aufgabe der x-Dynamik ist es L zu minimieren. Wäre λ konstant, würde

der Gradient von L ausreichen um dies zu bewirken. Da λ aber selbst eine Dynamik besitzt, wird diese durch den zusätzlichen Term $-(\nabla^2 L(x,\lambda))^{-1}\left(\sum_{i=1}^m \lambda_i g_i(x)\nabla g_i(x)\right)$ kompensiert. Die Minimierung von L erfolgt indem die Lyapunov-Funktion

$$V_1(x,\lambda) = \nabla L(x,\lambda)^\top \nabla L(x,\lambda) \tag{25}$$

abnimmt. Dies ist daran zu erkennen, dass die Lie-Ableitung entlang des Vektorfeldes von (24)

$$\begin{aligned}\dot{V}_1 =& \nabla L(x,\lambda)^\top \left(\nabla^2 L(x,\lambda)\dot{x} + \frac{\partial \nabla L(x,\lambda)}{\partial \lambda}\dot{\lambda}\right) \\ =& -\nabla L(x,\lambda)^\top \nabla^2 L(x,\lambda)\nabla L(x,\lambda)\end{aligned} \tag{26}$$

negativ semi-definit ist. Sofern die Lösung $(x(t),\lambda(t))$ existiert, konvergiert sie zu der Menge

$$N = \{(x,\lambda) \in \mathbb{R}^n \times \mathbb{R}^m_+ : x = \bar{x}(\lambda), \nabla L(\bar{x}(\lambda),\lambda) = 0\}. \tag{27}$$

Hierbei bezeichnet $\bar{x}(\lambda)$ den Minimierer von $L(x,\lambda)$ bezüglich x zu einem gegebenen λ. Die zugrundeliegende Idee ist in Abb. 1 illustriert.

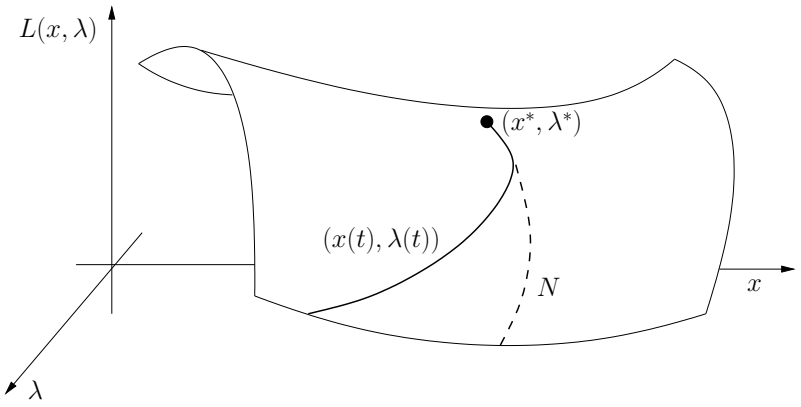


Abbildung 1: Konvergenz zur Menge N

Die Aufgabe der λ-Dynamik ist es die duale Funktion

$$V_2(\lambda) = L(\bar{x}(\lambda),\lambda) \tag{28}$$

auf der Menge N zu maximieren. Dies lässt sich daran erkennen, dass die Lie-Ableitung entlang des Vektorfeldes von (24), ausgewertet auf N

$$\begin{aligned}\dot{V}_2 &= \left(\nabla L(x,\lambda)|_{\bar{x}(\lambda)}\frac{\partial \bar{x}(\lambda)}{\partial \lambda} + \frac{\partial L(x,\lambda)}{\partial \lambda}\right)\dot{\lambda} \\ &= \sum_{i=1}^m g_i(\bar{x}(\lambda))^2\lambda_i\end{aligned} \tag{29}$$

positiv semi-definit ist. Die Schwierigkeit liegt nun darin die Existenz der Lösung sowie die Konvergenz zu einem Sattelpunkt von L zu garantieren. Dies kann unter folgenden Voraussetzungen gezeigt werden:

(A3) $f, g_i, i = 1, \ldots, m$ sind analytisch, strikt konvex und es gilt $\nabla^2 f(x) > 0$ für alle $x \in \mathbb{R}^n$.

(A4) Die Funktion $f_c(x) = f(x) - c^\top x, g_i,\ i = 1, \ldots, m$ besitzt ein Minimum für jedes $c \in \mathbb{R}^n$.

(A5) Die Slater Bedingung ist erfüllt.

Theorem 5. *Seien die Annahmen (A3) bis (A5) erfüllt. Dann konvergiert jede in $\mathbb{R}^n \times \mathbb{R}^m_{++}$ initialisierte Lösung $(x(t), \lambda(t))$ des Systems (24) zu einem Sattelpunkt (x^*, λ^*) der Lagrange-Funktion (2).*

Beweis. Für einen ausführlichen Beweis sei auf [7] verwiesen. □

Beide Systeme sind nur für strikt konvexe Probleme einsetzbar. Ein Vorteil der beiden Systeme ist, dass der Projektionsoperator eliminiert wurde, was zu einem glatten Vektorfeld führt. Wie sich weiter zeigt, hat das zweite System den Vorteil, dass es auch für nicht strikt konvexe Probleme, zu denen auch lineare Programme gehören, eingesetzt werden kann.

4 Lineare Programme

Eine direkte Anwendung von (24) auf lineare Programme ist nicht möglich, da $\nabla^2 L(x, \lambda) = 0$ ist. In diesem Abschnitt soll darauf eingegangen werden, wie das System für lineare Programme (10) angepasst werden kann. Die Idee ist einfach die Matrix $(\nabla^2 L(x, \lambda))^{-1}$ wird durch die Einheitsmatrix zu ersetzen.

Damit ergibt sich das System:

$$\dot{x} = -c - \sum_{i=1}^{m} \lambda_i a_i - \sum_{i=1}^{m} \lambda_i a_i (a_i^\top x - b_i) \tag{30a}$$

$$\dot{\lambda}_i = \lambda_i (a_i^\top x - b_i), i = 1, \ldots, m. \tag{30b}$$

Erstaunlicherweise führt diese einfache Modifikation auf keine Einschränkung im Konvergenzverhalten. Ganz im Gegenteil, das System (30) konvergiert immer zu dem Sattelpunkt von L und somit zu der Lösung des linearen Programms, sofern die Lösung des linearen Programms eindeutig ist. Unter folgender Annahme kann Konvergenz gezeigt werden:

(A6) Es existiert genau ein Sattelpunkt (x^*, λ^*) von (11).

Theorem 6. *Sei Annahme (A6) erfüllt. Dann konvergiert jede in $\mathbb{R}^n \times \mathbb{R}^m_{++}$ initialisierte Lösung $(x(t), \lambda(t))$ des Systems (30) zu dem Sattelpunkt der Lagrange-Funktion (11). Ferner ist der Sattelpunkt exponentiell stabil.*

Beweis. Der Beweis ist in drei Schritte unterteilt. Im ersten Schritt wird die Stabilität, im zweiten Schritt die asymptotische Stabilität und im dritten Schritt die exponentielle Stabilität des Sattelpunktes bewiesen.

Schritt 1: Es wird folgende positiv definite und radial unbeschränkte Lyapunov-Funktion betrachtet:

$$V(x,\lambda) = \frac{1}{2}(x-x^*)^\top(x-x^*) + \sum_{i\in A_c} \lambda_i - \lambda_i^* - \lambda_i^* \ln\frac{\lambda_i}{\lambda_i^*} + \sum_{i\in I_c}\lambda_i - \sum_{i=1}^m \lambda_i(a_i^\top x^* - b_i) \quad (31)$$

mit $A_c = \{i : \lambda_i^* \neq 0\}$ und $I_c = \{i : \lambda_i^* = 0\}$. Dann ist die Lie-Ableitung von V entlang des Vektorfeldes von (30)

$$\begin{aligned}
\dot V =& (x-x^*)^\top\dot x + \sum_{i\in A_c}(\lambda_i-\lambda_i^*)\frac{\dot\lambda_i}{\lambda_i} + \sum_{i\in I_c}\dot\lambda_i - \sum_{i=1}^m \dot\lambda_i(a_i^\top x^* - b_i)\\
=& -(x-x^*)^\top\nabla L(x,\lambda) + \sum_{i=1}^m(\lambda_i-\lambda_i^*)(a_i^\top x - b_i) - (x-x^*)^\top\sum_{i=1}^m\lambda_i a_i(a_i^\top x - b_i)\\
& - \sum_{i=1}^m \lambda_i(a_i^\top x - b_i)(a_i^\top x^* - b_i) \qquad (32)\\
=& L(x^*,\lambda) - L(x,\lambda^*) - \sum_{i=1}^m\lambda_i(a_i^\top x)^2 + \sum_{i=1}^m\lambda_i a_i^\top x b_i + \sum_{i=1}^m\lambda_i a_i^\top x^* a_i^\top x\\
=& L(x^*,\lambda) - L(x,\lambda^*) - \sum_{i=1}^m\lambda_i(a_i^\top x - b_i)^2.
\end{aligned}$$

Da es sich bei (x^*,λ^*) um den Sattelpunkt von L handelt, gilt $L(x^*,\lambda) \leq L(x,\lambda^*)$, somit ist $\dot V \leq 0$ und damit ist der Sattelpunkt (x^*,λ^*) stabil.

Schritt 2: Um nun asymptotische Stabilität des Sattelpunktes festzustellen wird zunächst gezeigt, dass $(x(t),\lambda(t))$ in die Menge

$$M = \left\{(x,\lambda)\in\mathbb{R}^n\times\mathbb{R}^m_+ : \lambda_i(a_i^\top x - b_i) = 0, i = 1,\dots,m; \right.$$
$$\left. \sum_{i=1}^m \lambda_i a_i + c = 0; b^\top\lambda = b^\top\lambda^*\right\} \qquad (33)$$

konvergiert.

Mit $\dot V = 0$ folgt nun zuerst, dass $\lambda_i(a_i^\top x - b_i) = 0, i = 1,\dots,m$ und daraus wiederum, dass die Grenzmenge $\Omega(x(t),\lambda(t))$ von $(x(t),\lambda(t))$ initialisiert in $\mathbb{R}^n\times\mathbb{R}^m_{++}$ eine Untermenge von $M_1 = \{(x,\lambda) : \lambda_i(a_i^\top x - b_i) = 0, i = 1,\dots,m\}$ ist. Ferner ist der Sattelpunkt (x^*,λ^*) stabil und somit ist $(x(t),\lambda(t))$ beschränkt, also ist die invariante Menge $\Omega(x(t),\lambda(t))$ kompakt. Zusätzlich muss $\Omega(x(t),\lambda(t))$ auch eine Untermenge von $M_2 = \{(x,\lambda) : \sum_{i=1}^m\lambda_i a_i = -c\}$ sein. Dies kann durch einen Widerspruch gezeigt werden: Sei $(x^0(t),\lambda^0(t))$ eine in $\Omega(x(t),\lambda(t))\backslash M_2$ initialisierte Trajektorie. Da $\Omega(x(t),\lambda(t)) \subseteq M_1$ und $\Omega(x(t),\lambda(t))$ invariant ist, gilt also $\dot x^0(t) = -c - \sum_{i=1}^m\lambda_i^0(t)a_i$ und $\dot\lambda_i^0(t) = 0, i = 1,\dots,m$. Also ist $\lambda^0(t) = \lambda^0(0)$ und damit $\dot x^0(t) = -c - \sum_{i=1}^m\lambda_i^0(t)a_i = -c - \sum_{i=1}^m\lambda_i^0(0)a_i \neq 0$. Dies widerspricht der Beschränktheit von $\Omega(x(t),\lambda(t))$. Also ist $\Omega(x(t),\lambda(t)) \subseteq M_1\cap M_2$.

Ferner folgt aus $\dot{V} = 0$, dass $L(x^*, \lambda) = L(x, \lambda^*)$ also

$$c^\top x^* + \sum_{i=1}^{m} \lambda_i (a_i^\top x^* - b_i) = c^\top x + \sum_{i=1}^{m} \lambda_i^* (a_i^\top x - b_i). \tag{34}$$

In $\Omega(x(t), \lambda(t))$ gilt $\sum_{i=1}^{m} \lambda_i a_i + c = 0$ und mit (14a) folgt aus (34) dass $b^\top \lambda = b^\top \lambda^*$. Mit dieser letzten Eigenschaft gilt also $\Omega(x(t), \lambda(t)) \subseteq M$. Demnach konvergiert also $(x(t), \lambda(t))$ in die Menge M.

Es wird nun gezeigt, dass M nur den Sattelpunkt (x^*, λ^*) enthält. Sei $(\tilde{x}, \tilde{\lambda}) \in M$. Unter der Annahme (A6) und mit Lemma 1 ist die Lösung des dualen Problems (12) eindeutig. Somit ist $\tilde{\lambda}$ eine zulässige Lösung des dualen Problems und es gilt $b^\top \tilde{\lambda} = b^\top \lambda^*$. Aufgrund der Eindeutigkeit folgt dass $\tilde{\lambda} = \lambda^*$ ist. Es lässt sich zeigen (s. [15], S. 63), dass wenn x^* die eindeutige Lösung von (10) ist und λ^* die eindeutige Lösung von (12) ist, dass das Gleichungssystem $a_i^\top \tilde{x} - b_i = 0, i \in A_c$ die Lösung $\tilde{x} = x^*$ eindeutig bestimmt[1]. Also enthält M unter der Annahme (A6) nur den Sattelpunkt. Somit ist dieser asymptotisch stabil.

Schritt 3: Um die exponentielle Stabilität des Systems zu beweisen, wird das System linearisiert. Die Linearisierung des Systems (30) um (x^*, λ^*) hat die Form

$$\begin{bmatrix} \dot{\Delta x} \\ \dot{\Delta \lambda} \end{bmatrix} = \Gamma \begin{bmatrix} \Delta x \\ \Delta \lambda \end{bmatrix}. \tag{35}$$

Sei $\Delta\lambda_A = \left[\ldots, \Delta\lambda_i, \ldots\right]$, $i \in A_c$ und analog $\Delta\lambda_{\bar{A}} = \left[\ldots, \Delta\lambda_i, \ldots\right]$, $i \in I_c$. Dann ist (35) ein System in Kaskadenform

$$\begin{bmatrix} \dot{\Delta x} \\ \dot{\Delta \lambda}_A \end{bmatrix} = \tilde{\Gamma} \begin{bmatrix} \Delta x \\ \Delta \lambda_A \end{bmatrix} + \begin{bmatrix} \Gamma_u \\ 0 \end{bmatrix} \Delta\lambda_{\bar{A}} \tag{36a}$$

$$\dot{\Delta\lambda}_{\bar{A}} = \Gamma_{\bar{A}} \Delta\lambda_{\bar{A}} \tag{36b}$$

mit Systemmatrix

$$\tilde{\Gamma} = \begin{bmatrix} \Gamma_1 & \Gamma_2 \\ \Gamma_3 & 0 \end{bmatrix} \tag{37}$$

wobei

$$\Gamma_1 = -\sum_{i \in A_c} \lambda_i^* a_i a_i^\top \tag{38}$$

$$\Gamma_2 = \left[\ldots, -a_i, \ldots\right], i \in A_c \tag{39}$$

$$\Gamma_3 = \left[\ldots, \lambda_i^* a_i, \ldots\right]^\top, i \in A_c \tag{40}$$

$$\Gamma_u = \left[\ldots, -a_i(a_i^\top x^* - b_i) - a_i, \ldots\right], i \in I_c \tag{41}$$

$$\Gamma_{\bar{A}} = \mathrm{diag}[\ldots, a_i^\top x^* - b_i, \ldots], i \in I_c. \tag{42}$$

[1]Das heißt, für die Matrix $\Gamma_2 = [a_i], i \in A_c$ gilt: $\Gamma_2 \in \mathbb{R}^{n \times n}$ und $\mathrm{Rang}(\Gamma_2) = n$. Da nun $\tilde{\lambda} = \lambda^*$ ist, folgt aus der Komplementaritätsbedingung $(a_i^\top \tilde{x} - b_i) = 0, i \in A_c$ und somit $\tilde{x} = x^*$.

Zunächst ist offensichtlich, dass $\Gamma_{\bar{A}}$ eine Diagonalmatrix mit strikt negativen Einträgen ist. Man kann daraus schließen, dass System (36b) exponentiell stabil ist.

Es wird nun das System (36a) untersucht. Zunächst sei festgestellt, dass wie oben Γ_2 vollen Rang besitzt. Weiterhin kann die Matrix $\tilde{\Gamma}$ folgendermaßen transformiert werden:

$$\bar{\Gamma} = \begin{bmatrix} I & 0 \\ 0 & X^{-1} \end{bmatrix} \tilde{\Gamma} \begin{bmatrix} I & 0 \\ 0 & X \end{bmatrix} = \begin{bmatrix} \Gamma_1 & \Gamma_2 X \\ -X\Gamma_2^\top & 0 \end{bmatrix} \tag{43}$$

wobei $X > 0$, $X = \text{diag}[\ldots, \sqrt{\lambda_i^*}, \ldots]$, $i \in I_A$ so dass $\Gamma_2 X X^\top = -\Gamma_3^\top$. Die Matrix Γ_2 besitzt vollen Rang und $\Gamma_1 < 0$.

Mit Lemma 2 kann gefolgert werden, dass $\bar{\Gamma}$ und damit auch $\tilde{\Gamma}$ Hurwitz sind. System (36a) und (36b) sind in Kaskadenform und somit ist das Gesamtsystem (35) exponentiell stabil. □

Bemerkung 1. *Falls der Sattelpunkt nicht eindeutig ist geht aus dem Beweis hervor, dass alle Sattelpunkte stabil sind und ferner, dass $\lambda(t)$ immer zu der Menge der Lösungen des dualen Problems konvergiert und die Dualitätslücke $c^\top x + b^\top \lambda$ verschwindet. Die Konvergenz zu einem Sattelpunkt ist genau dann gegeben, wenn die Grenzmenge von $(x(t), \lambda(t))$ mindestens einen Sattelpunkt enthält. Dies ist genau dann der Fall, wenn $x(t)$ in die Menge der Lösungen des primalen Problems konvergiert.*

5 Simulationsergebnisse

In diesem Abschnitt sollen die theoretischen Ergebnisse mit Hilfe von Simulationen verdeutlicht werden. Als konkretes Beispiel wird ein lineares Programm untersucht, das dazu dient die Konvergenz des AHU-Systems (3) mit dem hier vorgestellten System (30) zu vergleichen.

Das Optimierungsproblem hat die Form:

$$\begin{aligned} \min_x \; & -2x_1 - x_2 \\ \text{s.t.} \; & \underbrace{\begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \end{aligned} \tag{44}$$

Die Anfangsbedingungen wurden wie folgt gewählt: $x_1(0) = x_1^{AHU}(0) = 2.8$, $x_2(0) = x_2^{AHU}(0) = 2.5$, $\lambda_1(0) = \lambda_1^{AHU}(0) = 1$, $\lambda_2(0) = \lambda_2^{AHU}(0) = 1$, $\lambda_3(0) = \lambda_3^{AHU}(0) = 1$. Hierbei sind alle dem AHU-System zugehörigen Variablen mit hochgestelltem „AHU“ bezeichnet.

Die Linearisierung von (30) um den Sattelpunkt $x^* = [1, 0]^\top$, $\lambda^* = [0, 1, 2]^\top$ ergibt

$$\Gamma = \left[\begin{array}{cc|ccc} -2 & -2 & 0 & 0 & -1 \\ -2 & -3 & 0 & 1 & -1 \\ \hline 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 \end{array}\right]. \tag{45}$$

Im Falle des AHU-Systems (3) ist zu erkennen, dass eine „Linearisierung“ auf

$$\Gamma^{AHU} = \left[\begin{array}{cc|ccc} 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ \hline -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{array}\right] = \begin{bmatrix} 0 & A \\ -A^\top & 0 \end{bmatrix} \tag{46}$$

führt. Man erkennt an der schiefsymmetrischen Struktur, dass solch eine Matrix rein imaginäre Eigenwerte hat. Streng genommen ist hier jedoch eine Linearisierung von (3) um (x^*, λ^*) aufgrund des Projektionsoperators nicht zulässig.

Die Verläufe von $x(t)$ und $\lambda(t)$ sind für die Systeme (3) und (30) in Abb. 2 und 3 abgetragen. Offensichtlich ist bei diesem Optimierungsproblem keine Konvergenz des AHU-Systems (3) zu erkennen, wohingegen die Trajektorien des Systems (30) zu (x^*, λ^*) konvergieren.

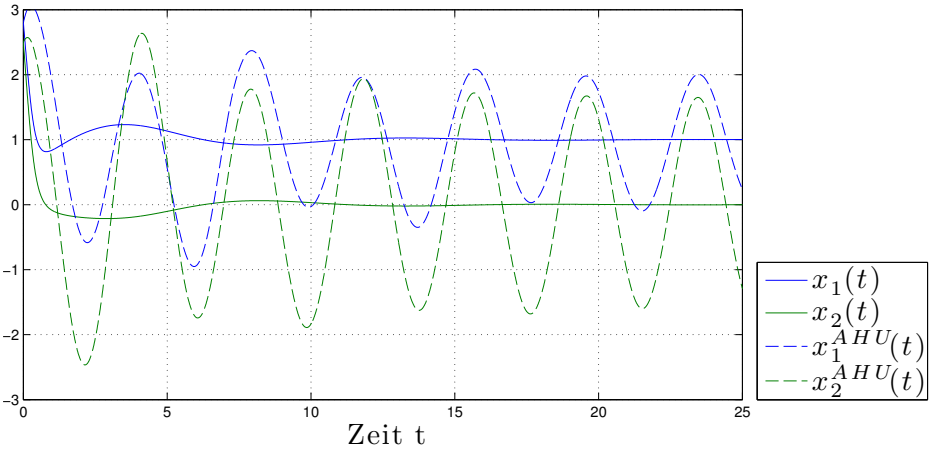


Abbildung 2: Trajektorien von (3a) und (30a)

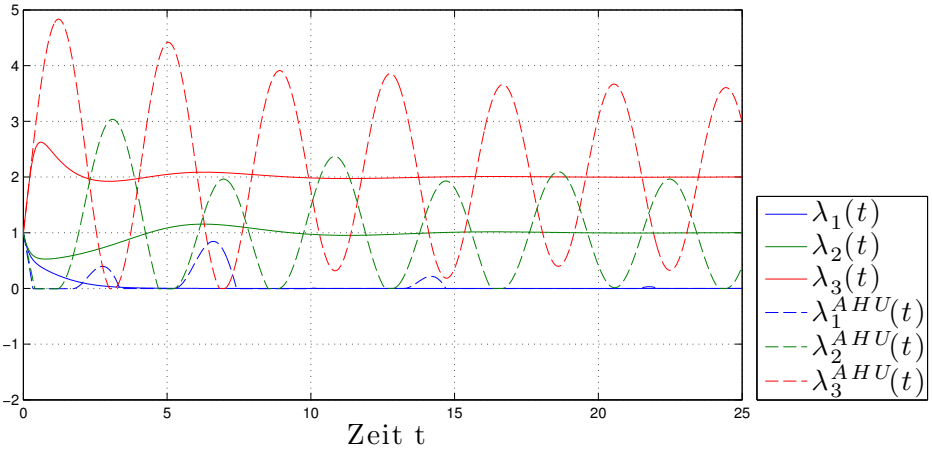


Abbildung 3: Trajektorien von (3b) und (30b)

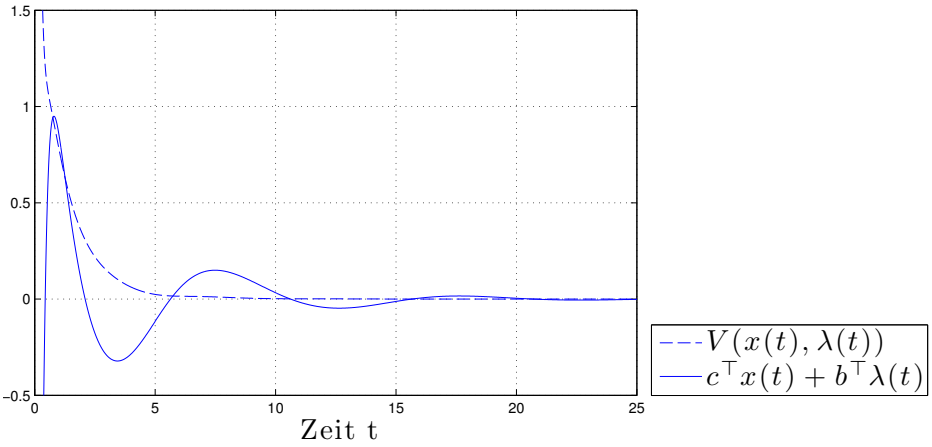


Abbildung 4: Lyapunov-Funktion (31) und Dualitätslücke von (30)

In Abbildung 4 ist die Lyapunov-Funktion (31) und die Dualitätslücke $c^\top x(t) + b^\top \lambda(t)$ für die Trajektorien des Systems (30) abgetragen. Die Lyapunov-Funktion ist strikt monoton fallend, und die Dualitätslücke verschwindet für hinreichend große Zeiten.

6 Zusammenfassung und Ausblick

In dieser Arbeit wurden zwei nichtlineare Systeme vorgestellt, deren Trajektorien zu einem Sattelpunkt der Lagrange-Funktion eines Optimierungsproblems konvergieren.

Der Hauptbeitrag dieser Arbeit ist ein nichtlineares System, das zum Lösen von linearen Programmen eingesetzt werden kann. Der Vorteil des Systems gegenüber dem AHU-System liegt in seiner Einfachheit. Es wird kein Projektionsoperator benötigt, der die Analyse erschwert. So ist es beispielsweise möglich, das System ohne Einschränkungen zu linearisieren. Mit Hilfe einer Lyapunov-Funktion wurde in Theorem 6 globale Stabilität und mit Hilfe der Linearisierung die exponentielle Stabilität des Sattelpunkt eines linearen Programms gezeigt. Das glatte Vektorfeld erlaubt unter anderem auch die Erweiterung auf semi-definite Programme, bei denen es sich bei den Lagrange-Multiplikatoren um Matrizen handelt.

Es ist geplant zu untersuchen, wie sich das System zur verteilten Optimierung einsetzen lässt. Ein Ansatz für diese Aufgabe wäre die Kopplungen zwischen den Variablen zu minimieren, um bestmögliche Ergebnisse zu erzielen. Ferner ist es interessant zu untersuchen, welche quantitativen Konvergenzeigenschaften das System besitzt und wie sie verbessert werden können.

7 Danksagung

Die Autoren bedanken sich für die Unterstützung der Deutschen Forschungsgemeinschaft, die diese Arbeit im Rahmen des Emmy-Noether Programms (Projekt: Novel Ways in Control and Computation) unterstützt hat.

8 Anhang

Die Erweiterung des Systems (30) auf lineare Programme mit Gleichungsnebenbedingungen

$$\begin{aligned} &\inf_x c^\top x \\ &\text{s.t. } a_i^\top x \leq b_i, i = 1, \ldots, m, \\ &\quad h_j^\top x = d_j, j = 1, \ldots, l \end{aligned} \tag{47}$$

mit $c, a_i, h_j \in \mathbb{R}^n$, und $b \in \mathbb{R}^m$, $d \in \mathbb{R}^l$ kann einfach durchgeführt werden. Die Lagrange-Funktion zu (47) ist gegeben durch:

$$L(x, \lambda, \nu) = c^\top x + \sum_{i=1}^{m} \lambda_i (a_i^\top x - b_i) + \sum_{j=1}^{l} \nu_j (h_j^\top x - d_j) \tag{48}$$

mit $x \in \mathbb{R}^n, \mu \in \mathbb{R}_+^m, \nu \in \mathbb{R}^l$ und dem dualen Problem:

$$\begin{aligned} &\sup_{\lambda,\mu} \; - b^\top \lambda - d^\top \nu \\ &\text{s.t. } \sum_{i=1}^{m} \lambda_i a_i + \sum_{j=1}^{l} \nu_j h_j = -c, \\ &\quad \lambda \geq 0. \end{aligned} \tag{49}$$

Die Dualitätslücke verschwindet genau wie im Fall (13)

$$c^\top x^* = -b^\top \lambda^* - d^\top \nu^*. \tag{50}$$

Das System hat nun die folgende Form:

$$\dot{x} = - c - \sum_{i=1}^{m} \lambda_i a_i - \sum_{i=1}^{l} \nu_i h_i - \sum_{i=1}^{m} \lambda_i a_i (a_i^\top x - b_i) - \sum_{j=1}^{l} h_j (h_j^\top x - d_j) \tag{51a}$$

$$\dot{\lambda}_i = \lambda_i (a_i^\top x - b_i), i = 1, \ldots, m \tag{51b}$$

$$\dot{\nu}_j = h_j^\top x - d_j, j = 1, \ldots, l. \tag{51c}$$

Es ist anzumerken, dass es durch die zusätzliche KKT-Bedingung

$$h_j^\top x - d_j = 0, j = 1 \ldots, l \tag{52}$$

keine Forderung über das Vorzeichen der Multiplikatoren ν_j gibt. Dadurch gibt es weder eine Einschränkung für die Anfangsbedingungen von $\nu_j(t)$ noch ist eine Modifikation der ν-Dynamik notwendig. Unter folgender Annahme kann Konvergenz gezeigt werden:

(A7) Es existiert genau ein Sattelpunkt (x^*, λ^*, ν^*) von (48).

Theorem 7. *Sei Annahme (A7) erfüllt. Dann konvergiert jede in $\mathbb{R}^n \times \mathbb{R}^m_{++} \times \mathbb{R}^l$ initialisierte Lösung $(x(t), \lambda(t), \nu(t))$ des Systems (51) zu dem Sattelpunkt der Lagrange-Funktion (48). Ferner ist der Sattelpunkt exponentiell stabil.*

Beweisskizze. Der Beweis kann analog zu dem von Theorem 6 durchgeführt werden. Es wird folgende positiv definite und radial unbeschränkte Lyapunov-Funktion betrachtet:

$$\begin{aligned} V(x,\lambda,\nu) =& \frac{1}{2}(x-x^*)^\top(x-x^*) + \sum_{i\in A_c} \lambda_i - \lambda_i^* - \lambda_i^* \ln\frac{\lambda_i}{\lambda_i^*} + \sum_{i\in I_c}\lambda_i \\ & - \sum_{i=1}^m \lambda_i(a_i^\top x^* - b_i) + \frac{1}{2}(\nu-\nu^*)^\top(\nu-\nu^*) \end{aligned} \tag{53}$$

mit $A_c = \{i : \lambda_i^* \neq 0\}$ und $I_c = \{i : \lambda_i^* = 0\}$. Man erhält für die Lie-Ableitung von V entlang des Vektorfeldes (51)

$$\dot{V} = L(x^*,\lambda,\nu) - L(x,\lambda^*,\nu^*) - \sum_{i=1}^m \lambda_i(a_i^\top x - b_i)^2 - \sum_{j=1}^l (h_j^\top x - d_j)^2. \tag{54}$$

Analog zur Konvergenz von (30) zur Menge (33), kann Konvergenz des Systems (51) zur Menge

$$\begin{aligned} M = \Bigg\{ & (x,\lambda,\nu) \in \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^l : \lambda_i(a_i^\top x - b_i) = 0, i = 1,\dots,m; \\ & h_j^\top x - d_j = 0, j = 1,\dots,l; \sum_{j=1}^l \nu_j h_j + \sum_{i=1}^m \lambda_i a_i + c = 0; \\ & b^\top\lambda + d^\top\nu = b^\top\lambda^* + d^\top\nu^* \Bigg\} \end{aligned} \tag{55}$$

gezeigt werden. Aus der Eindeutigkeit der Lösung folgt, dass M nur den Sattelpunkt (x^*, λ^*, ν^*) enthält. Somit kann die Konvergenz zu einem Sattelpunkt der Lagrange-Funktion (48) und damit zur Lösung von (47) festgestellt werden.

Die Linearisierung des Systems (51) um (x^*, λ^*, ν^*) hat die Form

$$\begin{bmatrix} \dot{\Delta x} \\ \dot{\Delta\lambda} \\ \dot{\Delta\nu} \end{bmatrix} = \Gamma \begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta\nu \end{bmatrix}. \tag{56}$$

Sei $\Delta\lambda_A = \big[\dots, \Delta\lambda_i, \dots\big]$, $i \in A_c$ und analog $\Delta\lambda_{\bar{A}} = \big[\dots, \Delta\lambda_i, \dots\big]$, $i \in I_c$. Dann ist (56) ein System in Kaskadenform

$$\begin{bmatrix} \dot{\Delta x} \\ \dot{\Delta\lambda}_A \\ \dot{\Delta\nu} \end{bmatrix} = \tilde{\tilde{\Gamma}} \begin{bmatrix} \Delta x \\ \Delta\lambda_A \\ \Delta\nu \end{bmatrix} + \begin{bmatrix} \Gamma_u \\ 0 \\ 0 \end{bmatrix} \Delta\lambda_{\bar{A}} \tag{57a}$$

$$\dot{\Delta\lambda}_{\bar{A}} = \Gamma_{\bar{A}} \Delta\lambda_{\bar{A}} \tag{57b}$$

mit Systemmatrix

$$\tilde{\tilde{\Gamma}} = \begin{bmatrix} \tilde{\Gamma}_1 & \Gamma_2 & \Gamma_4 \\ \Gamma_3 & 0 & 0 \\ -\Gamma_4^\top & 0 & 0 \end{bmatrix} \tag{58}$$

wobei

$$\tilde{\Gamma}_1 = -\sum_{i \in A_c} \lambda_i^* a_i a_i^\top - \sum_{i=1}^{l} h_i h_i^\top \tag{59}$$

$$\Gamma_4 = \left[\ldots, -h_i, \ldots \right], i = 1, \ldots, l \tag{60}$$

und Γ_2 aus (39), Γ_3 aus (40), Γ_u aus (41) und $\Gamma_{\bar{A}}$ aus (42). Weiterhin kann die Matrix $\tilde{\tilde{\Gamma}}$ folgendermaßen transformiert werden

$$\bar{\bar{\Gamma}} = \begin{bmatrix} I & 0 & 0 \\ 0 & X^{-1} & 0 \\ 0 & 0 & I \end{bmatrix} \tilde{\tilde{\Gamma}} \begin{bmatrix} I & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & I \end{bmatrix} = \begin{bmatrix} \tilde{\Gamma}_1 & \Gamma_2 X & \Gamma_4 \\ -X\Gamma_2^\top & 0 & 0 \\ -\Gamma_4^\top & 0 & 0 \end{bmatrix} \tag{61}$$

wobei $X > 0$, $X = \text{diag}[\ldots, \sqrt{\lambda_i^*}, \ldots]$, $i \in I_A$ so dass $\Gamma_2 X X^\top = -\Gamma_3^\top$. Mit Lemma 2 kann gefolgert werden, dass $\bar{\bar{\Gamma}}$ und damit auch $\tilde{\tilde{\Gamma}}$ Hurwitz sind. System (57a) und (57b) sind in Kaskadenform und somit ist das Gesamtsystem (56) exponentiell stabil.

Literatur

[1] K. J. Arrow, L. Hurwicz, and H. Uzawa. *Studies in Linear and Non-Linear Programming*. Stanford University Press, 1958.

[2] E. N. Barron, R. Goebel, and R. R. Jensen. Best response dynamics for continuous games. *Proceedings of the American Mathematical Society*, 138(3), 2010.

[3] D. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, 1982.

[4] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1995.

[5] D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, 1997.

[6] G. W. Brown and J. von Neumann. Solutions of games by differential equations. *Contributions to the Theory of Games, Annals of Mathematics Study*, 1(24), 1950.

[7] H.-B. Dürr and C. Ebenbauer. A smooth vector field for saddle point problems. *Accepted at the 50th Conference on Decision and Control in Orlando, USA*, 2011.

[8] K. H. Elster. *Nichtlineare Optimierung*. Harri Deutsch, 1978.

[9] D. Feijer and F. Paganini. Stability of primal-dual gradient dynamics and applications to network optimization. *Automatica*, 46(12), 2010.

[10] J. Hofbauer. Best response dynamics for continuous zero–sum games. *Discrete and Continuous Dynamical Systems*, 6(1), 2006.

[11] T. Kose. Solutions of saddle value problems by differential equations. *Econometrica*, 24(1), 1956.

[12] A. Nedić and A. Ozdaglar. Subgradient methods for saddle-point problems. *Journal of Optimization Theory and Applications*, 142(1), 2009.

[13] R. T. Rockafellar. *Convex Analysis (Princeton Mathematical Series)*. Princeton University Press, 1970.

[14] R. Srikant. *The Mathematics of Internet Congestions Control*. Birkhäuser, 2003.

[15] R. J. Vanderbei. *Linear Programming, Foundations and Extensions*. Kluwer Internation Series, 1997.

Finite State Machines Bring High Frequency Adaptive Switching Control to Power Electronics

Karel Jezernik

University of Maribor, FERI, Smetanova 17, 2000 Maribor, Slovenia

karel.jezernik@uni-mb.si

Abstract : In this paper, the design of current control loop and selection of the switching logic for three phase switching converters and AC electrical machines are developed. Main design specifications are robustness to load electrical parameters, fast dynamical response, reduced switching frequency and simple hardware implementation. To meet previous specifications, a discrete-event type controller is proposed, designed as finite-state automaton and implemented with a FPGA device. After general introduction, system analysis is preformed, control target are specified and the proposed DES strategy is presented and discussed. Further, actual controller architecture is based on FPGA Spartan 3E. Experimental results are represented using a Brushless AC motor as the converter load. However, this does not limit the wider applicability of the proposed controller that is suitable for different types of AC loads (rectifier, inverter). Switching strategy implemented with state transition diagram provides minimum number of switching of three-phase inverter that is confirmed trough simulation and experimental results. The switching among these modes is governed by the supervisory control approach. This example demonstrates interpreted design and rapid prototyping of our methodology which are besides theory valuable for practice.

1 Introduction

Power converters may be viewed as a set of voltage (or current) sourced subsystems interconnected through switches. The objective of the switches, which actuate as lossless transformers, is to allow the transfer of energy from one subsystem to another. The subsystems consist of passive elements (like inductors, capacitors, and resistances), power sources, and a load where the desired energy is delivered. Because of the presence of discontinuous elements the behaviour of power converters consists of several modes, corresponding to different circuit topologies, i.e. two-level VSI presented in this paper with 2^3 modes, [1].

Design of current control system for switching power converters and electrical machines is conventionally performed in two steps. One is related to PWM pattern selection and the second is the design of a control action to ensure the desired dynamical behaviour of the closed-loop system. In such an approach the dynamics of the system, and consequent the

design of the controller, is treated separately from the operation of the switches. The switching pattern design is usually performed as an open-loop system, based on the assumption that the average of the discontinuous outputs of the inverter switching matrix, repeats the references with design error. The control plant dynamics, determined by the energy storing capabilities of the load and internal L, C components is used to determine the switching frequency, but not actual switching pattern [1]. Discrete event sequential control approach offers a possibility to solve problems, the desired closed-loop dynamics and the PWM pattern selection, in the same framework.
Hysteresis control is a fast and robust control approach but it is mostly applied only to simple converter systems. Considering such controllers as discrete-event systems opens a more systematic view and enables the controller design even for nonlinear systems. The goal of this paper is to continue to a more systematic design of hysteresis controllers from the view point of discrete-event systems (DES) so that it will be possible to benefit from their high robustness and dynamic response. Also DES method strives to reduce switching number of inverter which directly affect to inverter losses. A very interesting item is the controller hardware: event-driven controllers can advantageously be realized on Field Programmable Gate Array (FPGA) with low effort.
FSM design is the technique for the design of clocked sequential circuits, which perform some repetitive actions [4]. State diagrams and state tables are the best representations of FSM and are used in this study. State diagram is the graphical representation of FSM. It shows the actual flow of data and flow paths from inputs to outputs. According to the problem definition, state diagrams can be easily designed. So we designed FSM for implementing the voltage source inverter (VSI) operation using VHDL [5]. We propose FSM because of easy implementation into VHDL with sophisticated tools available in ISE Xiling software. We develop two different FSM for design of DES current controller and supervisory controller of BLAC motor control [6].
FSM approach introduces an extended principle of the IF statement. Main parts of such statement are event, condition and action. FPGA allows that the operation (action) is completed immediately after changes of inputs (condition) are detected (event), without one cycle time delay. On the FPGA all process are executed in parallel, so time period of each process are independent of number of processes.
After a general introduction, system analysis is performed, control targets are specified, and the proposed control strategy is presented and discussed. Simulation and experimental results are presented using a BLAC motor, where main attention is reducing switching number of inverter.

2 DES Control System Model

DES systems can be viewed as a large collection of systems of various classes. A hierarchical structure arises when a logical control unit governs such a system by issuing logic decisions. This leads to the system framework shown below in figure which clearly illustrates this architecture (Fig. 1), [3].

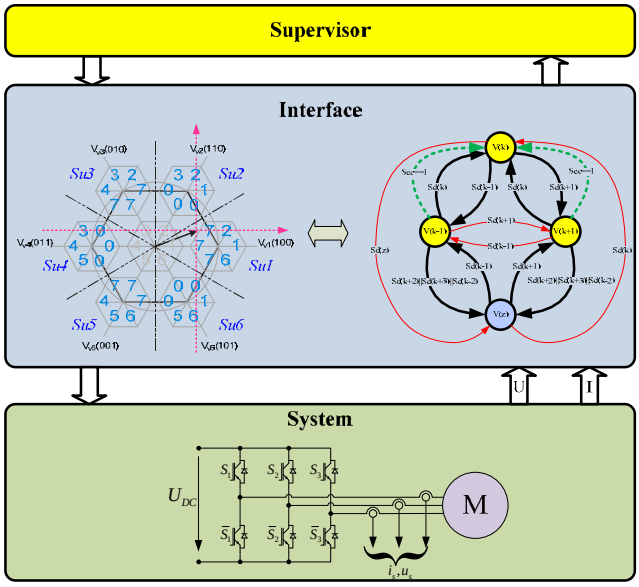


Fig. 1. Control architecture.

The top layer, which is a discrete event system, can use different types of description language such as finite state machines, fuzzy logic, Petri nets, etc. The bottom layer is a continuous system, and is usually the physical system. The interface plays the role of facilitating communication between the two different layers by means of translating signals between them. As the techniques for control design and analysis are well developed for the continuous and discrete systems, the design of the interface is of utmost importance because it determines the way in which the combined system behaves. For most system designs, the logic unit and continuous unit are usually designed separately and then combined together by an interface. Another approach is converting the whole system to be purely logic or continuous for design.

2.1 Continuous system

The BLAC motor combines many of the advantages of the permanent excited AC motor and the synchronous motor Fig. 2. The BLAC motor needs low reactive current, which is very similar to the DC motor, and the current is proportional to the torque. The shaft torque is therefore easy to estimate by detecting a three-phase current of BLAC motor [7]. The BLAC motor is the combination of a permanent excited synchronous motor and a three-phase inverter. The BLAC motor dynamics are described with differential equation for motor current

$$\frac{di_{si}}{dt} = \frac{1}{L_s}\left(u_i - R_s\, i_{si} - e_{si}\right); \quad i = 1,2,3 \tag{1}$$

where i_{si} denotes three phase stator currents, u_{si} phase voltage, e_{si} EMF voltage, R_s is resistance and L_s is inductance of motor windings. Developed torque of BLAC motor is

$$T_e = \frac{\sum_{i=1}^{3} e_{si} i_{si}}{\omega}. \tag{2}$$

The mechanical equation is

$$\frac{d\omega}{dt} = \frac{1}{J}\left(T_e - T_L\right) \tag{3}$$

where ω is shaft speed, T_e and T_L are electromechanical and load torque, respectively.

The considered control problem of current control of BLAC motor is the tracking of a three phase current reference signal. After the current control error is defined as $\triangle \boldsymbol{i}_s = \boldsymbol{i}_s - \boldsymbol{i}_s^d$, (1) can be rewritten in error form

$$\frac{d\triangle \boldsymbol{i}_s}{dt} + \frac{R_s}{L_s}\triangle \boldsymbol{i}_s = \frac{1}{L_s}\left(\boldsymbol{u}_s(\boldsymbol{V}_k) - \boldsymbol{e}_s\right) \tag{4}$$

which contains all the disturbances action on the system. Signal of current $\boldsymbol{i}_s$, speed ω and electromotive force $\boldsymbol{e}_s$ are continuous. The input voltage vector $\boldsymbol{u}_s(\boldsymbol{V}_k)$ is discontinuous and is result of switch positions structure in VSI inverter Fig. 3, [7].

The commutation of BLAC motor depends on the position of the rotor. The angle between the magneto motive forces of the stator and the rotor is fixed to 90◦, so the motor produces maximum torque and needs low reactive current.

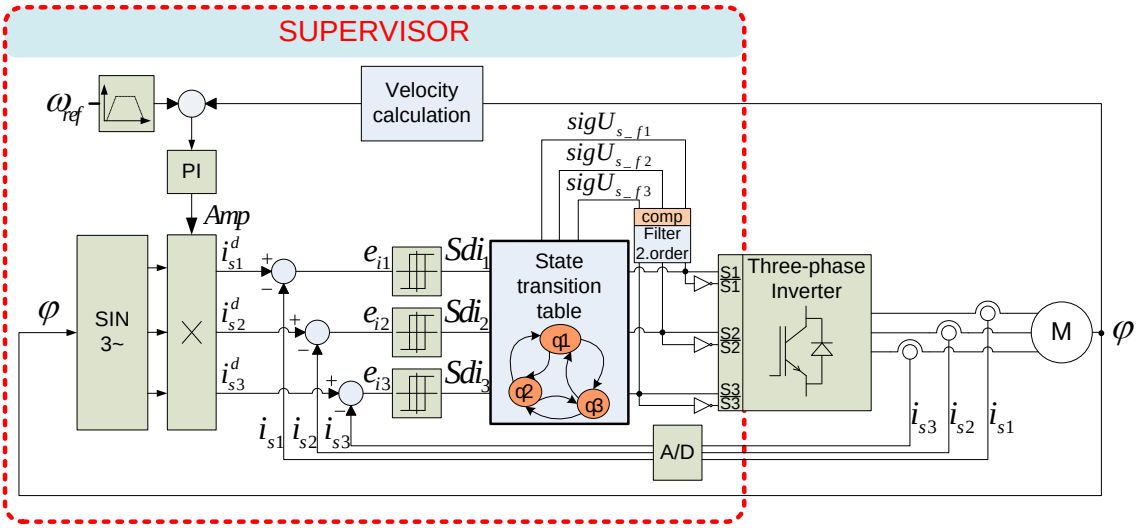


Fig. 2: General control scheme of the BLAC motor.

Fig. 2 shows the architecture of speed / current regulation of BLAC motor implemented into a FPGA circuit. The considered control task tracks a three-phase current reference signal, whose amplitude is determined by the output of proportional-integral (PI) speed controller. Current component i_d and i_q are not measured directly. Like in the classical field oriented control (d-q model), d-axes is set to zero (provided that correct rotor position is known), therefore the q-axes current component i_q is equivalent with the amplitude of the reference current (Fig. 2 - *Amp*).

2.2 Interface

2.2.1 Matrix Model of VSI Inverter

The input voltage vector $\boldsymbol{u}_{si}(\boldsymbol{V}_k)$ of the three-phase inverter connects an output of DES controller with continuous plant of BLAC motor windings (4). The basic principle of the current control is to manipulate the input voltage vectors $\boldsymbol{u}_s(\boldsymbol{V}_k)$ so that the desired current is produced by inverter. This is achieved by choosing an inverter switch combination $\boldsymbol{S}_k$ that drives the stator current vector by directly applying the appropriate inverter voltages $\boldsymbol{u}_s(\boldsymbol{V}_k)$ to the BLAC machine windings Fig. 1. The switch position of the three-phase inverter are described using the logical variables $\boldsymbol{V}_k$, dependent if switch $\boldsymbol{S}_k$ is ON or OFF. Each variable corresponds to one phase of inverter (Fig. 3). Three phase inverter can produce 2^3 voltage vector combinations; two of them are zero vectors and 6 active vectors.

Table 1 Control input voltage vector $\boldsymbol{u}_s(\boldsymbol{V}_k)$ definition.

Voltage vector	Hexagon	S_1	S_2	S_3	u_{s1}	u_{s2}	u_{s3}
V_0		0	0	0	0	0	0
V_1		1	0	0	$\frac{2}{3}U_{DC}$	$-\frac{1}{3}U_{DC}$	$-\frac{1}{3}U_{DC}$
V_2		1	1	0	$\frac{1}{3}U_{DC}$	$\frac{1}{3}U_{DC}$	$-\frac{2}{3}U_{DC}$
V_3		0	1	0	$-\frac{1}{3}U_{DC}$	$\frac{2}{3}U_{DC}$	$-\frac{1}{3}U_{DC}$
V_4		0	1	1	$-\frac{2}{3}U_{DC}$	$\frac{1}{3}U_{DC}$	$\frac{1}{3}U_{DC}$
V_5		0	0	1	$-\frac{1}{3}U_{DC}$	$-\frac{1}{3}U_{DC}$	$\frac{2}{3}U_{DC}$
V_6		1	0	1	$\frac{1}{3}U_{DC}$	$-\frac{2}{3}U_{DC}$	$\frac{1}{3}U_{DC}$
V_7		1	1	1	0	0	0

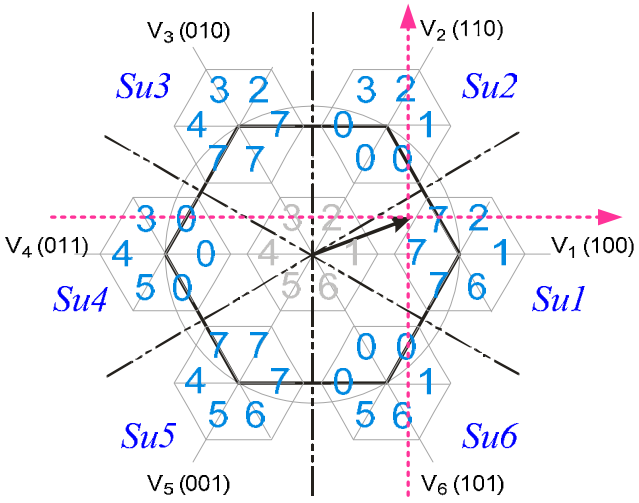


Fig. 3: Stator voltage $\boldsymbol{u}_s(\boldsymbol{V}_k)$ of three phase inverter

The energy flow between the input and output side of the three-phase inverter is controlled by switching matrix. By introducing the binary variables S_k that are "1" if particular switch S_k is ON and "0" if switch S_k is OFF, the behaviour of the switching matrix can be described by the three dimensional vector $\boldsymbol{u}_s(\boldsymbol{V}_k) = U_{DC}\boldsymbol{L}\boldsymbol{S}_k$, where matrix $\boldsymbol{L}$ and vector $\boldsymbol{S}(S_1, S_2, S_3)$ are defined as [8]

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}, \qquad (5)$$

$$\boldsymbol{S}^T = \begin{bmatrix} S_1 & S_2 & S_3 & \overline{S}_1 & \overline{S}_2 & \overline{S}_3 \end{bmatrix}$$

Relation (5) is true for the two levels VSI inverter circuit [7]. It essentially shows that this particular switching matrix is able to generate three independent control actions denoted as the components S_1, S_2 and S_3 of the discontinuous control vector

$$\boldsymbol{u}_s(\boldsymbol{V}_k) = U_{DC}\left[S_1 + S_2 e^{j2\pi/3} + S_3 e^{-j2\pi/3} \right] \qquad (6)$$

2.2.2 Discrete Event Controller

The controller is a discrete event system (DES) which is modelled as a finite state machine (FSM) or finite state automaton. FSM represents a behavioural model composed of a finite number of states, transition between this states and actions, similar for a flow graph in which one can inspect the way logic runs when certain conditions are met.

State transition techniques have shown the greatest potential as a synthesis technique for sequential DES control. They have been used for design and analysis of digital computer systems and have recently been adapted to industrial sequential control. To reap the real benefits of state transitions techniques, a new type of architecture must be developed for FPGA controllers. Response time is expected to be several orders of magnitude faster with the state machine and state table architecture presented in this paper.

A FSM on Fig. 4 represent a state transition diagram of VSI current controller. The states $\boldsymbol{V}(k)$ represent the input voltage vector $\boldsymbol{u}_s(\boldsymbol{V}_k)$ of three-phase inverter in Fig. 3, for each voltage sector of BLAC motor voltage $\boldsymbol{u}_s$. $\boldsymbol{V}(k)$ represents active voltage vector, $\boldsymbol{V}(k-1)$ is previous active voltage vector $\boldsymbol{V}(k+1)$ is possible future active voltage vector and $\boldsymbol{V}(z)$ produced zero voltage vectors of VSI inverter. Connections between voltage vectors (state of FSM in Fig. 4) are developed with the use of EVENT-CONDITION-ACTION (ECA) rules, and shown in Table 2.

Table 2 Event Condition Action (ECA) – rules.

Events	Sdi_0, Sdi_1,…Sdi_7
Conditions	Su1, Su2,…Su6 $V_{0(k-1)}$, $V_{1(k-1)}$,… $V_{7(k-1)}$
Actions	V_0, V_1,…V_7

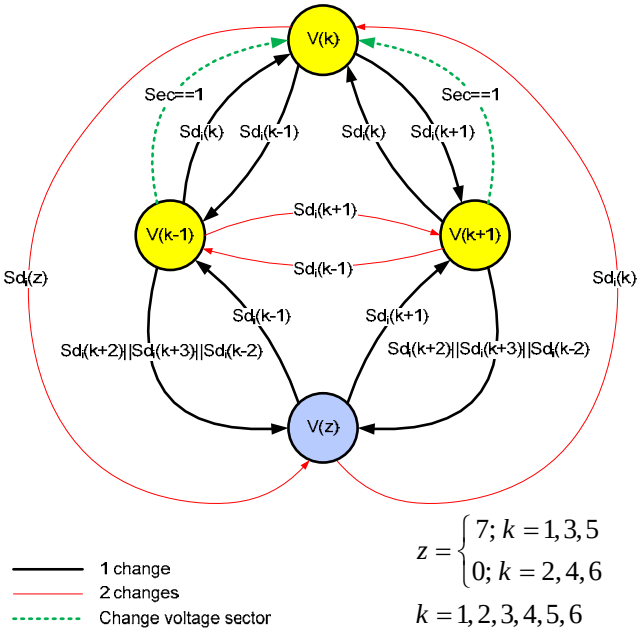


Fig. 4: State transition diagram

The action could be performed when event is recognized and the conditions are met. To consider a current controller as a discrete-event dynamical system [9], it allows focusing in much more details on the switching actions and will enable a better understanding of the controller design. A DES system reacts only if an event is recognized. To control the current $\boldsymbol{i}_s$, the sector of drive voltage $\boldsymbol{u}_s$ is recognize first, and based on the known sector, the input voltage vector $\boldsymbol{u}_s(\boldsymbol{V}_k)$ (the transistor switching pattern) for the current control is selected, respecting the current control error $\Delta \boldsymbol{i}_{si}$ (4).

Considering space vector representation of the stator voltage $\boldsymbol{u}_s(\boldsymbol{V}_k)$ (6), the voltage is represented as vector rotating around the origin, six active switching vectors of the three-phase transistor inverter represent the six active output voltage vectors denoted as $\boldsymbol{V}_1 ... \boldsymbol{V}_6$; $\boldsymbol{V}_0$ and $\boldsymbol{V}_7$ are two zero voltage vectors. According to signs of the phase voltages u_{s1}, u_{s2} and u_{s3}, the phase plane is divided into six sectors denoted by Su1 … Su6 (Fig. 3).

In regards to the situation (Fig. 3), the stator voltage space vector $\boldsymbol{u}_s$ is in sector Su1. In this sector, logical voltage vectors $\boldsymbol{V}_0$, $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, $\boldsymbol{V}_6$ and $\boldsymbol{V}_7$ are selected for the current control. $\boldsymbol{V}_0$, $\boldsymbol{V}_7$ are two zero vectors, while $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, $\boldsymbol{V}_6$ are three nearest adjacent live output voltage vectors to this sector. With the use of the DES theory, five output voltage vectors $\boldsymbol{V}_0$, $\boldsymbol{V}_1$, V_2, $\boldsymbol{V}_6$ and $\boldsymbol{V}_7$ are recognized as discrete states of the system. Events represent allowed transition among the discrete states i.e. allowed switching.

The proposed logical event-driven BLAC current control can be realized in the form of state transition table (Table 3), where sign of current control error are presented by $sign(\Delta \boldsymbol{i}_s)$ and currently active voltage sector is presented by $sign(\boldsymbol{U}_{s\_f})$. Three bits value of $sign(\boldsymbol{U}_{s\_f})$ consists of signs of three filtered stator phase voltage of BLAC motor.

To further improve the presentation, active voltage vectors are marked, in Table 3 with a blue

background. Because the transition between inverter switch states is performed by switching only one inverter leg, switching frequency is reduced. The zero voltage control vector can be consciously used to reduce the transistor's switching frequency [10].

Tabela 3 State transition table.

$sign\ U_{s\_f}$ / $sign\ \Delta i_s$		Su_1	Su_2	Su_3	Su_4	Su_5	Su_6
		100	**110**	**010**	**011**	**001**	**101**
Sdi_0	**000**	V_7	V_0	V_7	V_0	V_7	V_0
Sdi_1	**100**	V_1	V_1	V_7	V_0	V_7	V1
Sdi_2	**110**	V_2	V_2	V_2	V_0	V_7	V_0
Sdi_3	**010**	V_7	V_3	V_3	V_3	V_7	V_0
Sdi_4	**011**	V_7	V_0	V_4	V_4	V_4	V_0
Sdi_5	**001**	V_7	V_0	V_7	V_5	V_5	V_5
Sdi_6	**101**	V_6	V_0	V_7	V_0	V_6	V_6
Sdi_7	**111**	V_7	V_0	V_7	V_0	V_7	V_0

2.2.3 Supervisor

The controller and plant can not communicate directly in a control system, because they each utilize a different type of symbols. The objective is develop methodologies that, given the system description and performance specifications, extracts discrete-event controller that supervise the plant to guarantee, that this specifications are satisfied, Fig. 1.
Current amplitude, motor speed, DC link and motor temperature are monitored by the supervisor. The supervisor can affect the parameters and the reference values when any of the measured values is critical. If the system overloads, the supervisor immediately initiates the procedure to protect the system. Protection algorithm starts with modifying the parameters and reference values. When modifying of the parameters and reference values cannot solve the problem, the supervisor affects the state of the inverter directly shown as state diagram on Fig. 5.

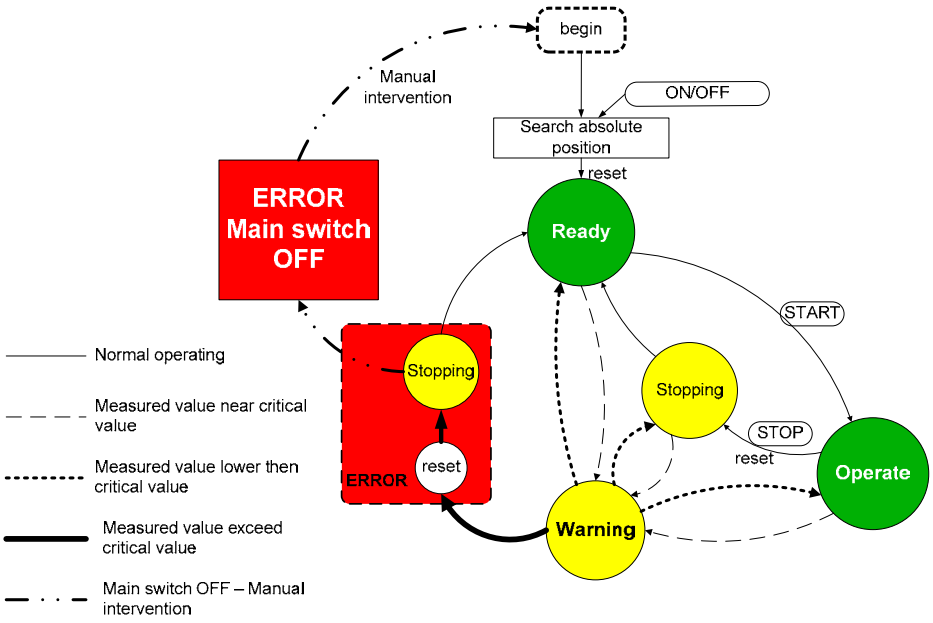


Fig. 5: FSM – Supervisor.

The steering function determines the converter's operation mode. The three main possible modes are of concern: **Ready**, **Operate** or **Error**. They are considered as discrete states of the monitoring function. Initially, the **Begin** mode is active. Turn the main switch (ON/OFF) to ON, enable search for absolute position of the rotor that is term for the transition to state **Ready**. Inverter voltage is ensured in this state. From the state **Ready, it** is possible to start the system (**Operate**) by pressing the START button. Operate mode allows the change of all free parameters and reference values. Stopping the system is possible by pressing the STOP button. **Stopping** mode resets all of the parameters (puts them on reference values) and puts the current and speed reference to zero. If rotor is stopped, the system reverts to the **Ready** mode. In these three states Ready, Operate and Stopping, the system constantly monitors values of current, inverter voltage, and temperature and rotor speed. If any of these values are outside the permitted quantities, the system continues in **Warning** mode. **Warning** mode is intended to alert the user that the system is in a limit range of the safe system activity. In the state **Error,** it is the first, to initiate the RESET, which causes the state **Stopping**. If stopping system is successful then the system returns to the state **Ready**, otherwise turn off the main switch. For system reboot, servicer should eliminate the error which causes a critical state of the system.

3 Implementation and Protection Issue

A low cost Xilinx Spartan 3 FPGA that contains 1.2 M logical gates and includes a 50 MHz oscillator has been used as a target component for the implementation of the controller. The architecture of each control algorithm is designed according to an efficient methodology that offers considerable design advantages such as reusability, reduction of development time and optimization of the consumed resources. Each control algorithm is partitioned into elementary modules, which are easier to develop and are more functional, Fig. 6.

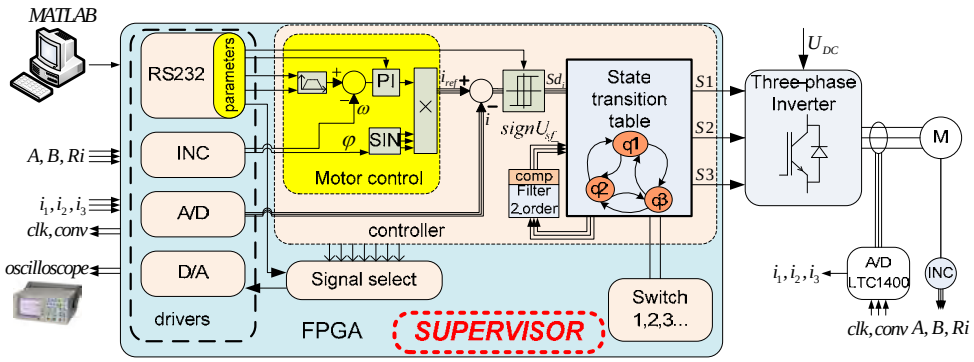


Fig. 6: FPGA controller of the BLAC motor.

As shown in Fig. 6, Xilinx Spartan XC3S1200E is used to implement the BLAC motor controller. This motor control Intellectual Property (IP) is divided into three parts. The first part is the driver's part with ADC and DAC management modules, incremental module for speed and position measurement and RS 232 module for connection of host PC equipment with Matlab/Simulink program environment. The second part includes the PI-speed motor controller. Output of speed controller is a measure for the desired torque of BLAC motor that multiplies the three-phase reference currents of machine. The currents phases are depend on

rotor position. The third part includes the state transition table with DES controller and an average output voltage of BLAC motor synchronization – voltage sector selection; Fig. 6. Drivers are made to incorporate every external device. A "signal select" output is implemented onto FPGA for monitoring and external supervisor activity.

4 Results

In all simulation and experimental results, the BLAC motor from Maxon, type EC32 with parameters $R_s = 0{,}56\ \Omega$, $L_s = 0.09$ mH, $J = 20$ gcm^2, $\Psi_{PM} = 0.0205$ Vs, $t_M = 6.6$ ms was used. The FPGA controller has the diagnostic features that are necessary for drive installation, test problem detection and elimination. The control algorithm is executed every 2.5 μs, and the switching frequency of three-phase inverter was set with the tolerance band of the hysteresis current control error. The control of both, the speed and applied torque, is possible, thus hardware in the loop operation can also be performed.
Simulation results show that DES control successfully reduces the numbers of switching. Fig. 7 shows reference and actual motor's speed, phase currents, voltage vectors and switching number. To keep a similar ripple current, the DES control performed by measuring current every 2.5 μs and the DES every 5 μs. At time 0.2 s external load was added. Switching number is introduced with four different diagrams. "*1 switch*" presents switching number for case when active voltage vector is changed with change state of only one switching element in VSI inverter leg. ("*2 switches*" - two switching elements, "*3 switches*" - three switching elements) "*All switches*" presents total switching number of all three switching elements. For comparisons conventional Hysteresis control, for changing active voltage vectors in most cases need to change all three switching elements. In DES control we strive to use only active voltage vectors where only one switching element change state.
The experimental results are illustrated in Fig. 8 and Fig. 9. The results are recorded with an oscilloscope and consequently the measured results are contaminated with noise.
Table 4 show that DES control reduces number of changing active voltage vectors by three times. In DES control are more then 90% of changing active voltage vectors made with only one switching element. Hysteresis control use 80% of changing active voltage vectors with three switching elements. Overall, this presents for 90% less switching on three-phase inverter of DES control.

Table 4 Maximum switching number (Fig. 7, Fig. 9).

	1 switch	2 switches	3 switches	Change active vectors	**All switches**
Hysteresis	805	17332	81499	99636	**279966**
DES (2,5 μs)	27216	224	0	27440	**27664**

Experimental results (Fig. 10) show reference and actual motor's speed, currents i_α, i_β and control input $\boldsymbol{u}_s(\boldsymbol{V}_k)$. Tracking the reference speed in both methods is practically the same. Advantage of DES control is in using active voltage vectors that are used much more orderly. For clarity, only active voltage vectors without zero vectors (V_0, V_7) are presented.

The switching number of each switching element increase with low speed, because of using more active voltage vectors which need three changing of switching elements in VSI inverter leg (Fig. 10). In DES control switching number is almost independent of speed.

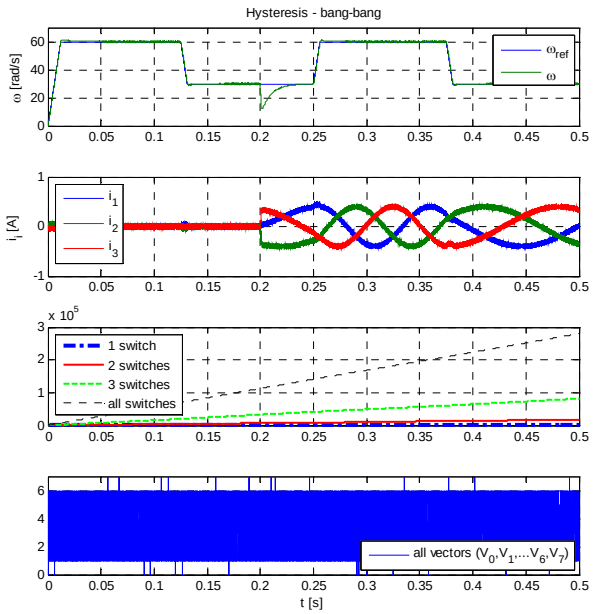


Fig. 7: Hysteresis control (Ts=5 μs).

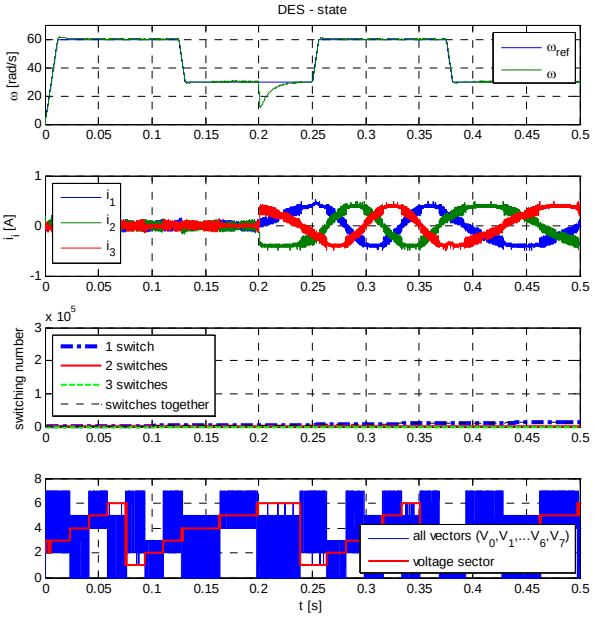


Fig. 8: DES control (Ts=5 μs)

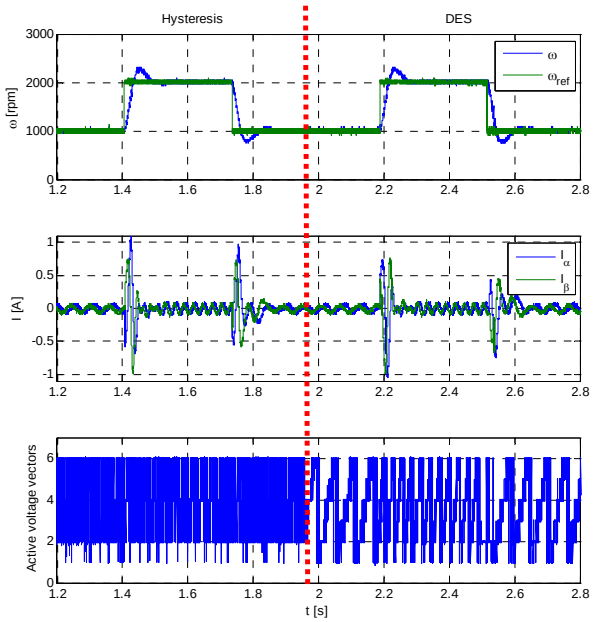


Fig. 9: Experimental results: speed tracking.

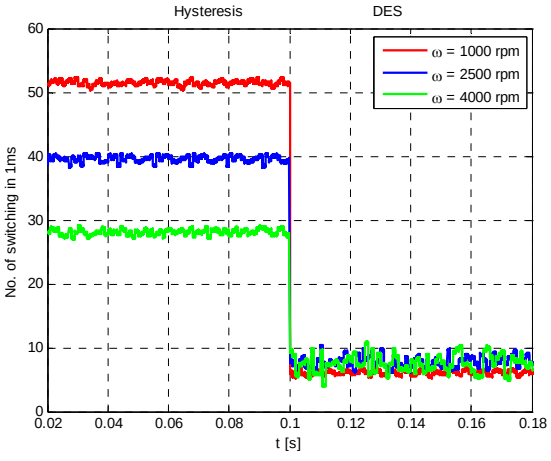


Fig. 10: Experimental results: switching number of each switching element at constant speed.

5 Conclusion

In this study, the DES control of BLAC motor (problem of speed and current control) has been addressed. Two layers of control architecture combine continuous time and discrete event system. DES part is implemented with FSM under ECA rules. Events identification, including other conditions, is most important for execute action. The nonlinear behaviour of the system limits the performance of conventional hysteresis controllers that are used for comparison. The simulation and experiments confirm the potential of the presented DES approach: switching number of three-phase inverter is significantly reduced (about 90%) and also traditional coding efforts are reduced. Formal mathematical background of the proposed approach and its correspondence to the conventional control system theory opens further possibilities for the design, simulation, and formal analysis of the discrete event systems. Interface is due to the simplicity of implementation on FPGA and efficiency operation in the form of FSM was successfully applied. The proposed approach offers a promising technique for design of complex and timely critical algorithms.

6 References

[1] Buisson J, Richard P.Y, Cormerais H (2005), "On the stabilization of switching electrical power converters". In *Hybrid Systems: Computation and Control*, pp.184-197, Springer, Berlin.

[2] Boldea I, Tutelea N.L (2009), *Electric machines: steady state, transients, and design with Matlab*. CRC Press, Taylor and Francis Group.

[3] Koutsoukos X.D, Antsaklis P.J, Stiver J.A, Lemmon M.D (2000), "Supervisory control of hybrid systems". *Proceedings of the IEEE*, vol.88, no. 7, pp. 1026-1049.

[4] Carmely T (2009), "Using Finite State Machines to Design Software". EE Times.

[5] Dubey R (2009), *Introduction to Embedded System Design Using Field Programmable Gate Arrays*. Springer – 1st edn., Verlag London.

[6] Kohavi Z, Jha N. K (2010), *Switching and finite automata theory*. Cambridge University Press

[7] Polič A, Jezernik K (2007), "Event-driven current control structure for a three phase inverter". *Int. Rev. Electrical Eng*., 2, (1), pp. 28-35.

[8] Šabanović A, Jezernik K, Šabanović N (2002), "Sliding modes applications in power electronics and electrical drives". In *Variable structure systems: towards the 21st century*, Springer, Berlin, pp. 223-251.

[9] Ramadge P. J. G, Wonham W.M (1989), "The control of discrete event systems". *Processing of the IEEE*, vol. 77, no.1, pp. 81-99.

[10] Monmasson E, Cirstea M.N (2007), "FPGA Design Methodology for Industrial Control Systems – A Review". *IEEE Trans. Ind. Electronics*, vol. 54, no. 4, pp. 1824-1842.

Der antike Mechanismus von Eleutherna

Untersuchung, Erläuterung und 3D-Simulation

Prof. D. Kalligeropoulos, Dr. S. Vasileiadou, Techn. Ing. A. Gkamaris

T.E.I. Piraeus

dkal@teipir.gr, svasil@teipir.gr, gkamaris.tasos@gmail.com

Kurzfassung

Bei den Ausgrabungen, die Prof. Petros Themelis im Jahre 1997 an der antiken Stadt Eleutherna bei Kreta unternahm, wurde ein eigenartiger, origineller im griechischen Raum, Mechanismus gefunden. Mit seiner Untersuchung wurden, im Jahre 2008, Prof. D. Kalligeropoulos und Dr. S. Vasileiadou beauftragt.

Die Schritte dieser interessanten Forschung waren die folgenden:

- Analyse durch systematische Untersuchung, vielfältige Photographierung, genaue Messung und Zeichnung aller Elemente des Mechanismus, sowie seiner beiliegenden Funden.
- Röntgenuntersuchung und axiale Tomographie für die Enthüllung der optisch nicht erreichbaren Gebiete des Mechanismus.
- Untersuchung der ähnlichen mit dem Mechanismus, parallelen archäologischen Funden.
- Zeichnerische Abbildung des gesamten Mechanismus und Erläuterung seiner Funktion, sowie der Funktion jedes von seinen einzelnen Elementen.
- Rekonstruktion des Mechanismus, in Form eines funktionellen eisernen Models im Maßstab 2:1.
- Und zu letzt, dreidimensionale digitale Simulation des Mechanismus und seiner Funktion.

1 Einführung

Die Stadt und ihre Ausgrabung

Die antike Stadt von Eleutherna ist auf einem Hügel am Fuß des Berges Ide, 30km südöstlich von der heutigen Stadt Rethymno im zentralen Kreta aufgebaut. Ihre Geschichte geht bis zur prähistorischen Periode zurück. Eine besondere Entwicklung zeigt die Stadt während der archaischen, der klassischen und besonders der hellenistischen Periode. Während der römischen Periode erscheinen eine besondere Bauaktivität und ein Aufschwung der Bevölkerung und des Reichtums der Stadt. Jedoch am Ende der römischen Periode, um 365 n.u.Z., wird die Stadt von einem großen Erdbeben ruiniert. Am Anfang der byzantinischen Zeit wird sie wieder rekonstruiert und besteht noch bis zum 8. Jh. Jedoch nach den wiederholten arabischen Angriffen und besonders nach dem neuen starken Erdbeben von 796 n.u.Z. wird sie endgültig verlassen.

Die Abteilung Geschichte und Archäologie der Universität Kreta unternimmt seit 1985 eine systematische Ausgrabung der antiken Eleutherna, mit Verantwortlichen die Professoren P. Themelis, A. Kalpaxis und N. Stambolidis.

Der Mechanismus von Eleutherna würde bei den Ausgrabungen des Archäologen Prof. Petros Themelis in der östlichen Seite der Stadt 1997 gefunden. Er wurde im Umfang des kleinen römischen Bades, Kleines Balaneion genannt, von den Trümmern des Daches eines Raumes, eventuell eines Lagers, bedeckt gefunden. Er wird um 365 n.u.Z. datiert. Ist also spätrömischer oder frühbyzantischer Zeit und wurde während des großen Erdbebens, das die Stadt erschütterte, verschüttert.

Der Mechanismus und seine Restaurierung

Der archäologische Fund war originell. Ein kleiner zylindrischer Körper, vom Rost besonders verdorben. Von einer Öffnung an seinem zerbrochenen Boden, konnte man die Existenz eines inneren Mechanismus, mit einer unerklärten Form und Funktion vermuten. Eine schwere eiserne Kette, vom Dach des Mechanismus abgerissen, wurde anbei gefunden. Und zu den anderen Funden der Ausgrabung im selben Raum gehörten eine Reihe von metallischen mit besonderer Fertigkeit geschmiedeten Gegenständen: viele Ackergeräte, Sicheln, Beile, eine eiserne Schaffschäre, eine Spange, ein eigenartiger Federmechanismus, ein eiserner Schlüssel,

ein silberner Ring und dazu manche metallische Bruchstücke, wie z.B. ein gebrochener Kettenring, die zum Mechanismus gehören könnten.

Nach der Restaurierung ist die Form des Mechanismus wesentlich klarer geworden. Die kupferne Außenfläche des zylindrischen Körpers enthüllte sich. Das Bruchstück mit der Kette und ihre Basis wurden am Dach des Mechanismus angeklebt. Die inneren Elemente des hohlen Körpers waren klarer erkennbar. Jedoch die Funktion blieb unerklärlich. Was war dieser Mechanismus? Diese Frage blieb unbeantwortet.

Bild 1: Der restaurierte Mechanismus von Eleutherna
(Im Archäologischen Museum von Rethymno aufbewahrt)

2 Analyse des Mechanismus

Die äußere Form

Im Februar 2007 wurde die Untersuchung des Mechanismus an Prof. D. Kalligeropoulos und Dr. S. Vasileiadou beauftragt. Beide waren Professoren der Antiken Griechischen Technologie, jedoch Ingenieure, Regelungstechniker aber keine Archäologen. Der restaurierte Mechanismus war im Keller des Archäologischen Museums von Rethymno aufbewahrt. Dort begann seine Untersuchung.

Die Frage über seine Funktion ist zurückgehalten. Zuerst untersuchte man seine äußere Form, sowie die Elemente die den Mechanismus zusammensetzen.

Die Kette und ihre Basis

Der Mechanismus besteht zunächst aus einer schweren, eisernen, geschmiedeten Kette, mit 10 Reifen und einer Gesamtlänge von circa 50cm. Sie hängt von einer starken, tempelförmigen, kreuzartigen, am Dach des zylindrischen Körpers des Mechanismus angepassten Basis ab. Die Basis hat eine Dicke von 7mm, Höhe 35mm und Breite 28mm. Die Reifen der Kette, von der Basis abgehend, werden sukzessiv kleiner und schmaler: Die Länge: 72-55mm, die Briete: 20-16.7mm, die Bleichbreite: 9-4mm, die Bleichdicke: 6-3mm. Unter den Beiliegenden des Mechanismus fand man ein Bruchstück, das eventuell ein ausgebrochenes Teil eines zusätzlichen, möglicherweise des letzten, Kettenreifes darstellte.

Fragen:

- Warum eine so starke Kette?
- Ihr freies Ende hing vom Zimmerdach oder war mit dem Mechanismus verbunden?
- Wie erklärt man die Reduktion der Dimensionen der Kettenreifen? Handelt es sich um eine Annäherung der gewünschten gesamten Kettenlänge oder der Dimensionen des letzten Reifes?
- Warum wiederum eine so starke Basis der Kette? War es eine Frage der Widerstandsfähigkeit des Mechanismus um ein großes Gewicht oder um eine besondere Kraft auszuhalten?

Der Körper des Mechanismus

Der Körper des Mechanismus ist zylindrisch, mit Durchmesser 67mm und Höhe 45mm, klein im Vergleich mit seiner Kette. Er besteht aus dem Dach, der zylindrischen Seitenfläche und seinem Boden.

Das Dach des Körpers

Das Dach ist kreisförmig, mit Durchmesser 67mm, und besteht aus zwei eisernen Platten, je 2mm dick. An der äußeren Platte war die Kettenbasis angepasst. Es gibt Spuren einer solchen Bolzenverbindung des größeren Teils der Basis mit dieser äußeren Platte. Die innere Platte scheint vom kleineren Durchmesser und an den inneren Elementen des Mechanismus angepasst zu sein. Zwischen den zwei Platten gab es wahrscheinlich einen kleinen Abstand,

nicht größer als 1mm. Das Bruchstück der äußeren Platte, das die Kettenbasis trug, ist während der Restaurierung exzentrisch am Dach des Mechanismus angepasst.

Frage:

– Warum diese exzentrische Position der Basis? Fehler oder was es vielleicht für die Funktion des Mechanismus sein Anhang oder die symmetrische Position der Basis nicht nötig?

Die zylindrische Seitenfläche des Körpers

Die zylindrische Seitenfläche des Körpers, mit einem Durchmesser von 67mm und eine Höhe von 45mm, bestehen aus einer eisernen 2mm dicken Platte und einer kupfernen Umhüllung, mit drei gleichbreiten, eingeschnitzten, dekorativen Kupferstreifen.

Die innere eiserne Platte hat einen senkrechten Spalt, eventuell beim Fall des Mechanismus wegen des Druckes an ihrer gehämmerten Schneide entstanden.

Bei der Untersuchung der Verbindung der Seitenfläche mit dem Dach des Mechanismus erkennt man daß die äußere Platte des Daches auf der eisernen Seitenplatte sitzt und auf ihr durch Hämmerung fest angepasst ist, während die innere Dachplatte die innere Seite der Zylinderfläche berührt. Entsprechendes gilt auch für die Platten des Bodens.

Frage:

– Weshalb die dekorative kupferne Umhüllung des Mechanismus? War es ein wertvolles Gerät? Oder war es nur ein Rostschutz?

Der Boden des Körpers

Der zerbrochene kreisförmige Boden des zylindrischen Körpers enthüllt seinen verborgenen internen Mechanismus. Die ausgerissenen Bruchteile seiner Platten machen mindestens den halben inneren Raum sichtbar. Der Boden besteht, wie das Dach, aus zwei Metallplatten, 2mm dick jede einzelne. Jedoch im Gegenteil zum Dach, sind diese beiden Bodenplatten nicht los sondern fest miteinander verbunden, wahrscheinlich gehämmert. Zwei äußere Bruchstücke passen an den Öffnungen des Bodens und vervollständigen ihn teilweise.

Auf der Außenseite des Bodens sind noch fünf (E1 bis E5) metallische Geschwülste deutlich zu erkennen. Die zwei äußere von denen (E2 bis E4) sind Verlängerungen und Lageransätze

der entsprechenden inneren Achsen des Mechanismus. Ein kupferner Ring im Umkreis des Bodens erlaubt die Annahme, daß der gesamte Boden mit einer Kupferplatte bedeckt war.

3 Der Innenraum des hohlen Körpers

Die Maschine

Aus den Öffnungen des Bodens sind im Innenraum des hohlen Körpers Teile eines komplexen Mechanismus zu erkennen: Senkrechte Achsen, unsymmetrische Scheidewände die den Raum in Kammern aufteilen und ein dünner, zackiger, um die zugänglichste jedoch unvollständige vertikale Achse gewickelter Fries.
Die Hinterseite des Körperraums bleibt unsichtbar und unzugänglich.

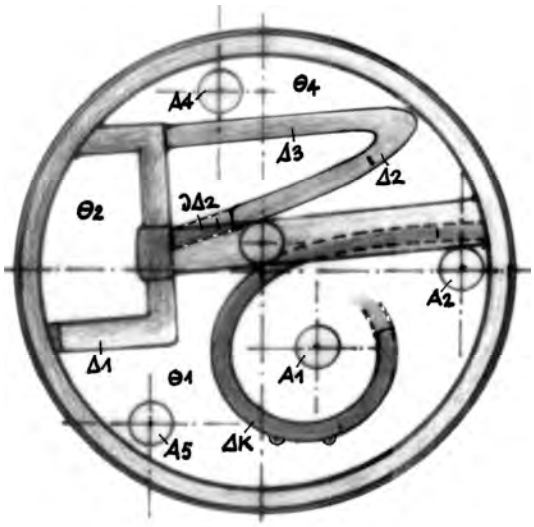


Bild 2: Schnitt im Innenraum des Körpers des Mechanismus
(Zylindrische Seitenfläche, Achsen, Scheidewände, Stab)

Die Achsen

Im Inneren des Körperraums sind zunächst vier Achsen (A1 bis A4) zu erkennen. Zu den Beiliegenden Funden befindet sich noch eine fünfte freie Achse (A5), die zum Mechanismus angepasst werden könnte. Die zwei externe, 6mm dicke Achsen A2 und A4 entsprechen den Geschwülsten E2 und E4 des Bodens. Die fünfte freie Achse A5 hat an ihrem freien Ende

auch eine Beule, die einer Geschwulst E5 am Boden des Mechanismus entsprechen könnte. Sie kann sogar, durch eine der Öffnungen des Bodens, im Inneren des Körpers angepasst werden.

Alle diese drei Achsen bilden ein gleichseitiges Dreieck und können somit als Stützachsen, die den Boden mit dem Dach des internen Mechanismus in einem festen Höhenabstand von circa 40mm festhalten, angedeutet werden.

Ungefähr im Zentrum des Innenraums stützt eine weitere, 6mm dicke senkrechte Achse A3, einen starken eisernen Stab δ, der den Boden des Mechanismus berührt. Und zuletzt erkennt man die vom Boden um 10mm kürzere Achse A1, im Zentrum des zylindrischen Frieses und von einer überlegenden Öffnung des Bodens von außen erreichbar.

Die Scheidewände

Eiserne, unsymmetrische, zwischen 2 und 4mm dicke, senkrecht zum Boden stehende Scheidewände unterteilen den Innenraum des Mechanismus in Kammern.

Die Scheidewand Δ1 ist π-förmig, 4mm dick, und an der Innenseite der zylindrischen Seitenfläche, sowie am Boden und Dach des Körpers fest angepasst. Sie bildet somit einen geschlossenen Raum Θ2 und stützt sich gleichzeitig auf den eisernen Stab δ. Die Scheidewände Δ2 und Δ3 sind dünner, 2mm dick, und sind schräg an der Scheidewand Δ1 der entgegengesetzten Innenseite der zylindrischen Seitenfläche und dem Stab δ angepasst. Δ3 ist jedoch kaum erreichbar.

Der Stab δ, 47mm lang, 6mm breit und 4mm dick, stützt sich auf der Achse A3 und der Scheidewand Δ1, dringt im geschlossenen Raum Θ2 und hat dort ein gehämmertes Ende. Zuletzt ist deutlich, um die unvollendete Achse A1, die zylindrische, oder genauer gesagt spiralförmige, kürzere Metallplatte ΔK zu erkennen. Sie ist nur 18mm hoch. Aus ihrer hellgrünen Farbe vermutet man daß sie aus Kupfer oder Bronze geschmiedet ist. Wahrscheinlich um eine bessere Elastizität zu erreichen. An ihrer oberen Kannte erkennt man zwei oder drei Zinken. Ihre hintere Kannte ist jedoch höher und an der zylindrischen Seitenwand wie auch an dem Stab δ angepasst.

4 Erläuterung der Funktion und totale Abbildung des Mechanismus

Röntgenuntersuchung des Mechanismus

Um die Suche nach der Form des Mechanismus zu vervollständigen war eine Untersuchung seiner optisch nicht zugänglichen Gebiete nötig.

Wir unternahmen zunächst eine Röntgenuntersuchung mit einem tragbaren Röntgenapparat im Archäologischen Museum von Rethymno. Wir röntgenten dann den Mechanismus im Röntgenlabor des Hauptkrankenhauses von Rethymno und untersuchten ihn zuletzt durch Axiale Tomographien im großen axialen Tomograph des selben Krankenhauses. Aus der Röntgenuntersuchung und der axialen Tomographie des Mechanismus bestätigten sich unsere ersten Haupthypothesen für seinen sichtbaren Teil und offenbarte sich, daß die Scheidewände $\Delta 2$ und $\Delta 3$ zu einer einzigen gekrümmten und elastischen Metallplatte, die die zentrale Achse A3 und den Stab δ berührte, gehörten.

So waren jetzt eine totale Rekonstruktion des Mechanismus und seine graphische Abbildung möglich. Und so könnte man jetzt mit der Frage über seine Funktion befassen.

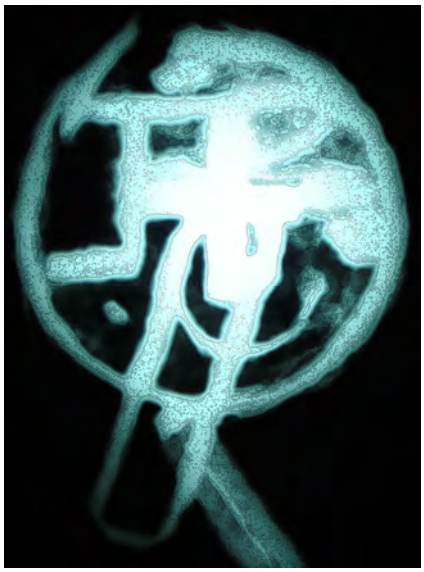


Bild 3: Axiale Tomographie des Mechanismus

Die zusammenhängenden archäologischen Funde

Die Untersuchung der zusammenhängenden archäologischen Funde, die mit dem vorliegenden Mechanismus Ähnlichkeiten erweisen, führte zu drei Beispielen:

1. Ein Mechanismus mit Kette, gefunden in den Ruinen eines römischen Palastes in Fishbourne von England, datiert um 290 n.u.Z., und vom Archäologen Barry Cunliffe als Vorhängeschloss (padlock) erkannt.
2. Ein Zylinderschloss vom 6sten Jh., gefunden in einer Kirche auf dem Berg Jelica von Westserbien zusammen mit anderen Ackergeräten sorgfältig versteckt.
3. Ein “antikes römisches Militärschloss aus Deutschland”, das im Katalog der englischen Gesellschaft Nokey mit der Bemerkung “sehr selten” zu finden ist.

Über die Funktion des Mechanismus

Die Ähnlichkeit dieser zusammenhängenden archäologischen Funde mit unserem Mechanismus, trotz ihrer ungeklärten Funktion, hat uns auf der Hypothese gelandet, daß es sich um ein Schloss handelt. Kompliziert, eigenartig, selten aber ... Schloss.

Auf Grund dieser Hypothese ist nun die Erläuterung der Funktion jedes einzelnen Elementes wie auch des gesamten Mechanismus die folgende.

Die Öffnungen des Bodens

Im Boden erscheinen drei Öffnungen ΑΠ1, ΑΠ2, ΑΠ5.

In der kreisförmigen Öffnung ΑΠ5 passt die freie Achse εΑ, die korrekt als die dritte Stützachse Α5 des Mechanismus gedeutet wurde.

Die quasi rechteckige Öffnung ΑΠ2, die zu der geschlossenen Kammer Θ2 der Scheidewand Δ1 führt, ist der Eingang des letzten Kettenreifs, der im Mechanismus eindringt um verschlossen zu werden. Hier passt genau der freie Reif εΚ, der eventuell den 11ten Reif der Kette ausmacht.

Die Öffnung ΑΠ1, gegenüber der kürzeren Achse Α1, hat die Form eines Schlüsselloches und ist der Eingang des Schlüssels.

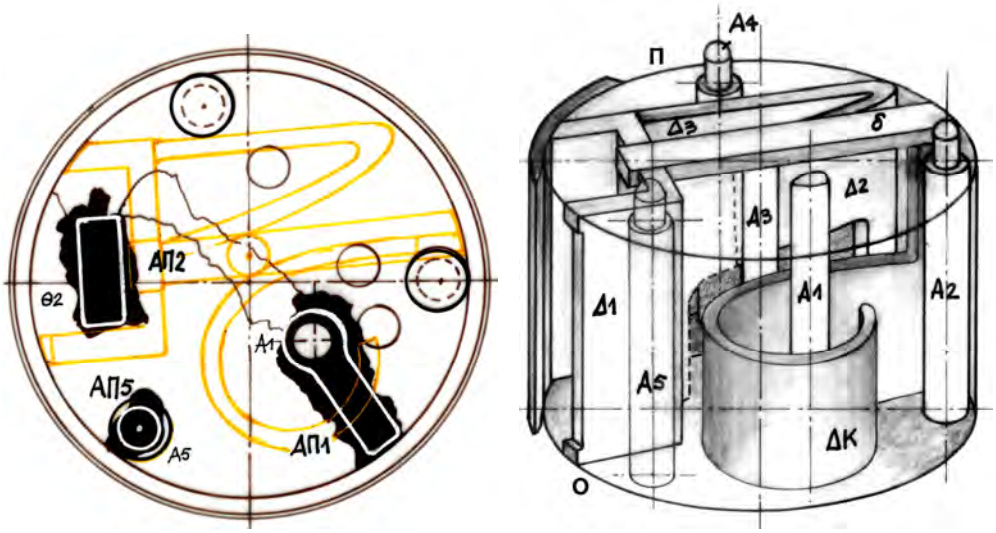


Bild 4: Abbildungen des Bodens und des Innenraums des Körpers
(Öffnungen des Bodens, Position der Achsen, der Metallfriesen und des Riemens)

Die Rolle der Achsen und des Stabes

Der Stab δ ist der Riegel des Schlosses. Er schleppt auf den Boden П des Mechanismus, dringt durch die rechteckige Öffnung der starken Scheidewand Δ1 in die Kammer Θ2 ein und verschließt somit den eingedrungenen Kettenreif. Er wird von oben von der Achse A3 gestützt und ist fest mit einem Ende des spiralförmigen Frieses ΔK verbunden.

Die Achsen A2, A4 und A5 verbinden, wie vermutet, symmetrisch den Dach mit dem Boden des Mechanismus, auf denen sie fest gelagert sind.

Die Achsen A1 und A3 sind fest am Dach des Mechanismus angepasst.

Die Rolle der anderen Metallplatten

Der spiralförmigen Fries ΔK bewirkt die Bewegung des Riegels δ, der mit dem Fries an seinem Ende fest verbunden ist. Wenn ein Schlüssel durch die Öffnung АП1 im Mechanismus eindringt, sich an den Zinken des Frieses ΔK anpasst, und sich um die Achse A1 dreht, zieht er den Riegel in den Raum Θ2 und verschließt das Schloss.

Die andere Metallplatte Δ2 bildet die Sicherung des Mechanismus, sie verhindert nämlich die Rückkehr des Riegels δ nach dem Verschloss. Diese Sicherung erfolgt wahrscheinlich durch

einen Schlitz oder eine Beule an der Seite des Stabes, bei dem das Ende der Platte Δ2 sich halten würde.

Der Schlüssel

Im selben Raum mit dem Mechanismus ist bei der Ausgrabung auch ein Schlüssel gefunden. Der Schlüssel ist 85mm hoch und 30mm breit. Er hat in seiner zylindrischen Achse eine Öffnung von 6-7mm Durchmesser (die Achse A1 ist 5.5-6mm dick), die Tiefe dieser Öffnung ist circa 40mm (während die Achse A1 30mm hoch ist), der Schlüssel hat im Abstand von 15mm von seiner Achse einen Einschnitt mit 14mm Tiefe und 3mm Breite (während der Abstand der Achse A1 und dem Fries ΔK auch 15mm und seine Dicke 2mm beträgt). Der Schlüssel passt genau und kann somit der eigene Schlüssel des Mechanismus sein. An seiner Innenseite wurde sogar ein goldener Überzug erkannt. Ein goldener Schlüssel? ...

Rekonstruktionen des Mechanismus

Funktionelle reale Rekonstruktion des eisernen Mechanismus

Der Mechanismus wurde im Maßstab 2:1 im April 2008 von dem begabten Schmied Dimitris Chatzis in Pyrgos der Insel Tinos nach den vorgegebenen Abbildungen treu rekonstruiert.

Bild 5: Die eiserne Rekonstruktion des Mechanismus

Digitale dreidimensionale Simulation des Mechanismus und seiner Funktion

Dr. Soultana Vasileiadou und Techn. Ing. Anastassios Gkamaris bearbeiteten sorgfältig eine exakte dreidimensionale elektronische Simulation des Mechanismus mit dem Programm 3D Studio Max. Dieses Programm ermöglichte auch die Simulation der vielfältigen Ansichten wie auch der speziellen Bewegungen jedes einzelnen Elementes und vervollständigte somit die optische Darstellung der gesamten Funktion des Mechanismus in einer digitalen Umwelt.

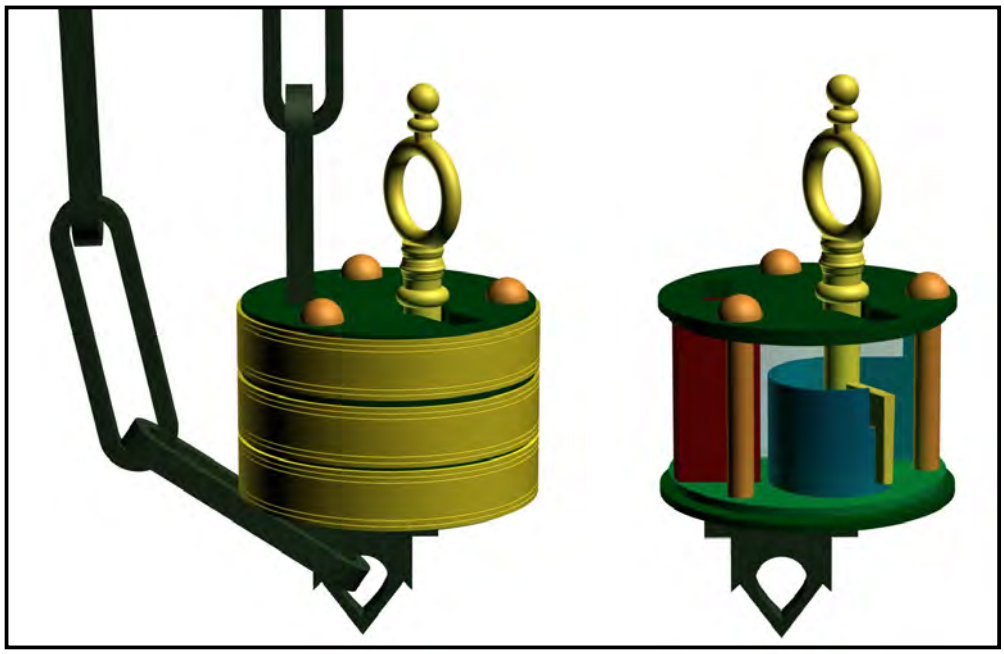


Bild 6: Dreidimensionale Abbildungen des Mechanismus

References

[1] Petros Themelis, Antike Eleutherna. Athen 2002, ISBN 960-214-033-X.

[2] Barry Cunliffe, *Fishbourne – A Roman Palace and its Garden*. Thames & Hudson, 1971.

[3] Mihailo Milinkovic, *Die byzantinische Höhenanlage auf der Jelica in Serbien– ein Beispiel aus dem*. Starinar. LI: 71-130, Beograd, 2002.

[4] Keyless Lock Store Nokey, Ancient Roman Key Gallery:
http://www.nokey.com/ankeymus.html

Odometriebasierte Fehlerdiagnose für quasiomnidirektionale mobile Radroboter

Theresa Rienmüller[*], Christoph Gruber, Michael Hofbaur
UMIT,
Institut für Elektrotechnik, Elektronik und Bioengineering,
Institut für Automatisierungs- und Regelungstechnik
Eduard-Wallnöfer-Zentrum 1a, Hall in Tirol
theresa.rienmueller@umit.at[†], christoph.gruber@umit.at
michael.hofbaur@umit.at

1 Einleitung

Der Einsatz eines mobilen Radroboters beinhaltet normalerweise die Vorgabe, eine gewisse Strecke abzufahren, bzw. ein bestimmtes Ziel zu erreichen. Um diese Aufgabe ohne direkte Einwirkung des Menschen erfüllen zu können, muss der Roboter einerseits wissen, wo er sich selbst befindet und andererseits selbstständig erkennen können, ob ein Fehler in seinem Selbstlokalisierungssystem oder gar ein mechanischer Defekt im Fahrwerk vorliegt. Die Wahrscheinlichkeit eines mechanischen Defekts ist zwar relativ gering, dennoch ist die Auswirkung auf die Aufgabe des Roboters mitunter fatal: Auf der einen Seite kann ein unentdeckter Ausfall zu weiteren Beschädigungen des Roboters oder gar zur Verletzung ihn umgebender Personen führen. Auf der anderen Seite kann die Mission ohne eine mögliche Rekonfiguration nicht mehr weiterverfolgt werden. Doch nicht nur schwerwiegende Defekte beeinflussen den Roboter auf dem Weg zu seinem Ziel. Mitunter reichen bereits kleine Störungen oder Fehler aus, um die Qualität der Schätzung der eigenen Position und Orientierung drastisch zu verschlechtern.

Ziel der vorliegenden Arbeit ist es, sowohl diese Störungen als auch die mechanischen Defekte zu erkennen und einer bestimmten Komponente im Roboter zuzuordnen. Das Hauptaugenmerk liegt dabei auf Robotern mit mehreren gelenkten Standardrädern. Solche Roboter haben den Vorteil, dass sie, nach Ausrichtung der einzelnen Räder, in jede beliebige Richtung fahren können. Die Herausforderung liegt hierbei in der Einhaltung der kinematischen Beschränkungen, das bedeutet in der exakten Abstimmung der Lenkwinkel und Raddrehzahlen. Anhand der von den Gebern gelieferten Messwerte muss wieder auf die Geschwindigkeit des Roboters geschlossen werden, um seine momentane Position bestimmen zu können. Da diese Abstimmung in einem realen System niemals perfekt sein wird, ist auch diese Umkehrung des

[*]Doctoral School Informations- und Kommunikationstechnologie,TUGraz, Rechbauerstrasse 12, 8010 Graz
[†]Korrespondenz bitte an diese Adresse

Problems nicht ganz einfach. Mit Hilfe der sich aus den Messwerten ergebenden Diskrepanzen in der Geschwindigkeitsschätzung des Roboters, sowohl einzelner Radpaare untereinander als auch gegenüber der Vorgabe, kann auf Defekte im Fahrwerk geschlossen werden.

Im nächsten Kapitel werden zunächst einige Arbeiten, die sich mit dem Thema der Fehlererkennung für mobile Roboter beschäftigen, vorgestellt. Danach wird auf die Besonderheiten der Kinematik quasiomnidirektionaler Roboter eingegangen und es werden verschiedene Betrachtungsweisen der kinematischen Gleichungen vorgestellt. Im vierten Kapitel werden schließlich mögliche Fehlerfälle vorgestellt und der Diagnosealgorithmus erläutert. Kapitel fünf beschreibt die verwendete Konfiguration unserer Roboterplattform und liefert experimentelle Ergebnisse.

2 Verwandte Arbeiten

Für die Aufgabe der Fehlererkennung von Radrobotern gibt es mehrere interessante Ansätze. Einerseits wurden modellbasierte Verfahren entwickelt, die für jeden betrachteten Fehlerfall ein dezidiertes Modell beinhalten und anhand dessen den Zustand des Roboters zu schätzen versuchen [12, 3]. Auf der anderen Seite wurden Algorithmen entwickelt, die auf dem Vergleich mehrerer Sensoren basieren [10, 2]. Keiner dieser Ansätze beschäftigt sich jedoch näher mit den kinematischen Beschränkungen des Roboters und den Auswirkungen eines Fehlers auf das gesamte Fahrwerk. Dies liegt zum Großteil darin begründet, dass es sich bei den betrachteten Roboterfahrwerken meist um einfachere Fahrwerke, wie beispielsweise solche mit Differentialantrieb, handelt.

3 Die Kinematik quasiomnidirektionaler mobiler Radroboter

Ein omnidirektionaler Roboter zeichnet sich dadurch aus, dass er zu jeder Zeit aus dem Stand in jede beliebige Richtung losfahren kann. Bei einem quasiomnidirektionalen Roboterfahrwerk ist für die Fahrt in eine beliebige Richtung i.A. eine vorherige Ausrichtung der Räder notwendig. In dieser Arbeit beschäftigen wir uns mit ebensolchen Robotern mit einer Radanzahl von $n \geq 3$. In Abbildung 1 ist beispielhaft ein Roboter mit $n = 3$ Rädern in seinem lokalen Koordinatensystem $\Sigma_R : \{0_R; x_R, y_R\}$ abgebildet. Die Anordnung der Räder wird dabei durch deren Aufstandspunkt in Polarkoordinaten in Σ_R beschrieben, wobei l_i den Abstand des Aufstandspunktes des i-ten Rades vom lokalen Koordinatenursprung und α_i den Winkel zur x_R-Achse beschreibt. Des weiteren spielt natürlich der Radradius r_i eine wichtige Rolle bei der Berechnung der Robotergeschwindigkeit, ebenso wie die mit der Zeit veränderlichen Größen β_i und $\dot{\varphi}_i$, der Lenkwinkel bzw. die Raddrehzahl.

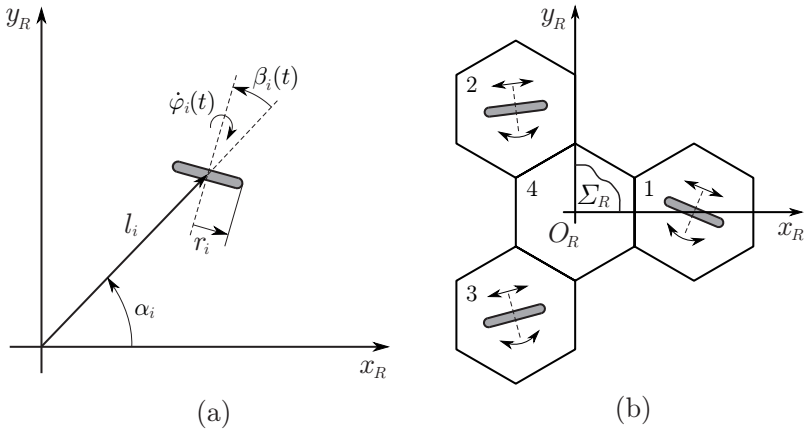


Abbildung 1: a) Parameter eines Rades im lokalen Koordinatensystem des Roboters. b) Der Roboter in seinem lokalen Koordinatensystem.

Die Geschwindigkeit des Roboters setzt sich aus der Bewegung in Richtung x_R und y_R zusammen (translatorische Komponente), sowie aus der Drehung um die Hochachse. Die Geschwindigkeit des Roboters relativ zu einem globalen Koordinatensystem, ausgedrückt im lokalen Koordinatensystem Σ_R wird damit zu:

$$\xi = \begin{bmatrix} \dot{x}_R & \dot{y}_R & \dot{\theta}_R \end{bmatrix}^T \tag{1}$$

Diese Bewegung kann auch als reine Drehbewegung um einen sogenannten Momentanpol beschrieben werden. Dieser Punkt berechnet sich im lokalen Koordinatensystem des Roboters anhand der Geschwindigkeitsinformation wie folgt:

$$MP_R = \begin{bmatrix} x_{MP} \\ y_{MP} \end{bmatrix} = \begin{bmatrix} -\dot{y}_R/\dot{\theta}_R \\ \dot{x}_R/\dot{\theta}_R \end{bmatrix} \tag{2}$$

Geometrisch betrachtet bedeutet das, dass sich alle Radachsen des Roboters in genau diesem Punkt schneiden (siehe Abb. 2).

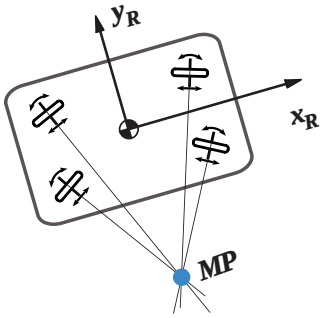


Abbildung 2: Der Momentanpol der Bewegung als Schnittpunkt der Radachsen.

Das kinematische Modell für einen solchen Roboter kann über die sogenannten Roll- und Gleitbedingungen abgeleitet werden [4]. Für ein Rad i des Roboters beschreibt die Rollbedingung

$$\begin{aligned} r_i\dot{\varphi}_i &= \begin{bmatrix} sin(\alpha_i+\beta_i) & -cos(\alpha_i+\beta_i) & l_i \cdot cos(\beta_i) \end{bmatrix} \cdot \xi \\ &= \mathbf{j}_i^T(\beta_i)\xi \end{aligned} \tag{3}$$

den Zusammenhang zwischen dem Lenkwinkel und der Raddrehzahl mit der Geschwindigkeit des gesamten Roboters. Diese Bedingung beschreibt die ideale Rollbewegung eines Rades normal zur Radachse. Zusätzlich dazu stellt die sogenannte Gleitbedingung sicher, dass die Geschwindigkeit des Roboters bezogen auf das betrachtete Rad keine Komponente quer zur Abrollebene, d.h. in Richtung der Radachse ausführt:

$$\begin{aligned} 0 &= \begin{bmatrix} cos(\alpha_i+\beta_i) & sin(\alpha_i+\beta_i) & l_i \cdot sin(\beta_i) \end{bmatrix} \cdot \xi \\ &= \mathbf{c}_i^T(\beta_i)\xi \ . \end{aligned} \tag{4}$$

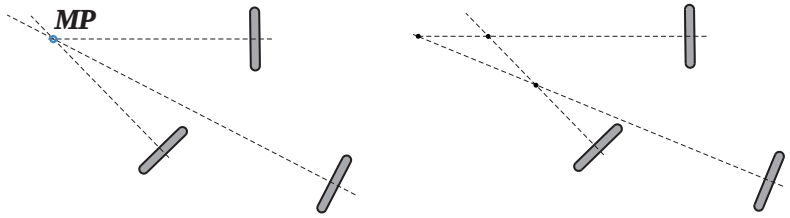
Da der Roboter üblicherweise über mehr als ein Rad verfügt, werden diese Gleichungen nun für alle n Räder zusammengefasst. Dabei werden die einzelnen Zeilenvektoren $\mathbf{j}_i^T(\beta_i)$ zeilenweise in die Matrix $\mathbf{J}_1(\beta)$ eingefügt. β beschreibt dabei den Spaltenvektor aller Lenkwinkel β_i. Auf dieselbe Art wird auch die Matrix für die Gleitbedingungen aufgebaut. Die Zeilenvektoren $\mathbf{c}_i^T(\beta_i)$ werden zur Matrix $\mathbf{C}_1(\beta)$ zusammengefasst. Daraus ergeben sich die Roll- und Gleitbedingungen in Matrizenschreibweise:

$$\mathbf{J}_2\dot{\varphi} = \mathbf{J}_1(\beta)\xi \tag{5}$$

$$\mathbf{0} = \mathbf{C}_1(\beta)\xi \tag{6}$$

$\mathbf{J}_2$ ist dabei eine konstante Diagonalmatrix mit den Radradien als Einträge in der Hauptdiagonale. $\mathbf{C}_1(\beta)$ und $\mathbf{J}_1(\beta)$ sind jeweils $n \times 3$ Matrizen. Anhand von Gleichung 6 erkennt man, dass die Gleitbedingung erfüllt ist, wenn $\xi \in null(\mathbf{C}_1(\beta))$. Koordiniert man die Lenkwinkel so, dass $rang(\mathbf{C}_1(\beta)) = 2$, dann gilt $dim(null(\mathbf{C}_1(\beta))) = 1$ und es verbleibt ein Freiheitsgrad für die Vorgabe von ξ.

Geometrisch betrachtet entspricht diese Bedingung dem Umstand, dass sich alle Radachsen des Roboters in genau einem Punkt schneiden (siehe Abb. 3a)). Dieser Punkt entspricht dann dem Momentanpol und die Bewegung des Roboters ergibt sich durch die Vorgabe der Rotationsgeschwindigkeit um diesen Punkt. Aufgrund von Ungenauigkeiten in der Abstimmung schneiden sich in Wirklichkeit nicht alle Radachsen in genau einem Punkt (siehe Abb. 3b). Der Momentanpol der *tatsächlichen* Bewegung des Roboters ist dadurch nicht mehr eindeutig durch den Schnittpunkt aller Achsen bestimmbar.

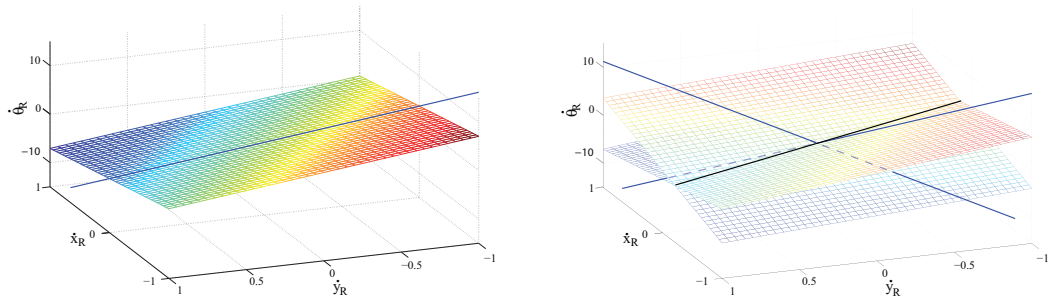


(a) Ideale Situation (b) Reale Situation

Abbildung 3: Schnittpunkt der Radachsen. (a) Ideale Situation: Alle Achsen schneiden sich in einem Punkt. (b) Reale Situation: Ein Rad ist nicht optimal ausgerichtet. Dies führt zu drei verschiedenen Schnittpunkten.

3.1 Der Geschwindigkeitsraum

Ein wesentlicher Nachteil der Betrachtung der Bewegung des Roboters mit Hilfe des Momentanpols in der (x_R, y_R)-Ebene stellt der Verlust der Geschwindigkeitsinformation dar, da dieser lediglich die Lage des Drehpunktes, nicht aber die Drehgeschwindigkeit um diesen Punkt definiert (siehe auch Gleichung 2). Aus der Lage des Momentanpols alleine kann die Bewegung des Roboters daher nicht mehr eindeutig rekonstruiert werden. Um eine Aussage über die möglichen Geschwindigkeiten, die ein Roboter mit einer gegebenen Radkonfiguration fahren kann, treffen zu können, wird daher der sogenannte Geschwindigkeitsraum eingeführt, dessen Achsen die entsprechenden Geschwindigkeiten darstellen: $\{0_\xi; \dot{x}_R, \dot{y}_R, \dot{\theta}_R\}$. Betrachtet man nun in diesem Raum welche Bewegungen der Roboter ausführen kann, erkennt man, dass alle möglichen Bewegungen des Roboters, für die die Gleitbedingung für *ein* gelenktes Standardrad (Gl. 4) nicht verletzt wird, einer Ebene in diesem Raum entsprechen. Bei einer Veränderung des Lenkwinkels β_i von Rad i, wird diese Ebene um eine Achse, die durch 0_ξ verläuft gedreht. Die Richtung dieser Achse wird durch die Koordinaten des Radaufstandspunktes in Σ_R bestimmt (siehe Abb. 4a)). Betrachtet man nun zwei unterschiedliche Räder, erhält man zwei Ebenen im Geschwindigkeitsraum und als mögliche Bewegung des Roboters die Schnittgerade der beiden Ebenen. Dieser Fall ist in Abbildung 4b) dargestellt. Auch hier gilt wieder, dass sich für $n \geq 3$ Räder im idealen Fall alle Schnittgeraden überlappen. Im Realfall ergeben sich natürlich für unterschiedliche Radpaare unterschiedliche Schnittgeraden. Im folgenden werden wir daher den Begriff der Geschwindigkeitskandidaten $\xi_{c,(k,l)}$ definieren, die sich jeweils anhand der Konfiguration zweier Räder k und l ergeben.



(a) Rad 1 (b) Rad 1 und Rad 2

Abbildung 4: a) Mögliche Bewegungen des Roboters, eingeschränkt durch die Gleitbedingung von Rad 1. Die blau eingezeichnete Gerade stellt dabei die Achse dar, um die die Ebene bei einer Veränderung des Lenkwinkels gedreht wird. b) Die zweite Ebene stellt die mögliche Bewegung des Roboters, eingeschränkt durch die Gleitbedingung von Rad 2 dar. Die blau eingezeichneten Geraden definieren die Rotationsachsen der Ebenen bei Veränderung des Lenkwinkels. Die Schnittgerade der beiden Ebenen ist schwarz dargestellt.

3.2 Vorwärtskinematik

Bei der Vorwärtskinematik geht es darum, anhand der gegebenen Konfiguration der einzelnen Räder $(\beta_i, \dot{\varphi}_i)$, die Geschwindigkeit des Roboters zu bestimmen. Unter der Annahme idealer Verhältnisse kann die Geschwindigkeit des Roboters mithilfe der Roll- und Gleitbedingungen (Gl.5-6) eindeutig bestimmt werden. Die Gleitbedingungsmatrix ist eine $n \times 3$ Matrix, die bei geeigneter Koordination der Lenkwinkel einen Rang von zwei hat. Es ist daher ausreichend, zwei linear unabhängige Gleichungen der Matrix zu betrachten, um die Einschränkung der Bewegung des Roboters durch die Gleitbedingung zu erfassen. Definieren wir nun eine reduzierte Gleitbedingungsmatrix $\mathbf{C}(\beta_k, \beta_l)$, die sich nur aus den linear unabhängigen Gleitbedingungen für die Räder k und l zusammensetzt. In weiterer Folge lässt sich nun nach [4] eine Matrix $\Sigma(\beta_k, \beta_l)$ definieren, deren Spalten den Nullraum von $\mathbf{C}(\beta_k, \beta_l)$ aufspannen:

$$span\{col(\Sigma(\beta_k, \beta_l))\} = null(\mathbf{C}(\beta_k, \beta_l)) \ . \tag{7}$$

Da die reduzierte Gleitbedingungsmatrix einen Rang von zwei hat, ist $\Sigma(\beta_k, \beta_l)$ ein Vektor. Mit diesem Zusammenhang kann die Robotergeschwindigkeit bereits eingeschränkt werden:

$$\xi = \Sigma(\beta_k, \beta_l) \cdot \eta \ . \tag{8}$$

Um nun den Faktor η zu bestimmen, ist eine weitere Gleichung vonnöten, da bisher nur zwei linear unabhängige Gleichungen für die Bestimmung des dreielementigen Vektors ξ herangezogen wurden. Daher wird nun Gleichung 8 in die Rollbedingung von Rad k[1] eingesetzt und

[1] Ohne Einschränkung der Allgemeinheit. Es kann natürlich auch die Rollbedingung von Rad l eingesetzt werden.

die resultierende Gleichung für die Vorwärtskinematik ergibt sich zu:

$$\xi = \Sigma(\beta_k, \beta_l) \frac{r_k \dot{\varphi}_k}{\mathbf{j}_k^T(\beta_k)\Sigma(\beta_k, \beta_l)} =: \xi_{c,(l,k)} \tag{9}$$

Man erkennt, dass in dieser Gleichung Singularitäten vorkommen. Bei quasiomnidirektionalen Fahrwerken treten diese nur dann auf, wenn: A) Der Momentanpol in den Aufstandspunkt eines der beiden Räder fällt, oder B) die Achsen der beiden Räder sich decken. Singularitäten des Typs A) sind strukturelle Eigenheiten quasiomnidirektionaler Fahrwerke, Singularitäten vom Typ B) verletzen die Forderung $rang(\mathbf{C}(\beta_k, \beta_l)) = 2$ und entstehen daher aus der speziellen Wahl eines Radpaars k, l [7].

4 Diagnose

Jedes Rad im Roboter verfügt sowohl über einen Lenk- als auch einen Radantrieb. Geber stellen hierbei die nötige Sensorinformation für die entsprechenden Werte zur Verfügung $(\tilde{\beta}_i, \tilde{\dot{\varphi}}_i)$. Im Folgenden werden nun die in dieser Arbeit berücksichtigten Fehler der einzelnen Komponenten näher beschrieben.

4.1 Mögliche Fehlerfälle

Zunächst unterscheiden wir drei unterschiedliche Arten von Fehlern, je nach Ort des Auftretens: a) Fehler in der Pfadplanung: Wir erlauben in Geometrie und Funktionalität veränderliche Fahrwerke. Ein Fahrbefehl des Pfadplaners kann daher für den momentan gültigen Betriebs- oder Fehlerzustand ungeeignet sein. b) Abweichungen im gewünschten Verhalten aufgrund mechanischer Defekte des Fahrwerks oder durch Einflüsse aus der Umgebung und c) Sensorfehler. In weiterer Folge unterscheiden wir verschiedene Defekte der einzelnen Komponenten. Da jedes Rad unseres Roboters mit Lenkantrieb und Radantrieb ausgestattet ist, ergeben sich pro Radeinheit zwei mögliche fehlerhafte Komponenten und die dazu passenden Sensorfehler. Außerdem kann eine Störung von außen das Verhalten des Fahrwerks beeinflussen. Tabelle 1 gibt einen Überblick über die betrachteten Fehlerfälle.

Komponente	Fehler	Simulation
Rad i	blockiert	$\dot{\varphi}_i = 0$
	tw. blockiert	$\dot{\varphi}_i$ wechselnd
	freilaufend	Motor OFF
Lenkung i	blockiert	Motor OFF
	Referenzierfehler (nicht messbar)	$\beta_i = \beta_i + \beta_{i,ref}$
	Offset (messbar)	$\beta_i = \beta_i + \beta_{i,off}$
Sensor i	Fehlfunktion	$y_i =$Wert
Störungen		

Tabelle 1: Mögliche Fehlerfälle der einzelnen Komponenten. Diese können als Einzel- oder Mehrfachfehler auftreten.

Diese Fehler können dabei einzeln oder auch in unterschiedlichen Kombinationen auftreten. Eine Störung von außen, wie etwa ein Zusammenstoß mit einem anderen Gegenstand, bewirkt i.A. eine Änderung des Verhaltens des gesamten Fahrwerks.

Analytische Redundanzbedingungen

Modellbasierte Diagnose erfordert in der Regel zwei Schritte [11]: Als erstes werden Unstimmigkeiten der aktuellen (gemessenen) Werte gegenüber den erwarteten Werten festgestellt. Solche Diskrepanzen werdenmittels sogenannter Residuen mathematisch erfasst: Sie werden im fehlerfreien Fall zu Null und weisen andernfalls auf mögliche Fehler im System hin [6]. Im nächsten Schritt muss dann anhand dieser Information und den daraus resultierenden Fehlerkandidaten eine Entscheidung getroffen werden, welcher Fehler aufgetreten ist bzw. welche Komponente im System fehlerhaft ist. Die im ersten Schritt erfassten Unstimmigkeiten erfordern einen Vergleich, wofür Redundanz vonnöten ist. Wie diese Redundanz entsteht ist dabei vom System und den vorhandenen Sensoren abhängig. Es gibt im Wesentlichen zwei Möglichkeiten: Einerseits kann diese Redundanz analytischer Natur sein und sich aufgrund der funktionalen Abhängigkeiten zwischen den Prozessvariablen ergeben, d.h. anhand der algebraischen oder zeitlichen Zusammenhänge im Modell. Auf der anderen Seite kann Redundanz auch dadurch entstehen, dass beispielsweise redundante Sensoren verwendet werden. In unserem Fall ist jede Radeinheit mit zwei Sensoren, jeweils einem für die Raddrehzahl und den Lenkwinkel ausgestattet. Die analytische Redundanz wird hierbei durch die Berechnung der Geschwindigkeit erzeugt: Jeweils zwei Lenkwinkel und eine Raddrehzahl sind ausreichend, um eine Schätzung für die Geschwindigkeit $\tilde{\xi}_{c,(k,l)}$ basierend auf den Messungen der Werte zweier Räder l und k des Roboters zu erhalten (siehe Gl. 9). Im Fall eines Roboters mit $n = 3$ Rädern ergeben sich so 6 Möglichkeiten, die Geschwindigkeit zu berechnen.
Die Residuen werden generiert, indem einerseits die Geschwindigkeitskandidaten mit dem Sollwert und andererseits jeweils zwei Geschwindigkeitskandidaten miteinander verglichen werden. Ist die Bedingung für einen Geschwindigkeitskandidaten erfüllt, ist das zugehörige

Residuum der Bedingung Null. Andernfalls ergibt sich ein Wert verschieden von Null[2]. Daher definieren wir die Residuen $r_{j,t}$ zum Zeitpunkt t als Boole'sche Fehlerindikatoren wie folgt

$$r_{j,t}(\xi_{Soll}, \xi_{Ist}) = \begin{cases} 0 & \xi_{Soll}, \xi_{Ist} \text{ stimmen überein} \\ 1 & \text{sonst} \end{cases} \tag{10}$$

4.2 Ablauf der Diagnose

Da wir einige verschiedene Arten von Fehlern mit unterschiedlichen Eigenschaften betrachten, wird ein mehrstufiger Ansatz verwendet, der diesen Unterschieden gerecht wird.

Schritt I: Überprüfung des Fahrbefehls
Unter der Annahme, dass der Sollwert für die Geschwindigkeit vom Pfadplaner korrekt berechnet wurde und fahrwerksimmanente Singularitäten vom Typ A) bereits berücksichtigt wurden, werden in diesem Schritt lediglich Singularitäten vom Typ B) ausgeschlossen. Durch diesen Schritt wird sichergestellt, dass für die Diagnose keine Geschwindigkeitskandidaten basierend auf zwei Rädern mit sich deckenden Radachsen herangezogen werden.[3]

Schritt II: Vergleich des Sollwertes mit den Geschwindigkeitskandidaten
Im nächsten Schritt werden nun anhand der *gemessenen* Werte der einzelnen Räder $(\tilde{\beta}_i, \tilde{\dot{\varphi}}_i)$, die Geschwindigkeitskandidaten berechnet. Anschließend werden die Residuen $r_{j,t}(\xi_{Soll}, \tilde{\xi}_{c,(k,l)})$ für alle $j = 1, ..., 2\binom{n}{2}$ Geschwindigkeitskandidaten nach Gleichung (10) gebildet.
Fasst man die Ergebnisse der Auswertungen aller Gleichungen zusammen, ergibt sich für jede fehlerhafte Komponente im System ein spezielles Muster von erfüllten/nicht erfüllten Bedingungen. Dieses wird als Fehlersignatur [5, 1]

$$S_t = \begin{bmatrix} r_{1,t} & r_{2,t} & \dots r_{q,t} \end{bmatrix} \ , \ q = 2 \cdot \binom{n}{2} \ , \tag{11}$$

bezeichnet und erlaubt eine Identifikation der fehlerhaften Komponente im System (Komponente bezeichnet dabei den mechanischen Teil und den zugehörigen Sensor.). Ist beispielsweise die Raddrehzahl von Rad 3 betroffen, erfüllen alle Geschwindigkeitskandidaten, deren Berechnung $\tilde{\dot{\varphi}}_3$ beinhaltet, die Redundanzbedingung nicht; Die Auswertung aller Redundanzbedingungen führt daher zu folgender Fehlersignatur:

$$S = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

und die Raddrehzahl von Rad 3 wird als fehlerhaft erkannt. Diese Situation ist auch in Abbildung 5 im Geschwindigkeitsraum dargestellt. Nur die Geschwindigkeitskandidaten, deren Berechnung $\tilde{\dot{\varphi}}_3$ beinhalten, zeigen deutliche Abweichungen vom gewünschten Verhalten.

[2] Durch numerische Ungenauigkeiten, ungenügendes Auflösungsvermögen, Messfehler und Rauschen können diese Bedingungen im Allgemeinen nicht exakt erfüllt werden. Es ergibt sich daher ein Wert verschieden von Null. Man definiert entsprechende Schwellwerte, nach deren Überschreitung die Bedingung als nicht mehr erfüllt gilt.

[3] In diesem Schritt kann auch eine Überprüfung des Geschwindigkeitssollwertes bezüglich der Singularitäten vom Typ A) und damit eine Validierung der vom Pfadplaner bestimmten Geschwindigkeitsvorgabe durchgeführt werden

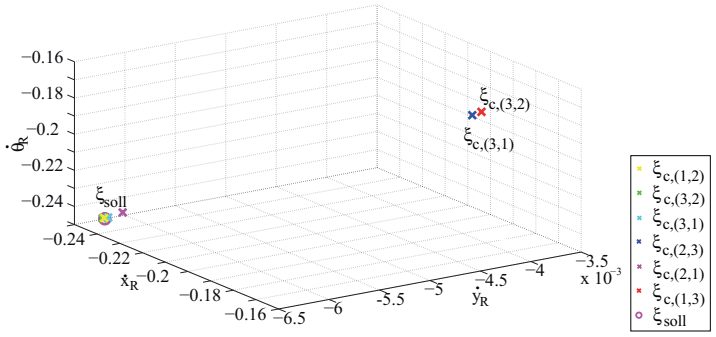


Abbildung 5: Vergleich der Geschwindigkeitskandidaten mit dem Sollwert im Geschwindigkeitsraum. Man erkennt, dass Kandidaten, deren Berechnung die Raddrehzahl von Rad 3 beinhalten, eine deutliche Abweichungen gegenüber dem Sollwert und den anderen Kandidaten aufweisen.

Die Auswirkungen eines Fehlers im Fahrwerk auf andere Komponenten sind abhängig von der Art des Fehlers. Das Blockieren eines Rades bewirkt beispielsweise auch eine deutliche Änderung der Drehzahlen der anderen Räder und damit eine Abweichung vom Sollwert. Dies wirkt sich natürlich auch auf die Auswertung der Residuen aus und die fehlerhafte Komponente kann auf diese Weise nicht mehr eindeutig bestimmt werden. Aus diesem Grund werden im nächsten Schritt die einzelnen Geschwindigkeitskandidaten miteinander verglichen.

Schritt III: Vergleich der einzelnen Geschwindigkeitskandidaten
In diesem Schritt werden Redundanzbedingungen für die einzelnen Geschwindigkeitskandidaten untereinander definiert, um die eigentlich fehlerhafte Komponente bestimmen zu können:

$$r_{j,t}(\tilde{\xi}_{c,(k,l)}, \tilde{\xi}_{c,(m,n)}) = \begin{cases} 0 & \tilde{\xi}_{c,(k,l)}, \tilde{\xi}_{c,(m,n)} \text{ stimmen überein} \\ 1 & \text{sonst} \end{cases} . \tag{12}$$

k,l,m,n sind dabei so zu wählen, dass alle möglichen Kombinationen von Komponenten abgedeckt werden. Hierbei wird von der Annahme ausgegangen, dass Komponenten, deren Geschwindigkeitskandidaten trotz eventuell vorhandener Abweichung vom Sollwert die Redundanzbedingung erfüllen, von der fehlerhaften Komponente betroffen sind, jedoch nicht die Ursache des Problems darstellen.

Schritt IV: Vergleich von $\dot{\Theta}_R$ mit einem Gyroskop
Als zusätzliches Messinstrument befindet sich ein Inertialnavigationssystem auf dem Roboter. Mithilfe des Gyroskops kann direkt ein Vergleichswert für die dritte Komponente des Geschwindigkeitsvektors $\dot{\Theta}_R$ gewonnen werden. Auf diese Weise ist es möglich Sensorfehler von Aktuatorfehlern zu unterscheiden. Eine deutliche Abweichung einer aus Messungen berechneten Geschwindigkeit in Kombination mit einem annehmbaren Messwert des Gyroskops liefert daher einen Hinweis auf einen Sensorfehler, wenn ein mechanischer Defekt, eine stärkere Beeinflussung des Fahrverhaltens bedeuten würde.

Auswertung der Residuen
Der Vergleich zweier berechneter Geschwindigkeitskandidaten $\tilde{\xi}_{c,(k,l)}$ miteinander oder mit dem Sollwert ist bei einem realen Roboter nicht ganz unproblematisch. In der vorliegenden Arbeit wurden für den Vergleich experimentell bestimmte Schwellwerte festgelegt, bei deren Überschreitung die Residen nach Gleichung 10 bzw. Gleichung 12 zu 1 werden. In weiterer Folge arbeiten wir an der Verwendung einer geometrisch fundierten Metrik, die berücksichtigt, dass es sich bei den Geschwindigkeiten im Geschwindigkeitsraum um translatorische und rotatorische Geschwindigkeiten handelt.

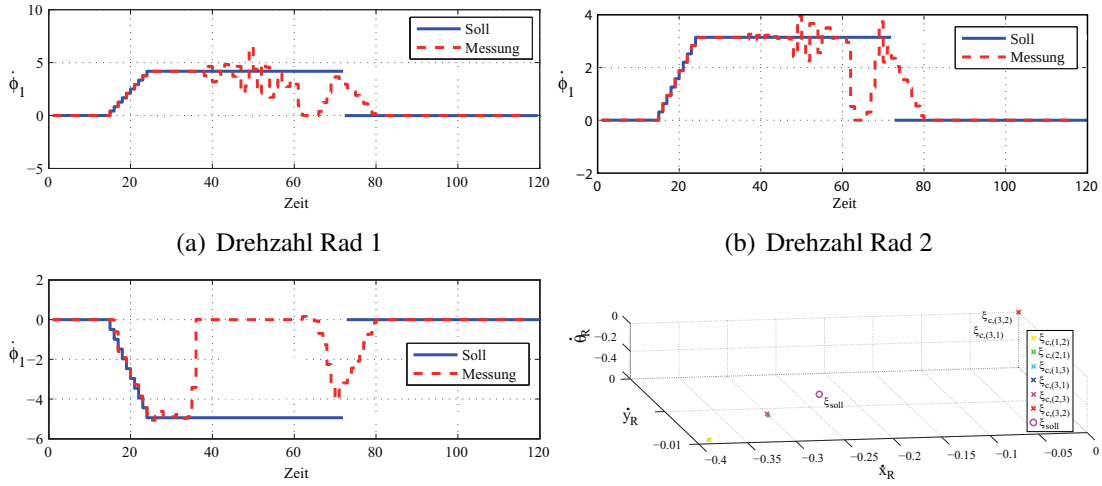
5 Testfahrten mit dem Roboter

Für die angeführten Testfahrten wird eine Anordnung unseres aus wabenförmigen Elementen aufgebauten Roboters [9, 8] mit 4 Modulen und 3 gelenkten Standardrädern verwendet (siehe auch Abb. 1). Der Roboter hat daher in diesem Fall sechs Komponenten, die fehlerhaft sein können, drei Radantriebe sowie drei Lenkantriebe mit den zugehörigen Sensoren.

Im folgenden Experiment erhielt der Roboter die Vorgabe, aus dem Stand bis zu einer Maximalgeschwindigkeit von $1.5m/s$ zu beschleunigen und sich dabei auf einem Kreisbogen von einem Meter Radius zu bewegen. Rad 3 ist dabei von Beginn an freilaufend. Dies resultiert in einer um den Sollwert schwankenden Drehzahl, wie in Abbildung 6c) erkennbar ist. Nach Auswertung der Residuen von Schritt II ergibt sich daher eine Fehlersignatur von

$$S = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

und die Drehzahl von Rad 3 wird als fehlerhaft erkannt.
Zum Zeitpunkt $t = 36$ wird in weiterer Folge Rad 3 blockiert. Der Roboter wird dadurch auf der Seite dieses Rades stark abgebremst; Die Drehzahlen der anderen Rädern weisen deutliche Abweichungen auf (siehe Abbildung 6a) und b)). Die Fehlersignatur der Vergleiche mit den Sollgeschwindigkeiten, erlaubt daher keine eindeutige Bestimmung der fehlerhafte Komponente. Betrachtet man nun aber in Schritt III die einzelnen Geschwindigkeitskandidaten, erkennt man, dass jene Geschwindigkeitskandidaten, die nur mit den Drehzahlen der Räder 1 und 2 berechnet wurden, die Redundanzbedingung nach Gl. (12) erfüllen. Abbildung 6d) verdeutlicht dies. Die Auswertung von Schritt III liefert daher die fehlerhafte Komponente $\dot{\varphi}_3$.



(a) Drehzahl Rad 1

(b) Drehzahl Rad 2

(c) Drehzahl Rad 3

(d) Geschwindigkeitskandidaten zum Zeitpunkt t=49

Abbildung 6: a) bis c) Verlauf der Drehzahlen der einzelnen Räder. Rad 3 ist bis $t = 36$ freilaufend. Dies resultiert in einer um den Sollwert schwankenden Geschwindigkeit. Danach wird Rad 3 blockiert; Alle Drehzahlen reagieren deutlich auf diesen Fehler. d) Geschwindigkeitskandidaten zum Zeitpunkt $t = 49$.

6 Zusammenfassung

Die vorliegende Arbeit stellt eine Methode zur Fehlererkennung für quasiomnidirektionale Radroboter vor. Basierend auf einer sorgsamen Analyse der kinematischen Beschränkungen eines solchen Roboterfahrwerks werden die Auswirkungen eines Fehlers auf das Fahrverhalten des gesamten Roboters betrachtet. Dies ist für quasiomnidirektionale Roboter von großer Bedeutung, da sich, bedingt durch die Überbestimmtheit eines solchen Fahrwerks, eine starke Kopplung des Verhaltens der Komponenten untereinander ergibt, welche für die Bestimmung der fehlerhaften Komponente im System wesentlich ist.

Diese Methodik stellt ein spezialisiertes Diagnoseverfahren dar; Jedoch kann durch die Berücksichtigung der besonderen Struktur des Problems durch explizite Anwendung der Fahrwerkskinematik ein entscheidender Vorteil bei der Erkennung der ursächlich fehlerhaften Komponente im Vergleich zu einer bloßen Betrachtung der Soll- und Istwerte der einzelnen Parameter gewonnen werden.

Literatur

[1] M. Bayoudh, L. Travé-Massuyès, and X. Olive. Hybrid systems diagnosis by coupling continuous and discrete event techniques. In *Proceedings of the IFAC World Congress, Seoul, Korea*, pages 7265–7270, 2008.

[2] J. Borenstein and L. Feng. Measurement and Correction of Systematic Odometry Errors in Mobile Robots. *IEEE Transactions Robotics & Automation*, 12(6):869–880, 1996.

[3] Mathias Brandstötter, Michael Hofbaur, Gerald Steinbauer, and Franz Wotawa. Model-based fault diagnosis and reconfiguration of robot drives. *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1203–1209, October 2007.

[4] Guy Campion, Georges Bastin, and Brigitte D'Andrèa-Novel. Structural properties and classification of kinematic and dynamic models of wheeled mobile robots. *IEEE Transactions on Robotics and Automation*, 12(1):47–62, 1996.

[5] V. Cocquempot, T. El Mezyani, and M. Staroswiecki. Fault detection and isolation for hybrid systems using structured parity residuals. In *IEEE/IFAC-ASCC : Asian Control Conference*, 2004.

[6] J. Gertler. A survey of analytical redundancy methods in failure detection and isolation. In *Preprints of the IFAC SAFEPROCESS Symposium*, pages 9–21, 1991.

[7] P. Robuffo Giordano, M. Fuchs, A. Albu-Schäffer, and G. Hirzinger. On the kinematic modeling and control of a mobile platform equipped with steering wheels and movable legs. In *Proceedings of the 2009 IEEE International Conference on Robotics and Automation (ICRA 2009)*, 2009.

[8] M. Hofbaur, M. Brandstötter, Ch. Schörghuber, and G. Steinbauer. On-line kinematics reasoning for reconfigurable robot drives. In *IEEE International Conference on Robotics and Automation (ICRA 10)*, 2010.

[9] Michael Hofbaur, Mathias Brandstötter, Simon Jantscher, and Christoph Schörghuber. Modular Re-configurable Robot Drives. *International Conference on Robotics and Automation and Mechatronics (RAM 2010)*, 2010.

[10] P. Sundvall and P. Jensfelt. Fault detection for mobile robots using redundant positioning systems. *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, pages 3781–3786, 2006.

[11] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Kewen Yin, and Surya N. Kavuri. A review of process fault detection and diagnosis part i: Quantitative model-based methods. *Computers and Chemical Engineering*, 27:293–311, 2003.

[12] Vandi Verma, Geoff Gordon, Reid Simmons, and Sebastian Thrun. Tractable particle filters for robot fault diagnosis: Real-time fault diagnosis. *Robotics & Automation Magazine*, 11(2):56 – 66, June 2004.

Stability analysis of linear switched systems utilising flow relations*

Kai Wulff† and Andreas Lorenz

†TU Ilmenau, Fachgebiet Regelungstechnik, Ilmenau
kai.wulff@tu-ilmenau.de

Abstract

In this paper we consider the stability of planar switched linear systems for arbitrary switching. We take an approach to establish algebraic necessary and sufficient conditions for stability that do not resort to any type of Lyapunov function. Instead we analyse the flow relations of the constituent systems and construct invariant sets using the solutions of the subsystems.

1 Introduction

In this note we consider stability properties of switched linear systems of the form

$$\dot{x}(t) = A_{s(t)}x(t),\ A_{s(t)} \in \mathcal{A} = \{A_1, \dots, A_N\} \subset \mathbb{R}^{n\times n} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ and for each time $t \in \mathbb{R}$ the matrix $A(t)$ equals exactly one matrix A_i belonging to the set $\mathcal{A}$. The switching between the system matrices A_i is described by a switching signal $s : \mathbb{R} \to \{1, \dots, N\}$, piecewise constant with finite numbers of discontinuities on any finite interval. The switching instances are denoted by $t_i, i \in \mathbb{N}$. We shall investigate the the asymptotic stability of (1) for arbitrary switching signals s.

The stability of the switched linear system (1) has been subject of extensive studies for more than two decades now. A comprehensive review on stability results obtained in the area can be found in [12, 8, 11] and most recently in [27, 13]. Also the relations to the absolute stability problem [14] have been well established, see e. g. [15]. Despite this large number of remarkable results, constructive necessary and sufficient conditions for the stability for arbitrary switching are scarce. The majority of constructive stability results consider the existence of a common quadratic Lyapunov function (CQLF), see [22, 21, 10, 32] to name only a few. However it is well known that the existence of a CQLF is sufficient but in general not necessary for the stability of (1), see e.g. [7]. Less conservative stability conditions can be found by investigating the existence of other

*This work was partially funded by the Studienstiftung des deutschen Volkes e. V.

types of Lyapunov functions. Considering other types of Lyapunov function such as piece-wise quadratic Lyapunov functions [9] or piece-wise linear Lyapunov functions [4, 23, 34, 36]. However, the majority of the results obtained are numerically costly or yield (still) conservative results for the general case. Necessary and sufficient conditions have been found in the form of powerful but non-constructive converse theorems [20, 7, 5, 19] that are not readily applicable in practice. In [3, 26] the stability boundary is characterised by means of the existence of a periodic solution. While the stability of the solution of (1) can be easily determined for periodic switching signals using Floquet theory, is remains an open problem how such worst case switching signal can be determined. The detailed description of the worst case switching signal in [3, 26] lead, however, to the derivation of numerical approaches that approximate the stability boundary arbitrarily close [33, 17].

In this paper we shall focus on the planar case where the system (1) switches arbitrarily between two subsystems. Indeed this case has been studied in a large number of contributions such as [24, 7, 6, 16], to mention only a few. However, constructive necessary and sufficient conditions for stability remain scarce. In [25, 3] the stability boundary is described by the existence (or absence for that matter) of a periodic solution. For the second order case it was found, that the significant periodic solution is induced by a switching rule with a period half of the period of the solution exhibiting two switches per period. The question of finding such switching rule remained unanswered. Nevertheless this characterisation of the switching rule is utilised in [33, 17] to formulate efficient algorithms that solve the stability problem numerically. Algebraic conditions are given in [31], where the eigenvalues of the two matrix products A_1A_2 and $A_1^{-1}A_2$ are related to the stability of (1). It is shown that $(\sigma(A_1A_2) \cup \sigma(A_1^{-1}A_2)) \cap \mathbb{R}^- = \emptyset$ constitutes a necessary and sufficient condition for the existence of CQLF for (1), where σ denotes the spectral operator. But again, these results is only sufficient for stability and therefore bears a certain amount of conservativeness. Some remarkable results on the second-order case can be found in [6, 16] with thorough treatment of the planar case yielding necessary and sufficient conditions for stability. However, the conditions derived are rather involved and consider various cases or complex formulations using generalised integrals which are not easily evaluated in practical control design. In [1] an explicit construction of the worst-case switching-signal as characterised in [25, 3] is given. This construction is used in [2] to derive algebraic necessary and sufficient conditions for quadratic and absolute stability for the planar case.

In this contribution we present an alternative approach to obtain a description of the worst-case switching sequence utilising the points of co-linear flow of the constituent systems. The solution of (1) for the constructed switching sequence provides a set that is suitable to determine the stability properties of (1). Subsequently we formulate a necessary and sufficient stability condition in terms of the spectrum of the matrix products A_1A_2 and $A_1^{-1}A_2$ following the approaches on CQLF results [31] and [18]. The obtained results are essentially equivalent to conditions found in [2], but allow for a more compact formulation.

The paper is structured as follows. In the following section we give some formal notation and concept of the solution of (1). In Section 3 we develop our approach of constructing invariant sets to establish stability of (1) and derive several results by

thorough investigation of the co-linear flow. Section 4 presents the main result with the necessary and sufficient condition for stability. The application of the main result is illustrated by several numerical examples in Section 5.

2 Preliminaries

In this note we consider the switched system (1) with $n = 2$ and $N = 2$. The LTI systems defined by the matrices $A_i \in \mathcal{A}$ are called subsystems or constituent systems. $\sigma(A)$ denotes the spectrum of A. For the stability of (1) for arbitrary switching we require the system matrices to be Hurwitz matrices, i. e. $\sigma(A_i) \subset \mathbb{C}^-$.

The solution of the switched system (1) for a given switching signal $s \in \mathcal{S}$ and initial state x_0 is given by

$$x(t; x_0, s) = e^{A_{s(t)}(t-t_l)} \prod_{j=1}^{l} e^{A_{s(t)}(t_j - t_{j-1})} x_0, \tag{2}$$

for $t \in [t_l, t_{l+1})$ where t_j denote the points of discontinuity of the switching signal s. For the constant switching signal $s(\cdot) \equiv i \in \{1, \ldots, N\}$ we denote the solution (2) of (1) by $x(t; x_0, i) = e^{A_i t} x_0$.

3 Co-Linear Flow And Invariant Sets

In this section we investigate the existence of co-linear flow for the constituent subsystems of (1). It turns out that the existence of points of co-linear flow is necessary (but not sufficient) for the instability of (1). However, the directions of co-linear flow can be used to construct (possibly) invariant sets for the solution of (1). The existence of such set guarantees stability for arbitrary switching.

We say the vector fields given by A_1 and A_2 have points of co-linear flow if there exists some vector $x \in \mathbb{R}^2 \setminus 0$ such that

$$\exists k \in \mathbb{R} \setminus 0 \ : \quad A_2 x = k A_1 x. \tag{3}$$

If A_1 is invertable we have

$$A_1^{-1} A_2 x = kx\,.$$

Thus, k is an eigenvalue and x is an eigenvector of the matrix product $A_1^{-1}A_2$. Therefore, the vector fields given by A_1 and A_2 have points of co-linear flow if and only if the matrix product $A_1^{-1}A_2$ has real eigenvalues. Depending on the algebraic and geometric multiplicity of the eigenvalues, we have one or two linearly independent eigenvectors denoted v_1, v_2 satisfying property (3). If $\sigma(A_1^{-1}A_2) \cap \mathbb{R}^- \neq \emptyset$ we say that the two vector fields have negative co-linear flow, if $\sigma(A_1^{-1}A_2) \subset \mathbb{R}^+$ the vector fields have positive co-linear flow.

In order to verify the invariancy of the sets constructed we define the following cones that describe the relation of the vector fields to each others. For every point $x \in \mathbb{R}^2$ we define the cones K_1 up to K_4 with respect to the flow given by the vector field of A_1.

Definition 1 *For the matrix $A_i \in \mathbb{R}^{2\times 2}$ and every point $x \in \mathbb{R}^2$ we define the following cones:*

$$\begin{aligned}
K_1^i(x) &:= \{\xi \in \mathbb{R}^2 \mid \xi = \ \ c_1 A_i x - c_2 x,\ c_1, c_2 \geq 0\} \\
K_2^i(x) &:= \{\xi \in \mathbb{R}^2 \mid \xi = -c_1 A_i x - c_2 x,\ c_1, c_2 \geq 0\} \\
K_3^i(x) &:= \{\xi \in \mathbb{R}^2 \mid \xi = -c_1 A_i x + c_2 x,\ c_1 \geq 0, c_2 > 0\} \\
K_4^i(x) &:= \{\xi \in \mathbb{R}^2 \mid \xi = \ \ c_1 A_i x + c_2 x,\ c_1 \geq 0, c_2 > 0\}.
\end{aligned}$$

Note that the cones K_1 and K_2 are closed while K_3 and K_4 are half-open sets. The union of K_1 and K_2 is an half-open plane which contains the origin. On the other hand, the union of K_3 and K_4 is an half-open plane excluding the origin. Therefore we can use this definition of cones to check whether the vector field A_2 the solution of the system (1) more away from the origin at the point x.

We shall now discuss some properties of the cones defined above in regard of the existence of co-linear flow. Assume there exist points $x_1, x_2 \in \mathbb{R}^2$ such that $A_2 x_1 \in K_1(x_1)$ and $A_2 x_2 \in K_4(x_2)$, then, by linearity, there must exist $x_\alpha = \alpha x_1 + (1-\alpha)x_2$ for some $\alpha \in [0, 1]$ such that $A_2 x_\alpha$ is exactly on the boundary of K_1. Since the boundary of K_1 is given by $A_1 x_\alpha$, x_α is a point of positive co-linear flow and satisfies (3) for some $k \in \mathbb{R}^+$. The same assertion holds for the cones K_2 and K_3. If we find two points for which the flow of A_2 is in K_2 and K_3, respectively, then there exists a point of negative co-linear flow such that (3) holds for some $k \in \mathbb{R}^-$. Furthermore, if there exist x_1, x_2 such that $A_2 x_1 \in K_3(x_1)$ and $A_2 x_2 \in K_4(x_2)$ then A_2 has a positive real eigenvalue with eigenvector x_α.

3.1 No Co-Linear Flow

We shall now consider the case where the vector fields A_1, A_2 have no points of co-linear flow, i. e. the matrix product $A_2^{-1}A_1$ has no real eigenvalues. Our aim is to construct a positive invariant set for the solutions (2) for all $s \in \mathcal{S}$ using the solutions of the constituent LTI systems.

Let one of the system matrixes A_i satisfy $\sigma(A_i) \in \mathbb{C}^- \setminus \mathbb{R}$. Then following set is well-defined for all $x_0 \in \mathbb{R}^2 \setminus \{0\}$, see Figure 3 for an illustration.

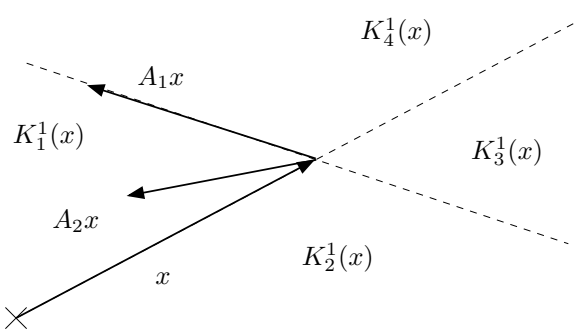


Figure 1: Definition of the cones $K_j^1(x)$ for a given $x \in \mathbb{R}^2$ with $A_2 x \in K_1^1(x)$

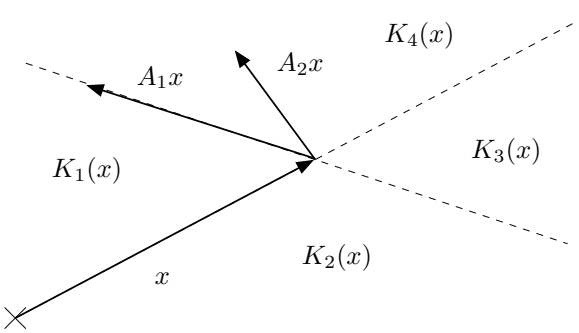


Figure 2: Definition of the cones for a given $x \in \mathbb{R}^2$ with $A_2 x \in K_4(x)$

Definition 2 *Let A_i be a Hurwitz matrix with $\sigma(A_i) \in \mathbb{C}^- \setminus \mathbb{R}$. We define:*

$$\Omega_i^*(x_0) := \left\{ \xi \in \mathbb{R}^2 \middle| \xi = \alpha x(t; x_0, i),\ t \geq 0,\ \alpha \in [0,1] \right\}.$$

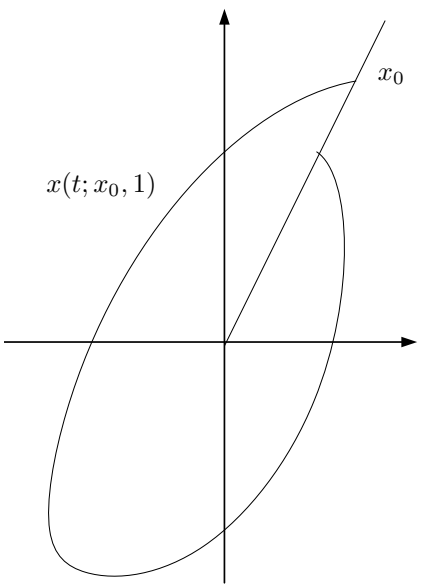


Figure 3: Construction of $\Omega_1^*(x_0)$

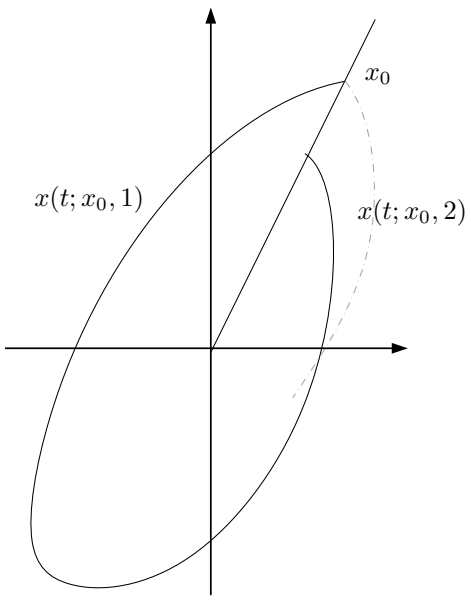


Figure 4: Construction of $\Omega^*(x_0)$

In order to establish stability via the set $\Omega_i^*(x_0)$ we have to show that it is a compact neighbourhood of the origin and invariant with respect to the solution (2).

Lemma 1 *$\Omega_i^*(x_0)$ is compact and a neighbourhood of the origin.*

Proof 1 *The set $\Omega_i^*(x_0)$ is defined as a closed set. Thus $\Omega_i^*(x_0)$ is compact if it is bounded for all x_0. Since A_i is Hurwitz the growth-rate ω_0 of the semi-group generated by A_i is negative. Therefore we can find $M \geq 1$ such that*

$$||e^{A_i t}|| \leq M||e^{\omega_0 t}|| \text{ with } \omega_0 = 0\ \forall\ t \geq 0\,. \tag{4}$$

Thus $||x(t; x_0, i)|| \leq M||x_0|| \forall t \geq 0$ and $\Omega_i^*(x_0)$ is compact.

Let $\tilde{\omega}_0 < \omega_0$ and choose M such that $||e^{A_1 t}|| \geq M||e^{\tilde{\omega}_0 t}||$, for all $t \geq 0$. Then there exists $\epsilon > 0$ such that the ϵ-ball B_ϵ around the origin satisfies $B_\epsilon \subseteq \Omega_i^*(x_0)$. Thus $\Omega_i^*(x_0)$ is a neighbourhood of the origin.

We shall now consider the invariance property of the set constructed using the flow defined by A_2 in regard to the cones $K_1^1, \ldots, K_4^1$ defined above. Since there does not exist any point of co-linear flow exactly one of the following in true:

$$\forall x \ : \ A_2 x \in \left(K_1^1(x) \cup K_2^1(x) \right) \tag{5}$$

$$\forall x \ : \ A_2 x \in \left(K_3^1(x) \cup K_4^1(x) \right). \tag{6}$$

Furthermore we note, that we require A_2 to be Hurwitz and thus there is no $x \in \mathbb{R}^2$ such that $A_2 x$ lies on the boundary of $K_3^1(x)$ and $K_4^1(x)$, but is either in the interior of $K_3^1(x)$ or in the interior of $K_4^1(x)$ for all $x \in \mathbb{R}^2$.

First we consider the case (5) where the vector $A_2 x$ belongs to $K_1^1(x)$ or $K_2^1(x)$ for all $x \in \mathbb{R}^2$. For this case we can construct sets that are positive invariant for the solutions of (1) using $\Omega_i^*(x_0)$. Together with Lemma 1 we establish the exponential stability of the system (1) for arbitrary switching.

Lemma 2 *Let $A_1, A_2 \in \mathbb{R}^{2\times 2}$ Hurwitz and $\sigma(A_1) \subset \mathbb{C}^- \setminus \mathbb{R}$. Then:*

$$[\forall x : A_2 x \in (K_1^1(x) \cup K_2^1(x))] \;\Rightarrow\; (1) \textit{ is exp. stable.}$$

Proof 2 *We consider first the case where we can choose x_0 such that $A_2 x_0 \in K_1^1(x_0)$ to construct $\Omega_1^*(x_0)$, see Figure 1. The boundary of $\Omega_1^*(x_0)$ is given by a part of the solution $x(t; x_0, 1)$ and $\{\alpha x_0\}$ for some $\alpha \in (0, 1]$. We shall now consider the flow of the constituent systems along those boundaries. For the part of the boundary which is defined by $x(t; x_0, 1)$ the flow $A_1 x$ of subsystem 1 is tangential to the boundary while the flow of subsystem 2 points into the set since $A_2 x \in (K_1^1(x) \cup K_2^1(x))$. Along αx_0 the flow of both subsystems point into the set. Thus $\Omega_1^*(x_0)$ is positive invariant for the solutions of (1). Since $\Omega_1^*(x_0)$ is compact and a neighbourhood of the origin (1) is stable for arbitrary switching. Furthermore, there exists no switching signal such that the solution of (1) stays on the boundary of $\Omega_1^*(x_0)$ for all $t \geq 0$ and thus (1) is exponentially stable for arbitrary switching.*

If there does not exist any $x_0 \neq 0$ such that $A_2 x_0 \in K_1^1(x_0)$, we have $A_2 x \in K_2^1(x)$ for all $x \in \mathbb{R}^2$. This implies that A_2 has no real eigenvector or $\sigma(A_2) \subset \mathbb{C}^- \setminus \mathbb{R}$. Consider now the set $\Omega^*(x_0) := \Omega_1^*(x_0) \cup \Omega_2^*(x_0)$ for some $x_0 \neq 0$, see Figure 2. $\Omega^*(x_0)$ is also compact and a neighbourhood of the origin since it is a union of sets with these properties. The boundary of $\Omega^*(x_0)$ is given by the solutions $x(t; x_0, 1)$ and $x(t; x_0, 2)$, for some $t \geq 0$. As $A_2 x \in K_2^1(x)$ the flow of subsystem 2 is pointing into the set $\Omega^*(x_0)$ along its boundary described by $x(t; x_0, 1)$. Note further that $A_1 \in K_1^2(x) \cap K_2^2(x)$ for all x. Thus the flow of subsystem 1 is pointing into the set $\Omega^*(x_0)$ along its boundary described by $x(t; x_0, 2)$. Further there is no switching signal resulting in a solution of (1) which stays on the boundary of $\Omega^*(x_0)$ for all $t \geq 0$. Hence, (1) is exponentially stable for arbitrary switching.

We shall now consider the remaining two cases, i. e. $A_2 x \in K_4^1(x)$ or $A_2 x \in K_3^1(x)$ for all x. Note that for both cases A_2 cannot have real eigenvalues as this implies $A_2 x = \lambda x \notin \left(K_3^1(x) \cap K_4^1(x)\right)$ for $\lambda < 0$.

The case that $A_2 x \in K_3^1(x)$ for all $x \in \mathbb{R}^2$ can be ruled out easily as the following lemma shows.

Lemma 3 *Let $A_1 \in \mathbb{R}^{2\times 2}$ Hurwitz and $\sigma(A_1) \subset \mathbb{C}^- \setminus \mathbb{R}$. Then there exists no Hurwitz matrix $A_2 \in \mathbb{R}^{2\times 2}$, such that $\forall x : A_2 x \in K_3^1(x)$.*

Proof 3 *We proof this lemma by contradiction. Assume that A_2 is Hurwitz and $\forall x : A_2 x \in K_3^1(x)$. Let $B = -A_2$. From linearity it follows that $\forall x : Bx \in K_1^1(x)$. By Lemma 2 the switched system (1) with $\mathcal{A} := \{A_1, B\}$ is exponential stable. Thus B is Hurwitz and A_2 is non-Hurwitz which contradicts our initial assumption.*

The remaining case where $A_2x \in K_4^1(x)$, can be mapped to Lemma 2 by interchanging the roles of the matrices A_1 and A_2.

Lemma 4 *Let $A_1, A_2 \in \mathbb{R}^{2\times 2}$ Hurwitz and $\sigma(A_1) \subset \mathbb{C}^- \setminus \mathbb{R}$. Then:*

$$[\forall x : A_2x \in K_4^1(x)] \;\Rightarrow\; (1) \textit{ is exp. stable.}$$

Proof 4 *Since $A_2x \in K_4^1(x)$ there is no $x \in \mathbb{R}^2$ such that $A_2x = \lambda x, \lambda \in \mathbb{R}$. Thus A_2 has non-real eigenvalues. Now swap the indices of the subsystems by defining $B_2 := A_1$ and $B_1 := A_2$ which yields $\forall x : B_2x \in (K_1^1(x) \cup K_2^1(x))$. By Lemma 2 the switched system is exponential stable.*

Remark: Note, that Lemmas 2 and 4 establish exponential stability of the switched system without resorting to any type of Lyapunov function. The proof of stability merely depends on the choice of the sets $\Omega_i^*(x_0)$.

If both matrices have real eigenvalues we can resort to existing results to establish stability of (1). If the matrix product $A_1^{-1}A_2$ do not have distinct, positive real eigenvalues then the matrix pencil $\sigma_\gamma(A_1 + \gamma A_2)$ is real for all $\gamma \in [0, \infty)$ (see [35],Th. 3.17). Furthermore, the pencil is strict Hurwitz since the matrix product has no negative real eigenvalues (see [35],Lem. 3.10). Thus we can assume that $\sigma_\gamma(A_1 + \gamma A_2) \subset \mathbb{R}^-$. In [30] a result is presented, which states that $\sigma_\gamma(A_1 + \gamma A_2) \subset \mathbb{R}^-$ implies the existence of an common quadratic Lyapunov function.

We obtain the following corollary.

Corollary 1 *Let $A_1, A_2 \in \mathbb{R}^{2\times 2}$ Hurwitz.Then:*

$$[\sigma(A_1^{-1}A_2) \subset \mathbb{C} \setminus \mathbb{R}] \Rightarrow (1) \textit{ is exp. stable}$$

3.2 Positive Co-Linear Flow

We consider now the case where the vectors fields of A_1 and A_2 have positive co-linear flow, i. e. the matrix product $A_1^{-1}A_2$ has positive real eigenvalues. We will examine the stability of the switched system (1) by constructing a set similar as before using a state-space partition defined by the points of co-linear flow. Using this partition we can define a state-dependent switching law that can be uniquely mapped to a switching signal s_{co}. The solution $x(\cdot; x_0, s_{co})$ is then used to construct the candidate for the invariant set.

For ease of exposition we shall assume that the matrix product is non-defective (diagonalisable).[1] Then $A_1^{-1}A_2$ has two linearly independent real eigenvectors v_1, v_2 which partition the state-space in four cones $C_1 \ldots C_4$ with disjoint interiors. Since the boundary of the cones C_j are defined by the points of co-linear flow we have

$$\forall\, j, \;\; \nexists x \in int(C_j) : \; A_2x = kA_1x \text{ with } k \in \mathbb{R}. \tag{7}$$

[1] It is not hard to show that the arguments still hold for the defective case.

It follows that for every cone C_j exactly one of the following is true:

$$A_2 x \in K_1^1(x) \cup K_2^1(x) \qquad \forall\, x \in int(C_j) \tag{8}$$

$$A_2 x \in K_3^1(x) \cup K_4^1(x) \qquad \forall\, x \in int(C_j). \tag{9}$$

Thus we can define the unique state-dependent switching law such that subsystem 1 is active within those cones C_j where (8) holds and otherwise subsystem 2 is active.

Let $x_0 = v_1$. Then the switching signal $s_{co} \in \mathcal{S}$ corresponding to the state-depending switching law is uniquely defined. We can then construct our candidate for the invariant set as follows (see Figure 5 for an illustration).

Definition 3 *Let $t' > 0$ such that $x(t'; x_0, s_{co}) = \gamma x_0$ with $\gamma > 0$ then*

$$\Omega(x_0) := \big\{ \xi \in \mathbb{R}^2 \big| \xi = \alpha x(t; x_0, s_{co}),\, t \in [0, t'],\, \alpha \in [0, 1] \big\}.$$

The construction of the set $\Omega(x_0)$ is similar to that of $\Omega_i^*(x_0)$ in Definition 2 except that the switching signal s_{co} is not constant and we only use a finite interval $t \in [0, t']$ of the solution (2). Using similar arguments as in Lemma 1 we can establish that the set $\Omega(x_0)$ is compact and a neighbourhood of the origin (the former is ensured as t' is finite).

Obviously the existence of t' with the property given in Definition 3 is crucial for the argument above. In the following we shall investigate whether such set exists and derive conditions that guarantee its invariance for the solutions of (1). We start considering (1) where at least one of the subsystems has non-real eigenvalues.

Lemma 5 *Let A_1, A_2 be two Hurwitz matrices in $\mathbb{R}^{2\times 2}$ with $\sigma(A_1) \in \mathbb{C}^- \backslash \mathbb{R}$ and $\sigma(A_1^{-1} A_2) \subset \mathbb{R}^+$. Then the set $\Omega(x_0)$ exists and is invariant for the solutions of (1) if and only if $\gamma \leq 1$.*

Proof 5 *For the initial condition $x_0 = v_1$ the switching signal s_{co} is uniquely defined by the state-dependent switching law defined above. The state-dependent switching law insures that none of the cones C_j is invariant for $x(\cdot; x_0, s_{co})$ for any $x_0 \in \mathbb{R}^2$. If A_1 and A_2 have non-real eigenvalues this fact is trivially true. In case that A_2 has real eigenvalues we note that A_2 is only active if (9) holds, i. e. the flow of subsystem 2 points further away from the origin that that of subsystem 1. Thus within those cones the solution of subsystem 2 is bounded from below by the solution of subsystem 1 and hence those cones are also invariant for the solution of subsystem A_2.*

Note further, that the flows of the subsystems have the same orientation (clockwise or counter clockwise) on all boundaries of the cones C_j. Therefore there exists a finite $t' > 0$ such that $x(t'; x_0, s_{co}) = \gamma x_0$ with $\gamma > 0$. Thus $\Omega(x_0)$ exists and is well defined.

To show the invariance of $\Omega(x_0)$ we first consider that part of the boundary $\delta\Omega(x_0)$ of the set which is given by $x(\cdot; x_0, s_{co})$. Clearly for those cones C_j for which subsystem 1 is active and hence (8) holds, the flow of subsystem 2 points into $\Omega(x_0)$. For the remaining cones C_j (9) holds. In fact, we have $A_2 x \in K_4^1(x)$ since A_2 is Hurwitz. This however implies that $A_1 x \in K_1^2(x)$ and thus the flow of subsystem points into $\Omega(x_0)$ in those cones C_j where subsystem 2 is active. It follows that $\Omega(x_0)$ is invariant for the solutions (1) if and only if the flow of both subsystems point into $\Omega(x_0)$ along the remaining boundary given by αx_0 with $\alpha \in [1, \gamma]$. Since both subsystems have co-linear flow for αx_0 the set $\Omega(x_0)$ is invariant for all solutions of (1) if and only if $\gamma \leq 1$.

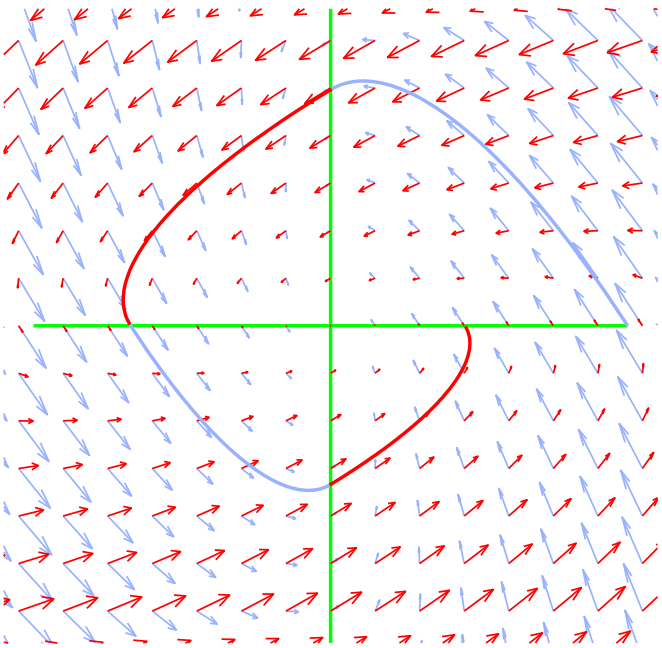


Figure 5: Illustration for the state-space dependent switching law defined by the points of co-linear flow (here along the coordinate-axis). The plotted trajectory is the solution $x(\cdot; x_0, s_{co}), t \in [0, t']$ defining the set $\Omega^*(x_0)$.

We shall now investigate the case where both matrices have real eigenvalues. The construction of the set $\Omega(x_0)$ in Definition 3 relies on the fact that none of the cones C_j are invariant for solutions of both subsystems. In other words the solutions of both subsystem in a specific cone did not tend to zero within this cone. Certainly this might happen, when both systems have real eigenvalues. However, if there exists an invariant cone for the solutions of both subsystems we can utilise the following result derived for Metzler matrices[2]:

Theorem 1 ([18]) *Let $A_1, A_2 \in \mathbb{R}^{2\times 2}$ be Hurwitz and Metzler matrices. There exists a CQLF for (1) if and only if $\sigma(A_1^{-1}A_2) \cap \mathbb{R}^- = \emptyset$.*

A particular property of Metzler matrices is that they define positive linear systems which have the positive orthant as an invariant cone. So if C_j is an invariant cone for our subsystems. Then we can use vectors that span C_j as a basis to simultaneously transform A_1 and A_2 such that the positive orthant is an invariant cone. It is not hard to show that the transformed matrices are Metzler matrices.

Corollary 2 *Let $A_1, A_2 \in \mathbb{R}^{2\times 2}$ be Hurwitz and $\sigma(A_1^{-1}A_2) \cap \mathbb{R}^- = \emptyset$. If there exists an invariant cone for the solutions of both subsystems, then there exists a CQLF for (1).*

To show stability for our case with real eigenvalues shall use Corollary 2. If there is no invariant cone the set $\Omega(x_0)$ can be constructed. Obviously, if there is a cone C_j such

[2]The matrix A is called Metzler matrix if the off-diagonal entries of A are non-negative.

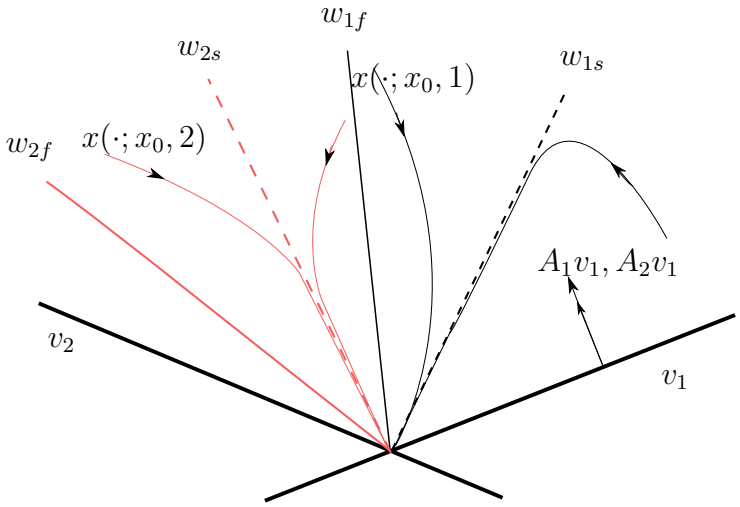


Figure 6: Illustration for the configuration of the eigenvectors w_i of the matrices A_i with respect to the cones C_j.

that the flow at the boundaries have opposite orientation in the sense of the associated cross-product of v_i and $A_1v_i = A_2v_i$ the above result guarantees stability. Thus we may only consider the case where the flow at all boundaries C_j have the same orientation.

Note that any two adjacent cones C_i and C_j contain all eigenvectors of the system matrices A_1, A_2. Note further, that the orientation of the flow changes only along the eigenvectors of each constituent system. Since the flow at the boundaries have the same orientation each cone contains either both or none eigenvector of the two system matrices. Hence, we can distinguish two possible cases:

i) Each cone contains both eigenvectors of one subsystem.

ii) One cone C_j contains all eigenvectors of both subsystems.

Clearly, in the first case for each cone exists exactly one subsystem having no eigenvector in that cone, see Figure 6. Hence, there is always a unique switching function, constant in the whole cone, providing a solution that leaves each cone in a finite time and thus the set Ω is well defined.

Lemma 6 *Let A_1, A_2 be two Hurwitz matrices in $\mathbb{R}^{2\times 2}$ with real eigenvalues such that $\sigma(A_1^{-1}A_2) \subset \mathbb{R}^+$ and the eigenvectors of A_1 and A_2 lie within different cones C_j defined by the vectors of co-linear flow. Then the set $\Omega(x_0)$ exists and is invariant for the solutions of (1) if and only if $\gamma \leq 1$.*

In the second case *ii*) there is no constant switching function which yields to an solution that leaves the cone in a finite time (since all eigenvectors are contained) and therefore the set Ω does not exists. In this case however we can show that there exists an invariant cone.

Lemma 7 *The switched system (1) with A_1, A_2 Hurwitz matrices with real spectrum. The system admits an common positive invariant cone if and only if all eigenvectors of*

both constituent systems are contained in one cone C_j, given by the partition of co-linear flows.

Proof 6 *Consider a subsystem with two real eigenvectors denoted by f (fast) and s (slow) with the related eigenvalues λ_f, λ_s, such that $\lambda_f < \lambda_s$. Then it can easily be seen that the solution is orientated in such a manner that it points always away from the fast eigenvector and therefore points towards the slow eigenvector.*

Let v_1 be that boundary of the cone C_j where the flows point into C_j and follow the direction of this flow.Obviously, no fast eigenvector can occur at first, since the flow points away from the fast eigenvector. If we have both slow eigenvectors adjacent to each others they form an invariant cone and thus Corollary 2 guarantees stability.

The remaining configuration is that the slow eigenvector is followed by the fast eigenvector of the same system. However this contradicts the requirements (8) *and* (9).

3.3 Negative Co-Linear Flow

In this case the resulting switched system (1) is not exponential stable. This result is shown in [29] and can be summarised as follows

Theorem 2 *Let $A_1, \ldots, A_k$ be Hurwitz matrices in $\mathbb{R}^{n\times n}$. Suppose that there are non-negative real numbers $\alpha_1, \ldots \alpha_k$ not all zero, such that $\alpha_1 A_1 + \ldots \alpha_k A_k$ has an eigenvalue λ with non-negative real part. Then the associated switched system* (1) *is not exponentially stable for arbitrary switching signals.*

If $\sigma(A_1^{-1}A_2) \cap \mathbb{R}^- \neq \emptyset$ then there exists some x and $k \in \mathbb{R}^+$ such that $A_1 x = -kA_2 x$ and therefore $\hat{A} := A_1 + kA_2$ has one eigenvalue equal to zero. Hence, by theorem 2 the related switched system is not exponentially stable for arbitrary switching signals.

Corollary 3 *If the matrix product $A_1^{-1}A_2$ has at least one negative real eigenvalue, the related switched system* (1) *with $\mathcal{A} = \{A_1, A_2\}$ is* not exponentially stable *under arbitrary switching.*

4 Main Result

In the previous section we showed that the points of co-linear flow can be used to construct invariant sets for the solutions of the switched system. The conditions in Lemmas 2, 4 and 7 in terms of the spectrum of the matrix product $A_1^{-1}A_2$ guarantee the existence of such a set. For the cases considered in Lemma 5 and 6 require further investigations regarding properties of the solution of (1) for the switching signal s_{co}. Indeed for those cases this switching signal can be considered as the worst case switching signal as discussed in [25, 3].

Before presenting our main result we characterise this switching signal explicitly in terms of the switching instances.

The state-space dependent switching law defined in section 3.2 guarantees for the cases of Lemma 5 and 6 that none of the cones C_j are invariant for the solution $x(\cdot; x_0, s_{so})$.

Therefore the switching instances are given by the time that the active system stays within the respective partition before it reaches the switching surface. Since the subsystems are linear this dwell-time is constant for any initial state at the boundary of the cones. Consider the cones C_j where the subsystem A_j is active and let the boundaries of C_j be given by v_1 and v_2. As shown in Lemma 5 and 6 there exists a finite time t_j^* such that

$$kv_2 = x(t_j^*) = e^{A_j t_j^*} v_1, \tag{10}$$

for some $k > 0$. We can now use linear systems theory to determine the time t_j^* for the following cases (see Appendix):

- A_j is defective with real eigenvalues:

$$t_j^* = \frac{v_{21}v_{12} - v_{22}v_{11}}{v_{22}v_{12}} \tag{11}$$

- A_j is non-defective with real eigenvalues $\lambda_{1,2}$:

$$t_j^* = \frac{\ln(\frac{v_{21}v_{12}}{v_{22}v_{11}})}{\lambda_1 - \lambda_2} \tag{12}$$

- A_j has non-real eigenvalues with $\Im\{\lambda_1\} = \beta$:

$$t_j^* = \frac{\arccos\langle v_2, v_1\rangle}{\beta}, \tag{13}$$

where v_{i1} and v_{i2} are the components of v_i and $\langle\cdot,\cdot\rangle$ denotes the scalar product.

Note that the switching signal defined by the state-space partition of co-linear flow and the dwell-times given by (11)-(13) is periodic and has two switches per period. Thus it constitutes the worst case switching signal in the sense of [25, 3]. Exponential stability is then guaranteed if

$$\rho\left(e^{A_2 t_2^*} e^{A_1 t_1^*}\right) < 1$$

where $\rho(\cdot)$ denotes the spectral radius.

We can now formulate our main result giving a necessary and sufficient condition for the stability of (1).

Theorem 3 *Let $A_1, A_2 \in \mathbb{R}^{2\times 2}$ be Hurwitz. Then (1) is exponential stable for arbitrary switching if and only if one of the following holds:*

(i) $\sigma(A_1^{-1}A_2) \subset \mathbb{C} \setminus \mathbb{R}$

(ii) $\sigma(A_1^{-1}A_2) \subset \mathbb{R}^+$ *and* $\rho\left(e^{A_2 t_2^*} e^{A_1 t_1^*}\right) < 1$ *with t_i^* given in (11)-(13).*

Proof 7 *Sufficiency of (i) is given by Corollary 1. Sufficiency and necessity of (ii) for for the existence of an invariant set for the case that s_{co} exists is given by Lemma 5 and 6. Since we require $\rho\left(e^{A_2 t_2} e^{A_1 t_1}\right)$ to be strictly less than 1, no solution (2) remains on the boundary of $\Omega(x_0)$ and thus exponential stability is guaranteed. In case there exists an invariant cone for the solutions of (1) (see Lemma 7), $\sigma(A_1^{-1}A_2) \subset \mathbb{R}^+$ is already sufficient for stability for arbitrary switching thus checking $\rho\left(e^{A_2 t_2} e^{A_1 t_1}\right) < 1$ is redundant but must be satisfied. Necessity is insured by Theorem 2 and the Lemmas 5 and 6.*

5 Numerical Example

In this section we present two examples of the application for Theorem 3. Therefore, we consider the system (1) with the constituent subsystems given by

$$A_1 = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \; A_2 = \begin{pmatrix} -1 & -8 \\ 0.125 & -1 \end{pmatrix}.$$

Note, that this example from [7] with a parameter $\alpha = 8$ does not admit a common quadratic lyapunov function, since the matrix product $A_1 A_2$ has a negative real spectrum (see [28]).

Both systems has complex eigenvalues and are Hurwitz, hence we can apply the theorem 3. Therefore we consider the spectrum of the matrix product $A_1^{-1} A_2$, which turns out to be positive real. Let v_1, v_2 denote the eigenvectors of this product which are exactly the points of co-linear flow of the constituent subsystems. By theorem 3 we now have to investigate the stability of the System $\Sigma_{\mathcal{A},s_{co}}$. The involved vector fields and the points of co-linear flow are presented in figure 7. The application of the switching law provides that the first subsystem is active in the first and third quadrant and that every point on the main axis is a switching point.

We choose the initial condition $x_0 = [0.9948 \quad -0,1014]$ and consider the related switching function s_{co}. The switching instances are then given by $t_1 = t_1^*$ and $t_2 = t_1^* + t_2^*$, where t_1^*, t_2^* denotes the time spans for which system one or two is active, respectively. By equation (13) we obtain $t_1^* = t_2^* = 0.9885$. The solution over one half period of the switching function is given by $x(T/2; x_0, s_{co}) = e^{A_2 t_2^*} e^{A_1 t_1^*} x_0$. The spectral radius of the related transition matrix is $\rho(e^{A_2 t_2^*} e^{A_1 t_1^*}) = -0.6726$. Hence, we conclude by Theorem 3 that the system (1) with $\mathcal{A} = \{A_1, A_2\}$ is exponentially stable under arbitrary switching.

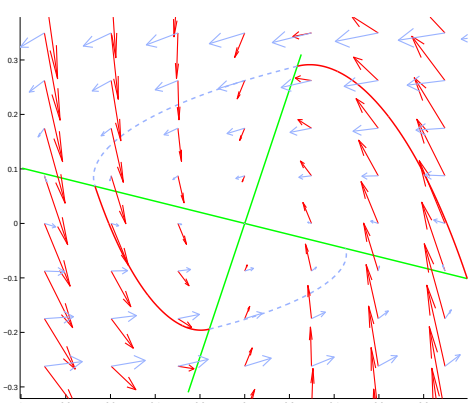


Figure 7: Ex 1. Vector fields $A_1 x$ (dark), $A_2 x$ (bright) and the solution $x(\cdot; x_0, s_{co})$, consisting of the solution to the related active subsystem

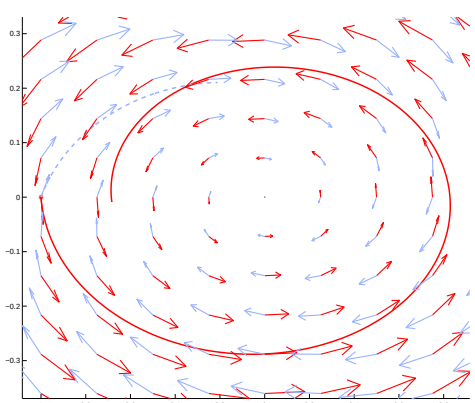


Figure 8: Ex 2. Vector fields $A_1 x$ (dark), $A_2 x$ (bright) and the solution of the constituent subsystem, both starting at the same initial state x_0

For the second example we consider (1) with

$$A_1 = \begin{pmatrix} -1.9 & -100 \\ 10 & -1.9 \end{pmatrix},\ A_2 = \begin{pmatrix} -9 & 100 \\ -10 & -9 \end{pmatrix},$$

where both subsystems has a non-real Hurwitz spectrum. The matrix product $A_1^{-1}A_2$ has no real eigenvalue, hence there does not exists any points of co-linear flow. We conclude by Theorem 3 that the switched system (1) with $\mathcal{A} = \{A_1, A_2\}$ is exponentially stable under arbitrary switching.

The involved vector fields and the solution of the related subsystems starting at x_0 are plotted in Figure 8. The figure shows the boundary of the positive invariant set for the flow of the switched system. Note, that this system admits a common quadratic Lyapunov function.

6 Conclusions

In this paper we present an approach for the stability analysis for switched linear systems. By considering the relation of the vector fields of the constituent systems to each others we derive stability conditions for the switched linear system for arbitrary switching without resorting to Lyapunov functions. We show that the existence of co-linear flow is necessary for the instability of the system. Furthermore we derive constructive necessary and sufficient stability conditions for the switched system that are readily applicable in practice.

References

[1] Moussa Balde and Ugo Boscain. Stability of planar switched systems: the nondiagonalizable case. *Commun. Pure Appl. Anal.*, 7(1):1–21, 2008.

[2] Moussa Balde, Ugo Boscain, and Paolo Mason. A note on stability conditions for planar switched systems. *International Journal of Control*, 82(10):1882–1888, 2009.

[3] N. E. Barabanov. On the Aizerman problem for third-order nonstationary systems. *Differential Equations*, 29(10):1439–1448, 1993.

[4] F. Blanchini. Nonquadratic Lyapunov functions for robust control. *Automatica*, 31(3):451–461, 1995.

[5] F. Blanchini and S. Miani. A new class of universal Lyapunov functions for the control of uncertain linear systems. *IEEE Transactions on Automatic Control*, 44(1):641–647, 1999.

[6] Ugo Boscain. Stability of planar switched systems: the linear single input case. *SIAM J. Control Optimization*, 41(1):89–112, 2002.

[7] W. P. Dayawansa and C. F. Martin. A converse Lyapunov theorem for a class of dynamical systems which undergo switching. *IEEE Transactions on Automatic Control*, 44(4):751–760, 1999.

[8] R. DeCarlo, M. Branicky, S. Pettersson, and B. Lennartson. Perspectives and results on the stability and stabilisability of hybrid systems. *Proceedings of the IEEE*, 88(7):1069–1082, 2000.

[9] M. Johansson and A. Rantzer. Computation of piecewise quadratic Lyapunov functions for hybrid systems. *IEEE Transactions on Automatic Control*, 43(4):555–559, 1998.

[10] D. Liberzon, J. P. Hespanha, and A. S. Morse. Stability of switched linear systems: a Lie-algebraic condition. *Systems & Control Letters*, 37(3):117–122, 1999.

[11] Daniel Liberzon. *Switching in Systems and Control*. Birkhäuser, 2003.

[12] Daniel Liberzon and A. Stephen Morse. Basic problems in stability and design of switched systems. *IEEE Control Systems Magazine*, 19(5):59–70, 1999.

[13] Hai Lin and Panos J. Antsaklis. Stability and stabilizability of switched linear systems: A survey of recent results. *Automatic control, IEEE Transactions on*, 54(2):308–322, February 2009.

[14] A. I. Lur'e and V. N. Postnikov. On the theory of stability of control systems. *Prikladnaya Matematika i Mekhanika*, 3(8):246–248, 1944. (in russian).

[15] Michael Margaliot. Stability analysis of switched systems using variational principles: an introduction. *Automatica*, 42(12):2059–2077, December 2006.

[16] Michael Margaliot and Gideon Langholz. Necessary and sufficient conditions for absolute stability: The case of second-order systems. *IEEE Transactions on circuits and systems*, 50(2):227–234, February 2003.

[17] Micheal Margaliot and Christos Yfoulis. Absolute stability of third-order systems – a numerical algorithm. *Automatica*, 42(10):1705–1711, 2006.

[18] Oliver Mason and Robert Shorten. Some results on the stability of positive switched linear systems. In *43rd IEEE Conference on Decision and Control*, Bahamas, 2004.

[19] Paolo Mason, Ugo Boscain, and Yacine Chitour. On the minimal degree of a common lyapunov function for planar switched systems. In *Conference on Decision and Control*, Atlantis, Bahamas, 2004.

[20] A. P. Molchanov and E. S. Pyatnitskii. Lyapunov functions that specify necessary and sufficient conditions of absolute stability of nonlinear nonstationary control systems, Parts I, II, III. *Automation and Remote Control*, 47:344–354, 443–451, 620–630, 1986.

[21] Y. Mori, T. Mori, and Y. Kuroe. On a wide class of linear constant systems which have a common quadraticLyapunov function. In *36th Conference on Decision and Control*, 1997.

[22] K. S. Narendra and J. Balakrishnan. A common Lyapunov function for stable LTI systems with commuting $\mathcal{A}$-matrices. *IEEE Transactions on Automatic Control*, 39(12):2469–2471, 1994.

[23] Andrezej Polański. On absolute stability analysis by polyhedral Lyapunov functions. *Automatica*, 36:573–578, 2000.

[24] E. S. Pyatnitskii. Criterion for the absolute stability of second order nonlinear controlled systems with one nonlinear, nonstationary element. *Automation and Remote Control*, 32:5–16, 1971.

[25] E. S. Pyatnitskii and L. B. Rapoport. Existence of periodic motion and tests for absolute stability of nonlinear nonstationary systems in the three-dimensional case. *Automation and Remote Control*, 52(5):648–658, 1991.

[26] E. S. Pyatnitskii and L. B. Rapoport. Periodic motion and tests for absolute stability of nonlinear nonstationary systems. *Automation and Remote Control*, 52(10):1379–1387, 1991.

[27] R. Shorten, F. Wirth, O. Mason, K. Wulff, and C. King. Stability criteria for switched and hybrid systems. *SIAM Review*, 49(4):545–592, 2007.

[28] R. N. Shorten and K. S. Narendra. Necessary and sufficient conditions for the existence of a common quadratic Lyapunov function for a finite number of stable second order linear time-invariant systems. *International Journal of Adaptive Control and Signal Processing*, 16(10):709–728, 2002.

[29] R. N. Shorten, F. Ó Cairbre, and P. Curran. On the dynamic instability of a class of switching systems. In *Proceedings of IFAC conference on Artificial Intelligence in Real Time Control*, Budapest, 2000.

[30] Robert N. Shorten and Kumpati S. Narendra. A sufficient condition for the existence of a common Lypunov function for two second order linear systems. In *Proceedings of the 36th Conference on Decision and Control*, San Diego, 1997.

[31] Robert N. Shorten and Kumpati S. Narendra. Necessary and sufficient conditions for the existence of a common quadratic Lyapunov functions for a finite number of stable second order linear time-invariant systems. *International Journal of Adaptive Control and Signal Processing*, 16:709–728, 2003.

[32] Robert N. Shorten and Kumpati S. Narendra. On common quadratic Lyapunov functions for pairs of stable LTI systems whose system matrices are in companion form. *IEEE Transactions on Automatic Control*, 48(4):618–621, 2003.

[33] K. Wulff and R. Shorten. On maximum sector bounds for absolute stability of single-input single-output systems. *International Journal of Control*, 80(6):1–11, 2007.

[34] K. Wulff, R. N. Shorten, and P. Curran. On the 45° region and the uniform asymptotic stability of classes of second order parameter varying and switched systems. *International Journal of Control*, 75(11):812–823, 2002.

[35] Kai Wulff. *Quadratic and Non-Quadratic Stability Criteria for Switched Linear Systems*. PhD thesis, Hamilton Institute, NUI Maynooth, 2005.

[36] Christos A. Yfoulis and Robert Shorten. A numerical technique for stability analysis of linear switched systems. In Rajeev Alur and George J. Pappas, editors, *Hybrid Systems: Computation and Control*, Lecture Notes in Computer Science, pages 631 – 645. Springer-Verlag, 2004.

Neues Regelkonzept für die dynamische Antriebsstrangprüfung

Robert Bauer
Kristl, Seibt & Co Ges.m.b.H.
Baiernstraße 122a, A-8052 Graz
robert.bauer@ksengineers.at

Zusammenfassung

In diesem Beitrag wird ein Überblick über verschiedene Methoden zur Regelung eines Antriebsstrangprüfstands gegeben und ein neues Konzept vorgestellt. Bei diesem neuen Konzept handelt es sich im Wesentlichen um eine Drehzahlregelung der Belastungsmaschinen, wobei allerdings erst eine besondere Maßnahme bei der Sollwertgenerierung den sinnvollen Einsatz in der Praxis ermöglicht.

1 Einleitung

Im Rahmen der Entwicklung von Antriebssträngen sind Tests am realen Objekt notwendig. Diese Tests können beispielsweise mit einem kompletten Versuchsfahrzeug durchgeführt werden, indem bestimmte Fahrmanöver auf einem passenden Untergrund absolviert werden. Hierfür muss zunächst der zu testende Antriebsstrang inklusive geeigneter Messtechnik in ein Fahrzeug verbaut und zum Testgelände transportiert werden. Bedenkt man den hohen Aufwand und die schlechte Reproduzierbarkeit, sind aber Versuche auf einem Prüfstand sinnvoller. Rollenprüfstände (Abb. 1, links) lassen aufgrund des großen Trägheitsmoments der Rollen keine dynamischen Versuche zu. Am besten sind Prüfstände geeignet, bei denen die Belastungsmaschinen direkt mit den Seitenwellen des Antriebsstrangs verbunden sind (Abb. 1, rechts). Diese mechanische Konfiguration ist bereits länger bekannt [1], die geeignete Regelung stellt aber nach wie vor eine Herausforderung dar.

Im Folgenden wird ein Überblick über bekannte Methoden zur Regelung dieser Prüfstandsart gegeben und ein neues Konzept vorgestellt. Beim neuen Konzept handelt es sich im Wesentlichen um eine Drehzahlregelung der Belastungsmaschinen, wobei die Sollwerte von mathematischen Modellen für Fahrzeug, Reifen und Rädern berechnet werden. Dadurch erfährt der Antriebsstrang die gleiche Belastung wie auf der realen Straße, und zwar entsprechend den aktuellen Pedalstellungen sowie dem gewählten Straßenzustand (trocken, nass, eisig, usw.).

Weiters erlaubt das neue Konzept den Einsatz von Belastungsmaschinen mit einem bis zu drei Mal größeren Trägheitsmoment als das der nachzubildenden Räder. Dadurch können Asynchronmaschinen eingesetzt werden, die etwa im Vergleich zu permanenterregten Synchronmaschinen einen geringeren Drehmomentrippel und günstigere Eigenschaften im Fehlerfall aufweisen sowie preiswerter sind.

Abbildung 1: Rollenprüfstand (links) und Antriebsstrangprüfstand (rechts), bei dem die Belastungsmaschinen direkt mit den Seitenwellen verbunden sind

2 Überblick über bekannte Regelungsmethoden

Ausgangspunkt ist ein Antriebsstrangprüfstand nach Abb. 2, bei dem die Prüfstandsregelung die gemessenen Drehzahlen $n_{\mathrm{XY,ist}}$ und Drehmomente $M_{\mathrm{XY,SW}}$ jeder Seitenwelle (XY=LV,RV,LH,RH) zur Verfügung gestellt bekommt und insbesondere die Luftspaltmomente $M_{\mathrm{XY,LS}}$ der Belastungsmaschinen vorgeben kann.

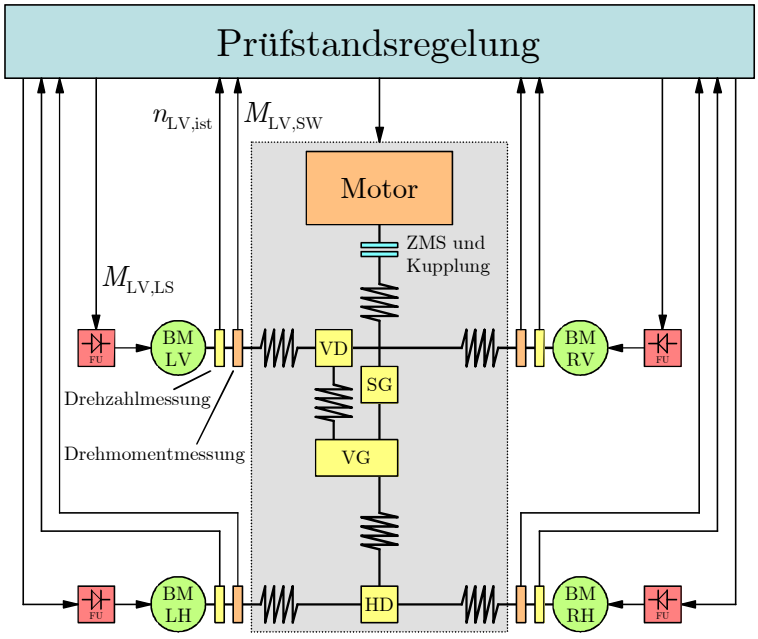


Abbildung 2: Prüfstand mit beispielhaftem Antriebsstrang inklusive Motor, Zweimassenschwungrad (ZMS), Kupplung, Schaltgetriebe (SG), Verteilergetriebe (VG), Vorderachsdifferential (VD) und Hinterachsdifferential (HD) sowie vier Belastungsmaschinen (BM, links/rechts und vorne/hinten) mit zugehörigem Frequenzumrichter (FU)

2.1 Vorbestimmte Drehzahlwerte

Bei dieser sehr einfachen und weit verbreiteten Methode werden vorbestimmte Drehzahlwerte als Sollwerte für eine Drehzahlregelung (Abb. 3) zusammen mit anderen aufgezeichneten Größen wie Pedalstellungen und gewählter Gang verwendet.

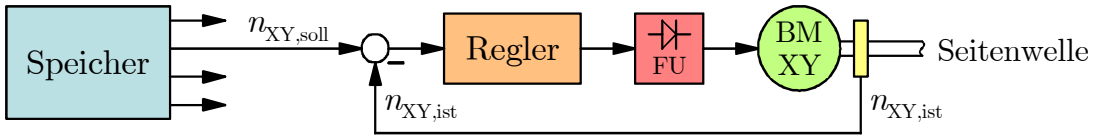


Abbildung 3: Drehzahlregelung mit vorbestimmten Sollwerten (XY=LV,RV,LH,RH)

Die vorbestimmten Werte stammten hierbei entweder von einer realen Testfahrt mit einem entsprechend ausgerüsteten Versuchsfahrzeug oder von einer Simulation mit mathematischen Modellen für Antriebsstrang, Räder, Reifen und Karosserie (z.B. [2-4]). Diese Methode weist allerdings eine Reihe von Nachteilen auf:

- Verhält sich der Antriebsstrang am Prüfstand nicht so wie im Versuchsfahrzeug oder in der Simulation, stimmen die vorgegebenen Drehzahlprofile nicht mit jenen überein, die sich im realen Fahrzeug auf der Straße ergeben würden. Ist beispielsweise die Leistung der Verbrennungskraftmaschine am Prüfstand geringer als im Versuchsfahrzeug, sind die im Voraus ermittelten Raddrehzahlen zu groß. Enthält der Antriebsstrang ein Automatikgetriebe, kann dieses zu anderen Zeitpunkten schalten als im Versuchsfahrzeug oder in der Simulation. Diese Problematik wird durch den zunehmenden Einsatz „intelligenter“ Antriebsstränge verschärft, die ihr Verhalten an den Fahrer, die aktuellen Straßenverhältnisse usw. anpassen und somit bewusst ändern.

- Die im Voraus ermittelten Zeitverläufe werden am Prüfstand als Sollwerte für Regelkreise mit unterschiedlicher Dynamik verwendet. Prinzipbedingt ergeben sich vor allem bei dynamischen Prüfungen Unterschiede in den zeitlichen Verläufen, die einer realistischen Prüfung entgegenstehen. Beispielsweise bleiben bei einem Anfahrtsszenario aus dem Stillstand die Belastungsmaschinen zu lange stehen, wodurch es im Antriebsstrang zu starken Drehmoment-Stoßbelastungen kommt, die im realen Fahrzeug auf der Straße nicht auftreten würden.

- Bei realen Drehzahlregelungen ergeben sich selbst bei Vorgabe identischer Sollwerte unterschiedliche Ist-Drehzahlen, die im Antriebsstrang zu unerwünschten Verspannzuständen führen können. Leicht einsichtig ist dieser Umstand bei einem komplett gesperrten Differential, bei dem bereits eine geringe Winkeldifferenz zwischen den Seitenwellen zu hohen Drehmomenten führt, die den Antriebsstrang sehr leicht schädigen können. Bei heute oft üblichen Sperrdifferentialen (die absichtlich eine leichte Sperrwirkung erzeugen) tritt dieser Effekt ebenfalls auf. Im realen Fahrzeug auf der Straße sorgen hingegen unterschiedlich stark schlupfende Räder für einen Ausgleich.

Zusammenfassend besteht das Hauptproblem dieser Methode darin, dass der Prüfstand nicht auf den Prüfling *reagiert* sondern nur in der vorbestimmten Weise *agiert*.

2.2 Klassische Straßenlastsimulation (RLS)

Bei der klassischen Straßenlastsimulation (road load simulation, RLS) wird die auf ein Fahrzeug wirkende Widerstandskraft mit Hilfe der sehr einfachen Formel

$$F_{\mathrm{RLS}} = R_0 + R_1 v + R_2 v^2 \tag{1}$$

mit nur drei konstanten Parametern R_0, R_1 und R_2 aus der aktuellen Fahrzeuggeschwindigkeit v berechnet. Zur Ermittlung der Fahrzeuggeschwindigkeit wird der Impulssatz

$$m\frac{dv}{dt} = -F_{\mathrm{RLS}} - mg\sin\gamma + \frac{M_{\mathrm{SW}}}{r_{\mathrm{dyn}}} \tag{2}$$

mit der Fahrzeugmasse m, der Erdbeschleunigung g, dem Steigungswinkel γ, der Seitenwellen-Drehmomentsumme

$$M_{\Sigma,\mathrm{SW}} = M_{\mathrm{LV,SW}} + M_{\mathrm{RV,SW}} + M_{\mathrm{LH,SW}} + M_{\mathrm{RH,SW}} \tag{3}$$

und dem dynamischen Radradius r_{dyn} verwendet. Aus der Fahrzeuggeschwindigkeit wird eine (einzige) Raddrehzahl

$$n_{\mathrm{soll}} = \frac{30}{\pi}\frac{v}{r_{\mathrm{dyn}}} \tag{4}$$

berechnet, die als Sollwert für die Drehzahlregelung aller Belastungsmaschinen dient (Abb. 4).

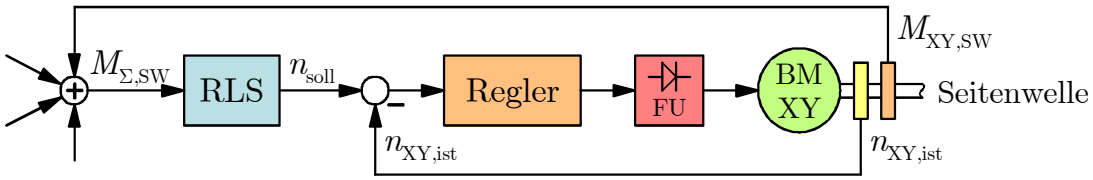


Abbildung 4: Klassische Straßenlastsimulation (XY=LV,RV,LH,RH)

Somit ist der Hauptnachteil der vorhergehenden Methode beseitigt, der Prüfstand reagiert auf den Prüfling. Allerdings weist auch diese Methode eine Reihe von Nachteilen auf:

- Durch die starre Kopplung der (einzigen) Raddrehzahl an die Fahrzeuggeschwindigkeit ist kein Radschlupf möglich, dynamische Fahrsituationen können daher am Prüfstand nicht nachgestellt werden.
- Die sich ergebende Drehmomentverteilung entspricht nicht unbedingt jener, die sich im realen Fahrzeug auf der Straße ergeben würde (vergl. Abschnitt 4.1).
- Wie bei der vorhergehenden Methode können wieder leicht unterschiedliche Ist-Drehzahlen bei Sperrdifferentialen zu unerwünschten Belastungen führen.

Zusammenfassend besteht das Problem dieser Methode darin, dass durch die Summenbildung die Reaktion auf die Momente der *einzelnen* Seitenwellen zu stark vereinfacht ist, um eine realistische Belastung des Antriebsstrangs zu ermöglichen.

2.3 Reifenmodell mit Drehmoment als Ausgang

Bei der vorhergehenden Methode wurde ein zu einfaches mathematisches Modell für die Reaktion auf den Antriebsstrang verwendet. Bei der ersten Methode im Abschnitt 2.1 wurden bereits detailliertere mathematische Modelle angesprochen, die dort zur Berechnung von vorbestimmten Drehzahlwerten verwendet werden. Es ist nun naheliegend, diese detaillierteren Modelle am Prüfstand in Echtzeit mitlaufen zu lassen, um so den Prüfling realistisch zu belasten.

Diese Idee wird in einer Patentschrift [5] aufgegriffen und folgenderweise umgesetzt: An jeder Seitenwelle wird die Raddrehzahl gemessen und dient als Eingangsgröße für ein die schlupfabhängige Reibung nachbildendes Reifenmodell. Jedes Reifenmodell steht mit dem Karosseriemodell in Verbindung und berechnet zusammen mit dessen Ausgangswerten eine vom Reifen auf die Fahrbahn übertragene Kraft (die wiederum als Eingangsgröße für das Karosseriemodell dient) und das zugehörige Drehmoment, das als Luftspaltmoment für die Belastungsmaschine an dieser Seitenwelle verwendet wird (Abb. 5).

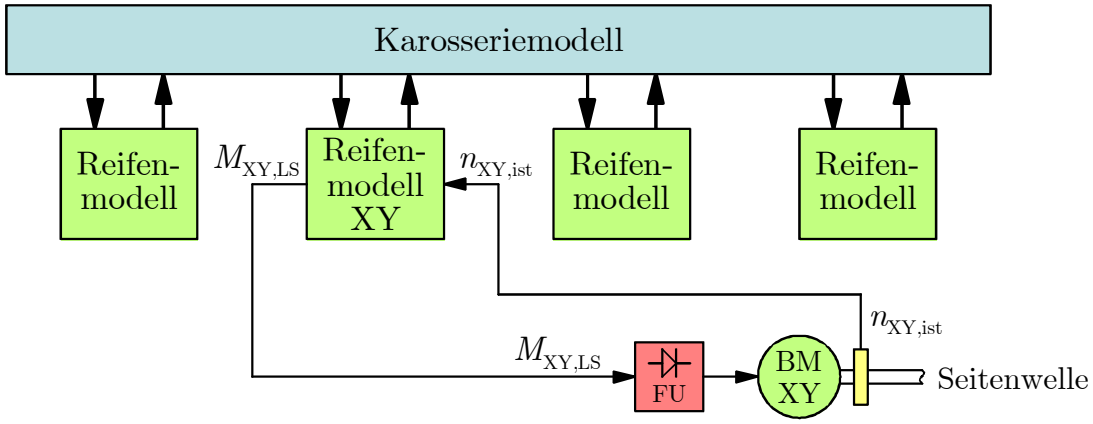


Abbildung 5: Prüfstand mit Karosserie- und Reifenmodellen (XY=LV,RV,LH,RH)

Diese Umsetzung weist allerdings eine Reihe von Nachteilen auf:

- Das Trägheitsmoment der Belastungsmaschine muss dem Trägheitsmoment des Rades (ca. 1-2 kgm^2) entsprechen. Da auch große Drehmomente gefordert sind, kommt für die Belastungsmaschine nach heutigem Stand der Technik nur eine permanenterregte Synchronmaschine mit sehr geringem Trägheitsmoment (ca. 1 kgm^2) in Frage. Dieser Maschinentyp hat aber typischerweise einen deutlich höheren Drehmomentrippel als beispielsweise Asynchronmaschinen, wodurch am Prüfstand Drehmomentschwankungen auftreten, die es im realen Fahrzeug auf der Straße nicht gibt. Weiters liegt bei einer sich drehenden permanenterregten Synchronmaschine an den Motorklemmen immer Spannung an, die im Fehlerfall gefährlich werden kann und schließlich ist dieser Maschinentyp wesentlich teurer als andere Bauformen mit vergleichbaren Leistungsdaten.

- Sollen Räder mit größerem Trägheitsmoment am Prüfstand nachgebildet werden, müssen die Belastungsmaschinen zusätzlich mit Schwungscheiben ausgerüstet werden.

- Die Belastungsmaschine müsste exakt das berechnete Soll-Luftspaltdrehmoment aufbringen. Aus regelungstechnischer Sicht ist dies schwierig, da dieses Drehmoment – im Gegensatz zum Drehmoment an der Seitenwelle – nicht gemessen werden kann und eine modellbasierte Schätzung immer leicht fehlerhafte Werte liefert. Da die Drehzahl dem integrierten Drehmoment proportional ist, integriert sich dieser Fehler auf.

Zusammenfassend besteht das Hauptproblem dieser Methode darin, dass zwar für Karosserie und Reifen mathematische Modelle verwendet werden, das Rad hingegen *physikalisch* mit Hilfe der Belastungsmaschine nachgebildet wird.

3 Neues Regelkonzept

Das Erkennen des Hauptproblems der vorhergehenden Methode im Abschnitt 2.3 führt zu folgender Idee: Es wird wieder ein Karosseriemodell am Prüfstand in Echtzeit verwendet, anstelle der Reifenmodelle werden aber Radmodelle eingesetzt, die die Reifenmodelle beinhalten. Nun dient das an jeder Seitenwelle gemessene Drehmoment als Eingangsgröße und man erhält eine Raddrehzahl, die als Sollwert für eine Drehzahlregelung verwendet wird (Abb. 6).

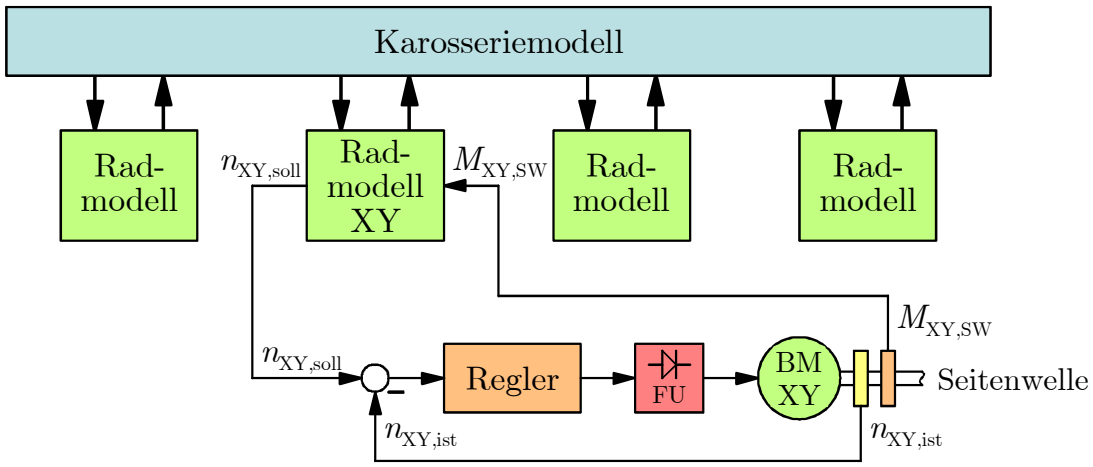


Abbildung 6: Neues Regelkonzept mit Karosserie- und Radmodellen (XY=LV,RV,LH,RH)

Mit dem neuen Konzept sind alle Nachteile der vorhergehenden Methoden eliminiert. Damit allerdings dieses Konzept auch in der Praxis funktioniert, muss noch eine besondere Maßnahme bei der Sollwertgenerierung getroffen werden, die im Widerspruch zur klassischen Modellbildung steht und im folgenden Abschnitt besprochen wird.

3.1 Trägheitskompensation mit Drehzahlregelung

Im Folgenden soll untersucht werden, wie ein Rad mit dem Trägheitsmoment J_R (Abb. 7, links) möglichst gut am Prüfstand nachgebildet werden kann.

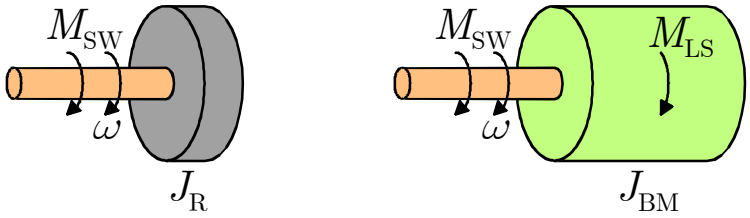


Abbildung 7: Wunschkonfiguration (links) und Verhältnisse am Prüfstand (rechts)

Hierzu wird vereinfachend angenommen, dass nur das Drehmoment M_{SW} über die Seitenwelle wirkt, also der Drallsatz

$$J_R \frac{d\omega}{dt} = M_{SW} \tag{5}$$

lautet. Auf ein reales Rad wirken natürlich noch andere Drehmomente, wie jenes, das der vom Reifen auf die Fahrbahn übertragenen Kraft oder der Rollreibung entspricht, diese zusätzlichen Drehmomente würden aber hier nur den Blick auf das Wesentliche erschweren. Die Übertragungsfunktion vom Drehmoment M_{SW} zur Ist-Drehzahl n_{ist} lautet daher idealerweise:

$$T_{ideal}(s) = \frac{n_{ist}(s)}{M_{SW}(s)} = \frac{1}{J_R s}\frac{30}{\pi} \tag{6}$$

Am Prüfstand befindet sich nun anstelle des Rades eine Belastungsmaschine mit dem Trägheitsmoment J_{BM} (Abb. 7, rechts), dafür kann zusätzlich das Luftspaltmoment M_{LS} aufgebracht werden. Bei der klassischen Trägheitskompensation mit Drehzahlregelung [6] wird zunächst mit Hilfe des Drallsatzes nach Gl. (5) eine Drehzahl n_{soll} berechnet, die anschließend als Sollwert für eine Drehzahlregelung verwendet wird (Abb. 8).

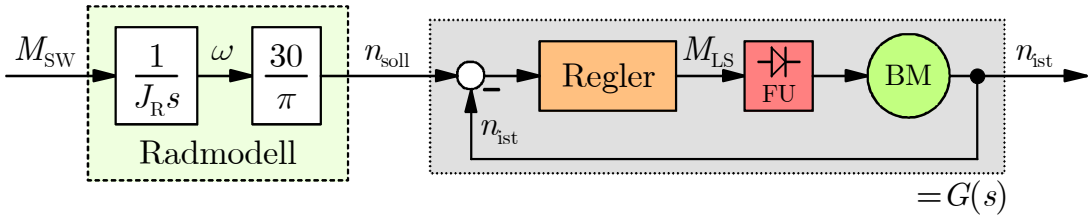


Abbildung 8: Regelungsschema der klassischen Trägheitskompensation mit Drehzahlregelung

Bei dynamischen Vorgängen sind Soll- und Istwert zwangsläufig nicht identisch (sonst wäre der Regler-Eingang Null), der Zusammenhang werde durch die Übertragungsfunktion

$$G(s) = \frac{n_{ist}(s)}{n_{soll}(s)} \tag{7}$$

beschrieben. Die Übertragungsfunktion vom Drehmoment der Seitenwelle zur Ist-Drehzahl lautet daher

$$T_{klassisch}(s) = \frac{n_{ist}(s)}{M_{SW}(s)} = \frac{1}{J_R s}\frac{30}{\pi}G(s) \tag{8}$$

und weicht offensichtlich vom Idealzustand nach Gl. (6) ab. Vereinfacht gesagt „hinkt“ das simulierte Rad dem realen Rad um die Dynamik der Drehzahlregelung hinterher, wodurch das simulierte Rad nicht das erwünschte Verhalten aufweist und in Kombination mit dem Antriebsstrang sogar instabil werden kann. Abhilfe schafft nun eine einfache Maßnahme, bei der im Radmodell *anstelle* des Drallsatzes die Übertragungsfunktion

$$P(s) = \frac{1}{J_R s} \frac{1}{G(s)} \tag{9}$$

eingesetzt wird (Abb. 9).

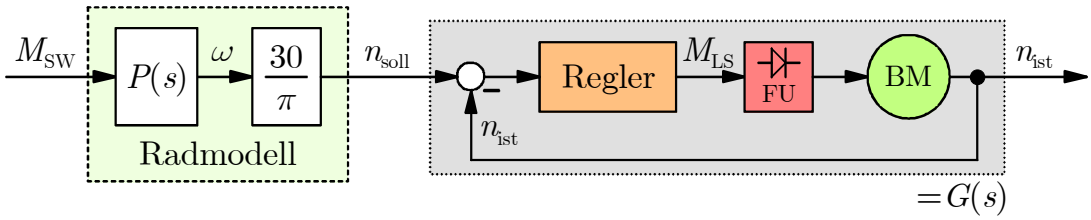


Abbildung 9: Trägheitskompensation mit neuer Sollwertgenerierung

Die Übertragungsfunktion vom Drehmoment der Seitenwelle zur Ist-Drehzahl lautet nun

$$T_{neu}(s) = \frac{n_{ist}(s)}{M_{SW}(s)} = \frac{1}{J_R s} \frac{1}{G(s)} \frac{30}{\pi} G(s) = \frac{1}{J_R s} \frac{30}{\pi} \tag{10}$$

und entspricht dem Idealzustand nach Gl. (6), das simulierte Rad verhält sich wie das reale Rad. Bemerkenswert ist die Tatsache, dass das neue Radmodell mit Gl. (9) *nicht* dem Drallsatz nach Gl. (5) entspricht und daher rein aus Sicht der Modellbildung *falsch* ist!

Anmerkung: Eigentlich stimmen Abb. 8 und 9 nicht ganz, da das Drehmoment der Seitenwelle nicht nur auf des Radmodell, sondern auch auf die Belastungsmaschine wirkt. Dieser zusätzliche Zweig würde aber die Überlegungen nur verkomplizieren und das Ergebnis unwesentlich verändern.

3.1.1 Beispiel

Ein kurzes Beispiel soll die Vorgangsweise verdeutlichen. Die Übertragungsfunktion der Drehzahlregelung laute

$$G(s) = \frac{b_0}{s + a_0} \tag{11}$$

und somit muss das Radmodell die Übertragungsfunktion

$$P(s) = \frac{1}{J_R s} \frac{1}{G(s)} = \frac{1}{J_R s} \frac{s + a_0}{b_0} = \frac{1}{J_R b_0} + \frac{a_0}{J_R b_0} \frac{1}{s} \tag{12}$$

enthalten, die von der Struktur einem PI-Regler entspricht. Man beachte, dass die Inverse von $G(s)$ alleine nicht realisierbar wäre, erst durch den Polüberschuß des Drallsatzes ist $P(s)$ realisierbar!

4 Experimentelle Ergebnisse

Das neue Regelkonzept wurde an einem Antriebsstrangprüfstand implementiert und anhand mehrerer Szenarien getestet. Als Prüfling wurde ein allradgetriebenes Fahrzeug (83 kW, 5-Gang-Schaltgetriebe) mit Torsen-Verteilergetriebe (nominelle Drehmomentverteilung vorne 50 %, hinten 50 %), einer Aufstandskraftverteilung von vorne 60 % zu hinten 40 % und Räder mit einem Trägheitsmoment von 1.5 kgm^2 eingesetzt. Als Belastungsmaschinen wurden Asynchronmaschinen (Nennleistung 219 kW, Nenndrehmoment 2091 Nm) verwendet, deren Trägheitsmoment mit 5 kgm^2 mehr als drei Mal so groß wie jenes der Räder ist.

4.1 Szenario 1: Gas geben

Bei diesem Szenario fährt man auf einer geraden, trockenen Straße mit einem bestimmten Gang (hier im ersten Gang), geht mit dem Fuß zunächst komplett vom Gaspedal und tritt es dann voll durch. Da die Kupplung immer geschlossen bleibt, können die im Antriebsstrang auftretenden Schwingungen hervorragend reproduziert und untersucht werden. Abb. 10 zeigt zum Vergleich die Drehmomente an den Seitenwellen mit der klassischen Straßenlastsimulation (links) und mit dem neuen Regelkonzept (rechts).

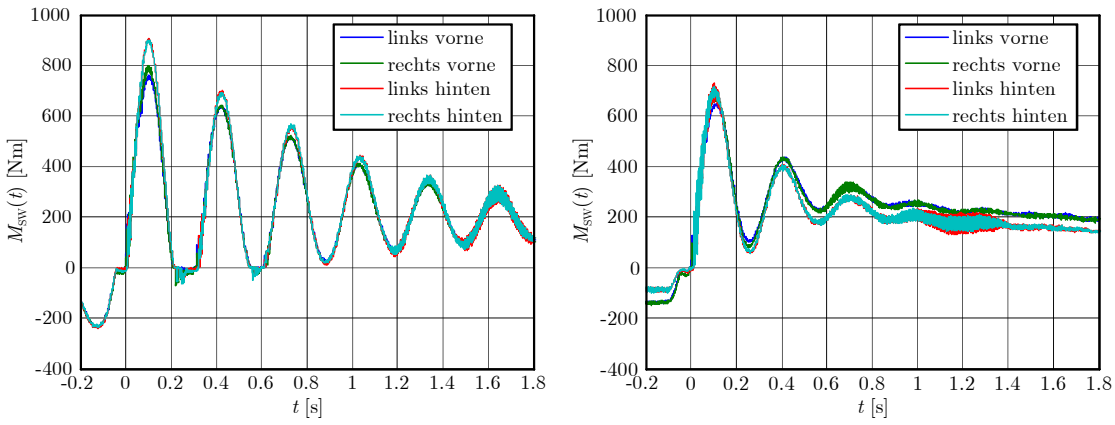


Abbildung 10: Drehmomente an den Seitenwellen beim Szenario „Gas geben“ mit klassischer Straßenlastsimulation (links) und mit neuem Regelkonzept (rechts)

Zwei Unterschiede sind deutlich erkennbar:

- Mit der klassischen Straßenlastsimulation sind die Schwingungen im Antriebsstrang sehr schwach gedämpft und völlig unrealistisch. Mit der neuen Methode wird hingegen der Antriebsstrang wie im realen Fahrzeug auf der Straße belastet.
- Die klassische Straßenlastsimulation führt im eingeschwungenen Zustand zu gleich großen Drehmomenten an allen Seitenwellen. Mit dem neuen Konzept erhält man hingegen eine Drehmomentverteilung wie auf der Straße: Das Verteilergetriebe würde zwar nominell die Drehmomente vorne und hinten gleich verteilen, aufgrund der unterschiedlichen Aufstandskräfte (vorne 60 %, hinten 40 %) würden die Hinterräder

aber mehr schlupfen und sich schneller drehen. Dies unterbindet das Verteilergetriebe durch entsprechend mehr Drehmoment an der Vorderachse.

Die unterschiedliche Drehmomentverteilung ist übrigens nicht anhand der Drehzahlen erkennbar, wie man in Abb. 11 (Detailausschnitt von Abb. 10 gegen Ende) gut erkennen kann. Dies führt zu einer weitere Erkenntnis: Auch das Nachfahren vorbestimmter Drehzahlwerte (Methode nach Abschnitt 2.1) würde bei diesem Verteilergetriebe *keinesfalls* zur gleichen Drehmomentverteilung wie auf der Straße führen!

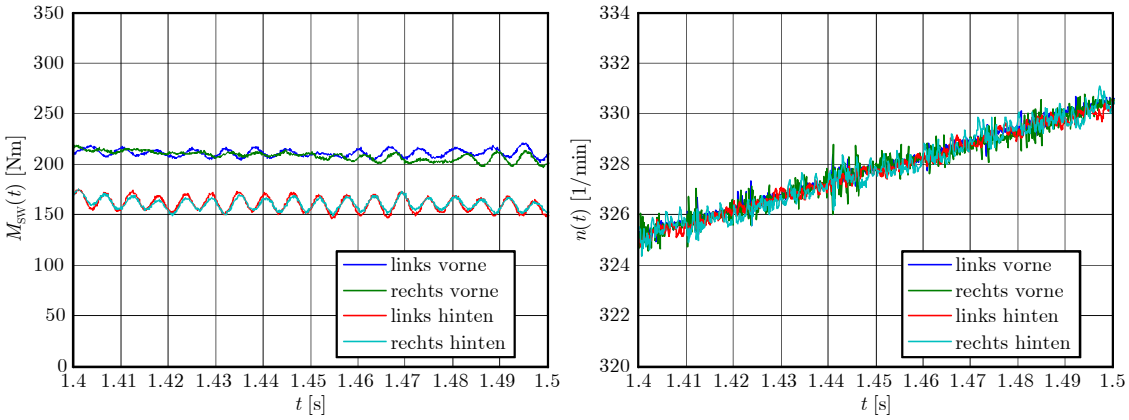


Abbildung 11: Drehmomente an den Seitenwellen (links) und Drehzahlen der Räder (rechts) mit neuem Regelkonzept (Detailausschnitt von Abb. 10 mit gleicher Zeitskala)

Abb. 12 zeigt einen weiteren Detailausschnitt von Abb. 10 um den Zeitpunkt Null. Man kann sehr gut erkennen, dass aufgrund der Spiele im Antriebsstrang das Drehmoment an der Hinterachse erst später als an der Vorderachse steigt (Abb. 12, links). Für eine realistische Belastung ist es nun entscheidend, dass sich die Drehzahlen der Räder *unterschiedlich* und zeitlich *exakt* passend ändern (Abb. 12, rechts).

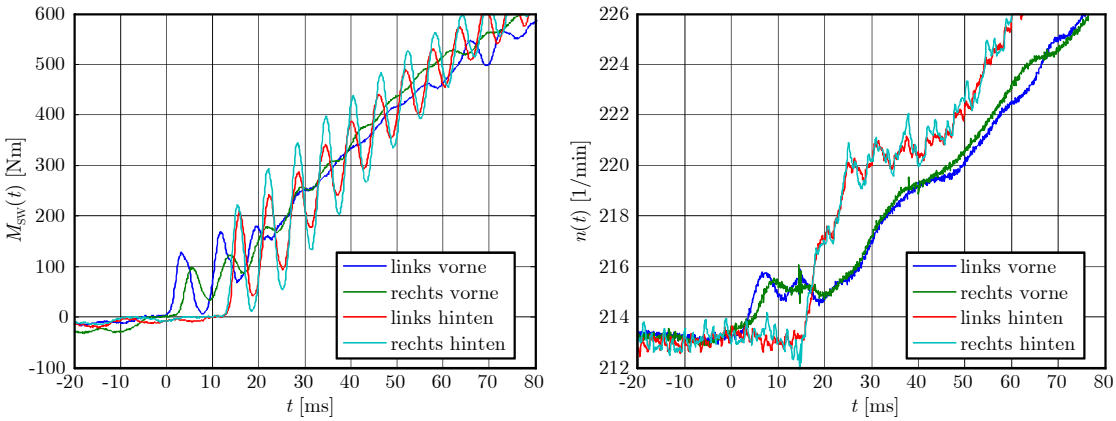


Abbildung 12: Drehmomente an den Seitenwellen (links) und Drehzahlen der Räder (rechts) mit neuem Regelkonzept (Detailausschnitt von Abb. 10 mit gleicher Zeitskala)

4.2 Szenario 2: Eisplatte

Bei diesem Szenario beschleunigt man auf einer geraden, trockenen Straße im zweiten Gang mit komplett durchgetretenem Gaspedal. Während des Beschleunigungsvorgangs fährt die rechte Fahrzeugseite über eine kurze, rutschige Stelle (z.B. eine Eisplatte, siehe Abb. 13).

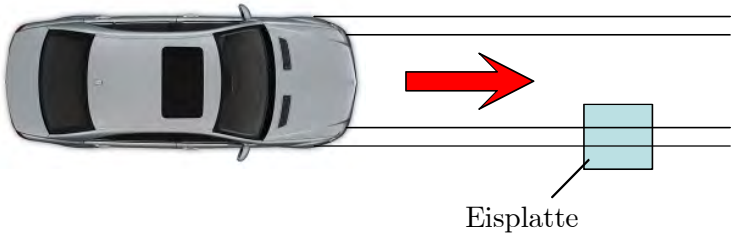


Abbildung 13: Fahrzeug mit Eisplatte auf der rechten Seite der Fahrspur

Zunächst dreht das rechte Vorderrad durch und wird gleich wieder von der trockenen Straße abgebremst, anschließend geschieht das gleiche mit dem rechten Hinterrad. Aufgrund der enormen Drehmomentstöße ist dieses Szenario eine Herausforderung – für den Antriebsstrang genauso wie für den Prüfstand. Abb. 14 (links) zeigt, dass mit dem neuen Konzept der Antriebsstrang wie erwartet belastet wird. Sehr gut sind auch die Reaktionen auf der linken Fahrzeughälfte aufgrund der Differentiale erkennbar.

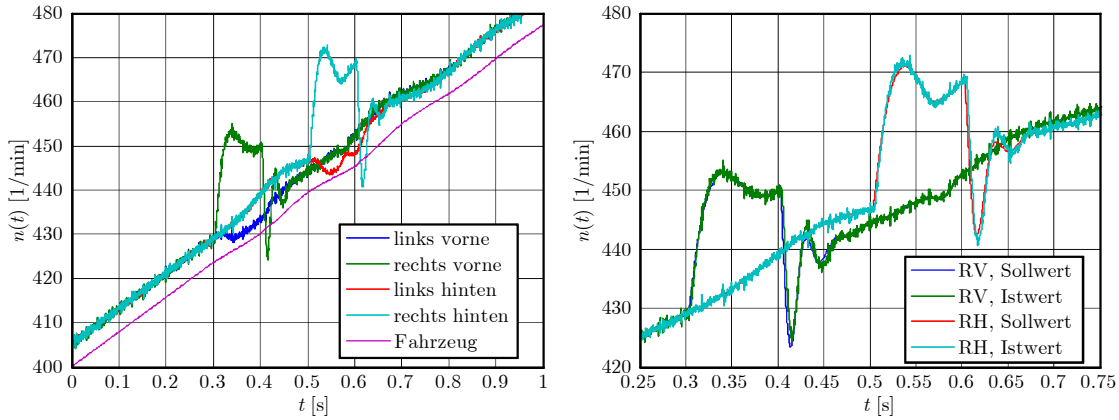


Abbildung 14: Drehzahlen der Räder (links, die Kurve mit der Bezeichnung „Fahrzeug“ entspricht der Drehzahl eines schlupffreien Rades) und Soll-/Istwertvergleich (rechts) beim Szenario „Eisplatte“

Die rechte Seite von Abb. 14 zeigt einen Vergleich der Soll- und Istdrehzahlen auf der rechten Fahrzeugseite. Obwohl das Trägheitsmoment der Belastungsmaschinen mehr als drei Mal so groß wie jenes der Räder ist, folgen sie dem geforderten Drehzahlverlauf hervorragend.

4.3 Szenario 3: Schneefahrbahn

Bei diesem Szenario fährt man auf einer geraden Schneefahrbahn im zweiten Gang. Zunächst beschleunigt man mit durchgetretenem Gaspedal, anschließend steigt man auf die Bremse. Abb. 15 zeigt die Drehzahlen der Räder am Prüfstand, man kann die druchdrehenden Räder in der Beschleunigungsphase sowie das Einsetzen des Anti-Blockier-Systems (ABS) in der Bremsphase sehr gut erkennen.

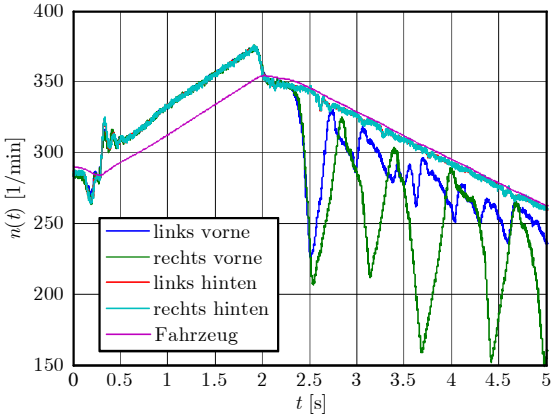


Abbildung 15: Drehzahlen der Räder beim Szenario „Schneefahrbahn“ (die Kurve mit der Bezeichnung „Fahrzeug“ entspricht der Drehzahl eines schlupffreien Rades)

Offensichtlich funktioniert das ABS beim Prüfling auf der rechten Vorderseite nicht richtig, da diese Drehzahl immer wieder zu stark absinkt. Wesentlich wichtiger ist aber die Feststellung, dass mit dem neuen Konzept das Anti-Blockier-System ohne jegliche Zusatzmaßnahme am Prüfstand getestet werden kann.

5 Zusammenfassung

Zunächst wurde ein Überblick über bekannte Methoden zur Regelung von Antriebsstrangprüfständen gegeben. Anschließend wurde ein neues Konzept vorgestellt, bei dem es sich im Wesentlichen um eine Drehzahlregelung der Belastungsmaschinen handelt, wobei die Sollwerte von mathematischen Modellen für Fahrzeug, Reifen und Rädern berechnet werden. Für einen erfolgreichen Einsatz in der Praxis muss allerdings noch ein besonderes Detail bei der Sollwertgenerierung berücksichtigt werden, das im Widerspruch zur klassischen Modellbildung steht. Das neue Konzept bietet im Vergleich zu bekannten Methoden mehrere Vorteile:

- Der Antriebsstrang wird wie im realen Fahrzeug auf der Straße entsprechend den individuellen Pedalstellungen des Fahrers und der gewählten Straßenbeschaffenheit (trocken, nass, eisig, usw.) belastet. Die Drehzahlen der einzelnen Räder reagieren jeweils zeitlich exakt passend zu den Drehmomenten an den Seitenwellen. Selbst bei drehmomentfühlenden Verteilergetrieben und Sperrdifferentialen stellt sich automatisch die gleiche Drehmomentverteilung wie auf der Straße ein.

- Das Anti-Blockier-System kann am Prüfstand ohne jegliche Zusatzmaßnahme getestet werden, für andere Fahrerassistenzsysteme müssen nur jene Sensoren simuliert werden, deren Messwerte am Prüfstand sinnlos wären (Beschleunigungssensoren, Gierratensensor, usw.).

- Unterschiedliche Trägheitsmomente der Räder können am Prüfstand rein durch Vorgabe über die Software und ohne physikalische Umbauten an den Belastungsmaschinen nachgebildet werden.

- Das Trägheitsmoment der Belastungsmaschinen darf bis zu drei Mal größer als jenes der Räder sein. Dadurch können Asynchronmaschinen als Belastungsmaschinen eingesetzt werden, die sich – im Vergleich zu permanenterregten Synchronmaschinen – durch einen sehr geringen Drehmomentrippel, ein günstigeres Verhalten im Fehlerfall und geringere Kosten auszeichnen.

- Die für die Regelung notwendigen Messgrößen (Drehzahlen und Drehmomente an den Seitenwellen) können direkt, sehr einfach und sehr genau gemessen werden. Da keine mathematischen Modelle zur Berechnung von nicht direkt messbaren Größen notwendig sind, können sich hier auch keine Modellfehler auswirken.

Abschließend wurde das neue Regelkonzept an einem Antriebsstrangprüfstand implementiert und anhand mehrerer Szenarien getestet, die Messwerte demonstrieren die Funktionstüchtigkeit des Konzepts in der Praxis.

Literatur

[1] Thun H.: *Prüfstand zum Testen des Antriebsstranges eines Fahrzeuges.* Europäische Patentschrift EP 0 338 373 B1, Europäisches Patentamt, 1989.

[2] Mitschke M., Wallentowitz H.: *Dynamik der Kraftfahrzeuge.* 4. Auflage, Springer, Berlin Heidelberg, 2004.

[3] Pacejka H.: *Tyre and Vehicle Dynamics.* Second Edition, Butterworth-Heinemann, Oxford, 2007.

[4] Rill G.: *Simulation von Kraftfahrzeugen*, Vieweg+Teubner, Regensburg, 1994.

[5] Germann S. et al.: *Verfahren zum Simulieren des Verhaltens eines Fahrzeugs auf einer Fahrbahn.* Europäische Patentschrift EP 1 037 030 B1, Europäisches Patentamt, 2000.

[6] Fegraus C., D'Angelo S.: *Inertia and Road Load Simulation for Vehicle Testing.* United States Patent US 4 161 116, United States Patent and Trademark Office, 1979.

Selbsteinstellender Stützregler zur Frequenzgangsmessung von Servoantriebsachsen bei externem Lastmoment

Joachim Weißbacher, Martin Horn, Jakob Rehrl
Alpen Adria Universität Klagenfurt
Institut für Intelligente Systemtechnologien
Lehrstuhl für Mess- und Regelungssysteme
Universitätsstraße 65-67, 9020 Klagenfurt
jweissba@edu.uni-klu.ac.at*

Zusammenfassung

Dieser Beitrag beschäftigt sich mit dem Entwurf eines selbsteinstellenden Reglers (STC) für Servoantriebsachsen, auf die konstante Lastmomente einwirken können. Dabei besteht das Ziel nicht darin, einen Regler zu entwerfen, der zu hoher Dynamik führt, sondern einen einfach zu parametrierenden „Stützregler“ zu finden, welcher den Einfluss des Lastmoments kompensiert. Somit bietet sich die Möglichkeit, im geschlossenen Kreis die Identifikation des Frequenzgangs der Strecke durchzuführen. Dieser kann als Grundlage für erweiterte Regelungskonzepte dienen. In einem ersten Schritt wird ein für den Reglerentwurf notwendiges, einfaches Modell abgeleitet, dessen Parameter unbekannt sind. Basierend auf diesem Modell wird der Regler nach geeigneten Kriterien entworfen, wobei sich zeigt, dass *ein* zu wählender Entwurfs- und *ein* unbekannter Schätzparameter dessen Verhalten vollständig beschreiben. Die Ermittlung des Schätzparameters wird durch einen klassischen Recursive-Least-Squares (RLS) Algorithmus durchgeführt. Des weiteren wird eine Hilfestellung bei der Wahl des Entwurfsparameters gegeben. Die Funktionstüchtigkeit des Regelungskonzepts wird abschließend anhand von Messungen an einem Testaufbau gezeigt.

1 Einleitung und Motivation

Ein Servoantrieb ist ein elektronisch geregelter Antrieb zur Umsetzung von mechanischen Bewegungen. Er besteht unter anderem aus einem Servoverstärker und einem Servomotor. Der Verstärker dient als Stellglied für den Motor, der einen Lagegeber besitzt, welcher wiederum seine Geberinformation dem Verstärker zur Verfügung stellt. Diese wird zum

*Korrespondenz bitte an diese Adresse

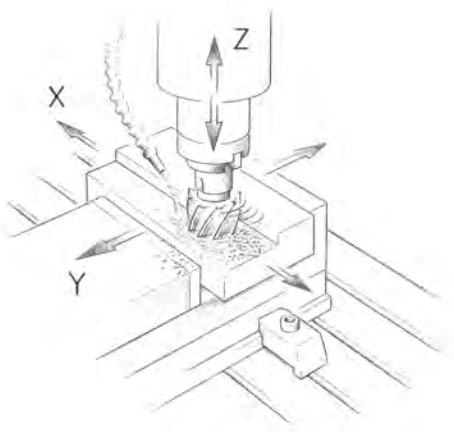


Abbildung 1: Beispiel einer Vertikalachse

einen für die Kommutierung des Stromes benötigt und zum anderen für die Regelung von Drehzahl und Lage. Ein Servoantrieb in Kombination mit der anzutreibenden Mechanik wird umgangssprachlich als Servoantriebsachse bezeichnet. Servoantriebe werden überall dort eingesetzt, wo Maschinenteile eine vorgegebene Bewegung ausführen müssen - so z.B. in Werkzeugmaschinen, Verarbeitungsmaschinen, Druckmaschinen, Robotern usw. [13]. Der Vorteil von Servoantrieben liegt darin, dass sie auf schnelle Drehzahländerung ausgelegt sind (kleines Massenträgheitsmoment, großes Spitzendrehmoment) und somit sehr dynamische und (bei entsprechend hoher Geberauflösung) auch sehr präzise Bewegungen ausführen können [16]. Voraussetzung dafür ist allerdings eine an die jeweilige Mechanik angepasste Reglerparametrierung. Genau darin liegt aber in der Praxis häufig die Schwierigkeit. Der Inbetriebnehmer der Maschine, oft eine Person mit geringem regelungstechnischen Wissen, muss durch intuitive Vorgehensweise (i.A. „Versuch und Irrtum“) zu einer zufriedenstellenden Reglereinstellung kommen. Erschwert wird die Situation noch, wenn aufgrund des mechanischen Aufbaus die Servoantriebsachse eine Vertikalachse ist. Eine solche ist dadurch gekennzeichnet, dass bei deaktivierter Regelung die Mechanik aufgrund der Gravitation beschleunigt wird. Ein typisches Beispiel einer Vertikalachse ist die z-Achse bei einer Fräsmaschine wie in Abbildung 1 dargestellt. Eine weitere Erschwernis ist dann gegeben, wenn durch die Mechanik eine Umsetzung einer Drehbewegung in eine translatorische Bewegung ausgeführt wird. In diesem Fall ist der Bewegungsbereich eingeschränkt. Das führt dazu, dass oft viele bekannte Methoden der Reglereinstellung (z.B. Sprungantwortmethoden [2]) nicht angewandt werden können.
Gewöhnlich wird mit der gefundenen Reglereinstellung die volle Dynamik der Maschine bei weitem nicht genützt. Das bedeutet in weiterer Folge geringere Produktivität und demzufolge Wettbewerbsnachteil für den Betreiber der Maschine.
Eine systematischere Vorgehensweise basiert auf der Messung und Auswertung des Frequenzgangs der Mechanik [10], [11], [14], [15]. Dazu wird die Mechanik mit einem Pseudo-Rausch-Binär-Signal (PRBS) angeregt und der Frequenzgang aus der Systemreaktion mit Hilfe der Fast-Fourier-Transformation (FFT) berechnet. Dieses nichtparametrische Modell dient dann als Grundlage zur Ermittlung der unbekannten Parameter des mechanischen Systems. Weiters wird in [11] erwähnt, dass durch Parametrierung des Drehzahl-

reglers der Frequenzgang auch im geschlossenen Kreis gemessen werden kann, wenn z.B. eine hängende Last auf den Antrieb wirkt. Allerdings wird keine Angabe gemacht, wie die Parameter des Drehzahlreglers zu wählen sind. Werden diese unpassend ausgewählt, kann dies aufgrund der fehlenden Positionsregelung zu großen Ausgleichsbewegung führen.
In [17] wird in einem ersten Schritt mittels Zweipunktregler und übergeordnetem Lageregler das Gesamtmassenträgheitsmoment der Servoantriebsachse bestimmt. Damit wird in einem zweiten Schritt ein provisorischer PI-Drehzahl- und P-Lageregler parametriert, so dass damit auch bei externem Lastmoment der Frequenzgang identifiziert werden kann. Weiters wird gezeigt, dass mit Hilfe der Methode der Harmonischen Balance Abschätzungen hinsichtlich der maximal zu erwartenden Drehzahl- sowie Positionsabweichungen gemacht werden können. Allerdings ist dafür eine Festlegung auf Minimal- und Maximalwert des Gesamtmassenträgheitsmoments notwendig. Während der Minimalwert durch die Motorträgheit bestimmt wird, ist der Maximalwert oft unbekannt und daher schwierig zu ermitteln.
Da das Gesamtmassenträgheitsmoment eine zentrale Größe bei der Parametrierung des Antriebsreglers darstellt, gibt es viele Arbeiten, welche sich mit seiner Online-Identifikation befassen [4], [18], [1], [8]. So wird z.B. in [18] ein konstant einwirkendes oder langsam veränderliches Lastmoment mit Hilfe eines Beobachters geschätzt und mit jenem verglichen, welches sich durch Mittelwertbildung über mehrere Bewegungsperioden ergibt. Durch den Einsatz von zwei adaptiven Reglern werden das geschätzte Massenträgheitsmoment und der geschätzte Reibungskoeffizient den tatsächlichen Werten nachgeführt. Voraussetzung für eine zufriedenstellende Funktion ist eine periodische Solldrehzahlvorgabe mit ausreichend Beschleunigungs- und Konstantdrehzahlphasen. Um aber die Drehzahltrajektorie so zu planen, dass keine Begrenzungen (Strom, Spannung) des Servoantriebs wirken, muss das Massenträgheitsmoment vorab bekannt sein.
In [1] wird eine Möglichkeit gezeigt, wie bei beschränktem Bewegungsbereich das Massenträgheitsmoment einer Servoantriebsachse bestimmt werden kann. Dazu werden dem positionsgeregelten Antrieb sehr kleine periodische Bewegungen vorgegeben. Aus dem Produkt der Motordrehzahl und des Integrals des Moments wird über je eine Periode gemittelt das Massenträgheitsmoment berechnet. Allerdings wird vorausgesetzt, dass der positionsgeregelte Antrieb bereits zufriedenstellend parametriert ist.
In [4] wird ein Algorithmus zur Identifikation des Massenträgheitsmoments basierend auf der Beobachtung des Positionsfehlersignals, welches durch einen Drehzahlbeobachter erzeugt wird und Information über die Abweichung des geschätzten vom tatsächlichen Trägheitsmoment enthält, gezeigt. Da der Drehzahlsollwert als Rechteckfunktion vorgegeben wird, bedeutet das für die Mechanik eine starke Belastung, was zu einer entsprechend unangenehmen Geräuschbildung führen kann. Weiters werden, wie auch in [8], keine Richtlinien zur Wahl darüber gegeben, wie die Amplitude und Frequenz des Sollwertsignals geeignet gewählt werden.
Zudem wird bei vielen der oben genannten Arbeiten von einer starren Mechanik ausgegangen, was aber in den meisten industriellen Anwendungen nicht der Fall ist.

In dieser Arbeit wird eine Möglichkeit aufgezeigt, wie mit Hilfe eines selbsteinstellenden Stützreglers der Frequenzgang einer Servoantriebsachse bei Einwirkung eines externen Lastmoments aufgenommen werden kann. Es wird besonderes Augenmerk auf

die Einfachheit der Bedienung (d.h. wenige Einstellparameter) sowie die ausschließliche Verwendung der vorgegebenen Kaskadenstruktur (industrieller Standard) gelegt. Weiters wird gezeigt, wie der bei der Identifikation zu erwartende Bewegungsbereich abgeschätzt und gezielt beeinflusst werden kann.
Die Arbeit ist wie folgt gegliedert: In Abschnitt 2 wird ein einfaches Modell für eine Servoantriebsachse aufgestellt und die Reglerparametrierung für den Fall bekannter Mechanikparameter und eines Entwurfsparameters durchgeführt. In Abschnitt 3 wird die Identifikation für die in Abschnitt 2 notwendigen Parameter aufgezeigt. Besonderes Augenmerk wird dabei auf die Wahl des Anregungssignals gelegt. In Abschnitt 4 wird eine Hilfestellung zur Wahl des Entwurfsparameters angegeben, der eine zentrale Größe bei der Parametrierung aus Abschnitt 2 darstellt. In Abschnitt 5 wird anhand experimenteller Ergebnisse die Funktionstüchtigkeit des vorgestellten Verfahrens gezeigt. Abschnitt 6 enthält eine Zusammenfassung der Arbeit.

2 Modellbildung und Reglerparametrierung

2.1 Modellbildung

In Anlehnung an die in Abbildung 1 dargestellte Vertikalachse wurde für experimentelle Versuche der in Abbildung 2 dargestellte mechanische Aufbau angefertigt. Er entspricht im Prinzip einem Riemenantrieb und ist so konstruiert, dass wesentliche mechanische Parameter einfach verändert werden können. So können die bewegte Masse und die Riemenspannung variiert werden. Dieser Aufbau deckt die wesentlichen bei einer Vertikalachse auftretenden Probleme ab. Zur Modellbildung wird von der Skizze in Abbildung 3 ausgegangen. Das Modell entspricht einem Zweimassenschwinger und wird durch die Gleichungen

$$J_m\ddot{\varphi} = M_M - R \cdot F_{CD}, \tag{1a}$$
$$m\ddot{x} = F_{CD} - F_L, \tag{1b}$$
$$F_{CD} = c\,(R\varphi - x) + d\,(R\dot{\varphi} - \dot{x}) \tag{1c}$$

beschrieben. Dabei ist J_m das Massenträgheitsmoment des Motors, M_M das Motormoment, F_{CD} ist die vom Riemen übertragene Kraft, m die translatorisch bewegte Masse, F_L die Lastkraft aufgrund der Gravitation, c die Drehfedersteifigkeit, d die Dämpfung und R der Radius der Umlenkrollen des Riemenantriebs. Weiters wird angenommen, dass nur die Motorposition φ gemessen werden kann. Durch Berechnung der Übertragungsfunktionen

$$\left.\frac{\mathscr{L}\{\varphi\}}{\mathscr{L}\{M_M\}}\right|_{F_L=0} = \frac{1}{s^2\,(J_m + mR^2)} \cdot \frac{s^2 m + sd + c}{s^2 \frac{J_m m}{J_m + mR^2} + sd + c}, \tag{2}$$

$$\left.\frac{\mathscr{L}\{\varphi\}}{\mathscr{L}\{F_L\}}\right|_{M_M=0} = -\frac{R}{s^2\,(J_m + mR^2)} \cdot \frac{sd + c}{s^2 \frac{J_m m}{J_m + mR^2} + sd + c} \tag{3}$$

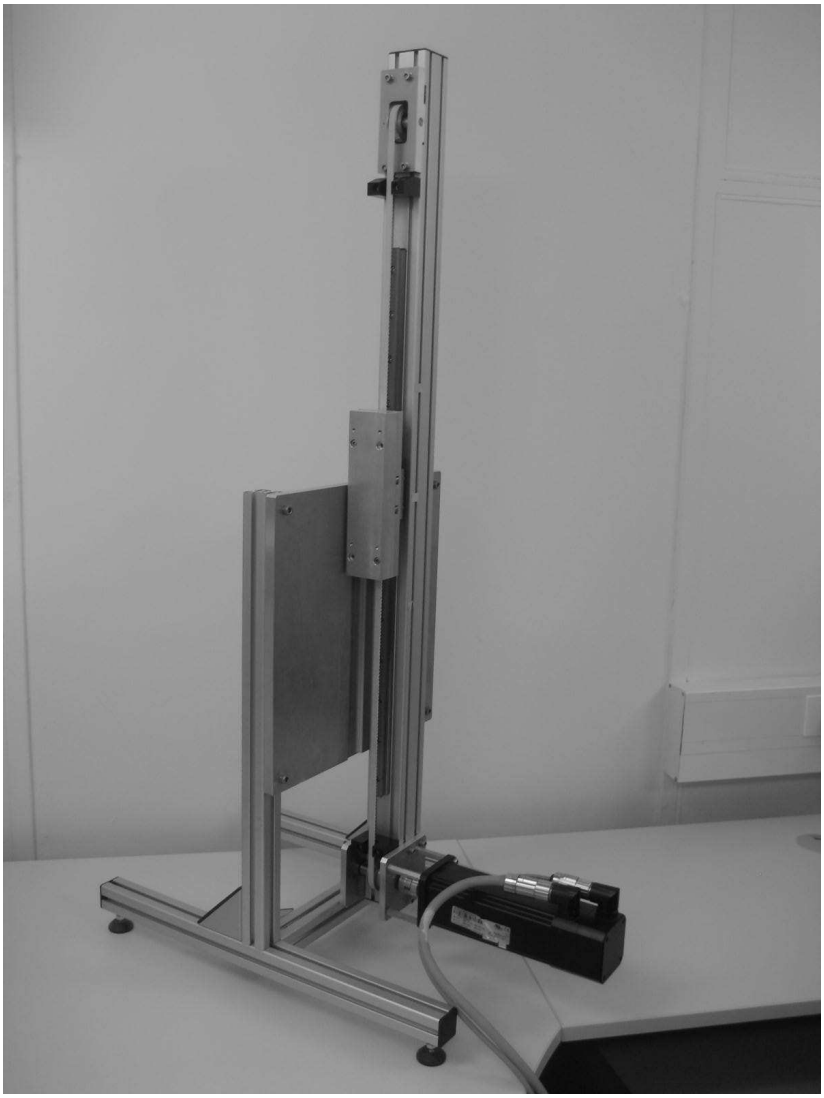


Abbildung 2: Foto des mechanischen Aufbaus

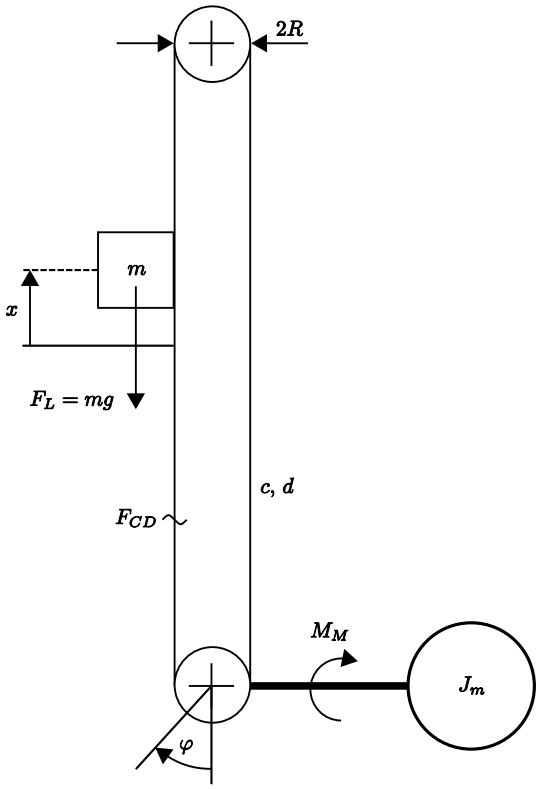


Abbildung 3: Skizze mechanischer Aufbau

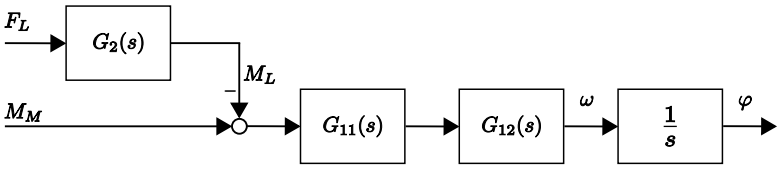


Abbildung 4: Strukturbild

und mit Hilfe der Abkürzungen

$$G_{11}(s) = \frac{1}{s\left(J_m + mR^2\right)}, \tag{4}$$

$$G_{12}(s) = \frac{s^2 m + sd + c}{s^2 \frac{J_m m}{J_m + mR^2} + sd + c}, \tag{5}$$

$$G_2(s) = R \cdot \frac{sd + c}{s^2 m + sd + c} \tag{6}$$

läßt sich das Strukturbild aus Abbildung 4 ableiten. $G_{11}(s)$ entspricht einem starren Antrieb mit dem Gesamtmassenträgheitsmoment $J = J_m + mR^2$, $G_{12}(s)$ beschreibt den flexiblen Anteil. Wird dieser mittels Normpolynomen 2. Ordnung [12] ausgedrückt, ergibt

sich die Übertragungsfunktion:

$$G_{12}(s) = \frac{\frac{s^2}{\omega_{0,Z}^2} + s\frac{2D_Z}{\omega_{0,Z}} + 1}{\frac{s^2}{\omega_{0,N}^2} + s\frac{2D_N}{\omega_{0,N}} + 1}. \tag{7}$$

Dabei entspricht $\omega_{0,Z}$ der Antiresonanzfrequenz und $\omega_{0,N}$ der Resonanzfrequenz des Zweimassenschwingers mit den zugehörigen Dämpfungsgraden D_Z und D_N, wobei für das Verhältnis von Zähler- und Nennerkennkreisfrequenz gilt [12]:

$$\omega_{0,Z} = \sqrt{\frac{J_m}{J_m + mR^2}}\,\omega_{0,N}. \tag{8}$$

Somit ist die Antiresonanzfrequenz niedriger als die Resonanzfrequenz und es gilt

$$\frac{1}{\omega_{0,Z}} > \frac{1}{\omega_{0,N}}. \tag{9}$$

Wird nun eine eingangsseitige Tiefpasserweiterung

$$\mathrm{F}(s) = \frac{1}{1 + sT} \tag{10}$$

durchgeführt (siehe Abbildung 5) und die Filterzeitkonstante so gewählt, dass

$$T \gg \frac{1}{\omega_{0,Z}} \tag{11}$$

erfüllt ist, kann unter Berücksichtigung von (9) der flexible Anteil $G_{12}(s)$ gegenüber $\mathrm{F}(s)$ vernachlässigt werden. $G_2(s)$ beschreibt die Auswirkung der abtriebsseitigen Lastkraft auf das antriebsseitige Lastmoment. Für den Fall einer konstanten Kraft $F_L = mg$ gilt

$$M_{L\infty} = \lim_{t\to\infty} M_L = \lim_{s\to 0} s \cdot G_2(s) \cdot \frac{mg}{s} = Rmg. \tag{12}$$

Somit kann unter der Voraussetzung einer konstanten Lastkraft und des eingangsseitigen Tiefpasses, dessen Grenzfrequenz hinreichend klein ist, das System (1) durch folgende Gleichungen hinreichend genau beschrieben werden:

$$J\ddot{\varphi} = M_M - M_{L\infty}, \tag{13a}$$

$$T\,\dot{M}_M = -M_M + \tilde{M}_M \tag{13b}$$

wobei

$$J = J_m + mR^2, \tag{14a}$$

$$M_{L\infty} = Rmg. \tag{14b}$$

Aus dem zugehörigen Signalflussplan in Abbildung 5 ist erkennbar, dass dieses vereinfachte System einem Doppelintegrator mit konstanter Eingangsstörung entspricht, wobei

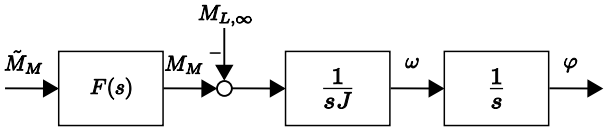


Abbildung 5: Signalflussplan vereinfacht

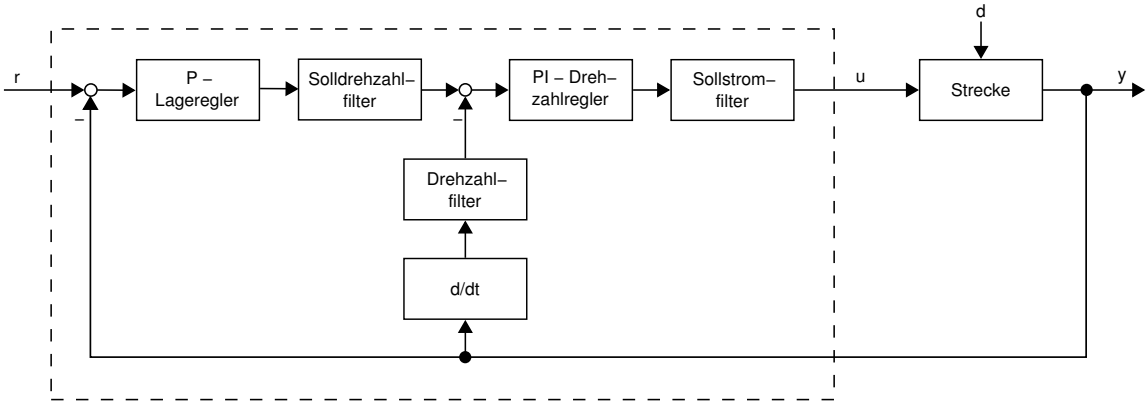


Abbildung 6: Vorgegebene Reglerkaskade

die Stellgröße tiefpassgefiltert wird.
Es sei an dieser Stelle erwähnt, dass bewusst keine drehzahlproportionale und statische Reibung in der Modellbildung berücksichtigt wurde. Der drehzahlproportionale Anteil wirkt sich erst bei höheren Drehzahlen stärker aus. Da es aber Ziel sein wird, während der Selbsteinstellung möglichst kleine Wege und somit kleine Drehzahlen zu erreichen, ist diese Vernachlässigung erlaubt. Der Reglerentwurf wird mit Hilfe linearer Methoden durchgeführt. Darin begründet sich auch die Vernachlässigung der (nichtlinearen) statischen Reibung. Die Erweiterung um einen Tiefpass bringt zusätzlich den Vorteil, dass sämtliche Totzeiten und die Dynamik des Stromregelkreises demgegenüber vernachlässigt werden können. Weiters wird eventuell vorhandenes Rauschen durch den Tiefpass geglättet und wirkt sich bei der Regelung geringer aus. Die getätigte Vereinfachung aufgrund des Tiefpasses ist auch dann zulässig, wenn das System ein Mehrmassenschwinger ist. Maßgeblich für die Wahl der Filterzeitkonstante ist dann nur die niedrigste Kennkreisfrequenz des mechanischen Systems.

2.2 Reglerparametrierung

Für die vereinfachte Strecke (13) aus Abbildung 5 kann nun ein Regler entworfen werden. Es wird davon ausgegangen, dass die Parameter J und T, welche die Strecke beschreiben, bekannt sind. Weiters wird die Reglerstruktur aus Abbildung (6) als gegeben vorausgesetzt. Diese entspricht einer klassichen Kaskadenstruktur [6] mit Lageregler und unterlagertem Drehzahlregler. Beide Regler sind als PI-Regler ausgeführt, können aber auch als P-Regler betrieben werden. Das Sollstromfilter besitzt eine modusabhängige Filtercharakteristik. Unter anderem kann dieses Filter als Tiefpass verwendet werden und so gemäß Gleichung (10) parametriert werden. Da nur die Motorposition gemessen wird, muss die

Drehzahl durch Differentiation, realisiert als Differenzenquotient mit nachgeschaltetem Tiefpassfilter, gewonnen werden. r entspricht der Führungsgröße, u der Stellgröße, d der Störgröße und y der Regelgröße. Im Folgenden wird die Wahl der Parameter für die Kaskade beschrieben.

2.2.1 Drehzahlregler

Da das konstante Lastmoment im Drehzahlregelkreis wirkt, soll auch der Drehzahlregler dessen Wirkung ausgleichen. Aus diesem Grund wird er als PI-Regler entworfen wobei als Einstellvorschrift das bekannte Verfahren des „symmetrischen Optimums“ [12] herangezogen wird. Dies hat folgende Gründe:

- Es sind keine Vorgaben wie Überschwingweite und Anstiegszeit notwendig. Die Spezifikation ergibt sich direkt aus den Streckenparametern.
- Wenige Streckenparameter sind notwendig, um daraus die Reglerparameter einfach berechnen zu können.

Jener Teil der Strecke (13), welcher als Grundlage zur Dimensionierung des Drehzahlreglers verwendet wird, besitzt die Gestalt:

$$G_v(s) = \left.\frac{\mathscr{L}\{\dot{\varphi}\}}{\mathscr{L}\left\{\tilde{M}_M\right\}}\right|_{M_{L\infty}=0} = \frac{1}{sJ(1+sT)}. \tag{15}$$

Daraus können für den Regler der Form

$$R_v(s) = k_v \frac{1+st_n}{st_n} \tag{16}$$

die Parameter k_v und t_n wie folgt berechnet werden:

$$k_v = \frac{J}{2T}, \tag{17}$$

$$t_n = 4T. \tag{18}$$

Der geschlossene Drehzahlregelkreis besitzt nach Kompensation des Zählerterms $1+s4T$ durch ein Solldrehzahlfilter die Form

$$T_v(s) = \frac{\dot{\varphi}_{ist}}{\dot{\varphi}_{soll}} = \frac{1}{1+s4T+s^2 8T^2+s^3 8T^3}. \tag{19}$$

2.2.2 Lageregler

Da der Drehzahlregelkreis ein Teil der Lagereglerstrecke ist, wird dessen Dynamik maßgeblich in den Entwurf des Lagereglers eingehen. Der zweite Teil der Strecke ist ein Integrator, welcher das Verhalten zwischen der Istdrehzahl und der Istposition beschreibt. Zusammengefasst ergibt sich somit als Strecke für den Lageregler die Übertragungsfunktion

$$G_p(s) = \frac{\varphi}{\dot{\varphi}_{soll}} = \frac{1}{s(1+s4T)} \tag{20}$$

wobei in (19) Terme 2. und 3. Ordnung vernachlässigt wurden [12]. Da in diesem Kreis keine weiteren Störungen auftreten können, reicht es aus, den Lageregler als P-Regler auszuführen. Als Einstellvorschrift wird in diesem Fall das „Betragsoptimum“ [12] verwendet. Für die Strecke (20) und den Regler

$$R_p(s) = k_p \tag{21}$$

kann der Parameter k_p wie folgt berechnet werden:

$$k_p = \frac{1}{8T}. \tag{22}$$

Mit Hilfe der Einstellvorschriften „Symmetrisches Optimum“ und „Betragsoptimum“ ist es möglich, die Reglerkaskade zu parametrieren. Die Parametrierung ist festgelegt durch die beiden Parameter Gesamtmassenträgheitsmoment J und Filterzeitkonstante T. Während die Filterzeitkonstante gewählt werden kann (Entwurfsparameter), muss das Gesamtmassenträgheitsmoment geschätzt werden. Auf die Möglichkeit der Schätzung von J und die Herleitung einer Hilfestellung zur Wahl von T wird im nächsten Abschnitt genauer eingegangen.

3 Identifikation

Die in Abschnitt 2.2 getätigte Annahme, dass das Gesamtmassenträgheitsmoment J vorab bekannt ist, wird hier fallen gelassen. Stattdessen wird jetzt versucht, einen Schätzwert $\hat{J}$ für den Parameter J zu ermitteln. Prinzipiell könnte das vereinfachte Modell (13a), (14) als Grundlage dienen, allerdings haben Messungen gezeigt, dass die Schätzung deutlich besser ausfällt, wenn zusätzlich noch die statische Reibung berücksichtigt wird. Aus diesem Grund wird für die Schätzung von folgender nichtlinearer Bewegungsgleichung ausgegangen:

$$J\ddot{\varphi} = M_M - M_{L\infty} - M_S \operatorname{sign}(\dot{\varphi}), \tag{23}$$

wobei M_S dem drehrichtungsabhängigen statischen Reibmoment entspricht.

3.1 Schätzer

Um eine zeitdiskrete Realisierung (Abtastzeit T_a) von (23) zu erhalten, werden die Größen $\varphi_k = \varphi(t_k)$, $\omega_k = \dot{\varphi}(t_k)$ und $\alpha_k = \ddot{\varphi}(t_k)$ mit

$$t_k = kT_a \qquad k = 0, 1, 2, \ldots \tag{24}$$

eingeführt. Mit Hilfe des Rückwärts-Differenzenquotienten können aus der gemessenen Größe φ_k die Größen ω_k und α_k berechnet werden. Bei konstanten Größen J, $M_{L\infty}$ und M_S soll für die zugehörigen Schätzwerte $\hat{J}$, $\hat{M}_{L\infty}$ und $\hat{M}_S$ in jedem Abtastzeitpunkt t_k gelten:

$$\hat{J}\alpha_k = M_{M,k} - \hat{M}_{L\infty} - \hat{M}_S \operatorname{sign}(\omega_k). \tag{25}$$

Es hat sich gezeigt, dass bei der Berechnung von α_k eine zusätzliche Tiefpassfilterung notwendig ist. Allerdings reicht dazu ein Tiefpassfilter mit einer Grenzfrequenz, die deutlich höher als die Grenzfrequenz des eingangsseitigen Tiefpasses ist, so dass dieses gegenüber dem eingangsseitigen Tiefpassfilter vernachlässigt werden kann. Wird für den Zeitbereich $[t_1 \dots t_N]$ Gleichung (25) ausgewertet, ergibt sich das folgende Gleichungssystem:

$$\mathbf{\Psi}\hat{\theta} = \mathbf{z} \tag{26a}$$

$$\begin{bmatrix} \alpha_1 & 1 & \operatorname{sign}(\omega_1) \\ \alpha_2 & 1 & \operatorname{sign}(\omega_2) \\ \vdots & \vdots & \vdots \\ \alpha_N & 1 & \operatorname{sign}(\omega_N) \end{bmatrix} \begin{pmatrix} \hat{J} \\ \hat{M}_{L\infty} \\ \hat{M}_S \end{pmatrix} = \begin{pmatrix} M_{M,1} \\ M_{M,2} \\ \vdots \\ M_{M,N} \end{pmatrix}. \tag{26b}$$

Mit den Größen

$$\hat{\theta} = \begin{pmatrix} \hat{J} & \hat{M}_{L\infty} & \hat{M}_S \end{pmatrix}^T, \tag{27}$$

$$\gamma_k = \begin{pmatrix} \alpha_k & 1 & \operatorname{sign}(\omega_k) \end{pmatrix}^T, \tag{28}$$

$$z_k = M_{M,k}, \tag{29}$$

kann mit Hilfe des Rekursiven-Least-Squares Algorithmus [9]

$$\mathbf{k}_k = \frac{\mathbf{P}_{k-1}\gamma_k}{1 + \gamma_k^T \mathbf{P}_{k-1}\gamma_k}, \tag{30a}$$

$$e_k = z_k - \gamma_k^T \hat{\theta}_{k-1}, \tag{30b}$$

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \mathbf{k}_k e_k, \tag{30c}$$

$$\mathbf{P}_k = \mathbf{P}_{k-1} - \mathbf{k}_k \gamma_k^T \mathbf{P}_{k-1} \tag{30d}$$

und den Startwerten

$$\mathbf{P}_1 = \lim_{\epsilon \to 0} \frac{1}{\epsilon}\mathbf{I} \qquad \hat{\theta}_1 = \mathbf{0}, \tag{31}$$

wobei $\mathbf{I}$ die Einheitsmatrix entsprechender Dimension darstellt, das Gleichungssystem (26a) zu jedem Abtastschritt „bestmöglich“ gelöst werden. Der erste Eintrag des Lösungsvektors entspricht dem gesuchten Schätzwert für J und kann als Eingabeparameter zur Berechnung der Reglerparameter verwendet werden.

3.2 Struktur des Gesamtsystems

Wird die Kombination aus Schätzer und Regler betrachtet, ergibt sich die Struktur aus Abbildung 7. Der Schätzer liefert einen Prozessparameter ($\hat{J}$) aus dem - zusammen mit dem noch festzulegenden Entwurfsparameter T - über das Symmetrische Optimum (SO) und das Betragsoptimum (BO) die entsprechenden Reglerparameter berechnet werden. In [3] wird eine solche Struktur als Self-Tuning-Controller (STC) bezeichnet. Voraussetzung für die Funktionstüchtigkeit des STC ist ein ausreichend anregendes Sollwertsignal $r(t)$. Aus folgenden zwei Gründen ist die Anregung des Systems über den Sollwert, wie in Abbildung 7 dargestellt, nicht zielführend:

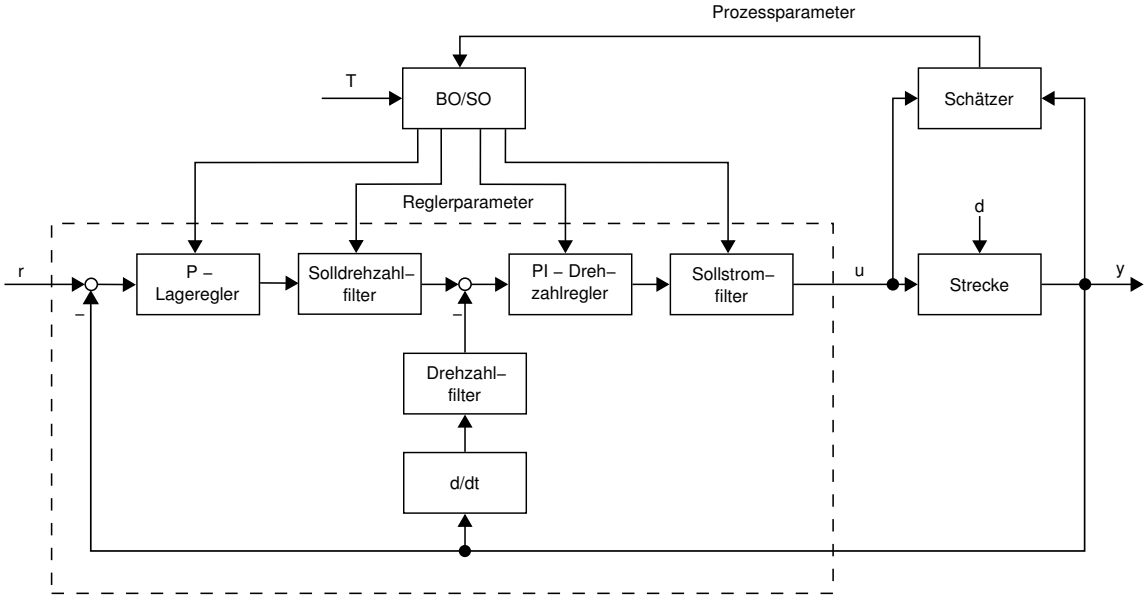


Abbildung 7: Struktur des Gesamtsystems

- Der zeitliche Verlauf ist bestimmt durch die maximal erlaubte Drehzahl und Beschleunigung. Die Positionstrajektorie müsste so geplant werden, dass die Beschleunigung einen vorgegebenen Wert nicht überschreitet um so sicher zu stellen, dass der Maximalstrom des Motors (Servoverstärkers) nicht überschritten wird. Da das Gesamtmassenträgheitsmoment J unbekannt ist, kann die maximale Beschleunigung *nicht* ermittelt und so die Trajektorie nicht geplant werden.

- Da die Drehzahlreglerverstärkung aus der ersten Komponente des Parametervektors $\hat{\theta}$ berechnet wird und dieser zu Beginn der Identifikation null ist, würde der Schätzalgorithmus nie starten.

Aus den genannten Gründen liegt es nahe, den Ort des Einwirkens des Anregungssignals zu verschieben. In Abbildung 8 ist die geänderte Struktur dargestellt. Diese bringt den Vorteil mit sich, dass anstatt einer Beschleunigung direkt ein Strom (dieser ist proportional dem Drehmoment) vorgegeben werden kann. Weiters können zu Beginn der Identifikation die Reglerparameter auf null stehen und die Anregung ist trotzdem vorhanden.
Nachdem nun die Struktur des Gesamtsystems festgelegt ist, müssen die Eigenschaften des Anregungssignals genauer untersucht werden.

3.3 Anregungssignal

Das Anregungssignal, welches einem Moment entspricht, muss so gewählt werden, dass die gesuchten Parameter richtig geschätzt werden. Es müssen ausreichend Beschleunigungs- und Bremsphasen zur Schätzung des Gesamtmassenträgheitsmoments sowie Richtungswechsel zur Schätzung der statischen Reibung vorhanden sein. Aus diesem Grund fällt die Wahl auf ein harmonisches Signal welches durch Amplitude und Frequenz bestimmt

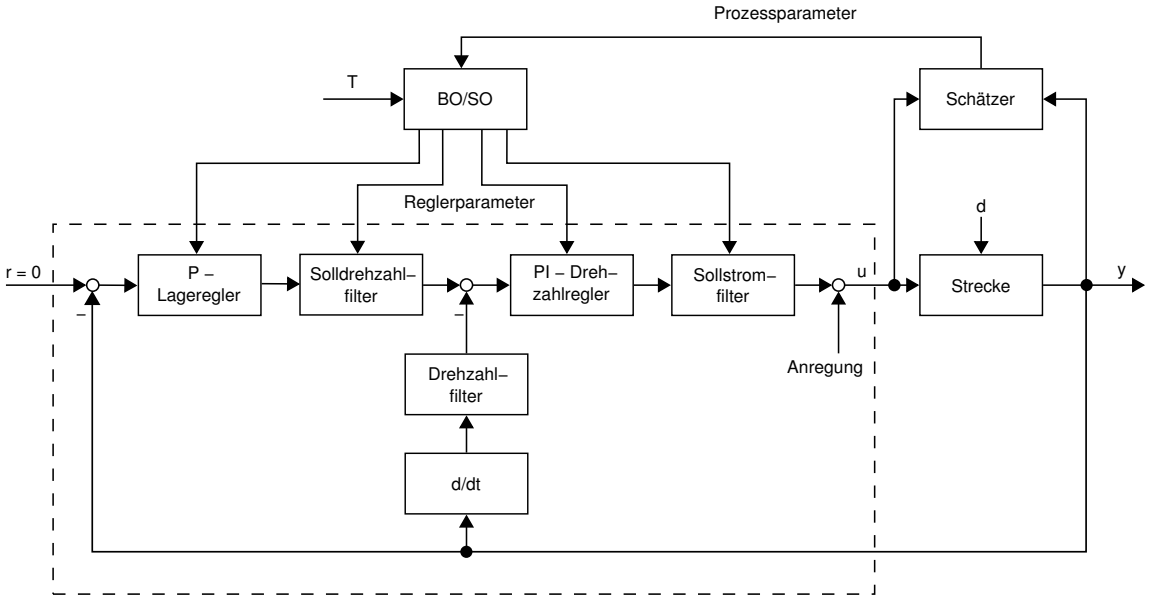


Abbildung 8: Struktur des Gesamtsystems bei verschobenem Eingriffsort des Anregungssignals

ist. Im Weiteren wird auf die geeignete Wahl dieser beiden Grössen sowie auf die Wahl der Anregungsdauer näher eingegangen.

3.3.1 Wahl der Amplitude

Die Anregungsamplitude orientiert sich am Nennmoment des Servomotors. Es wird davon ausgegangen, dass der Servoverstärker so groß gewählt ist, dass dieser den für das Nennmoment notwendigen Strom dauerhaft liefern kann.

3.3.2 Wahl der Frequenz

Da die Anregung an der selben Stelle angreift wie ein Störmoment, ist es sinnvoll, das Störverhalten des Systems zu untersuchen. Wird die Übertragungsfunktion $S(s)$ vom Störmoment d zur tatsächlichen Beschleunigung α gebildet, ergibt sich

$$S(s) = \frac{\alpha(s)}{d(s)} = \frac{1}{J} \frac{64s^3T^3\left(1+sT\right)}{64s^4T^4 + 64s^3T^3 + 32s^2T^2 + 8sT + 1}. \tag{32}$$

Die Störübertragungsfunktion ist abhängig von den Parametern J und T. Wird $J = 1$ und $T = 1$ gesetzt, besitzt der Frequenzgang die Gestalt aus Abbildung 9. Maximale Beschleunigung tritt dort auf, wo $|S\left(j\omega\right)|$ ein Maximum besitzt. Wird also jene Frequenz ω_{max} ermittelt, bei der das Maximum auftritt, entspricht diese der gesuchten Anregungsfrequenz. Sie ergibt sich zu

$$\omega_{max} = \frac{1}{2T}. \tag{33}$$

Somit ist die Anregungsfrequenz durch die Filterzeitkonstante T bestimmt und unabhängig vom Gesamtmassenträgheitsmoment J.

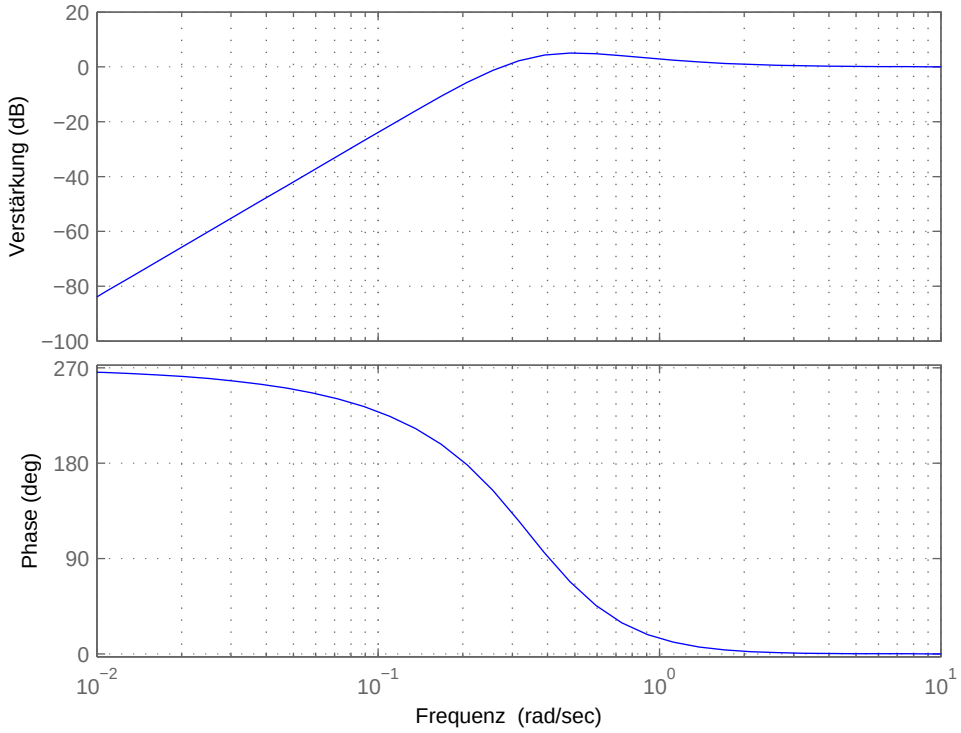


Abbildung 9: Bode-Diagramm der Störübertragungsfunktion $S(s)$

3.3.3 Wahl der Anregungsdauer

Um eine Abschätzung der minimalen Anregungsdauer zu erhalten, ist es sinnvoll, von der optimalen Lösung von Gleichung (26a) auszugehen [7]:

$$\hat{\theta} = \left(\mathbf{\Psi}^T\mathbf{\Psi}\right)^{-1}\mathbf{\Psi}^T\mathbf{z}. \tag{34}$$

Damit der Parametervektor berechnet werden kann, muss $\mathbf{\Psi}^T\mathbf{\Psi}$ invertierbar sein. Mit Hilfe von (26b) berechnet sich diese Matrix zu

$$\mathbf{\Psi}^T\mathbf{\Psi} = \begin{bmatrix} \sum_{k=1}^{N}\alpha_k^2 & \sum_{k=1}^{N}\alpha_k & \sum_{k=1}^{N}\alpha_k\text{sign}\left(\omega_k\right) \\ \sum_{k=1}^{N}\alpha_k & N & \sum_{k=1}^{N}\text{sign}\left(\omega_k\right) \\ \sum_{k=1}^{N}\alpha_k\text{sign}\left(\omega_k\right) & \sum_{k=1}^{N}\text{sign}\left(\omega_k\right) & \sum_{k=1}^{N}\text{sign}^2\left(\omega_k\right) \end{bmatrix}. \tag{35}$$

Wird unter der Voraussetzung eines harmonischen Anregungssignals und eines eingeschwungenen Systemzustands der Wert für N so gewählt, dass ein ganzzahliges Vielfaches der Anregungsperiode enthalten ist, dann gilt:

$$\sum_{k=1}^{N}\text{sign}\left(\omega_k\right) = 0, \tag{36a}$$

$$\sum_{k=1}^{N}\alpha_k = 0, \tag{36b}$$

$$\sum_{k=1}^{N}\alpha_k\text{sign}\left(\omega_k\right) = 0. \tag{36c}$$

In diesem Fall ergibt sich für (35) die Diagonalmatrix

$$\mathbf{\Psi}^T\mathbf{\Psi} = \begin{bmatrix} \sum_{k=1}^{N}\alpha_k^2 & 0 & 0 \\ 0 & N & 0 \\ 0 & 0 & \sum_{k=1}^{N}\text{sign}^2\left(\omega_k\right) \end{bmatrix}, \tag{37}$$

welche unter der Voraussetzung einer ausreichenden Systemanregung

$$\sum_{k=1}^{N}\alpha_k^2 > 0 \qquad N > 0 \qquad \sum_{k=1}^{N}\text{sign}^2\left(\omega_k\right) > 0 \tag{38}$$

leicht zu invertieren ist.
Aufgrund von Signalrauschen und der Tatsache, dass die Identifikation von einem stationären Systemzustand aus erfolgt, wird die Wahl für die minimale Anregungsdauer auf ein Vielfaches einer Anregungsperiode fallen.

Parameter	Einheit	Wert
Schaltfrequenz	kHz	20
Nennstrom	A	1.89
Nenndrehzahl	min^{-1}	3000
Widerstand	Ω	11.9
Induktivität	H	0.0365
Momentkonstante	Nm/A	1.45
Massenträgheitsmoment	kgm^2	0.00016

Tabelle 1: Antriebsparameter

4 Wahl der Filterzeitkonstante

Die Filterzeitkonstante T ist zentrale Größe bei der Parametrierung der Reglerkaskade und der Bestimmung der Anregungsfrequenz. Da die geeignete Wahl dieses Entwurfsparameters entscheidend für das zufriedenstellende Funktionieren des vorgestellten Konzepts ist, wird im Folgenden dazu eine Hilfestellung erarbeitet.

4.1 Untergrenze

Die Filterzeitkonstante muss so gewählt werden, dass (11) erfüllt ist. Für eine Vielzahl an Servoantriebsachsen wird diese Bedingung eingehalten, weil die Mechanik auf maximale Dynamik konstruiert ist und somit die Resonanz- und Antiresonanzfrequenz(en) hinreichend groß sind. Weiters muss die Filterzeitkonstante so groß gewählt werden, dass systembedingte Totzeiten (Abtastzeit, Positionsermittlung, Drehzahlermittlung, ...) und die Stromregelkreisdynamik im Vergleich dazu vernachlässigt werden können. Im Frequenzbereich heißt dies, dass die Durchtrittsfrequenz $\frac{1}{2T}$ des offenen Drehzahlregelkreises so klein gewählt werden muss, dass die zusätzliche Phasendrehung durch vorhandene Totzeiten die maximale Phasenreserve nicht erheblich verringert. Dann ist sichergestellt, dass der Frequenzgang des offenen Kreises den Verlauf aus Abbildung 10 aufweist. Um die zusätzliche Phasendrehung ermitteln zu können, wird für eine Kombination aus Servoverstärker und Motor mit den Daten aus Tabelle 1 der Frequenzgang bei abgekoppelter Mechanik aufgenommen. In Abbildung 11 ist in blau der gemessene Frequenzgang zwischen Sollmoment M_M und Istdrehzahl $\frac{\omega}{2\pi}$ dargestellt. Rot gibt die Approximation durch das einfache Modell aus (13a) wieder. Es ist erkennbar, dass die Betragskennlinie sehr gut mit der Messung übereinstimmt, die Phasenkennlinie hingegen für höhere Frequenzen stark abweicht. Wird die Phasenkennlinie durch

$$\varphi = -\frac{\pi}{2} - \omega T_{tot} \qquad \text{mit} \qquad T_{tot} = 0.00035s \tag{39}$$

approximiert, ergibt sich der Verlauf in schwarz, welcher das tatsächliche Verhalten deutlich besser wiedergibt. Zusätzliche Messungen haben gezeigt, dass der Wert dieser Totzeit im Wesentlichen von der Schaltfrequenz des Servoverstärkers abhängt und unabhängig von der angeschlossenen Mechanik ist. Somit ist es bei bekannter Schaltfrequenz des Servoverstärkers möglich, jene Frequenz anzugeben, bis zu der das Modell mit der Messung

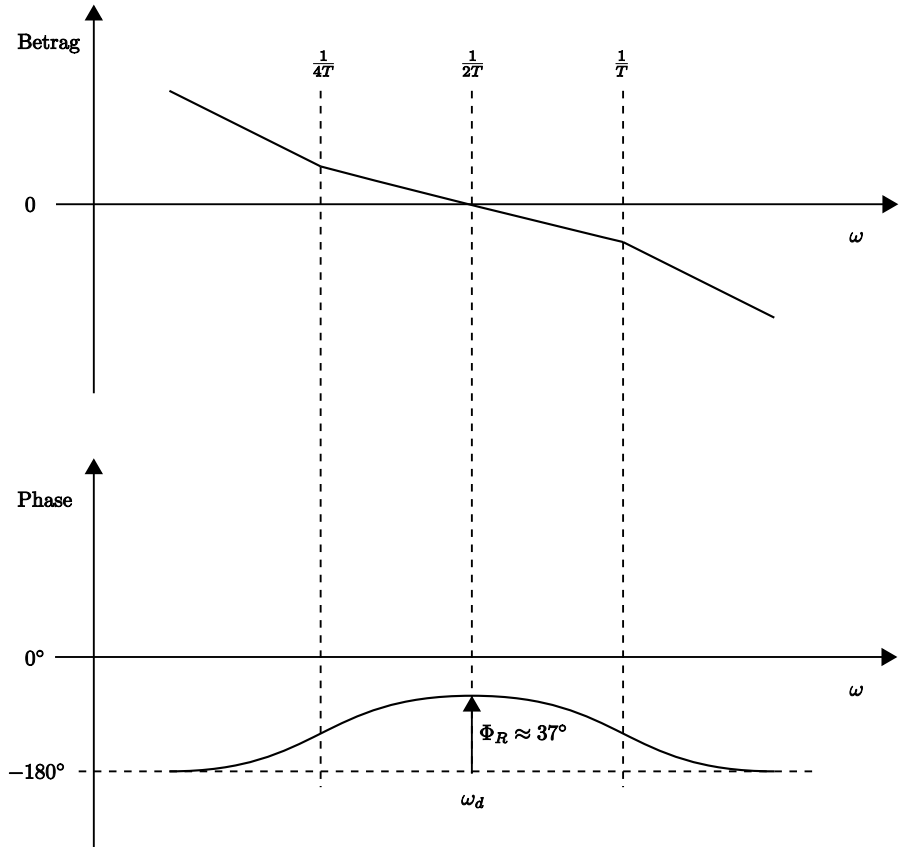


Abbildung 10: Frequenzgang des offenen Kreises, eingestellt mit dem Verfahren des symmetrischen Optimums

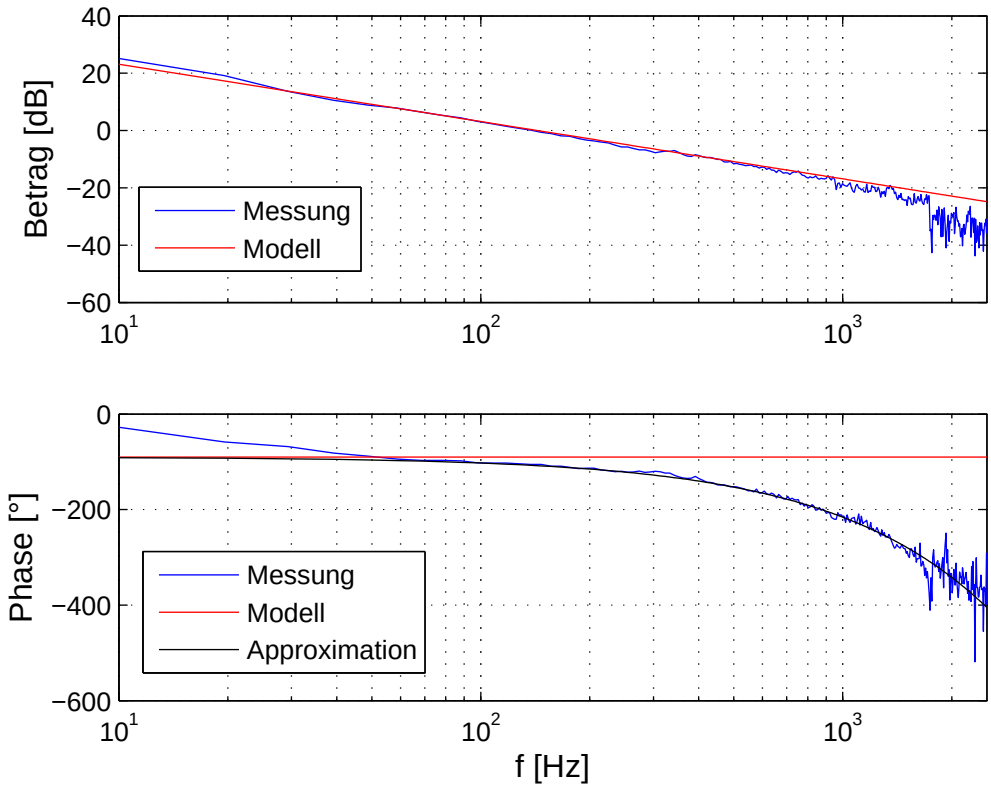


Abbildung 11: Frequenzgangsmessung Drehzahlregelkreis bei abgekoppelter Mechanik

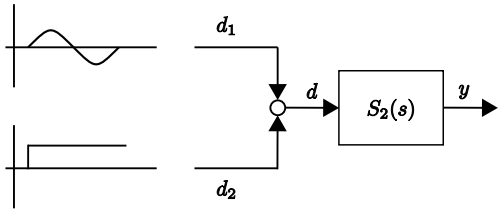


Abbildung 12: Strukturbild Störübertragungsfunktion

hinreichend genau übereinstimmt. Wird von einer 10% Abweichung vom theoretischen Wert $-\frac{\pi}{2}$ ausgegangen, muss folgende Gleichung erfüllt sein:

$$\varphi_{10} = -\frac{\pi}{2} - \omega_{10} T_{tot} = -1.1\frac{\pi}{2}. \tag{40}$$

Daraus folgt für die Obergrenze der Gültigkeit des einfachen Modells:

$$\omega_{10} = \frac{\pi}{20 T_{tot}}. \tag{41}$$

Da für die Durchtrittsfrequenz

$$\frac{1}{2T} < \omega_{10} \tag{42}$$

gelten muss, ergibt sich als Untergrenze für die Filterzeitkonstante:

$$T > \frac{1}{2\omega_{10}} = \frac{10 T_{tot}}{\pi}. \tag{43}$$

4.2 Obergrenze

Eine zu groß gewählte Filterzeitkonstante hat zur Folge, dass der Regler sehr träge auf das Lastmoment reagiert. Dies führt unter Umständen zu einer Ausgleichsbewegung, welche über den erlaubten Bewegungsbereich hinausgeht. Um eine Obergrenze für die Wahl der Filterzeitkonstante angeben zu können, ist es sinnvoll, das Störverhalten zu untersuchen. Wird die Übertragungsfunktion $S_2(s)$ vom Störmoment d zur Istposition y gebildet, so ergibt sich

$$S_2(s) = \frac{y(s)}{d(s)} = \frac{1}{J} \frac{64 s T^3 (1 + sT)}{64 s^4 T^4 + 64 s^3 T^3 + 32 s^2 T^2 + 8 sT + 1}. \tag{44}$$

Da das Anregungssignal d_1 an der selben Stelle angreift wie das Lastmoment d_2, lässt sich das Strukturbild aus Abbildung 12 ableiten. Die Gesamtstörung $d = d_1 + d_2$ wird von der Störübertragungsfunktion gefiltert und ergibt ein Antwortsignal y, welches in seiner Größe abgeschätzt werden kann. Mit Hilfe der Impulsantwort der Störübertragungsfunktion

$$s_2(t) = \mathscr{L}^{-1} \{S_2(s)\} \tag{45}$$

kann über die Faltungssumme

$$y(t) = \int_{\tau=0}^{\infty} s_2(t-\tau)\, d(\tau)\, d\tau \tag{46}$$

folgende Ungleichung abgeleitet werden [5]:

$$\|y\|_\infty \leq \|s_2\|_1 \cdot \|d_1 + d_2\|_\infty \tag{47}$$

Die 1-Norm der Impulsantwort $\|s_2\|_1$ ist abhängig von der Filterzeitkonstante T und dem Gesamtmassenträgheitsmoment J

$$\|s_2\|_1 = \|s_2(T, J)\|_1 \tag{48}$$

und kann auf numerischem Wege hinreichend genau berechnet werden. Für den Fall bekannter Motorträgheit J_m (für viele Servomotoren gegeben), kann folgende Abschätzung nach oben durchgeführt werden:

$$\|s_2(T, J)\|_1 \leq \|s_2(T, J_m)\|_1 . \tag{49}$$

Weiters lässt sich aufgrund der Tatsache

$$\|d_1 + d_2\|_\infty \leq \|d_1\|_\infty + \|d_2\|_\infty \tag{50}$$

der Einfluss der zwei Störterme getrennt abschätzen. $\|d_1\|_\infty$ entspricht der Amplitude des Anregungssignals (Moment) und ist bekannt (Abschnitt 3.3.1), $\|d_2\|_\infty$ entspricht dem Lastmoment und kann bei sinnvoll durchgeführter Antriebsauslegung mit $50 - 75\%$ des Nennmoments des Motors abgeschätzt werden.
Somit kann eine Obergrenze für die Filterzeitkonstante angegeben werden. Diese ist abhängig vom Nennmoment und Massenträgheitsmoment des Motors sowie dem erlaubten Bewegungsbereich.

4.3 Hilfestellung zur Wahl der Filterzeitkonstante

Zusammenfassend ergibt sich folgende Hilfestellung für die Auswahl der Filterzeitkonstante:

1. Bestimmung der Untergrenze mittels Gleichung (43).
2. Bestimmung von $\|d_1\|_\infty$ gemäß Abschnitt 3.3.1.
3. Bestimmung von $\|d_2\|_\infty$ aus dem Nennmoment des Motors.
4. Wahl einer (großen) Filterzeitkonstante T.
5. Berechnung der 1-Norm von $s_2(t)$ mit J_m und T.
6. Berechnung von $\|y\|_\infty = \|s_2\|_1 \cdot (\|d_1\|_\infty + \|d_2\|_\infty)$

7. Wenn $\|y\|_\infty$ größer als der erlaubte Bewegungsbereich ist, muss eine kleinere Filterzeitkonstante (größer als die Untergrenze) gewählt und mit 5 fortgesetzt werden. Ansonsten ist ein geeigneter Bereich (Untergrenze, Obergrenze) gegeben, aus dem ein Wert für T gewählt werden kann.

Es sei an dieser Stelle erwähnt, dass der abgeschätzte Bewegungsbereich mit dem erlaubten umso besser übereinstimmt, je genauer das tatsächliche Gesamtmassenträgheitsmoment bekannt ist. D.h. in Schritt 5 kann statt J_m der genauere Wert zur Abschätzung verwendet werden.
Wenn das Motormassenträgheitsmoment nicht bekannt ist, ist eine Abschätzung nicht möglich. Aber auch in diesem Fall kann das vorgestellte Konzept verwendet werden, es muss allerdings ausreichend Bewegungsbereich vorhanden sein.
Aufgrund von Reibung wird die tatsächliche Bewegung geringer ausfallen, als in der Abschätzung angenommen. Trotzdem macht es vom praktischen Gesichtspunkt aus Sinn, eine Bewegungsbereichsüberwachung durchzuführen. Wird der erlaubte Bewegungsbereich überschritten, wird eine von der Reglerparametrierung unabhängige Stillsetzung des Antriebs eingeleitet (so genannter Kurzschlusshalt).

5 Experimentelle Ergebnisse

In diesem Abschnitt wird die Funktionstüchtigkeit des vorgeschlagenen Konzepts gezeigt. Zu diesem Zweck wird von der Kombination Servoverstärker und Motor (Frequenzgang aus Abbildung 11) mit den Daten aus Tabelle 1 ausgegangen. An diesen Motor wird nun über eine Kupplung die Verbindung zur Mechanik hergestellt, so dass sich der Aufbau aus Abbildung 2 ergibt. Die vorgegebene Reglerstruktur ist so aufgebaut, dass der Drehzahlregler mit einer Abtastzeit von $200\mu s$ und der Lageregler mit $400\mu s$ arbeitet. Der zusätzlich notwendige Schätzalgorithmus und die Parameteradaption werden mit einer Abtastzeit $T_a = 400\mu s$ abgearbeitet.

5.1 Auswahl der Filterzeitkonstante

Die Grenzen ergeben sich entsprechend der Hilfestellung aus Abschnitt 4.3. Für einen Servoverstärker mit einer Schaltfrequenz von 20 kHz und einer Totzeit von 0.00035s ist die Untergrenze (43) durch

$$T > \frac{10T_{tot}}{\pi} = 0.0011\text{s} \tag{51}$$

gegeben. Die Filterzeit wird gleich der Obergrenze gewählt, so dass der erlaubte Bewegungsbereich 1/3 Motorumdrehung (in eine Richtung) entspricht. $\|d_1\|_\infty$ ergibt sich aus dem Produkt aus Nennstrom und Momentkonstante aus Tabelle 1 zu $\|d_1\|_\infty = 0.25 \cdot 1.89\text{A} \cdot 1.45\text{Nm/A} = 0.685\text{Nm}$ (als Motorstrom wird im vorliegenden Beispiel 25% des Nennstromes verwendet). Da die translatorisch bewegte Masse mit ca. $m = 1\text{kg}$ angegeben ist, kann daraus ein Schätzwert für das Lastmoment berechnet werden. Mit einem Radius der Umlenkscheiben von $R = 0.025\text{m}$ ergibt sich dafür ein Wert von ca. 0.25Nm

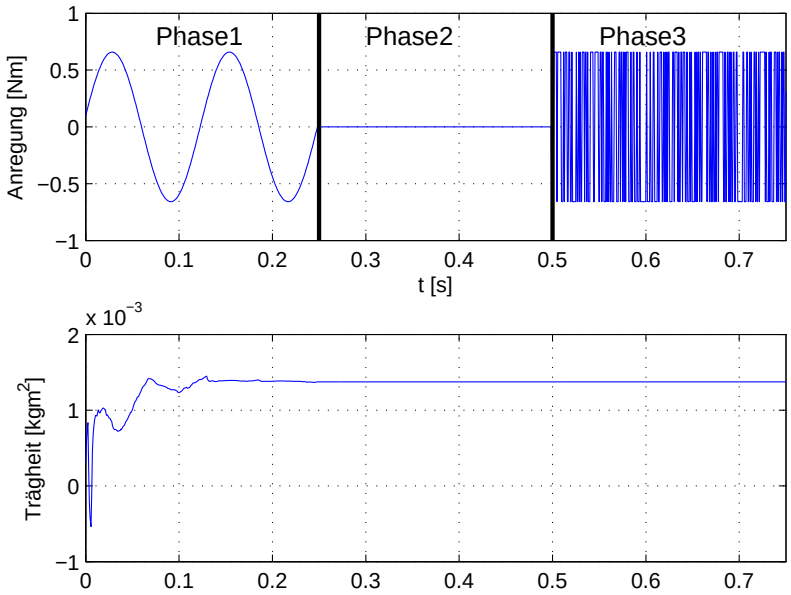


Abbildung 13: Anregungssignal und geschätztes Massenträgheitsmoment beim Identifikationsvorgang

und somit $\|d_2\|_\infty = 0.25$Nm. Weiters kann ein besserer Schätzwert für das Gesamtmassenträgheitsmoment berechnet werden: $J = J_m + mR^2 = 0.00016 + 1 \cdot 0.025^2 = 0.000785\text{kg}m^2$. Mit einem Wert von $T = 0.01$s ergibt sich für $\|s_2\|_1 = 2.1647$rad/Nm. Daraus kann $\|y\|_\infty = \|s_2\|_1 \cdot (\|d_1\|_\infty + \|d_2\|_\infty) = 2.1647 \cdot (0.685 + 0.25) = 2.02$rad ermittelt werden, was näherungsweise dem geforderten Wert von 1/3 Umdrehung entspricht.

5.2 Identifikationsvorgang

Der Identifikationsvorgang erstreckt sich über drei Phasen (siehe Abbildung 13, oben). In Phase 1 wird das System mit einem Sinus angeregt und die mechanischen Parameter werden geschätzt. Zeitgleich werden aufgrund der geschätzten Parameter die Reglerparameter nachgeführt. In Abbildung 13 ist erkennbar, dass der Schätzwert für das Massenträgheitsmoment nach genau einer Anregungsperiode das erste Mal einen stationären Wert erreicht. Nach einer Anregungsdauer von zwei Signalperioden wird in Phase 2 gewartet, bis die Ausregelzeit vorbei ist und das System sich wieder in Ruhe befindet. Aus Abbildung 14 ist erkennbar, dass sich sowohl Position als auch Drehzahl stationär einstellen und die angegebene vorausberechnete Grenze nicht überschritten wird. Untersuchungen mit der Übertragungsfunktion (44) haben ergeben, dass die Ausregelzeit hinreichend gut mit $t_{aus} \approx 25T = 25 \cdot 0.01 = 0.25$s festgelegt ist. Schließlich wird in Phase 3 die eigentliche Aufnahme des Frequenzgangs mit Hilfe eines Pseudo-Rausch-Binär-Signals (PRBS) durchgeführt.

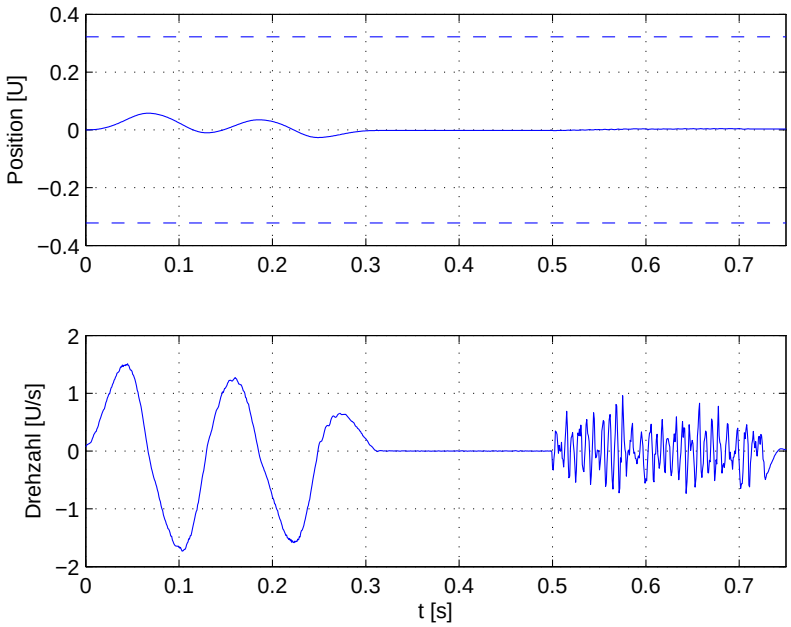


Abbildung 14: Position und Drehzahl beim Identifikationsvorgang

5.3 Frequenzgang der Mechanik

In vorliegendem Fall ergibt sich unter Anwendung obiger Prozeduren der Frequenzgang aus Abbildung 15. In blau dargestellt ist der gemessene Frequenzgang, in rot als Vergleich dazu der Frequenzgang des zugrundeliegenden einfachen Modells des Motors ohne Mechanik sowie in schwarz die Approximation der Phasenkennlinie durch ein Totzeitglied. Es ist erkennbar, dass die Betragskennlinie für tiefe Frequenzen aufgrund des Massenträgheitsmoments der angeschlossenen Mechanik deutlich abgesenkt wird. Weiters ist auch die typische Zweimassenschwinger-Charakteristik (Resonanz, Antiresonanz) erkennbar. Die Werte

$$\omega_{0,Z} = 2\pi \cdot 59\, rad/s \qquad \omega_{0,N} = 2\pi \cdot 137\, rad/s \tag{52}$$

können direkt aus dem Diagramm abgelesen werden. Bei einer Frequenz von ca. 1000Hz sind weitere Resonanzstellen erkennbar, welche aber wesentlich schwächer ausgeprägt sind. Die Phasenkennlinie weist bei tiefen Frequenzen eine deutlich geringere Phasendrehung auf, was auf die höhere Reibung zurückzuführen ist. Bei höheren Frequenzen passt der gemessene mit dem approximierten Phasenverlauf sehr gut überein. Daraus ist ableitbar, dass die Totzeit unabhängig von der Mechanik ist, was die ursprüngliche Annahme bestätigt (vergleiche Abbildung 11 mit Abbildung 15).
Anzumerken sei an dieser Stelle, dass die Bedingung $T = 0.01 \gg \frac{1}{\omega_{0,Z}} = 0.0027$ für diese Mechanik zwar erfüllt wird, dies aber nicht immer der Fall ist. Speziell bei Direktantrieben, bei denen aus Kosten- und Genauigkeitsgründen das Getriebe eingespart wird, verlagern sich die Resonanz- und Antiresonanfrequenzen hin zu niedrigeren Werten.

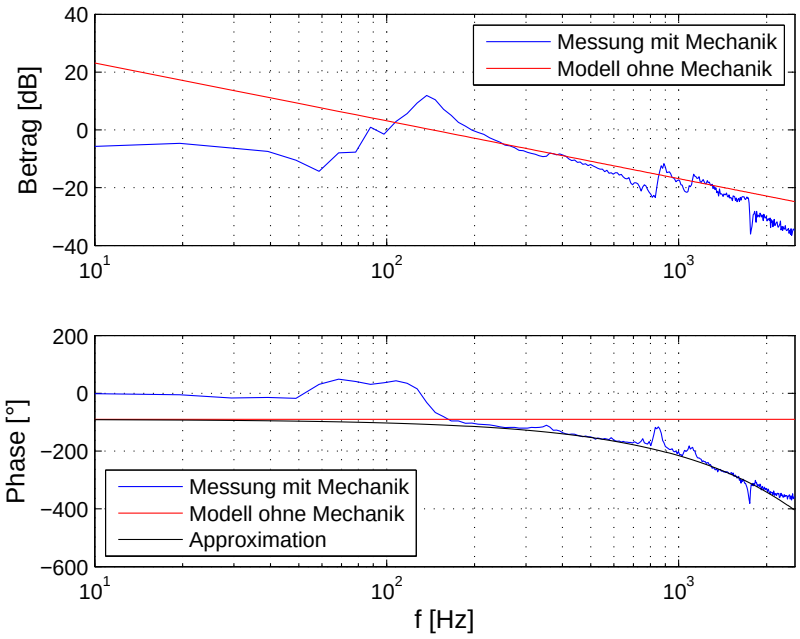


Abbildung 15: Frequenzgangsmessung Drehzahlregelkreis

6 Zusammenfassung

In dieser Arbeit wurde ein Konzept zur Identifikation des Frequenzgangs eines mechatronischen Systems vorgestellt, auf das eine konstante Störung wirkt. Dazu wurde in einem ersten Schritt ein einfaches Modell abgeleitet, dessen Parameter unbekannt sind. Für dieses Modell wurden unter der Annahme bekannter Streckenparameter die Reglerparameter der vorgegebenen Reglerstruktur berechnet. Die tatsächlichen Parameter der Strecke wurden durch einen RLS-Schätzer ermittelt. Im Hinblick auf industrielle Anwendbarkeit konnte erreicht werden, dass nur *ein* Entwurfsparameter notwendig ist. Für die Wahl dieses Parameters wurde eine Hilfestellung erarbeitet. An einem realen mechanischen Aufbau wurde das Konzept erfolgreich erprobt.
Ein Nachteil dieses Konzepts liegt derzeit noch darin, dass es bei sehr tief liegenden Resonanz- und Antiresonanzfrequenzen zu Stabilitätsproblemen kommen kann.

Die Autoren bedanken sich bei der Firma Bernecker + Rainer Industrie-Elektronik Ges.m.b.H. (www.br-automation.com) für die Bereitstellung der Hardwarekomponenten.

Literatur

[1] F. Andoh. Inertia identification method based on the product of the integral of torque reference input and motor speed. In *Proc. IEEE Int. Conf. Control Applications CCA 2008*, pages 1151–1158, 2008.

[2] K.J. Åström and T. Hägglund. *PID controllers: Theory, Design and Tuning*. Setting the standard for automation. International Society for Measurement and Control, 1995.

[3] V. Bobál. *Digital self-tuning controllers: algorithms, implementation and applications*. Advanced textbooks in control and signal processing. Springer, 2005.

[4] J.-W. Choi, S.-C. Lee, and H.-G. Kim. Inertia identification algorithm for high-performance speed control of electric motors. *Electric Power Applications, IEE Proceedings* -, 153(3):379 – 386, may 2006.

[5] J.C. Doyle, B.A. Francis, and A.R. Tannenbaum. *Feedback Control Theory*. Dover Books on Engineering. Dover Publications, 2009.

[6] O. Foellinger and F. Doerrscheidt. *Regelungstechnik: Einfuehrung in die Methoden und ihre Anwendung*. Huethig, 1994.

[7] R. Isermann. *Identifikation Dynamischer Systeme 1: Grundlegende Methoden*. Springer Verlag, 1992.

[8] Tae-Suk Kwon, Seung-Ki Sul, H. Nakamura, and K. Tsuruta. Identification of the mechanical parameters for servo drive. In *Industry Applications Conference, 2006. 41st IAS Annual Meeting. Conference Record of the 2006 IEEE*, volume 2, pages 905 –910, oct. 2006.

[9] L. Ljung. *System identification: theory for the user*. Prentice-Hall information and system sciences series. Prentice Hall PTR, 1999.

[10] M. Pacas and S. Villwock. Development of an expert system for identification, commissioning and monitoring of drives. In *Power Electronics and Motion Control Conference, 2008. EPE-PEMC 2008. 13th*, pages 2248 –2253, sept. 2008.

[11] M. Pacas, S. Villwock, and T. Eutebach. Identification of the mechanical system of a drive in the frequency domain. In *Industrial Electronics Society, 2004. IECON 2004. 30th Annual Conference of IEEE*, volume 2, pages 1166 – 1171 Vol. 2, nov. 2004.

[12] D. Schroeder. *Elektrische Antriebe - Regelung von Antriebssystemen*. Elektrische Antriebe. Springer, 2001.

[13] M. Schulze. *Elektrische Servoantriebe: Baugruppen mechatronischer Systeme*. HANSER VERLAG, 2008.

[14] S. Villwock, A. Baumuller, M. Pacas, F.-R. Gotz, Biao Liu, and V. Barinberg. Influence of the power density spectrum of the excitation signal on the identification of drives. In *Industrial Electronics, 2008. IECON 2008. 34th Annual Conference of IEEE*, pages 1252 –1257, nov. 2008.

[15] S. Villwock, M. Pacas, and T. Eutebach. Application of the welch-method for the automatic parameter identification of electrical drives. In *Industrial Electronics Society, 2005. IECON 2005. 31st Annual Conference of IEEE*, page 6 pp., nov. 2005.

[16] J. Weidauer. *Elektrische Antriebstechnik: Grundlagen, Auslegung, Anwendungen, Lösungen*. Publicis Corp. Publ., 2008.

[17] H. Wertz and F. Schutte. Self-tuning speed control for servo drives with imperfect mechanical load. In *Industry Applications Conference, 2000. Conference Record of the 2000 IEEE*, volume 3, pages 1497 –1504 vol.3, 2000.

[18] Sheng-Ming Yang and Yu-Jye Deng. Observer-based inertial identification for autotuning servo motor drives. In *Industry Applications Conference, 2005. Fourtieth IAS Annual Meeting. Conference Record of the 2005*, volume 2, pages 968 – 972 Vol. 2, oct. 2005.

Modeling and Control Issues in Systems Integration of Production Operations and Design of Continuous Processes

Dr. S. Vasileiadou, Prof. N. Karcanias
TEI Piraeus, CITY University London
svasil@teipir.gr, N.Karcanias@city.ac.uk

Abstract

The problem of System Integration in the context of an Industrial Enterprise is a complex problem with fundamental dimensions those of:

- Overall Process Operations,
- Overall System Design/Redesign, and
- Information and Data Organization and Software.

Each of the above has a multidisciplinary nature and it is frequently considered by the respective groups as representing the entirety of the problem.

The aim of this paper is to consider the general problem of integration from the Overall Process Operations aspects and focus on issues of Methodology, such as investigation of the "top-down" and "bottom-up" approaches, the Multi Modeling requirements, the Architecture of the overall Control Decision Making problem and address the problem of System Organization from its different aspects. We will show that Systems and Control concepts play a central role in the development of an overall integration methodology and associated techniques. This effort reveals the central role of Modeling and the significance of new important families of Systems, such as the Multilevel Hybrid Systems (Hierarchy of Operations), Time and Structure Evolving Systems (Design Problems), and Object Dynamic Systems (Data problems). The significance of the above families in the integration process and their link with generalized control problems are discussed.

1 Introduction

In any field of science or technology the scientists, experts or engineers have, on the one hand, the heavy mental load to learn about and keep up with their expertise, and on the other hand, they have the task to understand people from different disciplines or departments and to cooperate fruitfully with them. This two-fold action maybe is not strictly necessary in universities interdisciplinary cooperation but it is unavoidable for solving problems in industrial enterprises.

It is common in industrial projects people from one team to picture in clear and colorful detail everything in their knowledge base, but they are ignorant of results and have a hazy and colorless image of what happen in the field of another team. Project teams need a common framework, a common language so as to understand each other see the problem from a different point of view and reach integrated solutions on a new approach.

The need for communication is commanding not only between people from different teams or departments of the industrial enterprise, but also between “humans” and “machines”, between operational functions and between disciplines. Such an approach that breaks the traditional boundaries between technical and managerial disciplines, between operational function and even between software and data organization leads to the *System Integration*.

The problem of System Integration in the context of an Industrial Enterprise or in other words the *Enterprise Integration* is a complex problem. According to Brosey *et al*. [2] enterprise integration connects and combines people, processes, systems, and technologies to ensure that the right people and the right processes have the right information and the right resources at the right time.

The need of the *right information* requires a precise knowledge of the different activities in the enterprise. The need of the right information at the *right place* requires information sharing systems and capability of handling information transaction across different operating systems, heterogeneous hardware and software applications, and in general, heterogeneous environments that cross the process operations boundaries on a temporal basis. The need of the right information at the right place at the *right time* requires the up-date of the data created during the operations, as well as, the adaptation to any change originates from new demands, new technologies, new legislation etc. The need of coordinating people, processes, systems and technologies requires precise modeling of all the enterprise operations, which goes

beyond the exchange of information and information sharing, since it takes into account decision making capabilities and evaluation of operational alternatives.

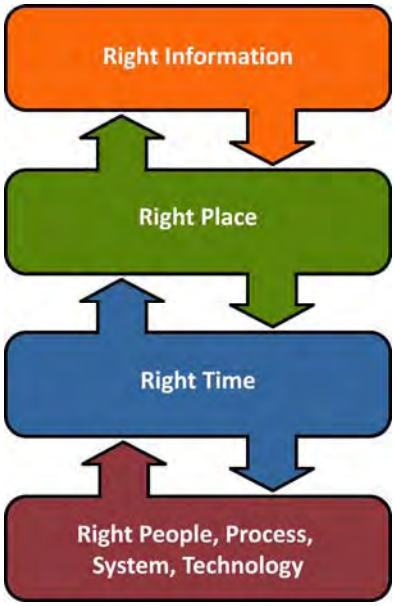


Figure 1: Enterprise Integration Overview

The consideration of the enterprise integration problem inevitably involves issues such as *system*, *modeling* and *control*. The integration concept itself is the bringing together of the existing and often disparate components of a system, i.e. the subsystems, into one system and ensuring that the subsystems function together as a system. On the other hand, the enterprise representation is above all a model of the structure, activities, processes, information, resources, people, behavior, goals etc. In addition, the better understanding, restructure or design enterprise operations various aspects should be assessed in the way of enterprise modeling. Last but not least, the concept of control has diversified applications that include financial, managerial, human behavior, environmental and safety aspects, as well as functional aspects of the industrial enterprise by means of the analysis and design of the proper system/controller that can self-regulate a process, with minimal human intervention, and keep its variables near set points.

2 Process Operations

In the broad framework of the enterprise activities there are functions related to the *strategy* of the enterprise, a task of top management, to the *operation* of the enterprise, where the production issues are placed, and finally to the *support* which rings about production process.

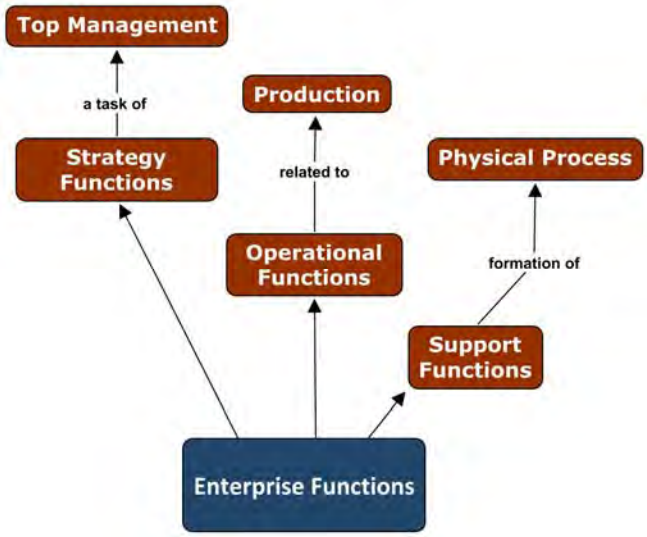


Figure 2: Nesting of Enterprise Activities – Functions

The organization of industrial enterprises has avoidably a multidimensional structure and the choice in visualizing depends on what one wishes to highlight. We will mainly focus on the area of *operational functions*, which are closely related to production and therefore have a precise modeling and control dimension. Let us sketch briefly the major groups of operational functions: A) *logistics*, B) *desired operations*, and C) *process control, supervision and information* [6].

A. The *logistic operational functions* define the outer layer of the operational hierarchy and here we may distinguish between: a1) the *planning* and a2) the *scheduling* activities. The planning operations look further into the future dealing with global issues and information about production and linking all parts of enterprise related to selling, producing, transporting, and acquiring raw materials of each product family. A nearer look into the future is a task of the scheduling operations, such as the utilities, storage, manpower, and conditions of a production site that result in a specific program of production.

B. The group of recipe operational functions aimed at *desired operations* is where the production runs and it is closely linked to the area of improvements, e.g. discovering the best operating point of continuous processes. Here we may distinguish between: b1) *recipe setting*, i.e. find the optimum prescription – recipe and specify the desired values in recipe, b2) *recipe initialization*, i.e. begin the production run, and b3) *recipe corrections*, i.e. insert the appropriate adjustments between actual set points and desired values. Desired operations, located between logistic control and process control and supervision, are the way of setting goals before going into all complexities of realization.

C. The *process control, supervision, and information functions* refer to the manipulation of process inputs, to the monitoring of process behavior, and to the on- and off-line measurements, respectively. Firstly, the most important activities related to *process control* are: c1) *quality control*, which is the function related to the quality of intermediate and final products according to the customer demands, legislation rules or environmental restrictions, c2) *regulatory, end point and sequence control*, which refer more than any other function to the direct influence on process control, e.g. keeping of on-line measured process variables near given set points (regulatory control), implementation of an optimum path in order a final value to be reached (end point control), opening/closing of valves or starting/stopping of pumps (sequence control), and c3) *emergency protection*, which is also a function of direct influence on the process operation by shutting it down or bringing it to a safe standby condition in case of danger or damage. On the other hand, of the type of *supervisory activities* there are: c4) the function of *state assessment*, which pronounces on the process state by means of easily observing displays of process variables and c5) the function of *off-normal handling*, i.e. in the case of off-normal deviation of one or more process variables the proper actions, such as diagnosis of derivation, correction of set points or replacement of faulty equipments, are taken by the operator. Last but not least, the *information functions* may be distinguished into: c6) *direct observations* by the operating personnel and *on-line measurements* by the automation system, c7) *quality analysis or off-line measurements*, the function of the laboratory analysis of a sample and the appropriate adjustments by the operator in the case the measured qualities deviate from target, and c8) *operation data base or real-time measurements*, the function of restoring real-time measurements for historic, statistic, optimization, finance and management purposes.

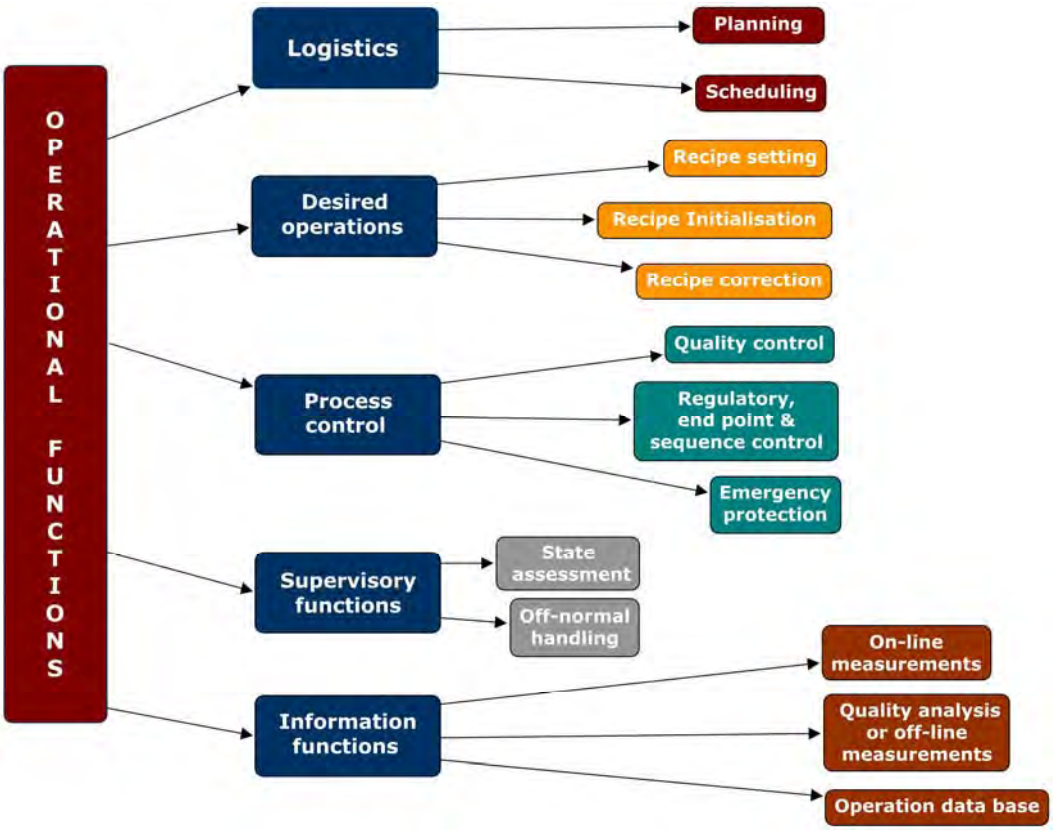


Figure 3: System Analysis of Operational Functions

The operational functions tuned to their respective goals need specific models that have to be adapted to the changes in the real world.

3 Modeling Operational Functions

The process models tailor to the specific tasks and conditions of each operational function. There is a remarkable abundance in their classification and the border lines between the different families are not always very clear. Some families of process models are: a) *black (empirical), white (fundamental or mechanistic) and grey models*, b) *on-line, off-line and line, support models*, and c) *static, dynamic models* [6]. The main characteristics of these categories are:

a) The *white models* are based on physical, chemical and/or biochemical principles and a complete set of equations is possible to be found, after a number of trials and errors, and also to be solved presuming that the process inputs (independent variables) are known. The development of such models is based on many assumptions and simplifications without

sacrificing model validity. Although they need long development time and their software implementations are sold for high prices, they are applicable over a relatively wide range of conditions and their poorly known parameters are relatively small in number. On the other hand, *black models* are based on experimental data and standard – not specific – relationships between input and output variables. They need little development time and effort, but the number of known parameters is rather large. The use of white models for the easy parts of a modeling problem and of black ones for the more difficult parts results in *white/black models*, whereas the knowledge of the ranges within which process variables will remain in practice turns the black models into *grey* ones.

b) If we distinguish models into those refer to the *process operations* and those to the *process design* we have the distinction between *off-line* and *on-line models*. If we classify models in terms of model output we have *line and support models*. Line models refer to the determination and realization of desired process conditions, e.g. set points for regulatory control loops, whereas support models provide information e.g. to control models.

c) When the interest of a process is restricted to the equilibrium point the proper model is a *static* one, which in addition can be augmented by dynamic equations. One step further we have the *dynamic models* with the major time constants of the process, suitable for example for setting recipes.

Another family of process models is that of *Boolean* and *quantitative models*. Both types are based on Boolean Logic and classical Set theory. The Boolean model is an information retrieval model that the information to be searched and the user's query are conceived as sets of terms. Retrieval is based on whether or not the information contain the query terms. On the other hand, qualitative models are formed by two or more propositions in disjunction or conjunction, signified in English by the words 'or' (or nor) and 'and' (or but), which compound statements either true if all their component propositions are true, or false if all their component propositions are false.

We will try to specify the different types of required models for the overall system by following a bottom-up direction of the operational functions hierarchy, i.e. by starting from the lower *level of data and measurements* (information functions) to the top business level. At the lowest data level the modeling problem focuses on the data reconciliation, since measurements are always contaminating by errors that result in unbalanced balances of

missing products or lost energy. A way out is to use dynamic process models that require information about component and energy holdups.

Concerning the next *level of supervisory functions*, the proper process models, which could secure the indication of any change in critical process variables by alarm signals, are developed by putting the process variables through filters corresponding to normal and off-normal condition. In addition, the correspondence of alarm set to overall process conditions, as well as the automatic detection of alarm signals and the auditory and visual display of them is succeeded by Boolean models set up by senior process operators. The off-normal handling function and especially the diagnosis part of it is mostly a task of quantitative models.

Going a level up we meet the *process control level* dealing with physical process input variables, such as flow, temperature, velocity etc. As flow rates and quantities vary, models have to be dynamic either linear (for narrow range variations) or nonlinear (for broad range variations). They are used to predict present or future qualities from previous operations. In the area of control there is a large variety of models depending on the nature of process, such as analytic, statistical, neural nets etc. The detailed examination of them is beyond the scope of this paper.

As we proceed upwards, the model development goes from detailed models on the unit level to simplified models on the plant level and to their further aggregation into total production models on the production site, business unit and possibly the enterprise level. As far as possible, it is desirable to derive one model form another. A model suitable for logistic and recipe functions may also supports the coordination between business units, production locations and plants. Such models are interrelated and it is tended to treat them as independently derived and not as byproducts of a unifying modeling process. Model compositions accompanied by simplifications is the dominant feature due to the process control hierarchy and this implies a nesting of models that is referred to as *embedding of function models*.

4 Control of Continuous Processes

The distinction between *continuous processes* and *batch processes* concerns how the production steps are related to each piece of process. In continuous processes each process performs only one step in the production process, whereas in batch processes several

production steps are executed in the same piece of process. In the former, where we place our investigation, there is a continuous inflow and outflow of materials, while in the latter the material output happens after the completions of the relevant production steps.

In order a process to be automatically controlled it must be provided with the correct inputs, which may be manipulated so as to bring the controlled output variables to the right value. If we place the process in the center of the operational activities of an enterprise, the following figure shows the interrelations between the process and the operational function, as they described previously.

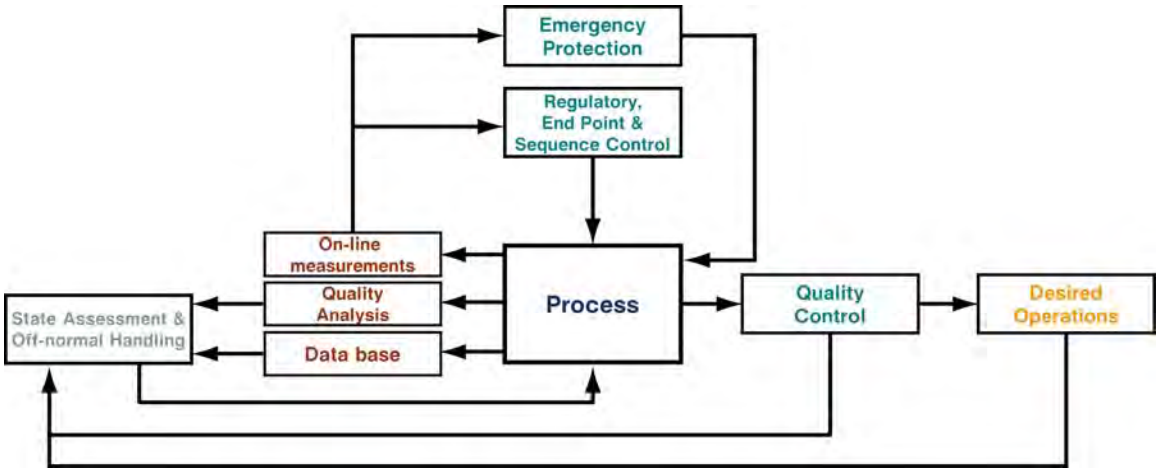


Figure 4: Interrelations of Process and Operational Functions

We adopt an input-output with internal variables description of processes as it is shown in figure 5, where $u_1,...,u_n$ are the independent manipulated variables called *system inputs*, $y_1,...,y_m$ are the independent controlled variables that can be measured and they are called *system outputs*, $d_1,...,d_n$ are the exogenous variables which cannot be manipulated, but they express the influence of external to the particular function variables and they are called *disturbances*, and $\underline{z}$ is the internal state vector expressing the variables involved in the particular process.

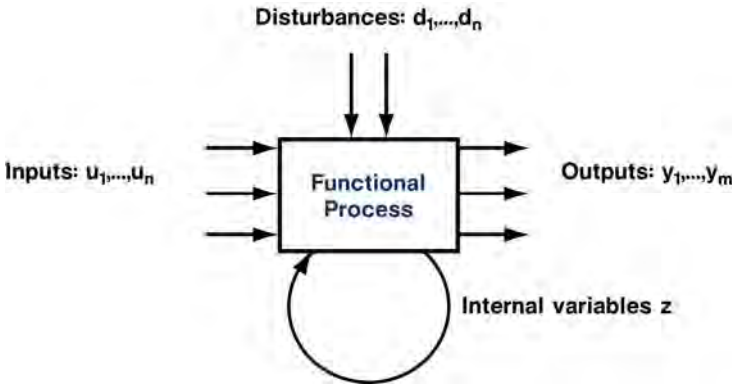


Figure 5: Generic description of a Process

The construction of a model that describes the relationship H between the vectors $\underline{u},\underline{d},\underline{y}$, for example $y = H(u,d)$, is major problem that involves a number of issues which may be classified as:

- For a given process establish a conceptual model for its role in the operational hierarchy.
- Define the vector of internal variable $\underline{z}$ associated with a given problem and determine its relationships to input, output vectors.
- Establish the relationships that exist between the alternative vectors $\underline{z}$ associated with problems of the operational hierarchy.
- Define the appropriate formal model.

The above generic steps are providing an approach, which however, involves many detailed modelling tasks. Typical problems here are issues such as classification of variables to inputs, outputs, disturbances, internal variables [2], specification of formal description H, definition of performance indices etc. When the classification of internal variables is completed, the key issue is the establishment of relationships between such variables; such relationships may be classified to implicit and explicit (oriented) forms respectively as:

$$M(u,y,d,z)\begin{cases} F(\underline{z},\underline{u},\underline{d})=0 & (1)\\ y=G(\underline{z},\underline{u},\underline{d}) & (2)\end{cases}$$

The nature of variables and the type of problem under consideration define the functions F,G. This model structure also shows how constraints $F(\underline{z},\underline{u},\underline{d})=0$ can be propagated from higher to lower levels. The selection of $\underline{z}$ implies that specifying F,G includes the modeling

of the interface of higher level operation to the level defined by $\underline{z}$. The model $M(u, y, d, z)$ in relations (1) & (2) will be referred to as a z-stage model. The specification of the overall control architecture involves the construction of the particular z-stage models and the definition of the relationships between them. The selection of the operational stage (i.e., logistics, scheduling, quality control, process control, state assessment, etc.) determines the nature of the internal vector $\underline{z}$ and thus also of the corresponding z-stage model.

It is worth pointing out that if we are interested in dynamics of the process, vector $\underline{z}$ is of a large dimension, it is considered as the composite vector of inputs, states, and outputs of the dynamic process and thus it is the process implicit vector. If, however, we go higher in the hierarchy and progressively consider issues of steady state optimisation, scheduling etc. then the corresponding implicit vector (composite vector of local inputs, internal variables, outputs) it has reduced dimensions. The dimensionality and nature of $\underline{z}$ depend on the problem under study. Describing the relationship between different stages internal vectors is an important problem and it is closely related to the hierarchical nesting of process operations. The interdependence of the different operational stages/functions is demonstrated by the following figure:

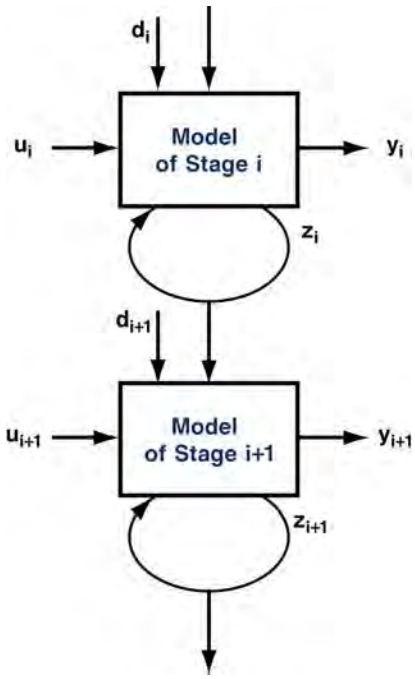


Figure 6: Models of successive operational stages/functions

A scheme such as the one described above is general and can be used to describe the meaning of the hierarchical nesting. Furthermore, such a scheme can be extended to describe relations between models associated with functions at the same level of the hierarchy [4]. The fact that each stage model in the hierarchy is of different nature than the others makes the overall system of hybrid nature and thus the theory of hybrid systems is crucial in the study of the control problems defined on the integrated hierarchy. On this multilayer structure we have two fundamental problems:

- Global Controllability Problem
- Global Observability Problem

The first refers to whether a high level objective can be realized within the existing constraints at each of the levels in the hierarchy and finally at lowest level, where we have the physical process. This is called global controllability, or realization of high level objectives throughout the hierarchy. This requires development of a multilevel hybrid theory and it can take different forms, according to the nature of the particular stage model. Global Controllability is central in the development of top-down approaches in the study of hierarchical organizations. The second problem refers to the property of being able to observe certain aspects of behavior of the production layer of the hierarchy by appropriate measurements, or estimation sub-processes which are built in the overall scheme. This is a Global Observability property and expresses the ability to define Model Based Diagnostics that can predict, evaluate aspects of emergent behavior of the process. Global Observability is a function of the structuring of the information system and it is linked to the bottom-up approach in the study of hierarchical organizations. The measurements, diagnostics defined on the physical process are used to construct the specific property functional models and thus Global Observability is linked to the quality of the respective functional model.

5 Conclusions

The paper has provided an overview of those technological aspects of systems integration, linked to Process Operations which have a Systems, Modeling and Control dimension. It identifies a range of new open issues of the Systems and Control type as well as new families of systems which are intimately linked to the new applications paradigm. The three central themes which emerge are the needs for control, the development of theory and methodology

for the different types of systems organizations, as well as developing some understanding for the different forms of evolving systems. In addition, integration of operations has many components and requires study of fundamental problems such as functional model, global controllability and observability. These problems have links between themselves and establishing such links is also a challenging problem that may be referred to as Process Operations Design. Of course, Process Operations are based always on a physical system, process. Establishing the links between Operational criteria (desirable goals) and Engineering Design Objectives - criteria, is a major challenge and it is referred to as Operations Design Interface problem. When operational objectives cannot be realized on the existing physical process, then the problem of Process Redesign arises and this is a problem that addresses together problems of Process Operations and Integration of Design simultaneously and can be considered within the current framework.

References

[1] A.C.P.M. Backx, *Engineering Aspects of Industrial Applications of Model-Based Control Techniques and System Theory, in Essays on Control: Perspectives in Theory and Its Applications*, pp. 79-109. Ed. H. L. Trentelman, L.C. Willems Birkhäuser Boston 1993.

[2] W. D. Brosey, R. E. Neal, D. Marks, *Grand Challenges of Enterprise Integration*. 8th IEEE International Conference on Emerging Technologies and Factory Automation, France, 2001.

[3] N. Karcanias, *Global Process Instrumentation: Issues and Problems of a Systems and Control Theory Framework*. Measurement, Vol. 14, pp 103-113, 1994.

[4] N. Karcanias, *Integrated Process Design: A Generic Control Theory/Design Based Framework*, Computers in Industry, Vol. 26, pp 291-301, 1995.

[5] N. Karcanias, M. Newby, I. Leontaritis, *Modeling of the Integrated Business and Operational Issues of Industrial Processes: A System-Based Approach.* Proceedings of ASI'98 Conference, Bremen, June 14-17.

[6] J. E. Rijnsdorp, *Integrated Process Control and Automation*. Elsevier Science Publishers, Amsterdam, 1991.

[7] F. Vernadat, *Enterprise Modeling and Integration: Principles and Applications*. Chapman & Hall, London, 1996.

Strategien zur Regelung von HCCI-Brennverfahren

Roland Karrelmeyer, Wolfgang Fischer
Robert Bosch GmbH, Robert-Bosch-Strasse 2, 71701 Schwieberdingen
Roland.Karrelmeyer@de.bosch.com

Kurzfassung

Hohe Verbrauchsvorteile im Sinne von CO_2-Reduktion bei gleichzeitig niedrigen Stickoxyd- und Partikelemissionen sind die charakteristischen Merkmale des HCCI - (**H**omogeneous **C**harge **C**ompression **I**gnition) Brennverfahrens bei Otto-Motoren. HCCI ist ein mageres Teillastbrennverfahren, welches aufgrund der niedrigen Stickoxyd-Emissionen mit einer konventionellen Abgasnachbehandlung auskommt. Demgegenüber stehen bei der Realisierung dieses Brennverfahrens spezielle Anforderungen an den Ventiltrieb, das Einspritzsystem, die Sensorik und an die Motorsteuerung. Das Fehlen eines direkten Verbrennungstriggers macht den Zeitpunkt der Selbstzündung sehr sensibel gegenüber Toleranzen der Aktorik, Veränderungen der Umgebungsbedingungen und Streuungen in der Kraftstoffqualität. Deshalb kommt der Steuerung und Regelung der Verbrennung eine entscheidende Rolle zu.

In dieser Arbeit werden Konzepte zur Regelung eines HCCI-Brennverfahrens mit interner Abgasrückführung, auch mit Trapping bezeichnet, vorgestellt. Hierbei dienen die Verbrennungslage und die pro Zylinder und Arbeitsspiel erzeugte Arbeit als charakteristische Größen des Verbrennungsprozesses und werden als Regelgröße verwendet. Diese sogenannten Verbrennungsmerkmale werden ihrerseits aus dem Brennraumdrucksignal gewonnen. Zur Beeinflussung der Verbrennung stehen die Stellgrößen des Luft- und Kraftstoffsystem zur Verfügung. Kernstück des Luftsystems ist ein von einer Nockenwelle angetriebener Ventiltrieb mit umschaltbarer Ventilhubkontur und Phasenverstellung für die Einlass- und Auslassventile. Hierdurch ist eine für alle Zylinder gleichermaßen wirksame Zylinderfüllungssteuerung möglich. Dage-

gen ist durch den Einsatz einer Benzindirekteinspritzung eine zylinderindividuelle Kraftstoffeinbringung realisierbar, womit eine zylinderindividuelle Stellgröße zur Beeinflussung der Verbrennung zur Verfügung steht. Ausgehend von diesen Regel- und Stellgrößen werden Konzepte zur Regelung des HCCI-Verbrennungsprozesses vorgestellt.

Die Arbeiten erfolgen an einem Versuchsmotor mit teilvariablem Ventiltrieb und einem System zur Kraftstoffdirekteinspritzung. Die aufgezeigten Regelkonzepte bestehen aus einer Kombination aus einer Vorsteuerung und einer zylinderindividuellen Brennmerkmalsregelung. Ergebnisse zur Betriebsartenumschaltung, zum Dynamikbetrieb und zur Stabilisierung der HCCI-Betriebsart werden ebenfalls vorgestellt. Zur Vermeidung von Verbrennungsaussetzern und extrem frühen und somit lauten Verbrennungen sowohl im Dynamikbetrieb als auch bei einem Betriebartenwechsel kommt speziell der Vorsteuerung eine besondere Bedeutung zu.

1. Einleitung

Ressourcenknappheit der Energieträger und zukünftige Gesetzgebung zum Klimaschutz sind die Triebfedern in der Motorenentwicklung. Hierbei stehen besonders die Reduzierung des CO_2-Ausstoßes und die der Schadstoffemissionen im Fokus der Entwicklungen. Neben anderen Technologien stellt das HCCI – Brennverfahren für Ottomotoren eine vielversprechende Alternative zur Erreichung der Ziele dar. Kernstück dieses Brennverfahrens ist die Selbstzündung eines mit Restgas angereicherten Kraftstoff-Luft-Gemisches [1]. Der angestrebte Verbrauchsvorteil wird durch einen thermodynamisch effizienteren Verbrennungsablauf sowie einen quasi ungedrosselten Motorbetrieb und eine somit magere Verbrennung erzielt. Bedingt durch die auftretende Raumzündung des Gemisches ohne ausgeprägte Flammenfront verbleibt die lokale Verbrennungstemperatur im Brennraum unter der kritischen NO_x-Bildungstemperatur. Daher kann im Unterschied zu anderen Magerkonzepten, wie z.B. Brennverfahren mit Ladungsschichtung, auf eine aufwendige und teure NO_x-Nachbehandlung verzichtet werden.

Das Fehlen eines direkten Verbrennungstriggers, wie es zum Beispiel der Zündfunke bei dem fremdgezündeten konventionellen Ottomotorenbetrieb ist, verbunden mit einer hohen Sensitivität hinsichtlich sich ändernder Umgebungsbedingungen schreibt der Steuerung und Regelung des HCCI-Brennverfahrens eine gewichtige Rolle zu. Vorausgesetzt wird hierbei eine gewisse Flexibilität in der Ansteuerung des Luft- und Kraftstoffpfads. Bezüglich des Kraftstoffpfades stellen die Komponenten der Benzindirekteinspritzung diese Flexibilität bereit. Volle Flexibilät im Luftpfad würde man durch den Einsatz einer vollvariablen Ventilsteuerung erhalten, denn hierbei könnten idealerweise Öffnungs-, Schließzeiten und der Hub der Ventile nahezu unabhängig voneinander vorgegeben werden. Durch tolerierbares Abrücken von der optimalen Verbrennungsführung kommen in einem ersten Schritt Nockenwellen-ventilsysteme in Betracht, die über Phasensteller und Hubumschaltung einen teilvariablen Eingriff auf den Luftpfad bereitstellen. Es wird ein Regelkonzept vorgestellt, das basierend auf einem Einlass- und Auslassnockenwellen-Phasensteller und einer Zweipunkt-Hubverstellung einen stabilen HCCI - Mode erzeugt.

Als Rückmeldung für die jeweiligen Regler werden zylinderindividuell die Verbrennungslage und die abgegebene Arbeit verwendet. Diese Verbrennungsmerkmale werden auf die Größen MFB50%[1] und NMEP[2] abgebildet, deren Berechnung aus dem Brennraumdruckverlauf erfolgt. Der MFB50% ist ein Maß für die Verbrennungslage und kennzeichnet den Punkt, bei dem etwa 50% des eingebrachten Kraftstoffes in Wärmeenergie umgesetzt wurde, wogegen der NMEP die geleistete Arbeit pro Arbeitszyklus widerspiegelt.

Aufgrund der physikalischen Randbedingungen ist der HCCI-Mode nur in einem beschränkten Betriebsbereich darstellbar. Daher ist in der Motorsteuerung eine Funktion zur momenten-neutralen Umschaltung zwischen verschiedenen Verbrennungsmodi zwingend erforderlich. Es wird auf die zu höheren Lasten erforderliche Modeumschaltung zwischen SI[3]- und HCCI-Mode eingegangen.

---

[1] MFB50% (**M**ass **F**raction **B**urned **50%)** bezeichnet den Punkt, bei dem 50% der eingebrachten Kraftstoffmasse umgesetzt wurde.

[2] NMEP (**N**et **M**ean **E**ffective **P**ressure) ist die auf das Hubvolumen bezogene pro Zylinder und Arbeitspiel geleistete Arbeit des Verbrennungsprozesses.

[3] SI (**S**park **I**gnition) bezeichnet den mittels Zündkerze fremdgezündeten Betrieb des Verbrennungsmotors.

2. Grundzüge des HCCI-Brennverfahrens

Die im Rahmen dieses Artikels aufgezeigten Ergebnisse basieren auf Untersuchungen an einem 4-Zylinder-Versuchsmotor, der mit einem Ventilsystem mit Phasensteller für Einlass- und Auslassnockenwelle und Ventilhubumschaltung ausgerüstet ist. Das System ermöglicht die Darstellung von Ventilstrategien, die eine für die Selbstzündung benötigte Enthalpie aus dem Abgas bereitstellen. Erreicht wird dieses durch eine negative Ventilüberschneidung, wie sie durch die Ventilhubverläufe in Abbildung 1 dargestellt ist.

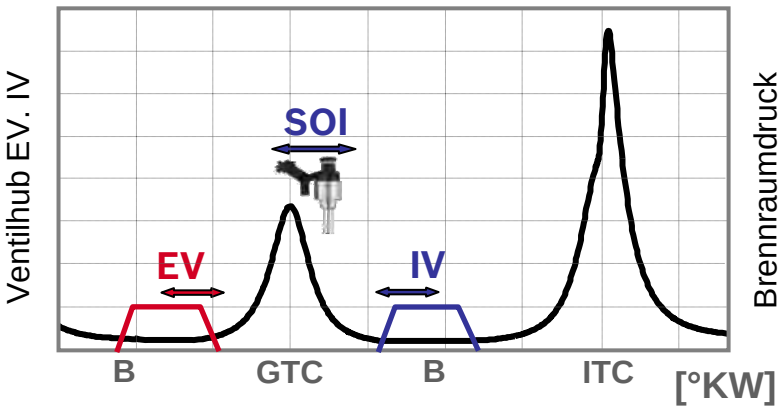


Abbildung 1: Ventil- und Einspritzstrategie für HCCI-Betrieb

Bei dieser Ventilstrategie wird durch frühes Schließen der Auslassventile EVC[4] und spätem Öffnen der Einlassventile IVO[5] erreicht, dass genügend heißes Restgas im Zylinder verbleibt, so dass nach dem Schließen des Einlassventils IVC[6] in der darauffolgenden Verdichtungsphase durch Kompression des Restgas-Kraftstoff-Luftgemisches ausreichend hohe Temperaturen zur Selbstzündung erzielt werden. Dabei wird das Restgas zuerst in Richtung des oberen Totpunktes der Gaswechselphase GTC[7] verdichtet und anschließend wieder expandiert. Um dabei Pumpverluste zu vermeiden, erfolgt das Öffnen der Einlassventile hinsichtlich GTC annähernd sym-

[4] EVC (**E**xhaust **V**alve **C**losing) bezeichnet den Kurbelwellenwinkel, bei dem das Auslassventil schließt.
[5] IVO (**I**nlet **V**alve **O**pening) bezeichnet den Kurbelwellenwinkel, bei dem das Einlassventil öffnet.
[6] IVC (**I**nlet **V**alve **C**losing) bezeichnet den Kurbelwellenwinkel, bei dem das Einlassventil schließt.
[7] GTC (**G**as flow **T**op dead **C**enter) bezeichnet den oberen Totpunk des Kolbens, an dem der Gaswechsel stattfindet.

metrisch zum Schließen der Auslassventile, siehe Abbildung 1. Im Bereich der Restgaskompression wird der Kraftstoff eingebracht. Die Einspritzung in das heiße Medium bewirkt eine gute Aufbereitung und Homogenisierung des Kraftstoffes. Geht man nun davon aus, dass bei ca. 1000K die Selbstzündung einsetzt, und dass man im Normalfall nach IVC keine Möglichkeit mehr hat das Gemisch zu beeinflussen, so steht der Ladungszustand insbesondere die Ladungstemperatur bei IVC in direktem Zusammenhang mit dem Einsetzen der Selbstzündung. Grundsätzlich kann festgehalten werden, dass mit steigender Ladungstemperatur bei IVC der Start der Verbrennung SOC[8] nach früh verschoben wird. Die Ladungstemperatur bei IVC, die sich direkt als Mischungstemperatur aus kalter Frischluft und heißem Restgas ergibt, wird direkt vom Mischungsverhältnis, also vom Restgasanteil bestimmt. Somit ist der Restgasanteil ein direkter Hebel zur Beeinflussung der Verbrennungslage. Etwas komplexer stellt sich der Einfluss des Einspritzbeginns SOI[9] auf die Verbrennungslage dar. Hier spielen thermodynamische Effekte, wie die Abkühlung durch Kraftstoffeinbringung und die Änderung des Polytropenexponenten nach Kraftstoffeinbringung eine wichtige Rolle. Sie bewirken in der Expansionsphase nach der Restgaskompression eine Änderung im Temperaturverlauf. Späte Einspritzlagen innerhalb der Restgaskompression bewirken ein Absenken der Ladungstemperatur bei IVC, welches dann zu späteren Verbrennungslagen führt. Letztendlich kann festgehalten werden, dass mit der Restgasmasse, also mit der Lage von EVC und mit dem Einspritz-Timing wichtige Hebel zur Verfügung stehen, um beim HCCI-Brennverfahren die Verbrennungslage zu steuern. Die Diagramme in der Abbildung 2 verdeutlichen diese Effekte, wobei von oben nach unten der Reihe nach über den Kurbelwellenwinkel der Brennraumdruck, die Wärmefreisetzung und die Temperatur dargestellt sind, siehe auch [2]. Aus den Diagrammen ist auch ersichtlich, dass sich mit zunehmender Frühverschiebung der Verbrennungslage die Dauer der Wärmefreisetzung verkürzt. Diese hat einen Anstieg im Druckgradienten des Verbrennungsdrucks zur Folge, welcher durch ein hartes dieselartiges Verbrennungsgeräusch am

[8] SOC (**S**tart **o**f **C**ombustion) bezeichnet den Kurbelwellenwinkel bezüglich des Zünd-OTs bei dem die Verbrennung einsetzt.
[9] SOI (Start of Injection) bezeichnet den Kurbelwellenwinkel, bei dem die Kraftstoffeinspritzung beginnt.

Prüfstand wahrnehmbar ist. Da es sich beim HCCI-Brennverfahren um ein mageres Brennverfahren handelt, ist in gewissen Bereichen genug Sauerstoff für die Verbrennung vorhanden. Damit kann über die eingebrachte Krafftstoffmenge direkt die abgegebene Leistung gestellt werden.

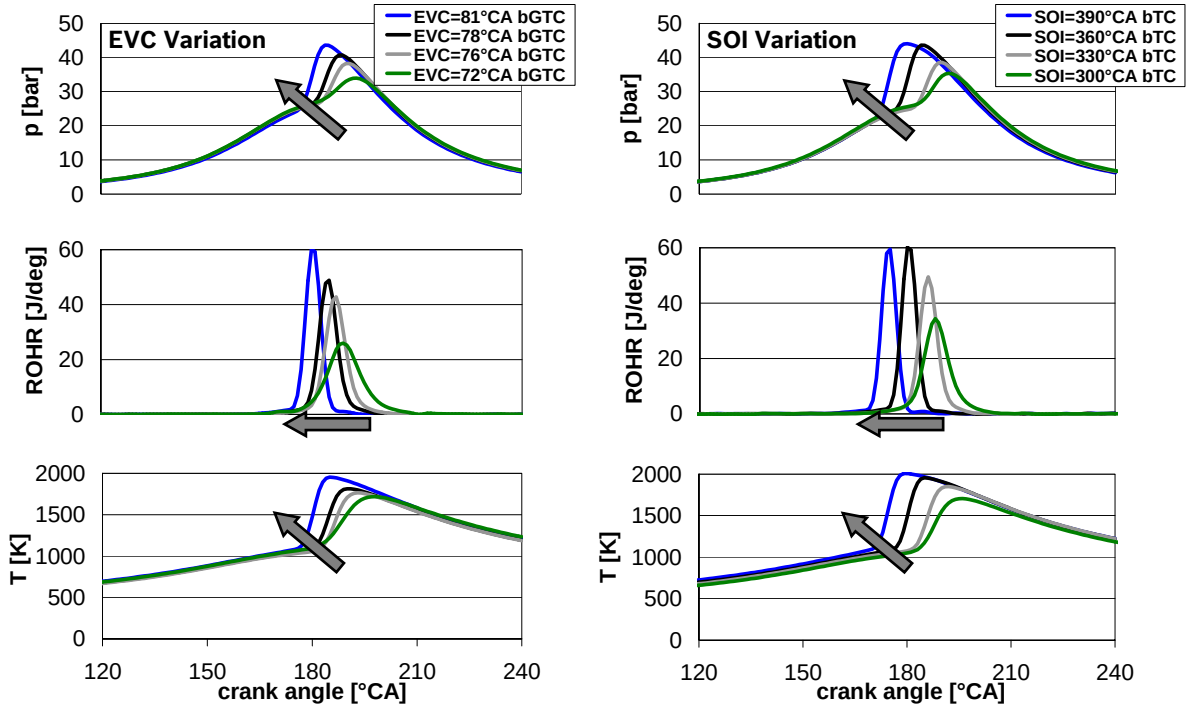


Abbildung 2: Einfluss von EVC und SOI auf den Brennverlauf

Neben der zuvor kurz skizzierten Basis-Einspritzstrategie kommen im HCCI-Betrieb abhängig von Drehzahl und Last auch Einspritzstrategien zum Einsatz, die Mehrfacheinspritzungen erfordern. Speziell bei Mehrfacheinspritzungen kommt der Genauigkeit bei der Mengendosierung durch die Kraftstoffinjektoren eine besondere Bedeutung zu. In Abbildung 1 ist eine Strategie mit Einfacheinspritzung dargestellt, die im Wesentlichen bei mittleren Lasten des HCCI-Kennfeldes verwendet wird.

Motoruntersuchungen am Prüfstand zeigen ein vielversprechendes Potenzial hinsichtlich Kraftstoffersparnis. Für einzelne HCCI-Betriebspunkte können Einsparungen von bis zu 30% im Vergleich zu homogenen, ins Saugrohr einspritzenden SI-Brennverfahren erzielt werden. Im Diagramm von Abbildung 3 ist das im jeweiligen Arbeitspunkt erreichte Einsparpotenzial angegeben.

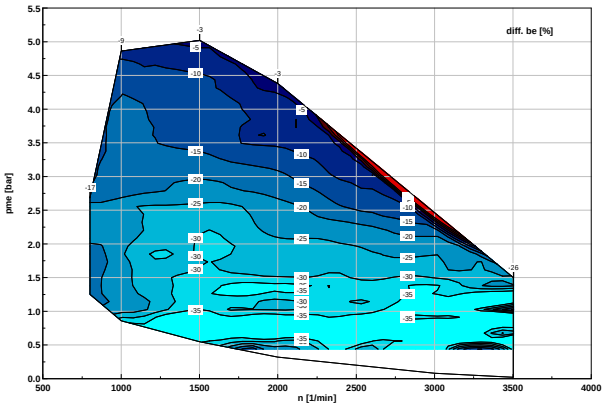


Abbildung 3: Einsparpotenzial HCCI gegenüber SI

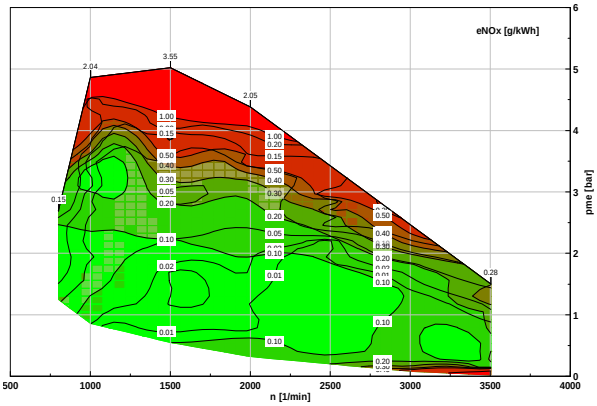


Abbildung 4: NOx-Rohemissionen

Abbildung 4 untermauert die Aussage der minimalen NOx Erzeugung beim HCC-Brennverfahren. Hier sind die NOx-Rohemissionen angegeben.

Die Realisierung der Betriebsart HCCI im realen Motorbetrieb stellt hohe Anforderungen speziell an den Ventiltrieb hinsichtlich Variabilität, Stellgeschwindigkeit und Stellgenauigkeit. Der am Versuchsmotor eingesetzte Ventiltrieb erfüllt diese Anforderungen. Er basiert auf einen jeweils auf die Einlass- und Auslassnockenwelle wirkenden auf Verstellgeschwindigkeit optimierten Phasensteller und eine durch schaltbare Rollenschlepphebel realisierte Ventilhubkontur-Umschaltung.

3. Regelungsplattform

Zur effizienten Entwicklung und Implementierung der zum stabilen HCCI-Betrieb benötigten Steuer- und Regelungsfunktionen ist eine flexible und leistungsstarke Hardwareplattform aufgebaut worden, die neben HCCI-spezifischen Funktionseinheiten auch die Basisfunktionalität einer Motorsteuerung abdeckt. Im Kern besteht diese Plattform aus zwei elektronischen Steuereinheiten, die über einen CAN[10]-Bus miteinander vernetzt sind. Die Topologie dieses Netzes mit den Komponenten dSPACE und MED zeigt Abbildung 5. Die jeweilige Funktionalität der einzelnen Komponenten ist durch ihren Zugriff auf den Motor gekennzeichnet. Die MED-Einheit realisiert den Zugriff auf den Kraftstoffpfad und auf die Drosselklappe. Die Phasenstellersysteme werden ebenfalls sowohl für die Einlass- als auch für die Auslassnockenwelle von

[10] CAN: **C**ontroller **A**rea **N**etwork

der MED angesteuert. Diese Konfiguration ermöglicht die Umsetzung der zuvor beschriebenen Steuerstrategie.

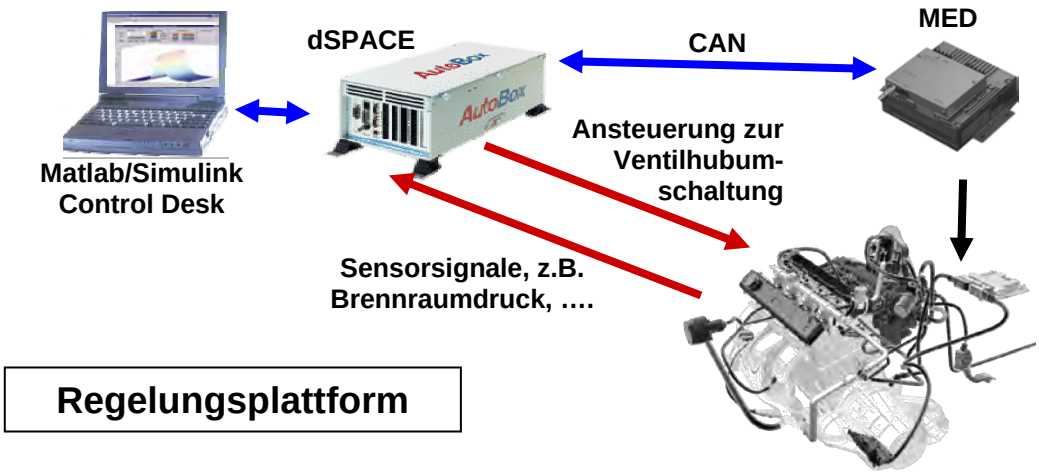


Abbildung 5: Topologie der Regelungsplattform

Kernstück der Plattform ist ein Rapid-Prototyping-System der Firma dSPACE. Dieses System besteht aus zwei PowerPC-Prozessorboards, dem Master- und dem Slaveboard, die über einen optischen Hochgeschwindigkeitslink miteinander kommunizieren.

Der Master übernimmt die Synchronisation mit dem Kurbelwinkel des Motors und realisiert eine winkelsynchrone Abtastung der Brennraumdrucksignale, die wiederum mittels eines Brennraumdrucksensors pro Zylinder gemessen werden. Ein das Mastersystem erweiterndes ADC[11]-Board stellt hierfür das entsprechende Prozessinterface zur Verfügung. Aus den abgetasteten Brennraumdruckverläufen werden nach FIR-Filterung und Sensorsignal-Offsetkorrektur die Hauptverbrennungsmerkmale MFB50% und NMEP extrahiert, die als Maß für die Verbrennungslage und die abgegebene Arbeit dienen. Bei der hier zugrunde liegenden Realisierung wird der MFB50% pro Zylinder und Zyklus aus einer vereinfachten Heizverlaufsrechnung ermittelt. Als weiteres Merkmal ist der maximale Druckgradient zu nennen, der

[11] ADC: **A**nalog-**D**igital-**C**onverter

besonders beim HCCI-Brennverfahren in Verbindung mit der Geräuschentwicklung zu betrachten ist.

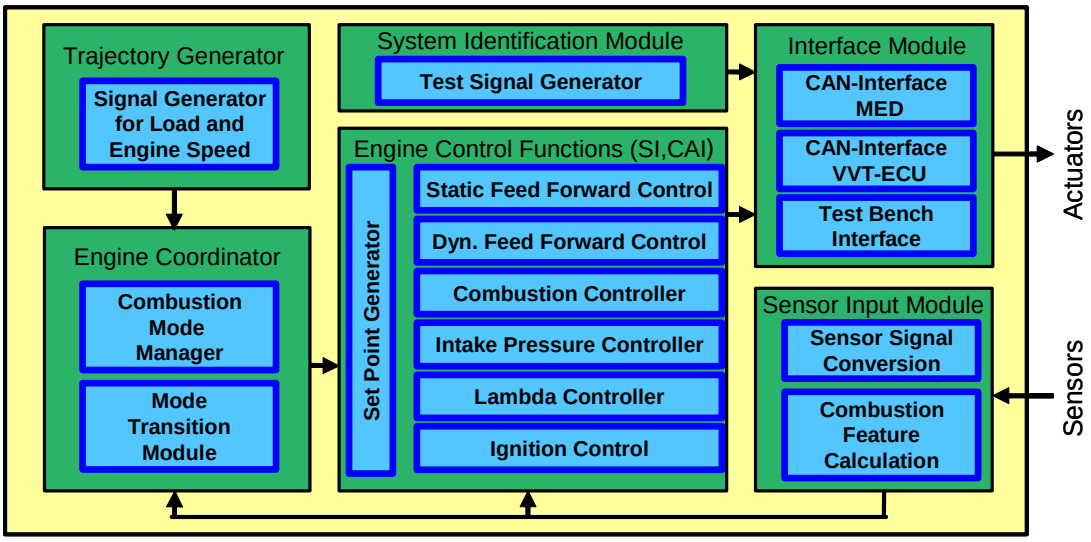


Abbildung 6: Überblick über die Plattform-Funktionseinheiten

Neben der Merkmalsberechnung obliegt dem Master die Erzeugung kurbelwellensynchroner Triggersignale, die auf der Slaveseite winkelsynchrone Prozessortasks auslösen, die wiederum die eigentlichen Reglerfunktionen beinhalten. Das slaveseitig eingesetzte CAN-Board übermittelt die aus den Reglertasks generierten Stellgrößen an die für den Kraftstoff- und Luftpfad zuständige Steuereinheit MED.

Es wurde schon darauf hingewiesen, dass es sich beim HCCI-Brennverfahren um ein Teillast-Brennverfahren handelt. Die Forderung eine Bewertung im gesamten Motorkennfeld vornehmen zu können, führt zwangläufig dazu, dass auch gewisse Teile des Motorenkennfelds im SI-Mode gefahren werden müssen. Diesen Teil übernimmt die MED, die hierfür entsprechend appliziert wurde. Da es das Ziel ist, eine entsprechende Bewertung bezüglich Verbrauch und Emissionen vornehmen zu können, ist es erforderlich, am Prüfstand Fahrzyklen, wie z.B. den NEDC [12], abzubilden. Dieses erfordert eine Erweiterung der Basisfunktionalität der Plattform. Neben der HCCI-Closed-Loop-Regelung sind Funktionseinheiten zur Umschaltung der Verbrennungsmodi und zur Ansteuerung des Motorprüfstands entwickelt und

[12] NEDC: **N**ew **E**uropean **D**riving **C**ycle

implementiert worden. Dadurch ist man in der Lage, Last- und Drehzahlprofile abzufahren.

Komplettiert wird die Plattform durch Funktionseinheiten, die eine Testanregung der unterschiedlichen Stellgrößen des Prozesses ermöglichen. Dieses ist besonders im Hinblick auf den Einsatz von Verfahren zur Regelstreckenidentifikation wichtig. Die so erhaltenen Streckenmodelle stellen die Grundlage sowohl für den Reglerentwurf als auch für die Entwicklung einer dynamischen Vorsteuerung dar, auf die im nachfolgenden Abschnitt näher eingegangen wird. Einen Überblick über die letztendlich implementierten Funktionseinheiten und ihre Verknüpfungen erhält man in Abbildung 6.

4. Modell der Regelstrecke

Grundlegend für eine regelungstechnische Analyse ist die Kenntnis der Regelstrecke „Verbrennung“, die sich hier als funktionale Abbildung der aktiv genutzten Stellgrößen eingespritzte Kraftstoffmenge (q), Einspritzwinkel (SOI) und dem Schließwinkel des Auslassventils (EVC) auf die Verbrennungsmerkmale NMEP, MFB50% und das messbare Verbrennungsluftverhältnis (λ) darstellt. Abbildung 7 verdeutlicht diesen funktionalen Zusammenhang mit expliziter Darstellung der durch die Abgasrückhaltung auftretenden Zyklus-zu-Zyklus-Kopplung.

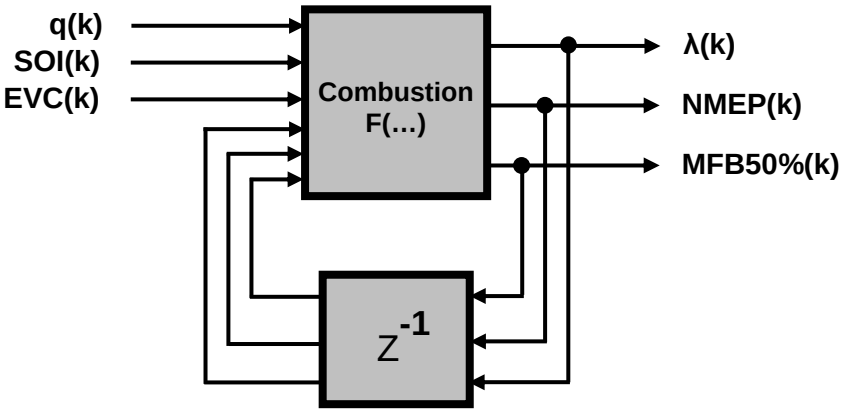


Abbildung 7: Modell der Regelstrecke

5. Systemidentifikation

Zur Identifikation der funktionalen Abhängigkeit (F) wird ein nichtlineares Datenbasiertes Verfahren eingesetzt. Dabei werden die Vorsteuerwerte der relevanten Stellgrößen an verschiedenen Betriebspunkten im stationären HCCI-Betrieb mit gleichverteiltem Rauschen angeregt und diese Stellgrößen sowie die entsprechenden Verbrennungsmerkmale zyklussynchron aufgezeichnet und einer Korrelationsanalyse zugeführt. Exemplarisch ist als Ergebnis der Korrelationsanalyse in der Abbildung 8 der Zusammenhang zwischen MFB50%(k) und q(k-1) bzw. in der Abbildung 9 der zwischen MFB50%(k) und q(k) gezeigt. Auf den ersten Blick mag die Abhängigkeit der Verbrennungslage von der Einspritzmenge des vorangegangen Zyklus aus Abbildung 8 unerwartet erscheinen. Sie erklärt sich aber direkt aus der Zyklus-zu-Zyklus-Kopplung, hervorgerufen durch das Rückhalten und Zwischenverdichten des Restgases, dessen Temperatur einen wesentlichen Einfluss auf die Lage der Selbstzündungsverbrennung ausübt. Die Temperatur des Restgases wird wiederum maßgeblich durch die im vorangegangenen Zyklus umgesetzte Einspritzmenge beeinflusst, siehe auch [3].

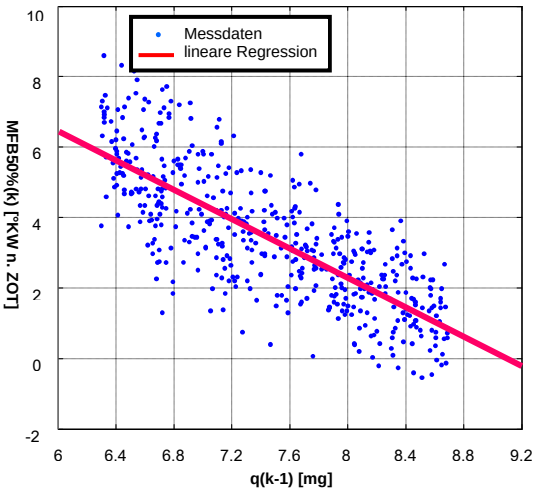


Abbildung 8: Zusammenhang zwischen MFB50%(k) und q(k-1)

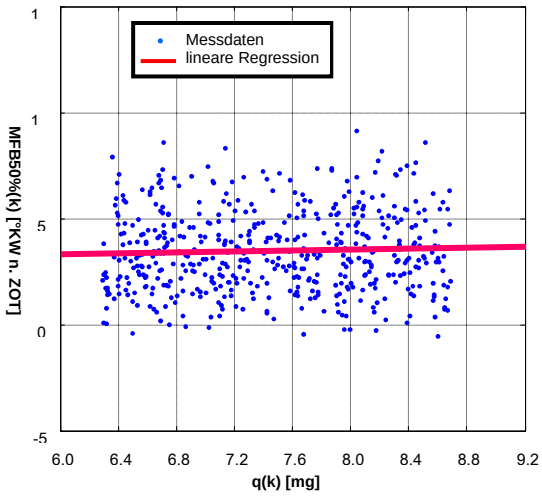


Abbildung 9: Zusammenhang zwischen MFB50%(k) und

6. Grundstruktur des Regelungskonzeptes

Abbildung 10 zeigt die in ihrer Grundform aus einer statischen Vorsteuerung und einem zylinderindividuellen Regler bestehende Struktur, die durch eine modellba-

sierte prädiktive Vorsteuerung erweitert wurde. Ihr obliegt u. a. die Steuerung von Modetransitionen, die speziell für den Wechsel zwischen SI- und HCCI-Mode verantwortlich ist.

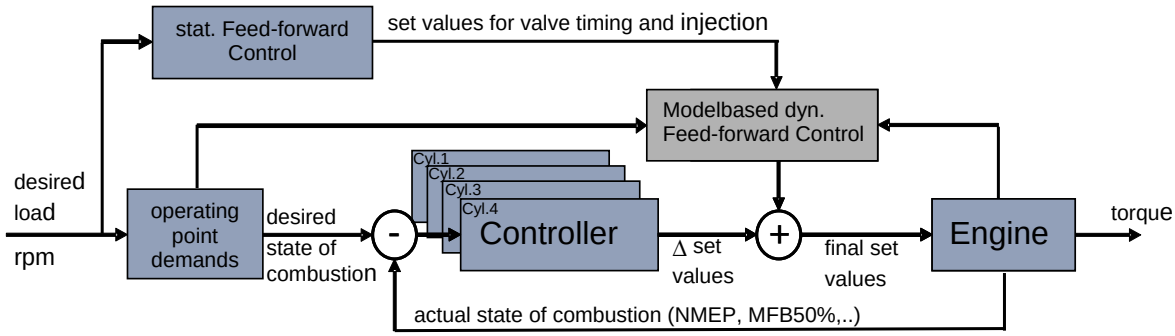


Abbildung 10: Struktur des Regelungskonzeptes mit Fokus auf der modellbasierten, prädiktiven Vorsteuerung

Die kennfeldbasierte Vorsteuerung von Einspritz- und Luftsystemstellern ist aus Ressourcengründen nicht zylinderindividuell ausgelegt, d.h. sie berücksichtigt bzgl. der betriebspunktabhängigen Stellgrößen entweder Mittelwerte über alle Zylinder oder optimal für einen spezifischen Zylinder ausgelegte Werte. Somit ist selbst im Nominalfall - dieser ist gegeben, wenn keine Umgebungs-/Kraftstoffqualitätseinflüsse, keine Aktor-Bauteiltoleranzen oder Alterungseffekte auftreten - eine Zylinderausgleichsregelung erforderlich.

7. HCCI-Verbrennungsregelung mit teilvariablen Ventiltrieb

Beim hier betrachteten teilvariablen Ventiltrieb besteht nicht die Möglichkeit, den Eingriff über das Ventilsystem auf die Verbrennung zylinderindividuell zu gestalten, denn die Verstellung der Nockenwellen über die Phasensteller wirkt auf alle Zylinder gleichermaßen. Daraus resultiert ausgehend von Abbildung 10 eine Reglerstruktur, die in Abbildung 11 dargestellt ist. Am Beispiel der Zylindergleichstellung bezüglich NMEP und MFB50% soll sie hier näher erläutert werden. Die Regelstrategie beruht auf einer Kaskaden-Regelung mit zwei hintereinander geschalteten Reglern. Ein Pri-

mär-Regler regelt den MFB50% und NMEP der einzelnen Zylinder durch den Einspritzbeginn (SOI) und die Kraftstoffmenge (q), die in den Brennraum eingespritzt wird. Dabei wird der Vorsteuerwert für den Einspritzzeitpunkt so gewählt, dass eine ausreichende Stellreserve bezüglich der beiden Stellbegrenzungen sichergestellt ist. Hier kommt zum Tragen, dass mit zunehmender Verschiebung der Einspritzung in Richtung Ende der Restgaskompression aufgrund des in Abschnitt 2 grob skizzierten Wirkungsmechanismus die Steuerbarkeit der Verbrennungslage über SOI abnimmt. Bei einer Verschiebung von SOI nach früh über GTC hinaus müssen zwei Fälle betrachtet werden. Kommt es im ersten Fall zu keiner Wärmefreisetzung im GTC, so bleibt der in Abschnitt 2 skizzierte Wirkmechanismus bestehen. Liegt aber im Brennraum ein Ladungszustand vor, der zu einer Wärmefreisetzung bei GTC führt, so drehen sich die Verhältnisse um. Daher wird der SOI bei der hier zugrunde liegenden Einspritzstrategie auf das Intervall [GTC, Ende Restgaskompression] begrenzt. Es wird angestrebt, dass der Vorsteuerwert von SOI annähernd in der Mitte dieses Intervalls liegt.

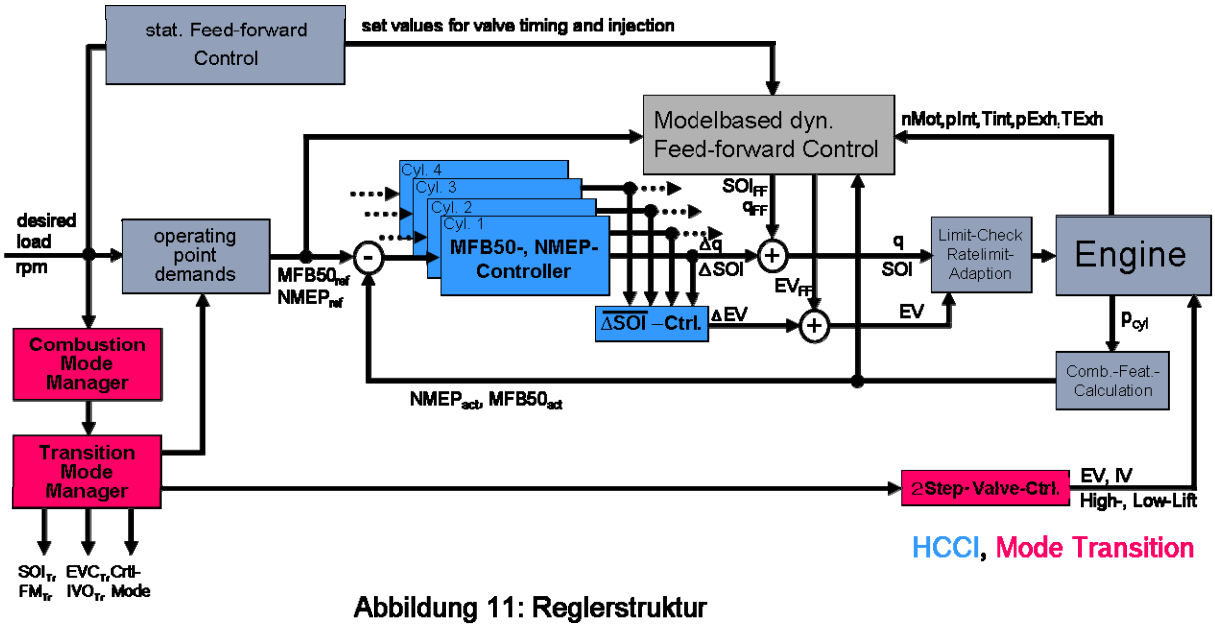


Abbildung 11: Reglerstruktur

Damit bei wechselnden Umgebungsbedingungen die Stellreserve erhalten bleibt, wird der zweite Stelleingriff, die Restgasmenge, verwendet. Dieses geschieht in der Art, dass mit einem Sekundär-Regler der Mittelwert der Einspritzzeitpunkte der einzelnen Zylinder auf dem gemeinsamen Vorsteuerwert gehalten wird. Dies

geschieht durch eine Verstellung von EVC über die Auslassnockenwelle, welches damit über die Änderung des Restgasanteils im Brennraum die Verbrennungslage aller Zylinder beeinflusst. Die Änderung des Einspritzzeitpunktes kann von Zyklus zu Zyklus erfolgen, was eine schnelle Reglerauslegung erlaubt. Damit dient der Primär-Regler neben der Zylindergleichstellung auch zur schnellen Regelung der Verbrennungslage im dynamischen Betrieb. Der Sekundär-Regler dient in erster Linie zur Kompensation von Motortemperaturänderungen beim Motor-Warmlauf oder von sich langsam ändernden Umweltbedingungen wie Luftdruck oder Außentemperatur. Beide Regelungen sind im Kern mit PID-Reglern bestückt. Die zuvor genannten Begrenzungen der Wirkmechanismen werden durch Stellgrößenbeschränkungen der Reglerausgänge berücksichtigt, die durch entsprechende Anti-Rest-Windup-Maßnahmen begleitet werden.

8. HCCI-Verbrennungsregelung – Ergebnisse am Versuchsmotor

Die HCCI-Closed-Loop-Regelung wurde in der zuvor beschriebenen Struktur an einem 4-Zylinder-Versuchsmotor am Prüfstand in Betrieb genommen. An dieser Stelle soll exemplarisch auf die erreichten Ergebnisse eingegangen werden. Dabei liegt der Fokus auf der Zylindergleichstellung bezüglich NMEP und MFB50% und der Betriebsartenumschaltung zwischen HCCI- und SI-Mode, sowohl im stationären wie auch im dynamischen Fall.

Die statische Vorsteuerung verwendet für den Luft- und Kraftstoffpfad gemeinsam für alle Zylinder nur einen Satz Vorsteuerwerte. Dieses hat zur Folge, dass sich aufgrund des unterschiedlichen thermodynamischen Verhaltens der einzelnen Zylinder eine Spreizung der Verbrennungsmerkmale NMEP und MFB50% bezüglich der einzelnen Zylinder ergibt. Die zylinderindividuelle Regelung von NMEP und MFB50% führt nach ihrer Aktivierung zu einer Gleichstellung der Zylinder, siehe hierzu Abbildung 12. Deutlich ist zu erkennen, dass durch die Gleichstellung des MFB50% auf ca. 8° nach Zünd-OT eine Reduzierung des maximalen Druckgradienten dp/dphi erreicht wird, welches sich deutlich in der Geräuschentwicklung bemerkbar macht.

Letztendlich können durch den Einsatz der zylinderindividuellen Regelung alle Zylinder in einem optimalen Arbeitspunkt betrieben werden, welches allein durch die

statische Vorsteuerung nicht möglich ist. Hier bestimmt der „schlechteste“ Zylinder den jeweiligen Arbeitspunkt der anderen.

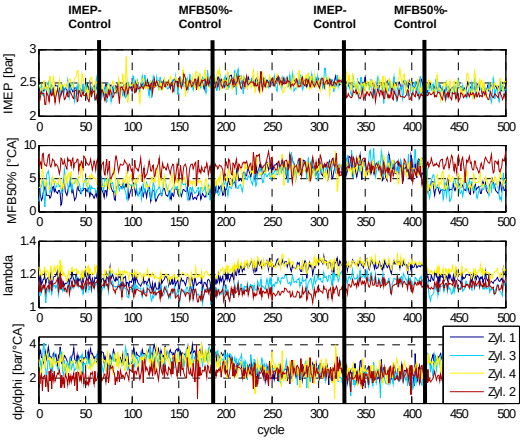


Abbildung 12: Zylindergleichstellung

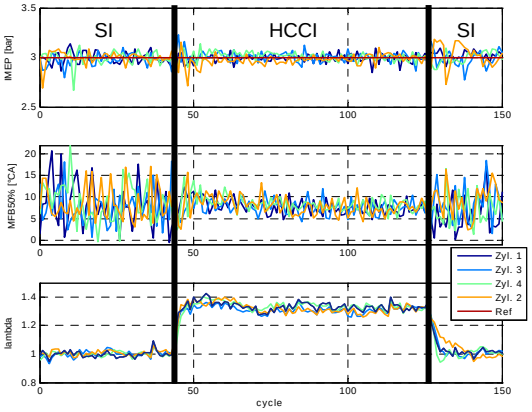


Abbildung 13: Stationärer Modewechsel

Der Betriebsbereich der HCCI-Verbrennung ist auf Grund der thermodynamischen Eigenschaften begrenzt. Besonders zu hohen Lasten ist eine Umschaltung des Verbrennungsmodus hin zum SI-Mode erforderlich. Abbildung 13 zeigt diesen Betriebsartenwechsel zwischen SI- und HCCI-Mode in dem stationären Arbeitspunkt 3 bar NMEP und einer Motordrehzahl von 2000 U/min. Deutlich sind die gegenüber dem SI-Mode stabilere Verbrennungslage und das erhöhte Lambda im HCCI-Mode zu erkennen.

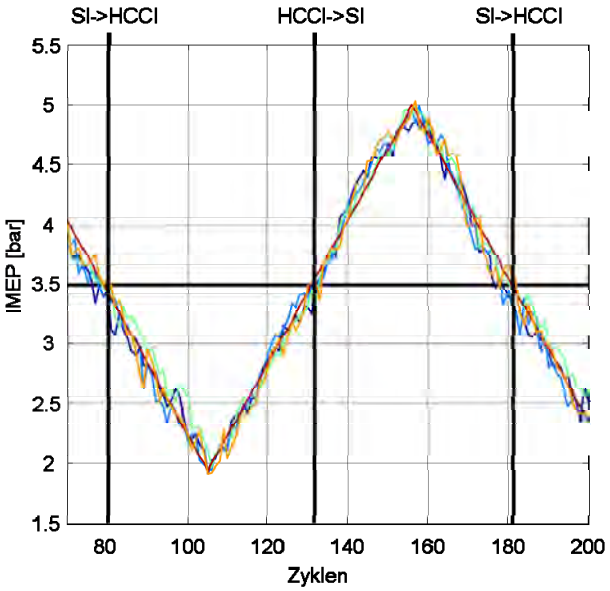


Abbildung 14: Lastrampe mit HCCI- / SI-Modeumschaltung

Einen Modewechsel im dynamischen Betrieb bei einer Motordrehzahl von 2000 U/min und rampenförmiger Erhöhung der Last von 2 bar auf 5 bar NMEP zeigt Abbildung 14. Der implementierte Betriebsartenkoordinator, vergleiche Abbildung 11, fordert bei 3,5 bar NMEP im Fall einer weiteren Lasterhöhung einen Modewechsel vom HCCI- in den SI-Mode an. Synchronisiert wird dieser Übergang vom Transitionsmodul,

welches die Reglerablösung und eine dynamische Korrektur steuert. Die dynamische Korrektur ist erforderlich, da von einem entdrosselten mageren Brennverfahren mit λ = 1.3 in ein gedrosseltes Brennverfahren mit λ = 1 umgeschaltet wird, und hierbei die Saugrohrdynamik eine wichtige Rolle spielt. Entsprechendes gilt im Fall der abfallenden Rampe für die Umschaltung vom SI- in den HCCI-Mode. Die Notwendigkeit auch hier eine dynamische Korrektur vornehmen zu müssen soll kurz erläutert werden. Im Vergleich zum HCCI-Mode stellen sich im SI-Mode aufgrund thermodynamischer Effekte wesentlich höhere Abgastemperaturen ein. Würde man eine im HCCI-Mode für diesen Arbeitspunkt erforderliche Restgasmenge direkt nach der Umschaltung vom SI-Mode einstellen, so kommt es aufgrund der wesentlich höheren Abgastemperatur der Restgasmenge zu weit nach früh verschobenen Verbrennungen. Bedingt durch die Zyklenkopplung kann es dann in den nachfolgenden Zyklen zu irregulären Verbrennungen bis hin zu Verbrennungsaussetzern kommen. Daher ist in den ersten Zyklen nach der Modeumschaltung die Restgasmenge zu verringern, um sie dann sukzessiv auf den für diesen Arbeitspunkt erforderlichen Wert zu bringen.

9. Zusammenfassung und Ausblick

Brennverfahren wie HCCI zeigen ein deutliches Potenzial zur Kraftstoffeinsparung bei Beibehaltung der bewährten Abgasnachbehandlung mittels 3-Wege Katalysator. Die Anforderungen an ein Ventilsystem zur Füllungssteuerung können mit einem relativ einfachen nockenwellengestützten System mit Phasenstellern und Ventilhubkontur-Umschaltung erfüllt werden. Aufgrund der Selbstzündung stellt das Brennverfahren hohe Anforderungen an die Motorsteuerung, speziell zur Beherrschung des dynamischen Motorbetriebs. Dies erfordert eine sehr genaue Rückmeldung über den Ablauf der Verbrennung. Basierend auf einer Rückmeldung mittels Brennraumdrucksensor wurde ein Konzept zur Steuerung und Regelung eines 4-Zylinder-Motors mit der Betriebsart HCCI dargestellt. Das für diese Aufgabe entworfene Rapid-Prototyping-System und die dafür entwickelte Basissoftware ermöglichen eine schnelle Umsetzung der geforderten Steuer- und Regelfunktionalität. Speziell

entwickelte und implementierte Funktionen zur Identifikation des Systemverhaltens bilden dabei die Basis zur Erlangung eines detaillierten Systemverständnisses.
Aufbauend auf der Systemidentifikation wurden Modelle entwickelt, die in Abhängigkeit des aktuellen Verbrennungsablaufes und den aktuell vorgegebenen Stellgrößen der Aktoren den Ablauf der folgenden Verbrennung prädizieren können. Dies stellt die Basis für eine dynamische Vorsteuerung dar, die insbesondere in der Last- und Drehzahldynamik sowie beim Betriebsartenwechsel benötigt wird. Ausgehend von einem teilvariablen Ventiltrieb wurde ein Regelungskonzept für die Verbrennungslage und Last präsentiert, mit dem die gewünschte Verbrennung auch bei sich ändernden Umgebungsbedingungen und Driften bzw. Toleranzen in der Aktorik sichergestellt wird.
Das präsentierte Steuer- und Regelkonzept bietet die Funktionalität für einen stabilen Motorbetrieb im relevanten HCCI-Bereich. Dieses beinhaltet sowohl den stationären und dynamischen HCCI-Betrieb als auch Umschaltvorgänge zwischen den verschiedenen Betriebsarten.
Weitere Arbeiten sind nötig. Sie zielen speziell auf die Erweiterung des HCCI-Bereichs und auf eine geräuschminimale Modeumschaltung. Daneben stehen auch Optimierungen bezüglich der Aktorik, der Sensorik und der Steuer- und Regelalgorithmen auf dem Programm. Letztendlich sei aber darauf hingewiesen, dass eine endgültige Bewertung des HCCI-Brennverfahrens nur in einem Fahrzeug erfolgen kann, wobei speziell die Fahrbarkeit und das Geräuschverhalten im Fokus stehen.

10. Verwendete Abkürzungen

HCCI:	**H**omogeneous **C**harge **C**ompression **I**gnition
MFB50%:	**M**ass **F**raction **B**urned 50%
NMEP:	**N**et **M**ean **E**ffective **P**ressure
SI:	**S**parc **I**gnition
EVC:	**E**xhaust **V**alve **C**losing
IVO:	**I**ntake **V**alve **O**pening
IVC:	**I**ntake **V**alve **C**losing
GTC:	**G**as Flow **T**op Dead **C**enter
SOC:	**S**tart **O**f **C**ombustion
SOI:	**S**tart **O**f **I**njection
NEDC:	**N**ew **E**uropean **D**riving **C**ycle

11. Literatur

[1] J.B. Heywood, Internal Combustion Engine Fundamentals, McGraw-Hill, 1988

[2] C. Sauer, A. Kulzer, M. Rauscher, J.-P. Hathout and M. Bargende, *Strategies for a CAI Gasoline Engine derived from Experiments and Simulation* ", in 7th International Stuttgart Symposium 2007

[3] R. Karrelmeyer, J. Häring, W. Fischer and J.-P. Hathout, "Closed-Loop Control of a 1-Cylinder Gasoline HCCI-Engine in Dynamic Operation", in "New Trends in Engine Control, Simulation and Modelling", 2006 IFP Paris

Modellierung einer Biomasse-Kleinfeuerungsanlage als Grundlage für modellbasierte Regelungsstrategien

Stefan Reiter[1,2,*], Markus Gölles[1], Thomas Brunner[1,3,4], Nicolaos Dourdoumas[2], Ingwald Obernberger[1,3,4]

[1] BIOENERGY 2020+ GmbH, Inffeldgasse 21b, 8010 Graz. E-mail: stefan.reiter@bioenergy2020.eu, markus.goelles@bioenergy2020.eu, thomas.brunner@bioenergy2020.eu

[2] Technische Universität Graz, Institut für Regelungs- und Automatisierungstechnik, Kopernikusgasse 24/I, 8010 Graz. E-mail: nicolaos.dourdoumas@tugraz.at

[3] Technische Universität Graz, Institut für Prozess- und Partikeltechnik, Inffeldgasse 21a, 8010 Graz E-mail: ingwald.obernberger@tugraz.at

[4] BIOS BIOENERGIESYSTEME GmbH, Inffeldgasse 21b, 8010 Graz

[*] Korrespondierender Autor

Zusammenfassung

Bei geeigneter Vorgabe der Einflussgrößen kann in modernen automatisch betriebenen Biomasse-Kleinfeuerungsanlagen ein emissionsarmer Feuerungsbetrieb bei gleichzeitig hohen Wirkungsgraden erzielt werden. Aufgrund des Verbesserungsbedarfs bei der Anlagenregelung kann dieses Potential jedoch noch nicht ausgeschöpft werden. Die Anwendung moderner Reglerentwurfsverfahren scheitert derzeit an der Verfügbarkeit adäquater mathematischer Modelle. In diesem Beitrag werden Modelle anhand einer marktverfügbaren automatisch betriebenen Biomasse-Kleinfeuerungsanlage (Hackgut befeuerter Kessel) für die relevanten Anlagenbereiche Brennstoffbett, Primär- und Sekundärverbrennungszone, Wärmeübertrager sowie Luft- und Brennstoffzufuhr hergeleitet. Bei der Verifikation zeigte sich eine sehr gute Abbildung der physikalischen Verhältnisse durch diese Modelle. Somit stellen sie eine geeignete Grundlage zur modellbasierten Regelung dieser Kleinfeuerungsanlage dar.

1 Einleitung

Um die immer strenger werdenden gesetzlichen Emissionsgrenzwerte für Biomasse-Kleinfeuerungsanlagen zu erfüllen, sind Hersteller gefordert, ihre Anlagen weiter zu entwickeln. Obwohl es in den letzten beiden Jahrzehnten einen deutlichen Entwicklungssprung gegeben hat, ist das volle Potential für einen emissionsarmen Feuerungsbetrieb bei hohen feuerungstechnischen Wirkungsgraden noch nicht ausgeschöpft. Dies gilt für moderne Biomasse-Feuerungsanlagen mit optimierten Feuerraumgeometrien, Verbrennungsluftführungs- und Luftstufungsstrategien, aufgrund des immer noch bestehenden Verbesserungsbedarfs bei der Anlagenregelung. Derzeit gibt es keine einheitliche Vorgehensweise zur Regelung von Biomasse-Kleinfeuerungsanlagen und es werden von den Herstellern zum Teil sehr unterschiedliche Strategien verfolgt. Sie basieren jedoch alle auf entkoppelten linearen Regelkreisen (meist PID-Regler). Diese berücksichtigen die Verkopplungen und teilweise stark nichtlinearen Zusammenhänge der einzelnen Prozessgrößen nur bedingt oder gar nicht. Das führt insbesondere bei Lastwechseln zu einem Abfall des Wirkungsgrades und einem Ansteigen der Emissionen. So endet beispielsweise der Wechsel von Volllast- zu Teillastbetrieb sehr häufig in einem „Stop&Go-Betrieb“ während der Teillastphasen.
Die Anwendung moderner Reglerentwurfsverfahren scheitert bis jetzt an geeigneten mathematischen Modellen. Für mittelgroße Biomasse-Feuerungsanlagen wurden entsprechende Modelle entwickelt und erfolgreich zum Entwurf einer modellbasierten Regelung verwendet. Bei deren experimenteller Verifikation zeigte sich eine deutliche Verbesserung des Betriebsverhaltens im Vergleich zur bisherigen Regelung [6].
Ziel dieser Arbeit ist, geeignete Modelle als Grundlage für einen Reglerentwurf für automatisch betriebene Biomasse-Kleinfeuerungsanlagen zu entwickeln. Im Mittelpunkt der Untersuchungen stehen Anlagen bis zu einem Leistungsbereich von 30 kW, wie sie üblicherweise zur Raumwärmeerzeugung in Ein- und Mehrfamilienhäusern eingesetzt werden. Die wesentlichen Unterschiede zu mittelgroßen Anlagen liegen in den verschiedenen Rosttechnologien (Flachrost statt Vorschubrost), den in Kleinanlagen zumeist gekühlten Feuerraumwänden, dem Fehlen einer Rauchgasrezirkulation sowie der unterschiedlich realisierten Primär- und Sekundärluftzufuhr (meist mit Klappen, jedoch ohne eigene Ventilatoren). Die Modellbildung wird anhand einer marktverfügbaren Biomasse-Kleinfeuerungsanlage zur Verfeuerung von Hackgut mit einer Nennleistung von 30 kW durchgeführt.

2 Anlagenbeschreibung

Die in Abbildung 1 dargestellte Anlage verfügt über einen kippbaren Rost sowie eine automatische Brennstoffzufuhr mit einer Förderschnecke, durch die das Hackgut von der Seite auf den Rost geschoben wird. Sie besitzt weiters eine gestufte Luftzufuhr, um einen emissionsarmen Betrieb zu erzielen. Dabei wird nur ein Teil der zur vollständigen Verbrennung nötigen Luft unter dem Rost als sogenannte Primärluft zugeführt. Die restliche Verbrennungsluft, die sogenannte Sekundärluft, wird über dem Rost mit einer möglichst hohen Geschwindigkeit durch die Sekundärluftdüsen eingebracht, um eine gute Durch-

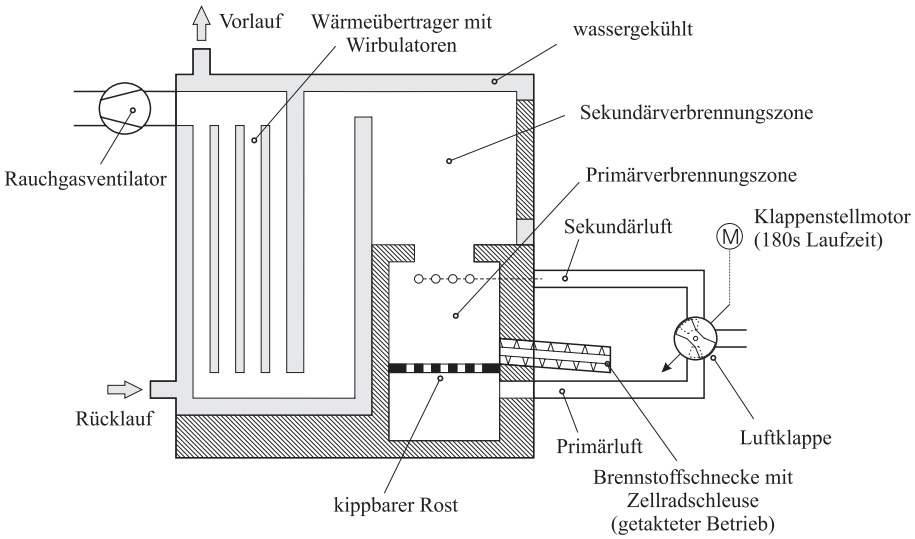


Abbildung 1 – Schematische Darstellung der Versuchsanlage (Hackgut-Rostfeuerung mit Warmwasserkessel)

mischung und somit eine emissionsarme Verbrennung zu erreichen. Die Variation der zugeführten Primär- und Sekundärluft wird über eine gemeinsame Luftklappe, sowie der Variation des Unterdruckes in der Feuerung über die Drehzahl des Rauchgasventilators erzielt. Durch den in der gesamten Anlage vorherrschenden Unterdruck wird sichergestellt, dass das Rauchgas die Feuerung nur durch den Kamin verlassen kann, wobei der Unterdruck ein zusätzliches Ansaugen von Umgebungsluft an undichten Stellen der Anlage bewirkt. Diese Luftströme werden als Falschluft bezeichnet und müssen, je nach Größenordnung und Relevanz für die Verbrennung, bei der Modellierung berücksichtigt werden. Um die bei der Verbrennung freigesetzte Energie nutzbar zu machen, wird diese über gekühlte Wände im Feuerraum sowie über einen Wärmeübertrager vom Rauchgas auf den Heizwasserkreis übertragen. Das abgekühlte Rauchgas verlässt die Anlage über den Kamin mit einer Temperatur, die typischerweise zwischen 80°C und 130°C liegt.

3 Modellbildung

Zur übersichtlichen Modellbildung erfolgt eine Unterteilung in drei relevante Anlagenbereiche: Brennstoffbett, Primär- bzw. Sekundärverbrennungszone und Wärmeübertrager. Letzterer beinhaltet sowohl die gekühlten Wände der Sekundärverbrennungszone als auch den eigentlichen Rauchrohr-Wärmeübertrager (siehe Abbildung 2). Zusätzlich ist die Entwicklung geeigneter Modelle zur Beschreibung der Luftzufuhr, des Falschlufteintrages, der Brennstoffzufuhr sowie des vom Rauchgasventilator erzeugten Unterdrucks im Feuerraum erforderlich. Diese Modelle sind zur gezielten Zufuhr der benötigten Stoffströme erforderlich. Eine besondere Herausforderung ist die Modellierung der zugeführten Primär- und Sekundärluft, da durch die gemeinsame Luftklappe eine starke Verkopplung besteht. In den folgenden Abschnitten erfolgt eine detaillierte Erläuterung der Modelle.

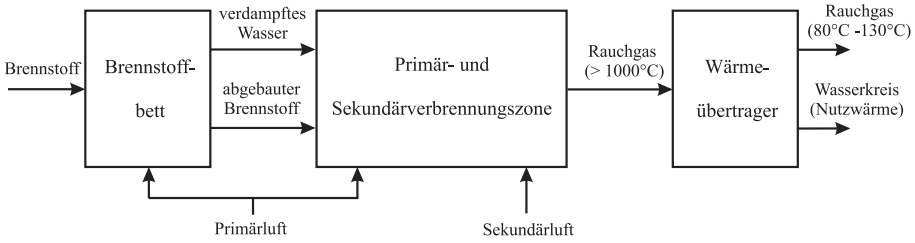


Abbildung 2 – Modellstruktur einer Biomasse-Kleinfeuerungsanlage

3.1 Brennstoffbett

Das Brennstoffbett stellt den wichtigsten Bereich einer Biomassefeuerungsanlage dar. Der in den Feuerraum eingebrachte Brennstoff wird aufgrund der vorherrschenden hohen Temperaturen erwärmt, was zum Verdampfen des im Brennstoff gebundenen Wassers (Trocknung) führt. Durch die weitere Erwärmung werden die flüchtigen Komponenten in die Gasphase freigesetzt, und die verbleibende Holzkohle wird mit der Primärluft verbrannt. Für das Abbrandverhalten von Biomassefeuerungsanlagen auf einem Rost existieren grundsätzlich zwei Ansätze: die Verbrennung von oben nach unten (Gegenstromverbrennung) und die Verbrennung von unten nach oben (Gleichstromverbrennung) [4, 7, 8]. In [4] wird gezeigt, dass typisches Hackgut mit einem Wassergehalt von 21 Gew.% auf dem Flachschubrost einer mittelgroßen Biomassefeuerungsanlage (Nennleistung 180 kW) eindeutig von unten nach oben abbrennt. Darüber hinaus wird ein einfaches mathematisches Modell vorgestellt, welches das Abbrandverhalten für einen Reglerentwurf hinreichend genau beschreibt. Bei der Gleichstromverbrennung wird davon ausgegangen, dass die zum Zünden des Brennstoffes an der Unterseite des Brenstoffbettes erforderliche Energie über den thermisch gut leitenden Rost aus dem Bereich, in dem die Verbrennung der Holzkohle stattfindet, bereitgestellt wird. Das in Kleinanlagen verwendete Hackgut ähnelt dem in [4] verwendeten, und der aus einem einzigen Element bestehende Rost verfügt über eine bessere Leitfähigkeit als ein aus einzelnen Elementen bestehender Flachschubrost. Aus diesem Grund wird für die betrachtete Anlage von einem Abbrand von unten nach oben ausgegangen und das in [4] vorgeschlagene Modell zur Beschreibung des Brennstoffbetts verwendet.

Dabei wird das Brennstoffbett in drei Zonen unterteilt. Der über die Brennstoffschnecke zugeführte, feuchte Brennstoffmassenstrom durchläuft zunächst eine sogenannte Totzone, in der er lediglich erwärmt wird. Danach erfolgt die Verdampfung des im Brennstoff gebundenen Wassers in der Verdampfungszone, und der, die Freisetzung der flüchtigen Komponenten sowie die Verbrennung der auf dem Rost verbleibenden Holzkohle beinhaltende, Abbau des trockenen Brennstoffs in der Abbauzone. Die mathematische Beschreibung der Verdampfung des Wassers sowie des Abbaus des trockenen Brennstoffes erfolgt dabei mit Hilfe je einer Massenbilanz für das Wasser in der Verdampfungszone $m_{\mathrm{W,Z}}$ und

den trockenen Brennstoff in der Abbauzone $m_{\mathrm{B,Z}}$

$$\frac{dm_{\mathrm{W,Z}}(t)}{dt} = -\underbrace{c_{\mathrm{Verd}} m_{\mathrm{W,Z}}(t)}_{\text{verdampftes Wasser}} + \underbrace{\frac{dm_{\mathrm{W}}(t - T_{\mathrm{t}}(t))}{dt}}_{\text{zugeführtes Wasser}} \tag{1}$$

$$\frac{dm_{\mathrm{B,Z}}(t)}{dt} = -\underbrace{c_{\mathrm{Abb}} m_{\mathrm{B,Z}}(t) [\dot{m}_{\mathrm{PL}}(t) + \dot{m}_{\mathrm{PL0}}]}_{\text{abgebauter trockener Brennstoff}} + \underbrace{\frac{dm_{\mathrm{B}}(t - T_{\mathrm{t}}(t))}{dt}}_{\text{zugeführter trockener Brennstoff}} . \tag{2}$$

Hierbei sind c_{Verd}, c_{Abb} und $\dot{m}_{\mathrm{PL0}}$ konstante Parameter, $\dot{m}_{\mathrm{PL}}$ der Primärluftmassenstrom, m_{W} die um eine Totzeit T_{t} verzögerte kumulierte Wassermasse und m_{B} der um eine Totzeit verzögerte kumulierte trockenen Brennstoff. Die kumulierte Wassermasse m_{W} und der kumulierte trockene Brennstoff m_{B} wurden zur mathematischen Vereinfachung eingeführt. Sie werden aus dem der Anlage zugeführten Massenstrom des im Brennstoff enthaltenen Wassers $\dot{m}_{\mathrm{W}}$, sowie dem zugeführten Massenstrom an trockenem Brennstoff $\dot{m}_{\mathrm{B}}$ berechnet:

$$m_{\mathrm{W}}(t) = \int_0^t \dot{m}_{\mathrm{W}}(\tau)\, \mathrm{d}\tau \tag{3}$$

$$m_{\mathrm{B}}(t) = \int_0^t \dot{m}_{\mathrm{B}}(\tau)\, \mathrm{d}\tau \tag{4}$$

Die Totzeit T_{t} hängt im Wesentlichen vom Wassergehalt des Brennstoffes und somit von der Wassermasse in der Verdampfungszone $m_{\mathrm{W,Z}}$ sowie dem Massenstrom an trockenem Brennstoff $\dot{m}_{\mathrm{B}}$ ab. Dabei führt ein höherer Wassergehalt des Brennstoffes zu einer größeren Totzeit, da die über den Rost transportierte Wärme aufgrund der größeren Wassermasse in der Verdampfungszone früher verbraucht wird. Eine Erhöhung des zugeführten trockenen Brennstoffes bewirkt hingegen eine Verringerung der Totzeit, da diese Zone schneller durchlaufen wird. Die Totzeit wird durch die Gleichung

$$T_{\mathrm{t}}(t) = c_{\mathrm{T,B}} \frac{m_{\mathrm{W,Z}}(t)}{\dot{m}_{\mathrm{B}}(t)} \tag{5}$$

mit dem konstanten Parameter $c_{\mathrm{T,B}}$ hinreichend genau beschrieben. Die Modellparameter c_{Verd}, c_{Abb}, $c_{\mathrm{T,B}}$ und $\dot{m}_{\mathrm{PL0}}$ wurden unter Minimierung eines quadratischen Gütekriteriums aus experimentell ermittelten Messdaten bestimmt. Hierbei wurden der Brennstoffmassenstrom $\dot{m}_{\mathrm{B}}$, der zugeführte Primärluftmassenstrom $\dot{m}_{\mathrm{PL}}$ und der Brennstoffwassergehalt sprungförmig variiert. Beispielhaft für die durchgeführten Versuche zeigt Abbildung 3 eine Gegenüberstellung der gemessenen und der mit dem Modell berechneten Massenströme an abgebautem trockenen Brennstoff sowie an verdampftem Wasser. Hierbei wurde der Brennstoffwassergehalt zu den Zeitpunkten $t = 11$ min bzw. $t = 49$ min für jeweils eine Minute von 46 Gew.% auf 66 Gew.% sowie auf 0 Gew.% sprungförmig verändert. Dabei blieb der Massenstrom des zugeführten trockenen Brennstoffes konstant. Die im eingeschwungenen Zustand auftretenden starken Schwankungen der gemessenen Werte resultieren aus der verhältnismäßig geringen Brennstoffmasse auf dem Rost im Vergleich zur Variation der Stückigkeit des Brennstoffes. Da die Schwankungen der Stückigkeit

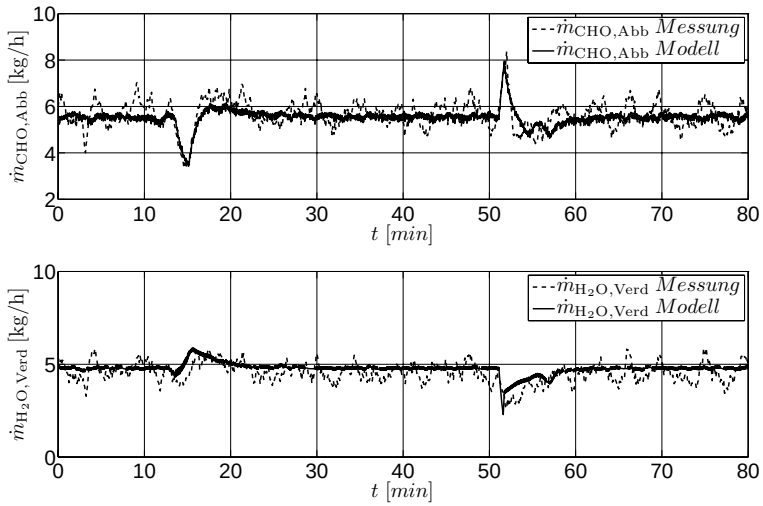


Abbildung 3 – Abbau- und Verdampfungsrate bei sprungförmigen Änderungen des Brennstoffwassergehaltes

nicht bekannt sind, können diese nicht durch das Modell abgebildet werden. Das Modell beschreibt das Brennstoffbett jedoch hinreichend genau, um einen modellbasierten Reglerentwurf durchzuführen.

3.2 Primär- und Sekundärverbrennungszone

In der gemeinsam betrachteten Primär- und Sekundärverbrennungszone erfolgt die Verbrennung des abgebauten, trockenen Brennstoffes (in Form von brennbarem Rauchgas) mit der zugeführten Primär- und Sekundärluft. Zusammensetzung und Menge des aus der Verbrennung resultierenden Rauchgases sowie dessen Temperatur bei vollständiger Verbrennung ohne Wärmeübertragung an die Umgebung (sogenannte adiabate Verbrennungstemperatur) werden mit Hilfe einer gewöhnlichen Verbrennungsrechnung [1, 5] berechnet.
In mittelgroßen Biomassefeuerungsanlagen (400 kW bis 10.000 kW) kommt es bei Laständerungen und insbesondere beim Hochfahren der Anlage zu einer deutlichen Abweichung zwischen adiabater Verbrennungstemperatur und Rauchgastemperatur am Wärmeübertragereintritt. Deren Ursache liegt in der Speicherwirkung der feuerfesten Auskleidung von Primär- und Sekundärverbrennungszone [1, 5]. Bei Kleinfeuerungsanlagen ist die auf die Kesselleistung bezogene Masse der feuerfesten Auskleidung jedoch um das Zehnbis Zwanzigfache geringer, womit eine deutlich geringere Speicherwirkung zu erwarten ist und deshalb vernachlässigt wird. Weiters bewegen sich die Umgebungsverluste von modernen Biomasse-Kleinfeuerungsanlagen lediglich im niedrigen einstelligen Prozentbereich der Kesselnennleistung und können deshalb bei regelungstechnischen Fragestellungen vernachlässigt werden. Diese beiden Feststellungen werden dadurch untermauert, dass die mit Hilfe der adiabaten Feuerraumtemperatur, des Rauchgasmassenstromes und der Rauchgastemperatur am Austritt aus dem Wärmeübertrager berechnete, rauchgasseitige Enthalpiestromdifferenz bei den durchgeführten Versuchen gut mit der gemessenen

wasserseitigen Leistung übereinstimmte.
Letztendlich können sowohl die Speicherwirkung der feuerfesten Auskleidung der Verbrennungszone als auch die Umgebungsverluste vernachlässigt werden. Somit geht man für die folgende Modellierung des Wärmeübertragers davon aus, dass nach der Verbrennung Rauchgas mit der adiabaten Verbrennungstemperatur zur Verfügung steht.

3.3 Wärmeübertrager

Die Aufgabe des Wärmeübertragers besteht darin, die bei der Verbrennung freigesetzte Wärme vom Rauchgas auf den Wasserkreis zu übertragen. Dies geschieht bereits in der gekühlten Sekundärverbrennungszone und in weiterer Folge im eigentlichen Rauchrohr-Wärmeübertrager. Da die Rückwirkung der Wasserseite aufgrund der vergleichsweise geringen Temperaturspreizung (~ 20 °C) im Vergleich zur Spreizung auf der Rauchgasseite (~ 800 °C) vernachlässigt werden kann, wird die Modellierung des Wärmeübertragers für den Rauchgas- und den Wasserteil getrennt durchgeführt. In [3] wird ein Ansatz zur Modellierung eines Rauchrohrwärmeübertragers vorgestellt, der auch für Biomasse-Kleinfeuerungsanlagen geeignet ist. Dabei wird der Rauchgasteil durch ein statisches Modell der Form

$$\dot{Q} = c_{\mathrm{WT}} \left[T_{\mathrm{RG}} - T_{\mathrm{W}}\right]^{q_1} \dot{m}_{\mathrm{RG}}^{q_2} \tag{6}$$

beschrieben. Hierbei entsprechen $\dot{Q}$ dem auf das Wasser übertragenen Wärmestrom, T_{RG} der Rauchgastemperatur beim Eintritt in den Wärmeübertrager, T_{W} der konstant angenommenen mittleren Wassertemperatur, $\dot{m}_{\mathrm{RG}}$ dem Rauchgasmassenstrom und c_{WT}, q_1 und q_2 den experimentell zu ermittelnden Modellparametern. Wie in Abschnitt 3.2 erwähnt, können die Speicherwirkung der feuerfesten Auskleidung der Primärzone und auch die Umgebungsverluste vernachlässigt werden. Somit wird die Rauchgaseintrittstemperatur gleich der adiabaten Verbrennungstemperatur

$$T_{\mathrm{RG}} = T_{\mathrm{ad}} \tag{7}$$

gesetzt. Die Wasserseite wird durch die gewöhnliche Differentialgleichung erster Ordnung

$$\frac{dT_{\mathrm{VL}}}{dt} = -\frac{\dot{m}_{\mathrm{W}}}{c_{\tau,\mathrm{W}}} T_{\mathrm{VL}} + \frac{\dot{m}_{\mathrm{W}}}{c_{\tau,\mathrm{W}}} \left[\frac{\dot{Q}}{\dot{m}_{\mathrm{W}} c_{\mathrm{W}}} + T_{\mathrm{RL}} \left(t - \frac{c_{\mathrm{T,W}}}{\dot{m}_{\mathrm{W}}}\right)\right] \tag{8}$$

beschrieben. Hierbei entsprechen T_{VL} der Vorlauftemperatur, $\dot{m}_{\mathrm{W}}$ dem Wassermassenstrom, c_{W} der spezifischen Wärmekapazitat von Wasser, T_{RL} der Rücklauftemperatur, $c_{\tau,\mathrm{W}}$ bzw. $c_{\mathrm{T,W}}$ den experimentell zu ermittelnden Modellparametern. Bei der in Abbildung 4 dargestellten experimentellen Verifikation des Modells wurden abwechselnd sprungförmige Änderungen der Rücklauftemperatur sowie des Wassermassenstroms vorgenommen. Hierbei zeigt sich auch bei großen Sprüngen eine sehr gute Übereinstimmung zwischen der gemessenen und der mit dem Modell berechneten Vorlauftemperatur.

3.4 Luft- und Brennstoffzufuhr

Um die bei einer modellbasierten Regelung benötigten Stoffströme gezielt zuführen zu können, ist der Entwurf weiterer Modelle zur Beschreibung der Luft- und Brennstoffzu-

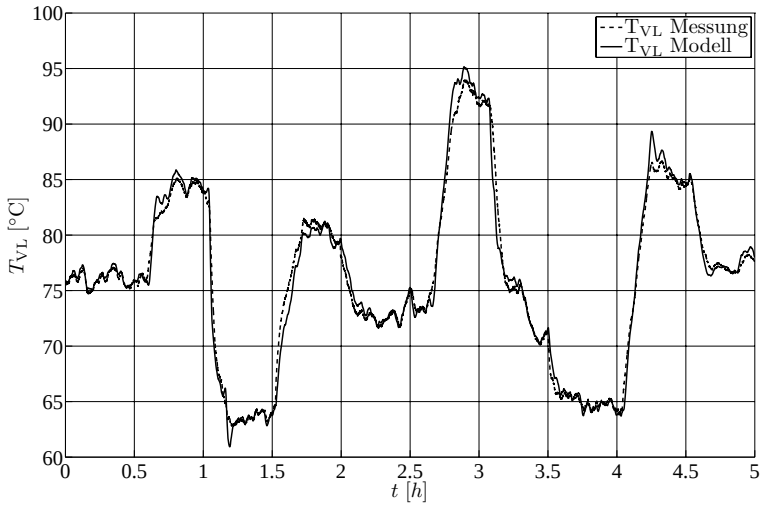


Abbildung 4 – Vergleich von gemessener und berechneter Vorlauftemperatur

fuhr erforderlich. Dazu gehören der Zusammenhang zwischen der Ansteuerung der Brennstoffzufuhr und dem daraus resultierenden Brennstoffmassenstrom, die Beschreibung der eigentlichen Luftzufuhr, die Abhängigkeit des Falschlufteintrages vom Feuerraumunterdruck sowie die Beschreibung der Druckanhebung des Rauchgasventilators, welche zur Vorgabe eines definierten Feuerraumunterdruckes nötig ist. Die Modellierungen der Luftzufuhr und des Falschlufteintrages basieren auf der Berücksichtigung der für die Druck- und Volumenstromverhältnisse wesentlichen Phänomene: dem Druckabfall in einem Rohr, dem Druckabfall an einer Blende sowie der Druckanhebung durch einen Ventilator [2]. Dabei wird der Zusammenhang zwischen Druckabfall Δp und Volumenstrom $\dot{V}$ in einem Rohr durch Gleichung (9) und an einer Blende durch Gleichung (10) beschrieben:

$$\Delta p = R_1 \dot{V}^{1,75} \quad \text{mit } R_1 = 0,242 \eta^{0,25} \rho^{0,75} \frac{l}{d^{4,75}} \tag{9}$$

$$\Delta p = R_2 \dot{V}^2 \quad \text{mit } R_2 = \frac{1}{2} \frac{\rho}{\alpha^2 {A_\mathrm{d}}^2} \tag{10}$$

Hierbei entsprechen l der Länge und d dem Durchmesser des Rohres, η der dynamischen Viskosität, ρ der Dichte des Fluids, α der als konstant angenommenen Durchflusszahl und A_d der Querschnittsfläche der Blende. Die Druckanhebung eines Ventilators wird durch

$$\Delta p = c_\mathrm{V} f_\mathrm{V}^2 \tag{11}$$

beschrieben, wobei c_V der Ventilatorkonstanten und f_V der Ventilatorfrequenz entsprechen. In den folgenden Abschnitten erfolgt eine detaillierte Erläuterung der vorgestellten Teilmodelle.

3.4.1 Luftzufuhr

Die Beeinflussung der zugeführten Primär- und Sekundärluft erfolgt durch Variation des vom Rauchgasventilator erzeugten Unterdrucks (Abschnitt 3.4.3) und der Position einer

gemeinsamen Luftklappe, mit der das Öffnungsverhältnis zweier Blenden für Primär- und Sekundärluft eingestellt werden kann. Zur Modellierung werden der gemeinsame Zuluftkanal bis zur Luftklappe sowie die Primärluftzufuhr als Druckabfall an einer Blende und die Sekundärluftzufuhr als eine Serienschaltung eines Rohres und einer Blende betrachtet (siehe Abbildung 5).

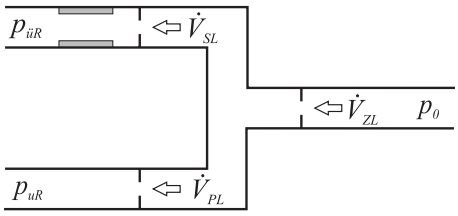


Abbildung 5 – Modellierung der Luftzufuhr

Die Zusammenhänge zwischen dem Druck unter dem Rost p_{uR} sowie über dem Rost $p_{\mathrm{üR}}$, dem Umgebungsdruck p_0, dem Primärluftvolumenstrom $\dot{V}_{\mathrm{PL}}$, dem Sekundärluftvolumenstrom $\dot{V}_{\mathrm{SL}}$ und dem Volumenstrom der gesamten Zuluft $\dot{V}_{\mathrm{ZL}}$ werden im Kaltzustand der Anlage durch

$$p_0 - p_{\mathrm{uR}} = R_{2,\mathrm{PL}}(\alpha)\dot{V}_{\mathrm{PL}}^2 + R_{2,\mathrm{ZL}}\dot{V}_{\mathrm{ZL}}^2 \quad (12)$$

$$p_0 - p_{\mathrm{üR}} = R_{2,\mathrm{SL}}(\alpha)\dot{V}_{\mathrm{SL}}^2 + R_{1,\mathrm{SL}}\dot{V}_{\mathrm{SL}}^{1.75} + R_{2,\mathrm{ZL}}\dot{V}_{\mathrm{ZL}}^2 \quad (13)$$

$$\dot{V}_{\mathrm{ZL}} = \dot{V}_{\mathrm{PL}} + \dot{V}_{\mathrm{SL}} \quad (14)$$

beschrieben. Dabei entsprechen $R_{2,\mathrm{PL}}(\alpha)$ und $R_{2,\mathrm{SL}}(\alpha)$ den von der Luftklappenstellung α abhängenden *blendigen Widerständen* für die Primär- und Sekundärluftzufuhr, $R_{1,\mathrm{SL}}$ dem als konstant angenommenen *rohrigen Widerstand* für die Sekundärluftzufuhr und $R_{2,\mathrm{ZL}}$ dem *blendigen Widerstand* für den gemeinsamen Zuluftkanal bis zur Luftklappe. Die Modellparameter $R_{2,\mathrm{PL}}(\alpha)$, $R_{1,\mathrm{SL}}$, $R_{2,\mathrm{SL}}(\alpha)$ sowie $R_{2,\mathrm{ZL}}$ wurden unter Minimierung eines quadratischen Gütekriteriums aus experimentell ermittelten Messdaten bestimmt. Da der zugeführte Primär- und Sekundärluftmassenstrom nicht getrennt voneinander gemessen werden können, war es erforderlich, die Versuche zur Parameterermittelung gestuft durchzuführen. Dabei wurden in je einem Versuch bei abgedichteter Primär- bzw. Sekundärluftöffnung sowie bei Originalkonfiguration der Feuerraumunterdruck und die Luftklappenstellung im gesamten Betriebsbereich variiert.
Abbildung 6 zeigt den gemessenen Volumenstrom $\dot{V}_{\mathrm{ZL}}$ sowie die mit dem Modell berechneten Volumenströme $\dot{V}_{\mathrm{PL}}$, $\dot{V}_{\mathrm{SL}}$ und die gesamte Zuluft $\dot{V}_{\mathrm{PL}}$+$\dot{V}_{\mathrm{SL}}$ bei der in Abbildung 7 dargestellten Variation der Drehzahl des Rauchgasventilators n_{RGV} sowie der Luftklappenstellung α. Dabei zeigt sich eine sehr gute Übereinstimmung zwischen gemessenem und berechnetem Volumenstrom der gesamten Zuluft im betrachteten Betriebsbereich. Zusätzlich liefern die dargestellten (berechneten) Volumenströme der Primär- und der Sekundärluft eine gute Abschätzung der grundsätzlichen Variationsmöglichkeiten der Luftzufuhr.
Die entwickelten Modelle ermöglichen somit eine Abschätzung der nicht getrennt messbaren zugeführten Primär- bzw. Sekundärluft und darüber hinaus eine entkoppelte, und in Folge dessen zielgerichtete Luftzufuhr.

Im Heißbetrieb der Anlage kommt es besonders in der hinter der Feuerraumauskleidung verlaufenden Sekundärluftzufuhr zu einer Erwärmung der zugeführten Luft. Dies erfordert eine Adaption der Modellparameter an die tatsächlich vorherrschenden Temperaturen im jeweiligen Anlagenabschnitt entsprechend ihrer Berechnungsvorschriften in (9) und (10). Aus Kostengründen verzichtet man dabei auf eine kontinuierliche Messung der entsprechenden Temperaturen. Stattdessen werden dem Lastzustand angepasste, konstant angenommene Temperaturen für die jeweiligen Anlagenabschnitte verwendet. Damit verbundene Modellfehler können in Kauf genommen werden, da sie durch die Verwendung integrierender Regler ausgeglichen werden.

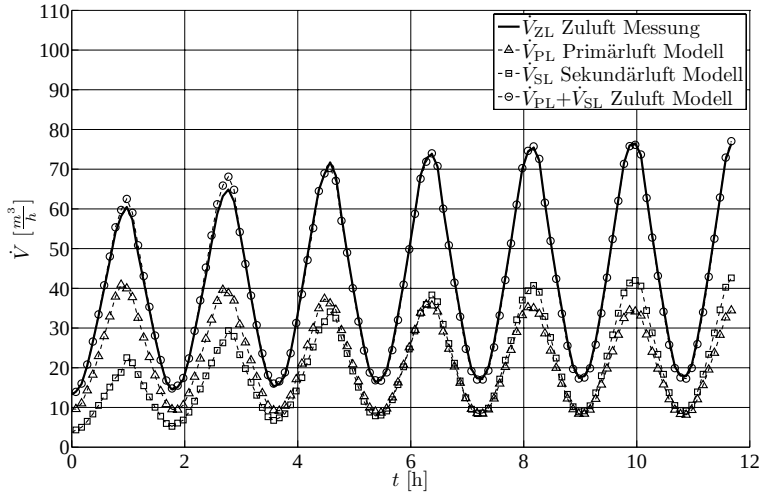


Abbildung 6 – Gegenüberstellung von gemessener und modellierter Luftzufuhr

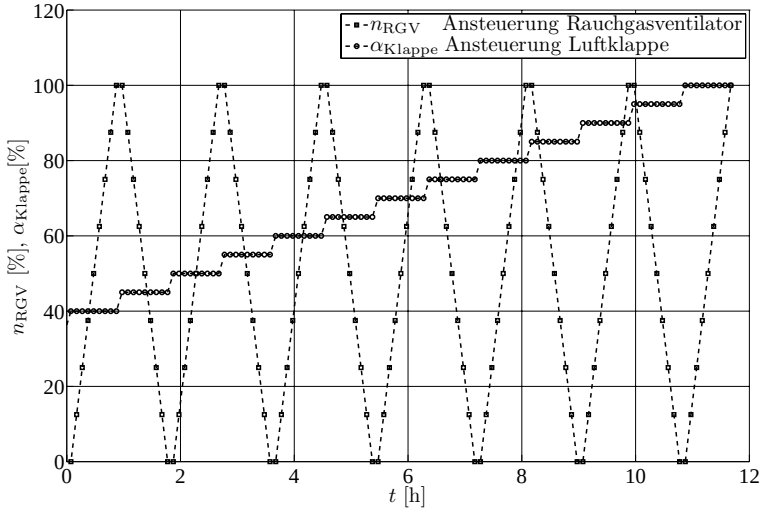


Abbildung 7 – Variation von Rauchgasventilatordrehzahl und Position der Luftklappe

3.4.2 Falschluft

Durch den in der gesamten Feuerungsanlage vorherrschenden Unterdruck kommt es an undichten Stellen der Anlage zum Ansaugen von sogenannter Falschluft, welche bei gut abgedichteten Anlagen üblicherweise 10 bis 20% der gesamten zugeführten Luft beträgt und bei der Modellbildung zusätzlich zur Primär- und Sekundärluft berücksichtigt werden muss.
Hierzu wurden in einem ersten Schritt die relevanten Falschluftquellen detektiert. Dabei zeigte sich, dass der Hauptteil der Falschluft über die Rostascheschnecke und weitere relevante Falschluftströme über die Brennstoffzufuhr, das Zündgebläse und den Aschebehälter angesaugt werden. Die Modellierung des Zusammenhanges der angesaugten Falschluft und des Feuerraumunterdrucks wurde analog zur Primär- bzw. Sekundärluftzufuhr durch geeignete Kombination von *rohrigen* und *blendigen Druckabfällen* für die jeweiligen Falschluftquellen getrennt durchgeführt. Für jede Falschluftquelle gilt

$$p_0 - p_{\mathrm{FR}} = R_1 \dot{V}_{\mathrm{FL}}^{1.75} + R_2 \dot{V}_{\mathrm{FL}}^2 \tag{15}$$

mit dem Umgebungsdruck p_0, dem Druck im Feuerraum p_{FR}, den Parametern R_1 und R_2 sowie dem Falschluftvolumenstrom $\dot{V}_{\mathrm{FL}}$. Zur Ermittlung von R_1 und R_2 für die verschiedenen Falschluftquellen wurde die Anlage so adaptiert, dass die gesamte Falschluft gemessen werden konnte. Durch schrittweises Abdichten der detektierten Falschluftquellen konnten die zur Parameterermittlung erforderlichen Messdaten durch Variation des Feuerraumunterdrucks und folglich des Falschluftstromes gewonnen werden. Basierend darauf wurden die Werte R_1 und R_2 für die jeweiligen Falschluftquellen mittels numerischer Optimierungsalgorithmen unter Minimierung eines quadratischen Gütekriteriums bestimmt. Abbildung 8 zeigt die berechneten Falschluftstöme für alle detektierten Falschluftquellen in Abhängigkeit des Feuerraumunterdrucks. Mit dem entworfenen Modell ist es nicht nur möglich, den gesamten Falschlufteintrag abzuschätzen, sondern ihn auch den unterschiedlichen Anlagenbereichen gemäß seiner Wirkung der Primär- oder Sekundärluft zuzuordnen. Dabei werden die über den Aschebehälter sowie die Rostascheschnecke angesaugten Falschluftstöme der Primärluft und die über die Brennstoffzufuhr sowie das Zündgebläse angesaugten Falschluftströme der Sekundärluft zugeordnet. Dies führt zu einer deutlichen Steigerung der erzielbaren Genauigkeit bei anderen Modellen (z.B. für das Abbrandverhalten der Biomasse auf dem Rost).

3.4.3 Rauchgasventilator

Die gezielte Vorgabe eines definierten Unterdrucks im Feuerraum ist eine wesentliche Voraussetzung, um mit Hilfe des in 3.4.1 entworfenen Modells für die Luftzufuhr einen gewünschten Primär- und Sekundärluftmassenstrom einstellen zu können. Dazu wird der Unterdruck über die Drehzahl des Rauchgasventilators, welcher sich am Ende der Anlage befindet, variiert. Die Druckanhebung Δp_{V} des Rauchgasventilators kann gemäß Gleichung (11) durch

$$\Delta p_{\mathrm{V}} = c_{\mathrm{V}} f_{\mathrm{V}}^2 \tag{16}$$

beschrieben werden, wobei c_{V} der Ventilatorkonstante und f_{V} der Ventilatorfrequenz entsprechen. Aufgrund der großen Temperaturunterschiede des Rauchgases am Ventila-

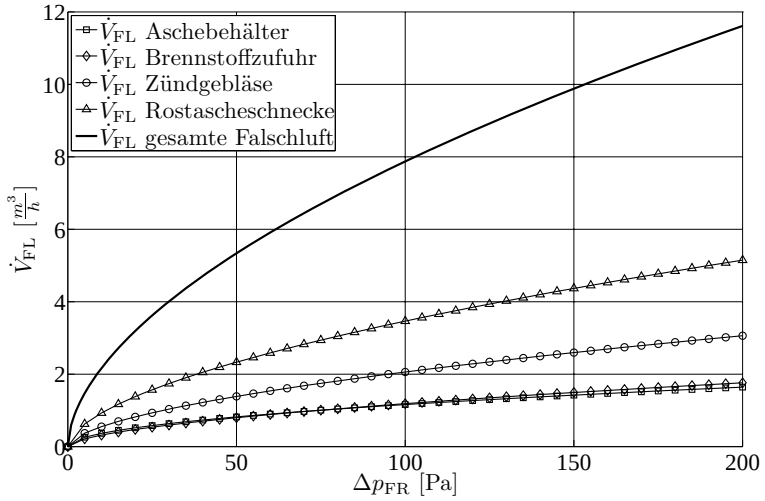


Abbildung 8 – Falschluftströme in Abhängigkeit des Feuerraumunterdrucks

tor wird die Abhängigkeit der Druckanhebung Δp_{V} des Ventilators von der Dichte des Rauchgases ρ gemäß

$$\Delta p_{\mathrm{V}} = c_{\mathrm{V0}} \frac{\rho}{\rho_0} f_{\mathrm{V}}^2 \tag{17}$$

berücksichtigt, wobei ρ_0 der Dichte des Gases entspricht, für das die Ventilatorkonstante ermittelt wurde (üblicherweise Umgebungsluft im kalten Anlagenzustand).

3.4.4 Brennstoffzufuhr

Die Variation des zugeführten Brennstoffmassenstromes bei der betrachteten Anlage erfolgt durch eine getaktete Ansteuerung der Brennstoffschnecke, wobei durch geeignete Wahl der Taktzeit die Gleichmäßigkeit der Verbrennung nicht beeinträchtigt wird. Die Abhängigkeit des zugeführten Brennstoffmassenstromes $\dot{m}_{\mathrm{B}}$ von der Einschaltzeit T_{e} und Ausschaltzeit T_{a} der Brennstoffzufuhr wird durch

$$\dot{m}_{\mathrm{B}} = \begin{cases} \dot{m}_{\mathrm{B,max}} \frac{T_{\mathrm{e}} - T_{\mathrm{min}}}{T_{\mathrm{e}} + T_{\mathrm{a}}}, & \text{für } T_{\mathrm{e}} \geq T_{\mathrm{min}} \\ 0, & \text{für } T_{\mathrm{e}} < T_{\mathrm{min}} \end{cases} \tag{18}$$

beschrieben. Der Parameter $\dot{m}_{\mathrm{B,max}}$ entspricht dem Brennstoffmassenstrom bei Dauerbetrieb der Schnecke mit Nenndrehzahl und ist im wesentlichen von der Schüttdichte des Hackgutes abhängig. Durch die elastische Aufhängung des antreibenden Motors kommt es beim Taktbetrieb der Schnecke zu Torsionseffekten, welche durch die Mindesteinschaltzeit $T_{\mathrm{ein,min}}$ berücksichtigt werden. Erst ab diesem Zeitpunkt beginnt die Brennstoffschnecke zu fördern. Mit Hilfe dieses sehr einfachen Modells wird das Verhalten der Brennstoffzufuhr näherungsweise beschrieben, wobei der Parameter $\dot{m}_{\mathrm{B,max}}$ für eine repräsentative durchschnittliche Schüttdichte des Hackgutes ermittelt werden muss.
Das Modell ist jedoch nicht in der Lage, die für Hackgut typische, aber gänzlich unbekannte Variation der Stückigkeit des Brennstoffs und die damit verbundene Schwankung

des vergleichsweise geringen Brennstoffmassenstromes abzubilden. Deshalb kommt konstruktiven Maßnahmen zur Vergleichmäßigung der Brennstoffzufuhr, wie z.B. der Steigung bzw. des Durchmessers der Brennstoffförderschnecken, insbesondere bei Biomasse-Kleinfeuerungsanlagen eine große Bedeutung zu.

4 Zusammenfassung

In diesem Beitrag wurden anhand einer marktverfügbaren Anlage einfache mathematische Modelle für eine automatisch betriebene Biomasse-Kleinfeuerungsanlage (im vorliegenden Fall ein mit Hackgut befeuerter Kessel) entwickelt. Die Modellbildung erfolgte aus Gründen der Übersichtlichkeit getrennt für die drei relevanten Anlagenbereiche Brennstoffbett, Primär- und Sekundärverbrennungszone sowie Wärmeübertrager. Dabei zeigte sich, dass für das Brennstoffbett und den Wärmeübertrager die für mittelgroße Rostfeuerungsanlagen existierenden Modelle verwendet werden können. Weiters zeigte sich, dass sowohl die Speicherwirkung der feuerfesten Auskleidung der Primärzone als auch die Umgebungsverluste vernachlässigt werden können. Deshalb werden die Primär- und Sekundärverbrennungszone durch eine gewöhnliche Verbrennungsrechnung ausreichend genau beschrieben. Es ist zu erwarten, dass die Speicherwirkung der feuerfesten Auskleidung und die Umgebungsverluste bei einer gut isolierten Biomasse-Kleinfeuerungsanlage mit gekühlter Sekundärverbrennungszone vernachlässigt werden können. Diese Annahme muss jedoch entsprechend der gezeigten Vorgehensweise verifiziert werden.
In einem weiteren Schritt wurden die Luft- und die Brennstoffzufuhr modelliert. Hierbei stellt insbesondere die gemeinsame Luftklappe für die Primär- und die Sekundärluftzufuhr eine besondere Herausforderung dar. Die Realisierung der Primär- und Sekundärluftzufuhr weicht allerdings je nach Hersteller zum Teil sehr stark von der betrachteten Konfiguration ab, weshalb die entsprechenden Modelle für jede Anlage individuell angepasst werden müssen.
Die entwickelten Modelle bilden die physikalischen Gegebenheiten trotz ihrer mathematischen Einfachheit sehr gut ab, und stellen eine geeignete Grundlage für einen modellbasierten Reglerentwurf dar.

Literatur

[1] Bauer R. *Modellbildung und modellbasierte Regelungsstrategien am Beispiel einer Biomasse-Feuerungsanlage*. Habilitationsschrift, Technische Universität Graz, 2009.

[2] Bauer R., Gölles M., Brunner T., Dourdoumas N., Obernberger I. Modellierung der Druck- und Volumenstromverhältnisse in einer Biomasse-Feuerung. *In - Automatisierungstechnik*, 55:404–410, August 2007.

[3] Bauer R., Gölles M., Brunner T., Dourdoumas N., Obernberger I. Modellierung des dynamischen Verhaltens der Wärmeübertragung in einem Rauchrohr-Wärmeübertrager. *In - Automatisierungstechnik*, 56:513–520, October 2008.

[4] Bauer R., Gölles M., Brunner T., Dourdoumas N., Obernberger I. Modelling of grate combustion in a medium scale biomass furnace for control purposes. *Biomass and Bioenergy*, 34(4):417–427, 2010.

[5] Gölles M. *Entwicklung mathematischer Modelle einer Biomasserostfeuerungsanlage als Grundlage für modellbasierte Regelungskonzepte.* Dissertationsschrift, Technische Universität Graz, 2008.

[6] Gölles M., Bauer R., Brunner T., Dourdoumas N., Obernberger I. Model based control of a biomass grate furnace. In *European Conference on Industrial Furnaces and Boilers*, April 2011. ISBN 978-972-99309-6-6.

[7] Thunman H, Leckner B. Ignition and propagation of a reaction front in cross-current bed combustion of wet biofuels. *Fuel*, 80:473–81, 2001.

[8] Thunman H, Leckner B. Co-current and counter-current fixed bed combustion of biofuel - a comparision. *Fuel*, 82:275–83, 2003.

Lockerung von Sensortoleranzen mittels regelungstechnischer Methoden für den Teilchenbeschleuniger CLIC

J. Pfingstner[1,2], J. Snuverink[2], D. Schulte[2], M. Hofbaur[3]

[1] Doctoral School Informations- und Kommunikationstechnik, TU Graz
[2] Europäische Organisation für Kernforschung (CERN), Genf
[3] Private Universität UMIT, Hall/Tirol

Kurzfassung

CLIC ist ein zukünftiger Teilchenbeschleuniger, der die Nachfolge des derzeit größten Beschleunigers (LHC) antreten könnte. Auf Grund der sehr kleinen verwendeten Teilchenstahlen, ist CLIC sehr empfindlich gegenüber Bodenbewegungen. Die Möglichkeit CLIC trotz der unumgänglichen Bodenbewegungen zu betreiben, wurde von vielen Fachleuten angezweifelt. In der vorgelegten Arbeit wird aber gezeigt, dass die am CERN entwickelten Bodenbewegungs-Gegenmaßnahmen die Bodenbewegungseffekte effizient unterdrücken und daher Bodenbewegungen kein Hindernis mehr für einen Betrieb von CLIC darstellen. Um dieses Ergebnis zu erzielen, wurde mit Hilfe teils schon vorhandener Codes ein Simulationsumgebung entwickelt, die das integrierte Verhalten von Beschleuniger, Strahl-Stahl-Effekten, Bodenbewegungen und Bodenbewegungs-Gegenmaßnahmen detailgenau nachbildet. Ein wichtiger Teil dieser Simulationen ist der sog. Linac-Regler dessen Design ebenfalls in dieser Arbeit präsentiert wird. Die Simulationsergebnisse zeigen, dass CLIC trotz Bodenbewegungen ausreichend elementare Teilchenkollisionen produzieren kann. Zusätzlich konnte durch den Entwurf des Linac-Reglers die spezifizierte Sensorauflösung von 10 nm auf 50 nm gelockert werden, was zu einer erheblichen Kostenreduktion führt. Auch die notwendigen Anforderungen an die Aktuatordynamik erwiesen sich als weniger anspruchsvoll als ursprünglich angenommen.

[1]Korrespondenz bitte an juergen.pfingstner@cern.ch

1 Einleitung

1.1 Der Teilchenbeschleuniger CLIC

Der derzeit größte Teilchenbeschleuniger, der Large Hardron Collider (LHC), wird voraussichtlich Ende 2011 die ersten aufregenden Ergebnisse präsentieren. Welche neuen Teilchen auch immer gefunden werden, einige ihrer Eigenschaften werden vom LHC nicht, oder nur ungenau gemessen werden können. Der Grund für diese Messlimitationen liegt in der Tatsache begründet, dass der LHC Protonen miteinander kollidiert. Protonen sind keine Elementarteilchen sondern bestehen aus Quarks und Gluonen, auch Partonen genannt. Zwar ist die Kollisionsenergie der Protonen genau bekannt, jedoch unterliegt die Energie der elementar interagierenden Partonen statistischen Fluktuationen. Die Interaktionsenergie einer Kollision der Partonen selbst ist daher nur mäßig genau bekannt, was zu Messungenauigkeiten führt.

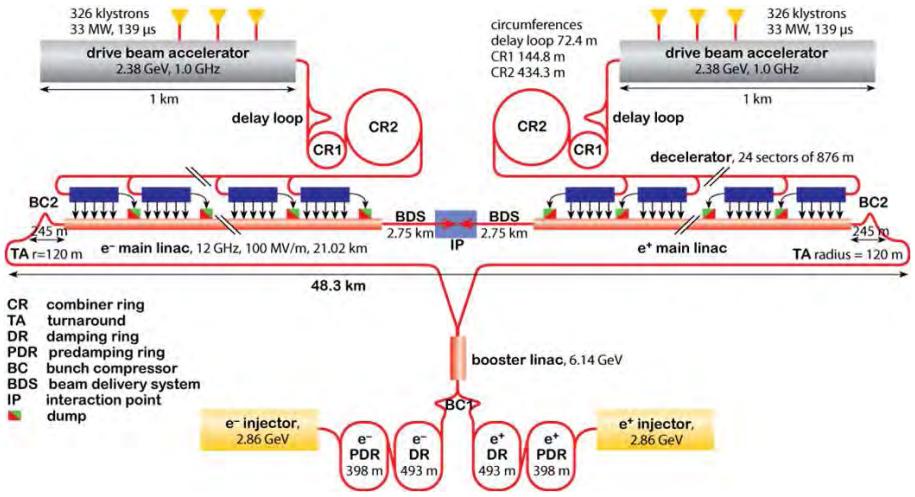


Abbildung 1: Struktur des CLIC-Komplexes. Aus Energieeffizienzgründen verwendet CLIC das sog. Zweistrahl-Beschleunigungskonzept, bei welchen die Elektron- und Positron-Hauptstrahlen (main beams) mit Hilfe eines Antriebsstrahls (drive beam) beschleunigt werden. Der obere Teil der Graphik zeigt die Erzeugung, Beschleunigung, Komprimierung und den Transport des Antriebsstrahls zum Hauptbeschleuniger (main linac), in dem die Energie des Antriebsstrahls auf den des Hauptstrahls übertragen wird. Der untere Teil der Graphik zeigt die Erzeugung und den Transport der Elektronen- und Positronenstrahlen. Wichtige Teile für diese Arbeit sind der Hauptbeschleuniger; das Beam Delivery System (BDS), in dem der Strahl für die Kollision konditioniert wird und der Kollisionspunkt (IP).

Eine zweite Messung mit einem komplementären Beschleunigertyp ist daher notwendig um die zugrundeliegenden Mechanismen der Teilchenproduktion vollständig zu verstehen. Der Vorschlag der Europäische Organisation für Kernforschung (CERN) für solch einen zukünftigen Beschleuniger heißt Compact Linear Collider (CLIC) [4]. CLIC kollidiert Elektronen und Positronen (beides elementare Teilchen) mit einer Kollisionsenergie von 3 TeV (Tera-Elektronenvolt). Die Verwendung dieser Teilchen hat allerdings einen entscheidenden Nachteil. Wird ein geladenes Teilchen durch ein Magnetfeld abgelenkt, gibt es tangential zur Bewegungstrajektorie elektromagnetische Strahlung, die sog. Synchrotronstrahlung, ab. Bewegt sich das Teilchen nahe der Lichtgeschwindigkeit ist die abgestrahlte Leistung $\propto 1/m_0^4$,

wobei m_0 die Teilchenmasse ist. Da Elektronen und Positronen ca. 2000-mal leichter sind als Protonen, können ersteres genannte nicht (mit vertretbaren Kosten) in einem Ringbeschleuniger auf die notwendige Energie beschleunigt werden. CLIC ist daher ein linearer Beschleuniger (siehe Abb. 1). Falls er gebaut werden sollte, wird er sich in einem ca. 50 km langen Tunnel 100 m unter der Erdoberfläche befinden.

Mit der Verwendung einer linearen Beschleunigungsstruktur ergibt sich eine zweite Problematik. Während beim LHC die zirkulierenden Strahlen sehr oft für Kollisionen wiederverwendet werden können, kann die Energie in den CLIC-Strahlen nur für eine Kollision genutzt werden. Um CLIC energieeffizient betreiben zu können muss daher die Anzahl der Strahlkollisionen, im Vergleich zum LHC, stark reduziert werden. Um diesen Nachteil wettzumachen, verwendet CLIC sehr kleine Strahlabmessungen am Kollisionspunkt. In vertikaler Richtung beträgt die Strahlgröße, ausgedrückt durch die Standardabweichung der Teilchenverteilung des Strahls, nur 1 nm. Detaillierte Informationen über CLIC können in [10] gefunden werden.

1.2 Die Problematik von Bodenbewegungen und Gegenmaßnahmen

Durch die Notwendigkeit der kleinen Strahlabmessungen am Kollisionspunkt wird CLIC äußerst empfindlich gegenüber Bodenbewegungen. Der grundlegende Mechanismus der Reduktion der Kollisionseffizienz durch Bodenbewegungen ist der folgende. Der CLIC-Strahl wird von Magneten entlang des Hauptbeschleunigers und des Beam Delivery Systems (BDS) bis an den Kollisionspunkt gelenkt. Bodenbewegungen verschieben diese Magneten, was zu einer ungewollten, transversalen Strahlablenkung führt. Diese Ablenkungen führen aber nicht dazu, dass der Strahl den Beschleuniger komplett verlässt, da die folgenden Magneten eine fokussierende Wirkung haben. An Stelle dessen beginnt der Strahl entlang des Beschleunigers zu oszillieren, was zu zwei Problemen am Kollisionspunkt führt. Erstens treffen sich die beiden Strahlen durch die zufälligen Oszillationen nicht mehr genau aufeinander, oder verfehlen sich sogar (Strahl-Strahl-Versatz δ). Zweitens kommt es durch die Strahloszillationen entlang des Beschleunigers zu einer Aufweitung des Strahls. Die Strahlgröße am Kollisionspunkt wächst ($\sigma_0^* + \Delta\sigma^*$) und die Teilchendichte sinkt. Es ist einsichtig, dass beide Effekt (Strahlversatz und Strahlvergrößerung) umso drastischere Auswirkungen haben, umso kleiner die nominelle Strahlgröße σ_0^* ist. Im Falle von CLIC würden Bodenbewegungen den Betrieb des Beschleunigers von der ersten Sekunden an unmöglich machen. Aus diesem Grund wurden in den letzten Jahren fünf Gegenmaßnahmen zur Unterdrückung von Bodenbewegungseffekten entwickelt (siehe Abb. 2), die im Folgenden kurz beschrieben werden.

Die erste Methode ist der sog. *Linac-Regler*. Seine Aufgabe ist es Strahloszillation entlang des Hauptbeschleunigers und des BDS zu unterdrücken. Um dies zu bewerkstelligen, werden die Strahloszillationen mittels 2122 sog. Strahlpositionsmonitoren (BPMs) gemessen. Der Linac-Regler verwendet diese Sensordaten um Stellgrößen für die 2104 Aktuatoren zu errechnen, welche die Oszillationen dämpfen und den Strahl in die Zentren der BPMs lenken. Es handelt sich also um ein Störungsunterdrückungsproblem. Als Aktuatoren werden sog. Tripods verwendet (siehe Abb. 3 und [3]), welche die QP-Magneten mechanisch verschieben.

Da der CLIC-Strahl eine gepulste Form hat, liefern die BPMs nur in der Taktfrequenz, d.h. alle 20 ms, Messdaten. Der Linac-Regler ist daher ein inhärent diskretes System mit einer Abtastrate von T_d=20 ms. Nach dem Shannonschen Abtasttheorem können Nutzsignale nur bis zu einer maximalen Frequenz von 25 Hz aufgelöst werden. Höhere Frequenzanteile werden zwar durch den Aliasing-Effekt auch in diesen 25 Hz Bereich gefaltet, haben aber aufgrund

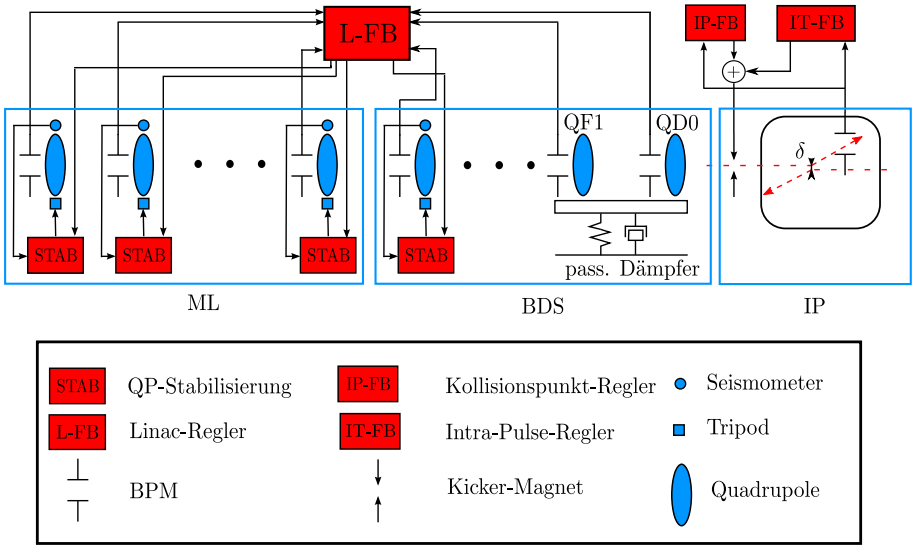


Abbildung 2: Übersicht der fünf Bodenbewegungs-Gegenmaßnahmen von CLIC. Der Linac-Regler dämpft die Strahloszillationen mit Frequenzen unter 1-4 Hz. Höhere Frequenzen werden mit der sog. QP-Stabilisierung bekämpft. Für die letzten beiden QP-Magneten vor dem Kollisionspunkt (QF1 und QD0) ist die QP-Stabilisierung nicht effizient genug und es muss ein passiver Bodenbewegungsdämpfer benutzt werden. Zusätzlich reduzieren das IP-FB und das IT-FB den Strahl-Strahl-Versatz δ am Kollisionspunkt.

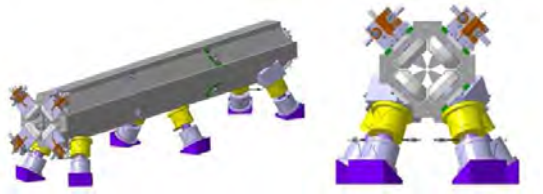


Abbildung 3: Ein Tripod hat Ähnlichkeit mit einer Steward-Plattform. Er besteht aus sechs Beinen die gemeinsam einen QP-Magneten tragen. Die Länge der Beine kann mittels piezo-elektrischen Aktuatoren verändert werden. Dadurch lässt sich der QP in vier Freiheitsgraden verschieben. Graphik mit freundlicher Genehmigung von K. Artoos.

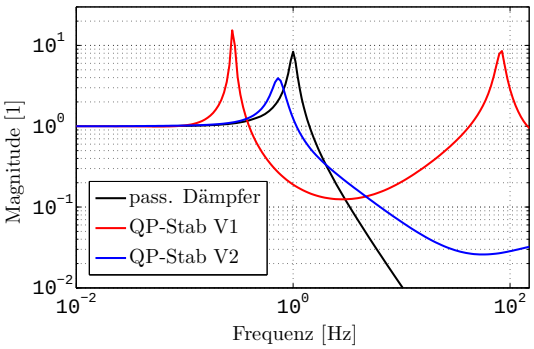


Abbildung 4: Absolutwerte von Übertragungsfunktionen in vertikaler Richtung: QP-Stabilisation Version 1 (rot) und Version 2 (blau), passiver Bodenbewegungsdämpfer (schwarz). Die QP-Stabilisation Version 1 und 2 unterscheiden sich durch den verwendeten Sensor, wobei Version 2 ein zukünftiges Projekt darstellt.

des stark abfallenden Bodenbewegungsspektrums wenig Auswirkung auf das Gesamtsignal. Aufgrund der aus regelungstechnischer Sicht notwendigen Überabtastung kann der Linac-Regler Bodenbewegungs-Effekte nur bis ca. 1-4 Hz effektiv unterdrücken.

Auf Grund dieser Limitation wurde die sog. *QP-Stabilisation* entwickelt. Zur Unterdrückung von höher-frequenten Strahloszillationen werden die QP-Magneten mechanisch stabilisiert. Zu diesem Zweck wird auf jedem QP ein Seismometer montiert, der die QP-Bewegungen im Bereich von 0.03 bis 150 Hz misst. Diese Sensordaten werden von einem Regelalgorithmus (siehe [3]) verwendet um wiederum denselben Tripod zu aktuieren, der auch schon vom Linac-Regler verwendet wurde. Die zugehörige Übertragungsfunktion der Bodenbewegungen auf den QP ist in Abb. 4 dargestellt.

Die letzten beiden QP-Magneten vor dem Kollisionspunkt (QF1 und QD0) sind besonders sensibel bezüglich transversaler Verschiebung. Da die QP-Stabilisation V1 Frequenzen um 0.3 und 70 Hz verstärkt reicht diese nicht aus um die Bodenbewegungen für QF1 und QD0 ausreichend zu dämpfen. Darum wurde in [5] ein *passiver Bodenbewegungsdämpfer* vorgeschlagen. Es handelt sich dabei um einen 110 Tonnen schweren Betonblock, der auf 10 pneumatischen Vibrationsisolatoren gelagert ist. Durch diesen Aufbau kann ein Tiefpass mit sehr niedriger Grenzfrequenz realisiert werden, der aber trotzdem robust gegenüber Störungen, die direkt auf das zu isolierende System (Betonblock) einwirken, ist.

Am Kollisionspunkt kann ein weiterer Effekt ausgenutzt werden um den kritischen Strahl-Strahl-Versatz δ zu reduzieren. Bei der Kollision der Elektron- und Positron-Strahlen beeinflussen sich die zugehörigen elektromagnetischen Felder gegenseitig. Im Speziellen kann durch Simulationen gezeigt werden [9], dass bei einem Strahl-Strahl-Versatz δ die Trajektorien der beiden Strahlen eine Winkeländerung erfahren. Für kleine δ ist diese Winkeländerung ein lineare Funktion von δ. In einem 3 m nach dem Kollisionspunkt positionierten BPM kann diese Winkeländerung in Form eine Trajektorienänderung detektiert werden. Da dieses Signal direkt proportional zum Strahl-Strahl-Versatz ist, wurden zwei dezidierte Regelalgorithmen entworfen. Der *Kollisionspunkt-Regler* (IP-FB) verwendet das δ-Signal um einen Kicker-Magneten zu aktuieren, der sich 3 m vor dem Kollisionspunkt befindet. Der zugehörige Regelalgorithmus [2] ist entworfen um δ zu minimieren. Da aber auch dieser Algorithmus den Beschränkungen durch die gepulste Form des CLIC-Strahls unterliegt, kann er wie der Linac-Regler nur Bodenbewegungen bis 1-4 Hz unterdrücken. Ein zweites System, der sog. *Intra-Puls-Regler* (IT-FB), kann diese Problematik teilweise umgehen. Das IT-FB benutzt zwar die gleichen Aktuatoren und Sensoren als das IP-FB, verwendet aber extrem schnelle Elektronikkomponenten (siehe [1]). Dadurch ist es möglich während eines Strahlpulses, der nur 156 ns lang ist, zu reagieren. Durch die anfallenden Verzögerungszeiten ist es aber nur möglich fünf Mal während eines Pulses die Kickerstärke zu verändern. Zurzeit wird das IT-FB als Reserve gehandhabt und ist daher nicht in den folgenden Simulationen berücksichtigt.

Diese Arbeit beschäftigt sich mit zwei Themen im Bereich der Bodenbewegungsbekämpfung für CLIC. Im Abschnitt 2 wird eine umfangreiche Simulationsumgebung vorgestellt, welche es ermöglicht die Auswirkungen der Bodenbewegungen und der Bodenbewegungs-Gegenmaßnahmen auf die Kollisionsqualität zu eruieren. Solch eine rechenintensive Herangehensweise ist notwendig, da es derzeit kein analytisches Modell gibt welches alle auftretenden Effekte ausreichend genau beschreibt. Basierend auf der vorgestellten Umgebung wir in Abschnitt 3 der Entwurf des Linac-Reglers präsentiert. Die begrenzte BPM-Auflösung (Sensorrauschen) wird sich dabei als eines der Hauptprobleme herausstellen. Die Simulationsergebnisse für das Gesamtsystem werden in Abschnitt 4 präsentiert und diskutiert.

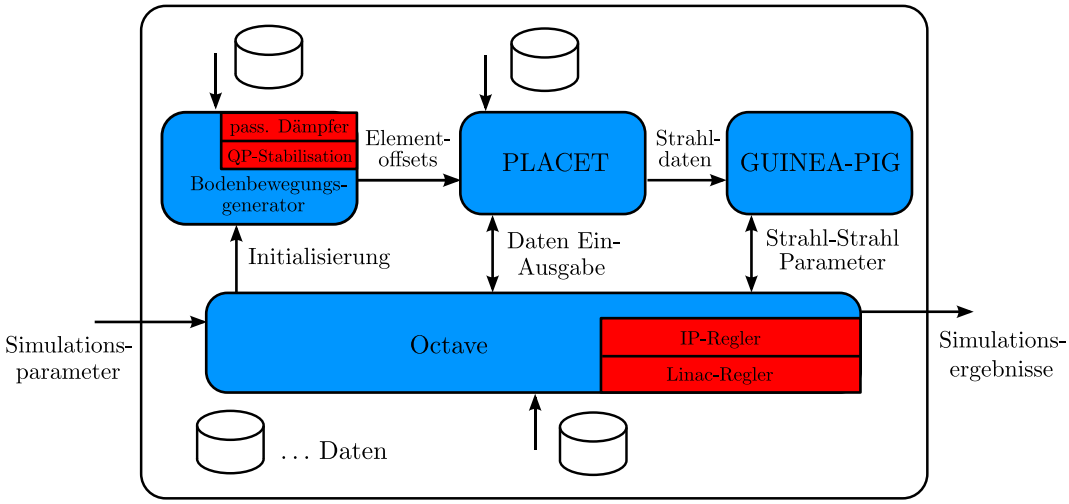


Abbildung 5: Die komplette Simulationsumgebung besteht aus vier Modulen: PLACET für das Strahl-Tracking, GUINEA-PIG zur Berechnung der Kollisionsparameter wie beispielsweise Luminosität, ein Bodenbewegungsgenerator und Octave um die einzelnen Simulationscodes zu verbinden und verwendete Algorithmen zu implementieren.

2 Simulationsumgebung

Die Qualität der Strahlkollisionen wird von Strahlphysikern mit Hilfe der sog. Luminosität gemessen. Die Abhängigkeit dieser Luminosität von den Bodenbewegungen stellt ein komplexes nichtlineares System mit einer hohen Anzahl von Eingangsvariablen dar. Obwohl einige vereinfachte Abschätzungen für die Auswirkungen einzelner Bodenbewegungs-Gegenmaßnahmen erstellt wurden, gibt es kein einfach zu handhabendes mathematisches Modell, welches alle Einflüsse ausreichend umfasst. Aus diesem Grund wurde eine Simulationsumgebung entworfen, welche verschiedene numerische Simulationscodes verbindet und erweitert. Eine Übersicht über das gesamte Simulationspaket ist in Abb. 5 zu sehen.

Die Hauptkomponente der Simulation ist der Strahl-Trackingcode PLACET [12]. PLACET ermöglicht es das Verhalten der Teilchenstrahlen entlang eines Beschleunigers zu berechnen. Dafür wird der Effekt aller Beschleunigerkomponenten mittels mathematischen Formalismen modelliert. PLACET ist darauf spezialisiert Imperfektionen der Komponenten (wie beispielsweise Veränderungen der Magnetpositionen) im Strahltracking zu berücksichtigen. Der Code ist in C++ geschrieben und wird üblicher Weise mittels der Skript-Sprache Tcl/Tk gesteuert. Es ist jedoch zusätzlich möglich innerhalb einer PLACET-Simulation eine Octave-Umgebung, ein frei erhältlicher Matlab-Clone, zu starten und notwendigen Berechnungen in Octave vorzunehmen.

Um realistische Komponentenverschiebungen für das Strahltracking in PLACET zu erzeugen wurde ein Bodenbewegungsgenerator in PLACET integriert (obwohl in Abb. 5 separat angeführt). Es handelt sich dabei um eine modifizierte Version eines bereits vorhandenen Generators [8]. Die entscheidende Modifikation besteht darin, das es nun möglich ist Filterfunktionen vorzugeben, mit denen die ursprünglichen Bodenbewegungen gefiltert werden. Dadurch kann der Einfluss des passiven Bodenbewegungsdämpfers und der QP-Stabilisation in die Simulationen integriert werden. Im Moment wurden zwei Gruppen von Bodenbewegungsmodelle implementiert. Für Bodenbewegungen bis zu einer Minute werden leicht

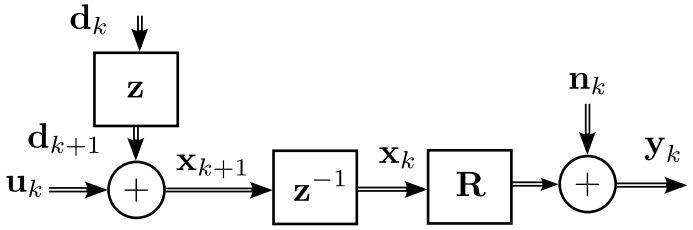


Abbildung 6: Struktur des diskreten Beschleunigermodells. Die Bodenbewegungen $\boldsymbol{d}_{k+1}$ und die Stellgrößen $\boldsymbol{u}_k$ verändern die QP Positionen $\boldsymbol{x}_{k+1}$, wobei k der Zeitindex ist. Die Bodenbewegungen $\boldsymbol{d}_k$ wirken eigentlich direkt auf $\boldsymbol{x}_k$. Um aber die Standardsystemdarstellung, in welcher die Störgröße am Eingang des zu regelnden System wirkt, verwenden zu können, wurde $\boldsymbol{d_k + 1 = zd_k}$ verwendet. Die QP-Verschiebungen $\boldsymbol{x}_k$ resultieren in Strahloffset-Messungen in den BPMs $\boldsymbol{y}_k$. Der Zusammenhang wird über die Matrix $\boldsymbol{R}$ hergestellt. Das Sensorrauschen ist mit $\boldsymbol{n}_k$ bezeichnet.

modifizierte Versionen der Modell von Seryi [13] verwendet. Für längere Simulationen findet das ATL-Gesetz [14] Verwendung.

Nachdem PLACET realistische Elektron- und Positron-Strahlen am Kollisionspunkt berechnet hat, können diese von GUINEA-PIG dazu verwendet werden um die zugehörige Luminosität und andere Strahl-Strahl-Parameter zu berechnen. GUINEA-PIG ist eine Simulationsumgebung [11], welche sehr komplexe physikalische Prozesse modelliert. Um die gegenseitige Strahl-Strahl-Ablenkung zu bestimmen, müssen die elektro-magnetischen Felder der beiden relativistischen Strahlen berechnet werden. Zusätzlich werden relevante Effekte der Quantenelektrodynamik (QED) und Quantenchromodynamik (QCD) berücksichtigt.

Auf Grund der Fähigkeit PLACETs während der Programmausführung eine Octave-Umgebung zu starten, kann die komfortable Skriptsprache von Octave für viele anfallende Aufgaben genutzt werden. In der präsentierten Simulationsumgebung werden beispielsweise die Steuerung des gesamten Programmablaufs und der verschiedenen Softwaremodule, die Verarbeitung von Simulationsergebnissen und die Ein- und Ausgaben in Octave durchgeführt. Da in Octave ein Vielzahl von hoch-optimierten numerischen Funktionen zur Verfügung stehen, wurden die verschiedenen Varianten der Regelalgorithmen des Linac-Reglers und des Kollisionspunkt-Reglers ebenfalls in dieser Umgebung implementiert.

3 Entwurf des Linac-Reglers

3.1 Systemmodell und Wahl der Regelerstruktur

Das zu regelnde System ist in Abb. 6 dargestellt. Die zu regelnde Größe ist der Strahloffset in den BPMs $\boldsymbol{y}_k - \boldsymbol{n}_k$, welcher von den Bodenbewegungen $\boldsymbol{d}_k$ angeregt wird. Da das System keine interne Rückkopplung hat, handelt sich um ein System mit endlicher Impulsantwort (FIR) mit 2104 Eingängen und 2122 Ausgängen. Die Problematik der Regelung dieses Systems ist nicht durch dessen komplexes dynamisches Verhalten, sondern durch dessen hohe Anzahl von Ein- und Ausgängen gegeben. Vereinfachungen sind notwendig um einen MIMO-Regler dieser Größe zu entwerfen. Ein weiterer zentraler Punkt ist die passende Berücksichtigung des komplexen Verhaltens der Bodenbewegungen als Vorgabe für den Reglerentwurf. Die Unterdrückung von Sensorrauschen hat sich als zusätzliche entscheidende Anforderung an

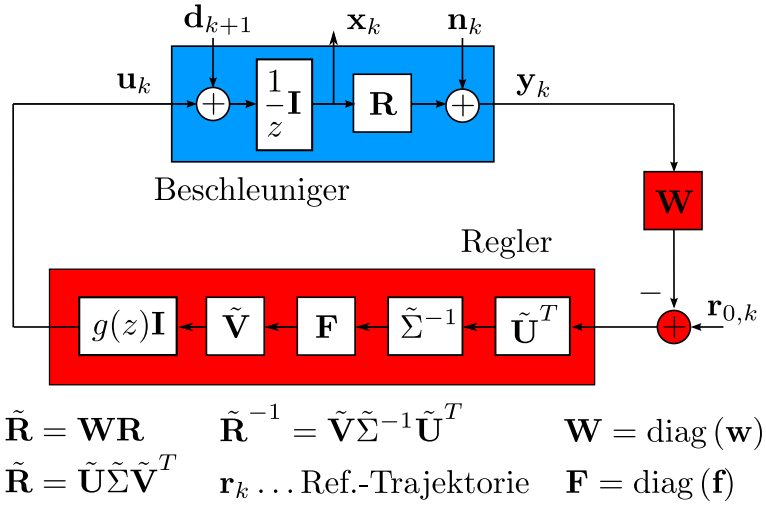


Abbildung 7: Struktur des gewählten Reglerkonzepts. Der Regler basiert auf der Singulärwertzerlegung der Matrix $\boldsymbol{WR}$, wobei $\boldsymbol{W}$ eingesetzt wird um die BPM-Messungen verschieden zu gewichten. Der Regler ist aufgeteilt in einen Teil der nur von der Eingangsrichtung der Messungen abhängt $\tilde{\boldsymbol{V}}\boldsymbol{F}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{U}}^T$ und einen der rein zeitabhängig ist $g(z)\boldsymbol{I}$, wobei $\boldsymbol{I}$ die Einheitsmatrix symbolisiert.

den Regler herausgestellt.

Um all die genannten Anforderungen zu erfüllen, wurde die Struktur eines Entkopplungsreglers gewählt. Für das Beschleunigersystem ist diese Entkopplung basierend auf der Singulärwertzerlegung (SVD) der Matrix $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ zweckmäßig, was zu einem sog. SVD-Regler führt (siehe Abb. 7 und [15]). Es gibt zwei Gründe für diese Wahl. Zum einen verlangen die Größe des Systems und die Komplexität der Bodenbewegungsspektren Vereinfachungen beim Entwurf. Diese ist beim SVD-Regler durch die Entkopplung des Problems in 2104 SISO-Schleifen gegeben. Zum anderen ist diese Entkopplung, auf Grund der einfachen Systemstruktur, für alle Frequenzen vollständig erfüllbar.

Damit die Entkopplung ersichtlich wird, ist es erforderlich nicht die ursprünglichen, sondern die koordinaten-transformierten Signale $\hat{\boldsymbol{u}}_k = \tilde{\boldsymbol{V}}^T\boldsymbol{u}_k$, $\hat{\boldsymbol{x}}_k = \tilde{\boldsymbol{V}}^T\boldsymbol{x}_k$ und $\hat{\boldsymbol{y}}_k = \tilde{\boldsymbol{U}}\boldsymbol{W}\boldsymbol{y}_k$ zu betrachten. Der Dach-Index kennzeichnet hier die Entkopplung. Auch die Eingangssignale müssen konsequenter Weise zu $\hat{\boldsymbol{d}}_{k+1} = \tilde{\boldsymbol{V}}^T\boldsymbol{d}_{k+1}$ und $\hat{\boldsymbol{n}}_k = \tilde{\boldsymbol{U}}^T\boldsymbol{W}\boldsymbol{n}_k$ transformiert werden. Diese Transformationen führen zu entkoppelten Regelkreisen. Es sei darauf hingewiesen, dass jeder dieser Kreise einem speziellen Aktuatorenvektor $\tilde{\boldsymbol{v}}_i$ und einem Messvektor $\tilde{\boldsymbol{u}}_i$ entspricht, welche durch die Spalten der Matrizen $\tilde{\boldsymbol{V}}$ und $\tilde{\boldsymbol{U}}$ gegeben sind. Die entkoppelten Regelkreise regeln daher jeweils ein Aktuator- und BPM-Muster, welches alle Aktuatoren und BPMs umfasst.

Trotz der Vereinfachung der Entkopplung gilt es noch immer 2104 SISO-Regler zu entwerfen. Um das System weiter zu vereinfachen wurde die Wahl getroffen nur eine frequenzabhängige Filterfunktion $g(z)$ zu wählen, welche für jeden entkoppelten Kanal verwendet wird. Um den unterschiedlichen Verhältnissen zwischen den Strahloszillationen- und Sensorrauschsignal in verschiedenen SISO-Schleifen Rechnung zu tragen, wurde ein zusätzlicher konstanter Verstärkungsfaktor f_i pro Kanal als Entwurfsparameter offen gelassen. Diese Verstärkungsfaktoren sind als Diagonalelemente in der Matrix $\boldsymbol{F}$ zusammengefasst. Es sei darauf hingewiesen, dass der daraus resultierende Regler einen rein frequenzabhängigen An-

teil besitzt und einen Anteil der rein von der Richtung des Messvektors abhängt. Da die Messungen an aufeinander folgenden räumlichen Punkten durchgeführt werden, jeweils dieselbe physikalische Größe gemessen wird und die Anzahl der Messwerte sehr groß ist, kann der Messvektor als ein räumliches Signal aufgefasst werden. In diesem Text wird daher von einem frequenz-abhängigen und einem räumlichen Filter die Rede sein. Um unterschiedliche BPM-Messungen verschieden zu gewichten wurde die diagonale Gewichtungsmatrix $\boldsymbol{W}$ eingeführt. Für den Regler muss daher nicht die SVD der Matrix $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ sondern des Systems $\boldsymbol{W}\boldsymbol{R} = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{V}}^T$ berechnet werden. Im Moment wird zwar für $\boldsymbol{W}$ nur die Einheitsmatrix verwendet, die Option $\boldsymbol{W}$ zu verändern wurde aber im Entwurf gelassen um zusätzliche freie Parameter für zukünftige Veränderungen zu habe.

Im Folgenden wird der Entwurf des frequenzabhängigen und des räumlichen Filters beschrieben. Vor allem das Design des räumlichen Filters wird von entscheidender Bedeutung sein, um den Einfluss des Messrauschens $\boldsymbol{n}_k$ ausreichend zu unterdrücken. Ein früherer Entwurf mit Hilfe von Kalman-Filtern als $g(z)$ und $\boldsymbol{F} = \boldsymbol{I}$ konnte diese Aufgabe nicht zufriedenstellend erfüllen (siehe [7]), da der räumliche Filter nicht adäquat war. Um das Messrauschen $\boldsymbol{n}_k$ möglichst wenig in das System einzukoppeln ist es zweckmäßig die Frobenius-Norm des räumlichen Filters $||\tilde{\boldsymbol{V}}\boldsymbol{F}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{U}}^T||_F$ zu minimieren. Der Grund dafür wird ersichtlich wenn man das Verhalten eines Zufallsvektors (Messrauschen) bei Multiplikation mit einer beliebigen Matrix $\boldsymbol{A}$ betrachtet. Sei $\boldsymbol{n}$ ein Vektor mit N Elementen, wobei die Elemente des Vektors $n_i \; \forall i = 1, \ldots, N$ diskrete, weiße, normalverteilte, stochastische Prozesse mit Varianz σ_n^2 sind. Dann ist der Erwartungswert der quadrierten l^2-Norm dieses Vektors zu jedem beliebigen Zeitpunkt gegeben durch

$$E\{||\boldsymbol{n}||_{l^2}^2\} = E\{\boldsymbol{n}^T\boldsymbol{n}\} = E\left\{\sum_{i=1}^{N} n_i^2\right\} = \sum_{i=1}^{N} E\left\{n_i^2\right\} = N\sigma_n^2. \tag{1}$$

Zur Berechnung der l^2-Norm des Vektors $\hat{\boldsymbol{n}} = \boldsymbol{A}\boldsymbol{n}$ ist es sinnvoll die Matrix $\boldsymbol{A}$ in ihrer singulärwertzerlegten Form $\boldsymbol{A} = \boldsymbol{U_A}\boldsymbol{\Sigma_A}\boldsymbol{V_A^T}$ zu verwendet. Sowohl $\boldsymbol{U_A}$ als auch $\boldsymbol{V_A}$ sind dabei orthonormale Matrizen. Zu bemerken ist, dass der Vektor $\tilde{\boldsymbol{n}} = \boldsymbol{V_A^T}\boldsymbol{n}$ ebenfalls aus mittelwertfreien, normalveteilten Zufallsvariablen mit einer Standardabweichung von $\sigma_{\tilde{n}} = \sigma_n$ besteht, was durch kurze Rechnung gezeigt werden kann. Es ist daher vorteilhaft $\hat{\boldsymbol{n}} = \boldsymbol{U_A}\boldsymbol{\Sigma_A}\tilde{\boldsymbol{n}}$ zu betrachten.

$$\begin{aligned} E\left\{||\hat{\boldsymbol{n}}||_{l^2}^2\right\} = E\left\{\hat{\boldsymbol{n}}^T\hat{\boldsymbol{n}}\right\} = E\left\{\tilde{\boldsymbol{n}}^T\boldsymbol{\Sigma}_A^T\boldsymbol{U}_A^T\boldsymbol{U}_A\boldsymbol{\Sigma}_A\tilde{\boldsymbol{n}}\right\} = \\ E\left\{\sum_{i=1}^{N_A}\tilde{n}_i^2 s_i^2\right\} = \sigma_n^2\sum_{i=1}^{N_A} s_i^2 = \sigma_n^2||A||_F^2, \end{aligned} \tag{2}$$

wobei s_i die Singulärwerte von $\boldsymbol{A}$ sind. Für unsere Anwendung bedeutet Gl. (2), dass eine möglichst gute Unterdrückung des Messrauschens durch den räumlichen Filter durch eine Minimierung der Frobenius-Norm $||\tilde{\boldsymbol{V}}\boldsymbol{F}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{U}}^T||_F = \sqrt{\sum_i f_i^2/s_i^2}$ erreicht werden kann (f_i und s_i bezeichnen hier die Diagonalelemente von $\boldsymbol{F}$ und $\boldsymbol{\Sigma}$). Wie dies bewerkstelligt werden kann ohne dabei die Bodenbewegungsunterdrückung negative zu beeinflussen wird im Folgenden gezeigt.

3.2 Frequenzabhängiger Entwurf

Zum Entwurf des frequenzabhängigen Filters des Linac-Reglers wurde die klassische loop-shaping Methode für diskrete Systeme gewählt. Der Gesamtfilter $g(z)$ besteht aus vier Einzelkomponenten

$$g(z) = I(z)T(z)P(z)L(z). \tag{3}$$

Den Hauptbestandteil des Entwurfs bildet der Integrator

$$I(z) = \frac{z}{z-1}. \tag{4}$$

In Falle der Verwendung des Verstärkungsfaktors $f_i = 1$, kann gezeigt werden, dass $g(z) = I(z)$ als Regler eine Störung d_k in nur einem Zeitschritt vollständig beseitigen kann, was einem Dead-Beat-Regler entspricht. Das ATL-Bodenbewegungsmodell, welches vor allem für längere Zeiträume Gültigkeit hat, modelliert Bodenbewegungen als Brownsche Bewegung, nimmt also voneinander unabhängige Bodenbewegungsinkremente an. In diesem Fall ist die Wahl $g(z) = I(z)$ sogar optimal in Bezug auf die Bodenbewegungsunterdrückung.

So gut der reine Integrator auch für die Bodenbewegungsunterdrückung geeignet ist, er verstärkt auch Messrauschen stark. Daher muss der Dead-Beat-Regler mit dem nachgeschalteten Tiefpass

$$T(z) = \frac{z\left(1 - e^{-\frac{T_d}{T_1}}\right)}{z - e^{-\frac{T_d}{T_1}}} \tag{5}$$
$$\text{mit} \quad T_d = 0.02\,\text{s} \quad \text{und} \quad T_1 = 0.1\,\text{s}$$

erweitert werden. Die Form von $T(z)$ wurde mit der sog. ζ-Transformation (siehe [6]) ermittelt. Es handelt sich dabei um eine Methode die ausgehend von einem kontinuierlichen System, das diskrete System des abgetasteten kontinuierlichen Systems bestimmt. Die Funktion $T(z)$ wurde ausgehend von einem kontinuierlichen Tiefpass erster Ordnung bestimmt. Der Wert der Zeitkonstante T_1 des kontinuierlichen Tiefpasses wurde durch Simulationen festgelegt. Wie in Abschnitt 4 gezeigt werden wird, führt diese Maßnahme zu einer Lockerung der Toleranzen der Sensorauflösung. Des Weiteren wird durch die künstliche Verlangsamung des Reglers durch $T(z)$ gezeigt, dass die Aktuatoren nicht so hoch-dynamisch sein müssen wie ursprünglich erwartet, was ebenfalls zu einer Lockerung der Aktuatorspezifikationen führt.

Auf Grund der Verwendung von verschiedenen Stabilisationssystemen für das Final Doublet (FD) und den restlichen Beschleunigers ist es notwendig den Filter $P(z)$ zum Entwurf hinzugeführt. Die Übertragungsfunktionen des passiven Dämpfers und der QP-Stabilisation unterscheiden sich über weite Frequenzbereiche stark voneinander (siehe Abb. 4). Dies resultiert in einer zusätzlichen Verschiebung des Final Doublets (FD) vom restlichen Beschleuniger. Besonders stark ist diese Verschiebung um 0.3 Hz da hier die QP-Stabilisation eine Verstärkung von über einen Faktor 10 aufweist und gerade hier das Bodenbewegungsspektrum hohe Komponenten hat. Die Verschiebung zwischen dem FD und dem restlichen Beschleuniger führt zu sekundären Effekten, die zu einem Strahlwachstum am Kollisionspunkt führen. Um auf die Verschiebungen schnell genug reagieren zu können ist es notwendig die Störunterdrückung

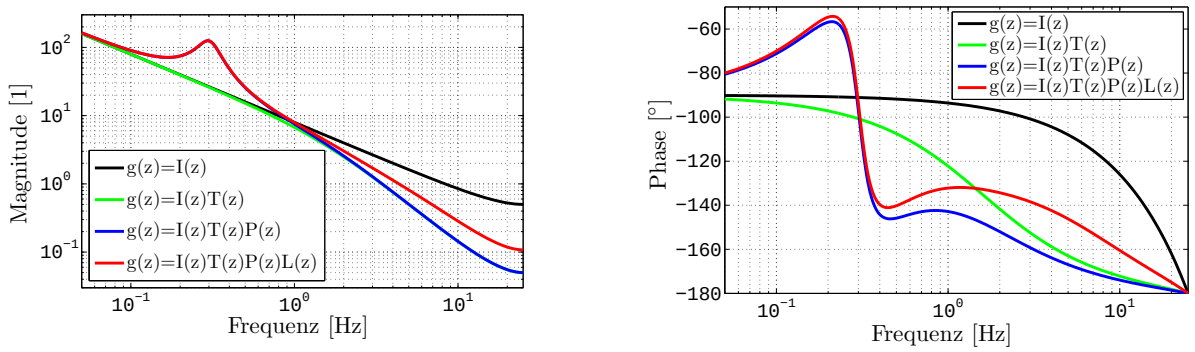


Abbildung 8: Betrag- (links) und Phasengang (rechts) des offenen Regelkreises. Der offene Regelkreis hat die Übertragungsfunktion $(\sigma_i z^{-1})(g(z) f_i \sigma_i^{-1})$, wobei in der Abbildung verschiedene $g(z)$ verglichen wurden. Es sei hier bemerkt, dass die abgebildeten Übertragungsfunktionen diskret sind und daher Symmetrieeigenschaften haben. Die Funktionen sind symmetrisch um 25 Hz und periodisch mit 50 Hz. Das heißt, dass Signale mit $n50$ Hz, wobei $n \in \mathbb{N}$, vom Regler gleich behandelt werden.

für Frequenzen um 0.3 Hz zu stärken. Dies wird durch das Element

$$P(z) = \frac{(1-n_1)(1-n_2)}{(1-z_1)(1-z_2)} \frac{(z-z_1)(z-z_2)}{(z-n_1)(z-n_2)} \tag{6}$$
$$\text{mit} \quad z_{1,2} = e^{(-1.43 \pm 2\pi i 0.2) T_d} \quad \text{und} \quad n_{1,2} = e^{(-0.3 + 2\pi i 0.3) T_d}$$

erreicht. Es handelt sich dabei um jeweils ein konjugiert komplexes Polpaar in Zähler und Nenner. Die Positionen der Pol- und Nullstellen wurde so gewählt, dass es zu einer Verstärkung bei 0.3 Hz mit einer passenden Überhöhungsbreite kommt. Die Positionierung der Pol- und Nullstellen wurde manuell durchgeführt anstelle komplexere Verfahren, wie die Bilinear-Transformation, zu verwenden. Die Schreibweise als e-Potenz wurde gewählt um einfach sicherstellen zu können, dass die Pole und Nullstellen sich immer innerhalb des Einheitskreises der z-Ebene befinden. Der konstante Faktor sichert, dass die Verstärkung von $P(z)$ außerhalb der Überhöhung gegen 1 geht.

Auf Grund der eingeführten Elemente ist die Phasenreserve des offenen Kreises, und damit das Stabilitätsverhalten nicht ausreichend. Das Lead-Glied

$$L(z) = \frac{(1-n_3)(z-z_3)}{(1-z_3)(z-n_3)} \tag{7}$$
$$\text{mit} \quad z_3 = e^{-17 T_d} \quad \text{und} \quad n_3 = e^{-38 T_d}$$

wird verwendet um diese Phasenreserve zu erhöhen. Auch hier wurden die Pol- und Nullstellen per Hand platziert. Wie aus Abb. 8 (rechts) ersichtlich ist, erhält man damit eine Phasenreserve von 36.3°. Wird der noch offene Parameter f_i kleiner 1 gewählt kommt es zu einer Verschiebung der Betragskennlinie nach unten wodurch die Phasenreserve weiter vergrößert wird. Dieses Verhalten ist ein entscheidender Vorteil dieses loop-shaping Entwurfs gegenüber anderen Designmethoden, welche meist die Phasenreserve nur für eine fixe Reglerverstärkung sicherstellen.

3.3 Räumlicher Entwurf

Der im vorherigen Abschnitt festgelegte frequenzabhängige Filter $g(z)$ wird für alle entkoppelten Regelschleifen verwendet. Zusätzlich gibt es einen offenen Verstärkungsfaktor f_i pro Regelschleife um den unterschiedlichen Stör- und Rauschsignale der Regelschleifen Rechnung zu tragen zu. Ziel ist es dabei f_i so zu wählen, dass die Kombination aus Messrauschen und Bodenbewegungsanregung ein möglichst kleines Ausgangssignal hervorrufen. Die Fourier-Transformation des Ausgangssignals $\hat{y}_i$ für die $i$te Regelschleife kann durch

$$\hat{Y}_i(i\omega) = \hat{S}_i(z = e^{i\omega T_d})\hat{D}_i(iw) - \hat{T}_i(z = e^{i\omega T_d})\hat{N}_i(iw) \tag{8}$$

berechnet werden. Dabei ist $\hat{S}_i(z = e^{i\omega T_d})$ die Störungungsübertragungsfunktion und $\hat{T}_i(z = e^{i\omega T_d})$ die Führungsübertragungsfunktion des entkoppelten Kreises. Die Eingangssignale dieses Kreises sind die Störung durch Bodenbewegungen $\hat{D}_i(i\omega)$ und das Messrauschen $\hat{N}_i(i\omega)$. Die Übertragungsfunktionen $\hat{S}_i(z)$ und $\hat{T}_i(z)$ sind gegeben durch

$$\hat{S}_i(z) = z\frac{\hat{G}_i(z)}{1 + \hat{G}_i(z)\hat{C}_i(z)} \quad \hat{T}_i(z) = \frac{\hat{G}_i(z)\hat{C}_i(z)}{1 + \hat{G}_i(z)\hat{C}_i(z)} \tag{9}$$
$$\text{mit} \quad \hat{G}_i(z) = \frac{\sigma_i}{z} \quad \text{und} \quad \hat{C}_i(z) = g(z)\frac{f_i}{\sigma_i}.$$

Dabei sind $\hat{G}_i(z)$ und $\hat{C}_i(z)$ die Übertragungsfunktionen des entkoppelten Beschleunigers und des Reglers. Die Nullstelle bei $z = 0$ in $\hat{S}_i(z)$ stammt von der speziellen Modellierung der Bodenbewegungen, wie in Abb. 6 ersichtlich ist. Um $\hat{Y}_i(i\omega)$ berechnen zu können, müssen auch noch die Fourier-transformierten Eingangssignale $\hat{D}_i(i\omega)$ und $\hat{N}_i(i\omega)$ bekannt sein.

Das Messrauschen $\hat{N}_i(i\omega)$ der entkoppelten Schleifen kann im Zeitbereich durch das ursprüngliche Messrauschen ausgedrückt werden als $\hat{\boldsymbol{n}} = \boldsymbol{U}^T\boldsymbol{W}\boldsymbol{n}$. Daraus ist ersichtlich, dass auch $\hat{\boldsymbol{n}}$ ein mittelwertfreier, normalverteilter Zufallsvektor ist. Der Varianzenvektor der Komponenten von $\hat{\boldsymbol{n}}$ ist durch

$$\boldsymbol{\sigma}_{\hat{n}} = \text{diag}\left(E\left\{\hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^{\boldsymbol{T}}\right\}\right) = \text{diag}\left(\boldsymbol{U}^T\boldsymbol{W}E\left\{\boldsymbol{n}\boldsymbol{n}^T\right\}\boldsymbol{W}^T\boldsymbol{U}\right) \tag{10}$$

gegeben. Es sei darauf hingewiesen, dass die BPM-Auflösungen entlang des Beschleunigers variieren. Dadurch ist $E\left\{\boldsymbol{n}\boldsymbol{n}^T\right\}$ nicht nur einfach eine Einheitsmatrix mit skalarem Multiplikator und es ist daher keine weitere Vereinfachung möglich. Die vorhandene Korrelation zwischen den Rauschsignalen $\hat{n}_i$ wird beim Entwurf vernachlässigt. Da nun die Varianzen bekannt sind kann $\hat{N}_i(iw)$ berechnet werden. Da es sich um einen weißen Zufallsprozess handelt, ist $\hat{N}_i(iw)$ eine konstante Funktion deren Wert so gewählt wird, dass das Spektrum die in Gl. (10) berechnete Varianz aufweist.

Die Erstellung eines Modells für die Bodenbewegungsanregung $\hat{D}_i(iw)$ ist schwieriger als für das Messrauschen. Obwohl es derzeit Bemühungen gibt dieses Spektrum aus den vorhandenen Bodenbewegungsmodellen abzuleiten, wurde für die hier präsentierte Arbeit ein Umweg gewählt. Durch Simulation wurden entsprechende zeitliche Signale generiert, gespeichert und Fourier-transformiert. Da es sich dabei um Zufallsprozesse handelt musste die Simulation mehrmals mit verschiedenen Zufallsgenerator-Anfangswerten durchgeführt werden. Die zugehörigen Spektren wurden anschließen gemittelt.

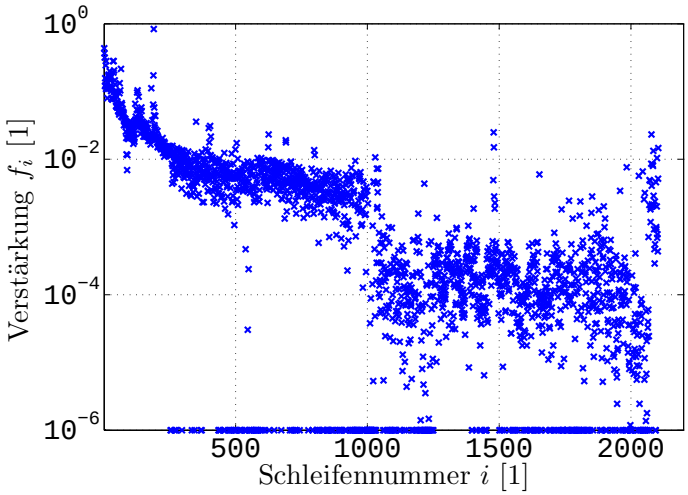


Abbildung 9: Verstärkungsfaktoren f_i der einzelnen, entkoppelten Regelschleifen. Bis auf eine Ausnahme sind alle 2104 f_i deutlich kleiner als 1. Die f_i wurden nach unten mit 10^{-6} künstlich begrenzt, um keine offenen Regelschleifen zu generieren.

Da nun alle notwendigen Übertragungsfunktionen und Eingangssignale bekannt sind, kann der optimale Wert für f_i gefunden werden. Dazu wird die L^2-Norm von $\hat{Y}_i(i\omega, f_i)$ minimiert.

$$\min_{f_i} ||\hat{Y}_i(i\omega, f_i)||_{L^2} = \min_{f_i} \int_{\omega=-\infty}^{+\infty} |\hat{Y}_i(i\omega, f_i)|^2 d\omega \tag{11}$$

Da vor allem $\hat{D}_i(iw)$ nicht in analytischer Form vorliegt, wurde das obige Integral nummerisch für jedes f_i bestimmt. Die Integration wurde dafür nur über einen ausreichend großen Frequenzbereich durchgeführt.

Die Ergebnisse können weiter verbessert werden, wenn in die Optimierung ein weiterer Punkt miteinbezogen wird. Wie bereits in der Einleitung erwähnt, tragen sowohl das Strahlwachstum als auch der Strahlversatz zum Luminositätsverlust bei. Der Linac-Regler beeinflusst beide Effekte mit der gleichen Übertragungsfunktion, da der Messvektoren einer Regelschleife $\boldsymbol{u}_i$ sowohl ein Strahlwachstum wie auch ein einen Strahlversatz verursacht. Beide Effekte werden vom Linac-Regler in der gleichen Art und Weise beeinflusst. Für den Strahlversatz kommt es aber zu einer zusätzlichen Multiplikation des Ausgangssignals mit der Übertragungsfunktion des IP-Reglers $O(z = e^{i\omega T_d})$. Für das Strahlwachstum ist daher das Spektrum $Y_i(i\omega)$ relevant, während der Strahlversatz durch $O(z = e^{i\omega T_d})Y_i(i\omega)$ bestimmt wird. Um diese beiden Signale zu einer gemeinsamen Kostenfunktion zu kombinieren, muss deren Relevanz für den zugehörigen Luminositätsverlust für jede Regelschleife bestimmt werden. Dazu wurde eine Simulation durchgeführt in welcher der Beschleuniger nach den Bodenverschiebungsvektoren $\boldsymbol{v}_i$ (Spalten der Matrix $\boldsymbol{V}$) der entkoppelten Regelschleifen ausgerichtet wurde. Der zugehörige Luminositätsverlust $\Delta\mathcal{L}_i$ entspricht dem Verlust durch Strahlwachstum und Strahlversatz. Als zweiter Schritt wurde der Strahlversatz dieser Strahlen künstlich zu null gemacht. Der zugehörige Luminositätsverlust $\Delta\mathcal{L}_{i,c}$ entspricht ausschließlich dem Luminositätsverlust auf Grund von Strahlwachstum. Die Spektren $\hat{Y}_i(i\omega, f_i)$ und $O(z = e^{i\omega T_d})Y_i(i\omega, f_i)$ können jetzt mit einer Kombination dieser charakteristischen Luminositätsverlustwerte gewichtet werden, um eine neue Kostenfunktion zu formen. Da der

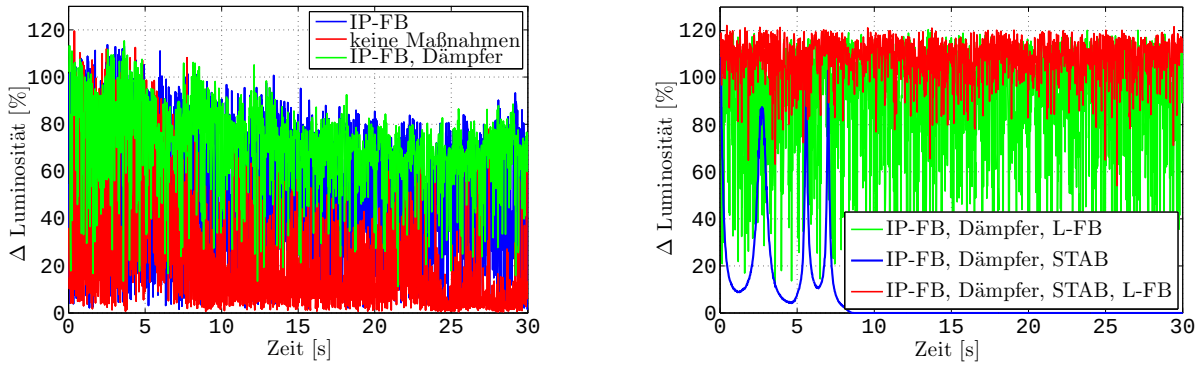


Abbildung 10: Luminositätsverlust mit verschiedenen Bodenbewegungs-Gegenmaßnahmen. In den abbgebildeten Kurven wurde keine Mittelung über verschiedene Anfangswerte des Zufallsgenerators durchgeführt, um ein besseres Gefühl für die Verhältnisse im Echtzeitbetrieb zu vermitteln. Die maximal erreichbare Luminosität ist 121 %. Der Grund dafür ist, dass CLIC beim Entwurf künstlich überdimensioniert wurde, um dem Luminositätsverlust durch Bodenbewegungen Rechnung zu tragen. Bodenbewegungen und andere dynamische Effekte können daher 21 % Luminositätsverlust verursachen und der Beschleuniger arbeitet noch immer mit der nominellen Luminosität von 100 %.

Luminositätsverlust quadratisch von Stahlwachstum und Strahlversatz abhängt, wurde das folgende modifizierte Minimierungsproblem formuliert.

$$\min_{f_i} \int_{\omega=-\infty}^{+\infty} \left[\frac{\Delta\mathcal{L}_{i,c}}{\Delta\mathcal{L}_i} |\hat{Y}_i(i\omega, f_i)|^2 + \frac{\Delta\mathcal{L}_i - \Delta\mathcal{L}_{i,c}}{\Delta\mathcal{L}_i} |O(z = e^{i\omega T_d})\hat{Y}_i(i\omega, f_i)|^2 \right] d\omega \tag{12}$$

Die dadurch berechneten Verstärkungsfaktoren f_i sind in Abb. 9 dargestellt. Der zugehörige räumliche Filter hat eine Frobenius-Norm von nur 14.1, im Vergleich zu 1478.1 für die vollständige Inversion von $\boldsymbol{WR}$.

4 Resultate

Um die Wirksamkeit der vier verschiedenen Bodenbewegungs-Gegenmaßnahmen und im Speziellen des Linac-Reglers zu verifizieren, wurden rechenintensive Simulationen mit der in Abschnitt 2 beschriebenen Simulationsumgebung durchgeführt. Zur Simulation von einer Sekunde Beschleunigerbetrieb wird auf einem durchschnittlichen PC eine Stunde Rechenzeit benötigt. Da die Ergebnisse abhängig von den Anfangswerten des verwendeten Zufallsgenerators sind, mussten mehrere Simulationen mit verschiedenen Angangswerten durchgeführt werden. Die Ergebnisse wurden anschließend gemittelt. Da die dafür benötigte Rechenleistung nicht von einem Rechner (in annehmbarer Zeit) aufgebracht werden kann, wurden die Simulationen am Cluster-System des CERN-Rechenzentrums durchgeführt.

Abbildung 10 zeigt die Auswirkung der verschiedenen Bodenbewegungs-Gegenmaßnahmen auf die Luminosität. Wie man erkennen kann, ist die Luminosität ohne Gegenmaßnahmen (rote Kurve links) schon von der ersten Sekunde an völlig unzureichend. Durch die Verwendung des IP-Reglers (blaue Kurve links, leider etwas verdeckt) werden die Strahlen für niedrige Frequenzen aufeinander gelenkt, was zu einer deutlichen Verbesserung des Ergebnisses führt. Trotzdem sind die hochfrequenten Komponenten noch viel zu stark. Eine

weitere Verbesserung bringt die Verwendung des passiven Dämpfers (grüne Kurve links), wodurch der Strahlversatz verursacht vom Final Doublet stark reduziert wird. Der verbleibende Strahlversatz wird durch die QP-Stabilisation (blaue Kurve rechts) reduziert. Obwohl die QP-Stabilisation hochfrequente Komponenten stark reduziert (Glattheit der entsprechenden Kurve) , verstärkt sie doch Frequenzen um 0.3 Hz stark, was zu einem Strahlwachstum führt und die Luminosität unbrauchbar macht. Dieses Wachstum kann aber durch den Linac-Regler effizient unterdrückt werden (rote Kurve rechts). Ohne QP-Stabilisation könnte der Linac-Regler das gewünschte Ergebnis aber nicht erzielen (grüne Kurve rechts).

Eine Mittelung über 20 Simulationen ergibt, dass die Kombination aller Gegenmaßnahmen zu einer durchschnittlichen Luminosität von 108.5 % führt, was deutlich über dem spezifizierten Wert von 100 % liegt. Ein wichtiger Punkt um dieses Ergebnis erzielen zu können, ist die Fähigkeit des Linac-Reglers das BPM-Messrauschen zu unterdrücken. In Abb. 11 werden verschiedene Regler miteinander verglichen. Der Dead-Beat-Regler (schwarze Kurve) ist sehr sensibel bezüglich Messrauschen und kann nicht verwendet werden. Die Tatsache, dass die entsprechende Kurve für sehr kleine Auflösungen nicht gegen null geht ist ein Artefakt der Simulation. Da der Simulations-Strahl im Vergleich zum echten CLIC-Strahl nur mit relative wenige Teilchen modelliert wird, kommt es zu Schottky-Rauschen in den BPM-Messungen. Der Dead-Beat-Regler ist so sensibel gegenüber Messrauschen, dass schon diese kleinen Fluktuationen die Luminosität in der Simulation stark beeinflussen. Durch die Verwendung des optimierten räumlichen Filters (rote Kurve) kann die Sensibilität entscheidend verringert werden. Die Verwendung des optimierten frequenzabhängigen Filters $g(z)$ (blaue Kurve) führt zu einer weiteren Robustifizierung des Designs. Die grüne Kurve zeigt schließlich, dass dieser Regler viel empfindlicher bezüglich der BPM-Auflösung im Beam Delivery System als im Hauptbeschleuniger ist.

5 Zusammenfassung und Ausblick

Der Teilchenbeschleuniger CLIC ist hoch sensibel gegenüber Bodenbewegungen. Aus diesem Grund wurde die Möglichkeit des Betriebs von CLIC von der Teilchenbeschleuniger-Fachwelt seit jeher in Frage gestellt. Die CLIC-Studie hat daher das Bodenbewegungsproblem als einen der kritischen Punkte genannt, welche in der derzeitigen Phasesphase verifiziert werden. Die vorgelegte Arbeit präsentiert einen entscheidenden Schritt in diesem Verifikationsprozess des Bodenbewegungsproblems.

In Abschnitt 2 wurde eine Simulationsumgebung vorgestellt, die es ermöglicht das komplexe Zusammenspiel von Beschleuniger, Strahl-Strahl-Effekten, Bodenbewegungen und den vier Bodenbewegungs-Gegenmaßnahmen zu modellieren. Dadurch wird es möglich die Gegenmaßnahmen zu testen und untereinander abzustimmen. Noch viel wichtiger ist aber, dass gezeigt werden konnte, dass die zu erwartende Luminosität unter Berücksichtigung von Bodenbewegungen durchschnittlich 108.5 % beträgt, was deutlich über der Spezifikation von 100 % liegt. Dieses Ergebnis zeigt, dass CLIC mit dem angenommenen Bodenbewegungsmodell, welches als eher pessimistisch bezeichnet werden darf, betrieben werden kann.

Ein essentieller Bestandteil zur Erreichung dieses Ergebnisses ist der Linac-Regler, dessen Entwurf in Abschnitt 3 vorgestellt wurde. Dieser Regler besteht aus einem frequenzabhängigen und einem räumlichen Filter. Der präsentierte Entwurf sichert nicht nur die Erreichung des Luminositätsziels von CLIC, sondern führt auch zu einer Lockerung der Sensortoleranzen.

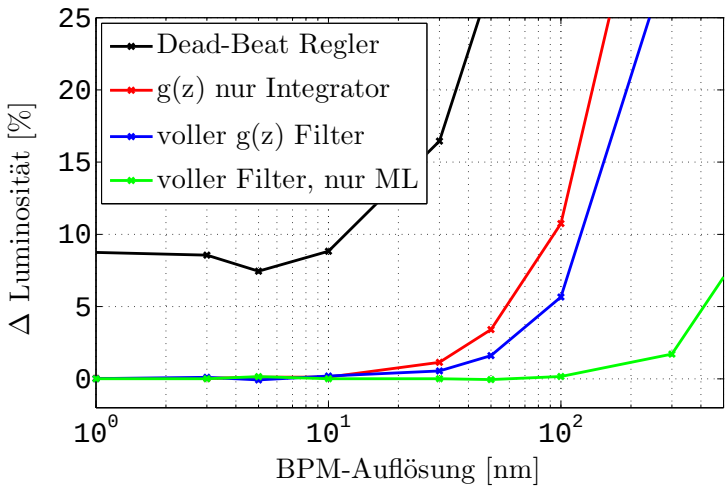


Abbildung 11: Luminositätsverlust auf Grund der BPM-Auflösung. Die schwarze Kurve entspricht der Verwendung eines Dead-Beat-Reglers, welcher durch $\boldsymbol{F} = \boldsymbol{I}$ und $g(z) = I(z)$ gegeben ist. Der zur roten Kurve gehörige Regler verwendet an Stelle der Einheitsmatrix die optimierten f_i für $\boldsymbol{F}$. Für die blaue Kurve wurde zusätzlich $g(z) = I(z)T(z)P(z)L(z)$ verwendet, was dem vollständigen Reglerentwurf entspricht. Auch für die grüne Kurve wurde der vollständige Regler verwendet, es wurden aber perfekte BPMs für das Beam Delivery System angenommen.

Die ursprünglichen verwendeten BPMs hatten eine Auflösung von 10 nm, was an der Grenze des heute möglichen liegt und mit erheblichen Kosten verbunden ist. Durch den Verbesserung des Linac-Reglers konnte diese Toleranz auf 50 nm vergrößert werden, wodurch Standard-BPMs verwendet werden können. Simulationen zeigen, dass für den Hauptbeschleuniger, zumindest aus der Sicht des Linac-Reglers, noch größere BPM-Auflösungen möglich sind. Des Weiteren konnte nachgewiesen werden, dass die Anforderungen an die Dynamik der Aktuatoren geringer ist als ursprünglich angenommen.

Zukünftig wird sich die Arbeit an einer verstärkten gegenseitigen Optimierung der einzelnen Gegenmaßnahmen orientieren. Des Weiteren werden noch detailliertere Robustheitsanalysen vorgenommen werden, obwohl bereits Arbeiten in diesem Bereich durchgeführt wurden. Ein zukünftiges Projekt ist auch die Untersuchung der Möglichkeit weniger Aktuatoren zu verwenden um weitere Kosten einzusparen.

Literatur

[1] P.N. Burrows et al. Performance of the FONT3 fast analogue intra-train beam-based feedback system at ATF. In *Proceedings of the 2006 European Particle Accelerator Conference (EPAC06)*, 2006.

[2] B. Caron, G. Balik, J. Snuverink, J. Pfingstner, and D. Schulte. Integrated simulation of ground motion mitigation techniques for the future compact linear collider (CLIC). *Nuclear Instruments and Methods in Physics Research Section A*, 2011. to be published.

[3] C. Collette et al. Nano-motion control of heavy quadrupoles for future particle colliders:

An experimental validation. *Nuclear Instruments and Methods in Physics Research A*, (643):95 – 101, 2011.

[4] John Ellis and Ian Wilson. New physics with the Compact Linear Collider. *Nature*, 409:431–435, 2001.

[5] A. Gaddi. Dynamic analysis of the final focusing magnets pre-isolator and support system. Technical Report LCD-Note-2010-11, CERN, 2010.

[6] Felix Gausch, Anton Hofer, and Kurt Schlacher. *Digitale Regelkreise*. Oldenbourg Verlag, 1993. ISBN: 3-486-22734-3.

[7] Jürgen Pfingstner et al. Adaptive Scheme for the CLIC Orbit Feedback. In *Proceedings of the 1st International Particle Accelerator Conference (IPAC10)*, 2010.

[8] Y. Renier, P. Bambade, and A. Seryi. Tuning of a 2D ground motion generator for ATF2. Technical Report LAL/RT 08-18, CARE/ELAN document-2008-005, ATF-08-10, LAL, 2008.

[9] J. Resta-López, P.N. Burrows, and G. Christian. Luminosity performance studies of the compact linear collider with intra-train feedback system at the interaction point. *Journal of Instrumentation*, 5(9), 2010.

[10] Hermann Schmickler et al. CLIC Conceptual Design Report. Technical report, CERN, 2011. to be published.

[11] Daniel Schulte. *Study of Electromagnetic and Hadronic Background in the Interaction Region of the TESLA Collider*. PhD thesis, Universität Hamburg, 1996.

[12] Daniel Schulte et al. The Tracking code PLACET. Technical report. https://savannah.cern.ch/projects/placet.

[13] Andrey Sery and Olivier Napoly. Influence of ground motion on the time evolution of beams in linear colliders. *Phys. Rev. E*, 53:5323, 1996.

[14] Vladimir Shiltsev. Observations of random walk of the ground in space and time. *Phys. Rev. Lett.*, 104(23):238501, Jun 2010.

[15] Sigurd Skogestad and Ian Postlethwaite. *Multivariable Feedback Control: Analysis and Design*. Wiley-Interscience, 2005. ISBN: 0-470-01168-8.

Normalformen für flache Systeme

K. Schlacher, M. Schöberl
Johannes Kepler Universität
Institut für Regelungstechnik und Prozessautomatisierung
Altenberger Straße 69, 4040 Linz
kurt.schlacher@jku.at*, markus.schoeberl@jku.at

Zusammenfassung

Können dynamische Systeme auf gewisse Normalformen transformiert werden, dann ist damit nicht nur die Flachheit dieser Systeme gezeigt, sondern die flachen Ausgänge lassen sich dann direkt ablesen. Während diese Transformationen für lineare Systeme sehr einfach zu berechnen sind, ist dieses Problem für Systeme impliziter gewöhnlicher Differentialgleichungen höchstens ansatzweise behandelt. Systeme, die affin in den Ableitungskoordinaten sind, nehmen hier eine Zwischenstellung ein. Dieser Beitrag stellt eine Normalform für diese Systemklasse vor, wobei gezeigt wird, wie mit der Hilfe Pfaff'scher Systeme prinzipiell überprüft werden kann, ob ein dynamisches System auf diese Form gebracht werden kann, und damit die flachen Ausgänge bestimmt werden können.

1 Einleitung

Seit der Einführung der Eigenschaft der Flachheit dynamischer Systeme vor ca. 20 Jahren, siehe z.B. [1] und die dort zu findenden Zitate, wurde diese Eigenschaft für verschiedenste Reglerentwürfe, sei es Trajektorienplanung, Störreglung, etc. ausgenutzt. Grob gesprochen erlaubt Flachheit in diesen Fällen die Rückführung nichtlinearer Probleme auf lineare. Auch wurde dieser Ansatz zumindest teilweise vom konzentriertparametrischen auf den verteiltparametrischen Fall übertragen. Aber das allgemeine Problem im konzentriertparametrischen Fall festzustellen, ob ein System flach ist, ist auch heute noch nicht zufriedenstellend gelöst. So findet man in [2] notwendige und hinreichende Bedingungen, allerdings sind diese nicht einfach anzuwenden. Denn sie erfordern die Lösung von Systemen partieller Differentialgleichungen, deren Lösbarkeit erst selbst nachgewiesen werden muss. Eine konstruktive Methode, dieses Problem zu lösen wird z.B. in [4] vorgestellt.

Dieser Beitrag stellt eine Normalform für die spezielle Klasse der AD-Systeme vor, dies sind nichtlineare Systeme, bei denen die Ableitungskoordinaten affin eingehen. Der Ansatz der sukzessiven Vereinfachung ist eine Spezialisierung von Ideen, die man im Beitrag [3] findet, wobei man sich jetzt auf eine Unterklasse nichtlinearer Systeme einschränkt, die sich als Pfaff'sche Systeme darstellen lassen. Dazu wird im zweiten Abschnitt der lineare und zeitinvariante Fall zur Motivation rekapituliert. Der allgemeine implizite Fall wird im dritten Abschnitt diskutiert. Die Unterklasse der Pfaff'schen Systeme zusammen mit etwas Notation und einigen Grundlagen wird im vierten Abschnitt vorgestellt. Das Verfahren des zweiten Abschnitts wird dann im fünften Abschnitt auf Pfaff'sche Systeme erweitert. Im sechsten Abschnitt findet der Leser ein einfaches Beispiel. Schlussendlich endet dieser Beitrag mit einigen Bemerkungen.

2 Der lineare Fall

Man wähle Koordinaten $x^1, \dots, x^n$ für den Zustand und $u^1, \dots, u^m$ für den Eingang, dann hat ein explizites lineares System die Form

$$x_t^{\alpha_x} \quad = \quad A^{\alpha_x}_{\beta_x} x^{\beta_x} + B^{\alpha_x}_{\alpha_u} u^{\alpha_u} \tag{1}$$

*Korrespondenz bitte an diese Adresse

mit $\alpha_x, \beta_x = 1, \ldots, n$, $\alpha_u = 1, \ldots, m$ und den Ableitungskoordinaten $x_t^{\alpha_x}$. Man stellt nun leicht fest, dass die Anzahl der Koordinaten, die zur Beschreibung des Systems verwendet werden, genau dann minimal ist, wenn das Gleichungssystem $B_{\alpha_u}^{\alpha_x} u^{\alpha_u} = 0$ nur die triviale Lösung besitzt. Dies gelte von jetzt an. Eine interessante Frage ist, ob man mit weniger Koordinaten auskommt. Hierzu löse man nur das Gleichungssystem $M_{\alpha_x}^i B_{\alpha_u}^{\alpha_x} = 0$, $i = 1, \ldots, n-m$ und erhält so das kleinere, allerdings implizite System

$$\begin{aligned} M_{\alpha_x}^i x_t^{\alpha_x} &= M_{\beta_x}^i A_{\alpha_x}^{\beta_x} x^{\alpha_x} \\ &= N_{\alpha_x}^i x^{\alpha_x} \end{aligned}$$

sowie mit der Lösung des Gleichungssystem $\bar{B}_{\alpha_x}^{\alpha_u} B_{\beta_u}^{\alpha_x} = \delta_{\beta_u}^{\alpha_u}$, $\beta_u = 1, \ldots, m$ ein System

$$\bar{B}_{\alpha_x}^{\alpha_u} x_t^{\alpha_x} = \bar{B}_{\alpha_x}^{\alpha_u} A_{\beta_x}^{\alpha_x} x^{\beta_x} + u^{\alpha_u} ,$$

das nach dem Eingang u^{α_u} einfach auflösbar ist. Kennt man nun die Funktionen $x^{\alpha_x}(t)$, dann ist $u^{\alpha_u}(t)$ einfach berechenbar. Für die weiteren Untersuchungen sind also Systeme der Art

$$M_{\alpha_z}^i z_t^{\alpha_z} = N_{\alpha_z}^i z^{\alpha_z} \tag{2}$$

mit $i = 1, \ldots, n$ und den Koordinaten $z^1, \ldots, z^{n_z}$ von Interesse. Um auszuschließen, dass das System (2) algebraische Gleichungen enthält, wird von jetzt an angenommen, dass das Gleichungssystem $\lambda_i M_{\alpha_z}^i = 0$ nur die triviale Lösung zulässt.

Die natürliche Frage analog zu oben ist, ob die Anzahl der gewählten Koordinaten minimal ist. Dies kann man nun sehr einfach beantworten, fasst man die Matrizen $\left[M_{\alpha_z}^i\right]$, $\left[N_{\alpha_z}^i\right]$ als lineare Abbildungen $M, N : \mathbb{R}^{n_z} \to \mathbb{R}^n$ auf. Gilt nämlich $\ker(M) \cap \ker(N) = S \neq \operatorname{span}(\{0\})$, dann sind redundante Koordinaten vorhanden. Dazu wähle man die Zerlegung $\mathbb{R}^{n_z} = R \oplus S$ mit den zugehörigen Basen $\left[R_{\alpha_r}^{\alpha_z}\right]$, $\left[S_{\alpha_s}^{\alpha_z}\right]$, $\alpha_r = 1, \ldots, n_r = n_z - n_s$, $\alpha_s = 1, \ldots, n_s = \dim(S)$. Die Transformation

$$z^{\alpha_z} = R_{\alpha_r}^{\alpha_z} r^{\alpha_r} + S_{\alpha_s}^{\alpha_z} s^{\alpha_s}$$

bringt das gewünschte Ergebnis

$$M_{\alpha_z}^i R_{\alpha_r}^{\alpha_z} r_t^{\alpha_r} = N_{\alpha_z}^i R_{\alpha_r}^{\alpha_z} r^{\alpha_r} .$$

Auch ist es von Interesse, ob das System (2) ein Untersystem enthält, das nach Variablen auflösbar ist, deren zugehörige Ableitungsvariablen nicht im System vorkommen. Die notwendige und hinreichende Bedingung hiefür ist, dass man einen Unterraum $U \subset \mathbb{R}^{n_z}$ findet, für den gilt $U \subset \ker(M)$ und $U \cap \ker(N) = \operatorname{span}(\{0\})$. Wieder führt eine Zerlegung der Art $\mathbb{R}^{n_z} = R \oplus S$ mit den zugehörigen Basen $\left[R_{\alpha_r}^{\alpha_z}\right]$, $\left[S_{\alpha_s}^{\alpha_z}\right]$, $\alpha_r = 1, \ldots, n_r = n_z - n_s$, $\alpha_s = 1, \ldots, n_s = \dim(S)$ und die Transformation

$$z^{\alpha_z} = R_{\alpha_r}^{\alpha_z} r^{\alpha_r} + S_{\alpha_s}^{\alpha_z} s^{\alpha_s}$$

zum gewünschte Ergebnis

$$M_{\alpha_z}^i R_{\alpha_r}^{\alpha_z} r_t^{\alpha_r} = N_{\alpha_z}^i R_{\alpha_r}^{\alpha_z} r^{\alpha_r} + N_{\alpha_z}^i S_{\alpha_s}^{\alpha_z} s^{\alpha_s} ,$$

denn das Gleichungssystem $B_i^{\alpha_s} N_{\alpha_z}^i S_{\beta_s}^{\alpha_z} = \delta_{\beta_s}^{\alpha_s}$ ist lösbar. Man erhält somit das einfach auflösbare System. Gilt noch $U = \ker(M)$, das System hat also keine redundanten Variablen, dann vereinfachen sich die Berechnungen noch ein wenig.

Mit diesen Überlegungen gelangt man zu einer einfachen Zerlegung des Systems (2), wobei wieder angenommen wird, dass es keine redundanten Koordinaten enthält. Dazu führe man neue Koordinaten $r^1, \ldots, r^{n_r}$ und $s^1, \ldots, s^{n_s}$ so ein, dass das System die nachfolgende Form

$$\bar{M}_{\alpha_r}^{\bar{i}} r_t^{\alpha_r} = \bar{N} r^{\alpha_r} \tag{3}$$

$$\hat{M}_{\alpha_r}^{\alpha_s} r_t^{\alpha_r} = \widehat{N}_{\alpha_r}^{\alpha_s} r^{\alpha_r} + s^{\alpha_s} \tag{4}$$

mit $\bar{i} = n - n_s$ annimmt. Hier wird $n_s > 0$ angenommen, gilt $n_s = 0$, dann ist das System (2) autonom, und das jetzt vorgestellte Verfahren ist nicht anwendbar. Kennt man nun eine Lösung von (3), dann kann

man die von (4) einfach durch Einsetzen bestimmen. Nun ist es durchaus möglich, dass das Teilsystem (3) redundante Koordinaten enthält, auch wenn dies auf (2) nicht zutrifft. Man erhält also eine weitere Vereinfachung durch Einführung neuer Koordinaten $y^1, \ldots, y^{n_y}$ und $\bar{z}^1, \ldots, \bar{z}^{n_z}$ so, dass gilt

$$\begin{aligned} \bar{M}^{\bar{i}}_{\alpha_{\bar{z}}} \bar{z}^{\alpha_{\bar{z}}}_t &= \bar{N}^{\bar{i}}_{\alpha_{\bar{z}}} \bar{z}^{\alpha_{\bar{z}}} && (5) \\ A^{\alpha_s}_{\alpha_{\bar{z}}} \bar{z}^{\alpha_{\bar{z}}}_t + B^{\alpha_s}_{\alpha_y} y^{\alpha_y}_t &= C^{\alpha_s}_{\alpha_{\bar{z}}} \bar{z}^{\alpha_{\bar{z}}} + D^{\alpha_s}_{\alpha_y} y^{\alpha_y} + s^{\alpha_s} \, . && (6) \end{aligned}$$

Offensichtlich kann man wieder die Lösung von (6) sehr einfach bestimmen. Dazu muss man nur die von (5) kennen und die Funktionen $y^{\alpha_y}(t)$ vorgeben. Wegen dieser Eigenschaft nennt man die Größen y^{α_y} auch *flache Ausgänge*. Wendet man diese Transformationen nun auf das System (5) an, vorausgesetzt dies ist möglich, dann kommt man zu einer weiteren Reduktion des Systems. Durch Wiederholung dieses Verfahrens gelangt man zu einer Reduktion der Anzahl der Systemgleichungen, wobei es nun zwei mögliche Abbruchskriterien im letzten Schritt gibt. 1) Das System (5) wird leer, es gilt $n_{\bar{z}} = 0$. Die Lösung des Systems (2) ist folglich durch die Vorgabe der flachen Ausgänge eindeutig bestimmt. Man sagt dann auch, das System (2) ist flach. 2) Das System (6) ist leer, es gilt $n_s = 0$. Das System (2) ist nicht flach, es ist sogar nicht erreichbar, da es ein autonomes Teilsystem enthält.

Systeme vom Typ (2) haben noch die Besonderheit, dass man sie immer auf Zustandsform bringen kann. So führt z.B. die Transformation

$$\begin{aligned} x^{\alpha_x} &= M^{\alpha_x}_{\alpha_z} z^{\alpha_z} \\ u^{\alpha_u} &= \bar{M}^{\alpha_u}_{\alpha_z} z^{\alpha_z} \end{aligned}$$

mit $\alpha_x = 1, \ldots, n_z$ auf die Zustandsform (1), wobei noch

$$A^{\alpha_x}_{\beta_x} M^{\beta_x}_{\alpha_z} + B^{\alpha_x}_{\alpha_u} \bar{M}^{\alpha_u}_{\alpha_z} = N^{\alpha_x}_{\alpha_z}$$

gilt. In analoger Weise lassen sich alle impliziten Systeme dieses Abschnitts auf Zustandsform bringen. Es muss jedoch betont werden, dass die Zustandsform für das hier vorgestellte Verfahren keine Bedeutung hat.

3 Der allgemeine implizite Fall

Es ist nun naheliegend zu versuchen, das obige Verfahren auf nichtlineare Systeme zu erweitern. Dazu wählt man wieder Koordinaten $x^1, \ldots, x^n$ für den Zustand und $u^1, \ldots, u^m$ für den Eingang und betrachtet ein explizites nichtlineares System der Form

$$x^{\alpha_x}_t = f^{\alpha_x}(t, x, u) \qquad (7)$$

mit $\alpha_x = 1, \ldots, n$ und den Ableitungskoordinaten $x^{\alpha_x}_t$. Bereits um festzustellen, ob die Anzahl der Koordinaten minimal ist, benötigt man den Satz über implizite Funktionen. Gilt für einen gewählten Punkt (t, x, u) die Bedingung $\operatorname{rank}\left(\left[\partial_{\alpha_u} f^{\alpha_x}(t, x, u)\right]\right) = m$, dann ist die Anzahl minimal, im anderen Fall sind weitere Untersuchungen notwendig. Diese lassen sich wieder relativ einfach durchführen, wenn $\operatorname{rank}\left(\left[\partial_{\alpha_u} f^{\alpha_x}(t, x, u)\right]\right) = \bar{m} < m$ in einer Umgebung eines Punktes t, x, u gilt. Ist die vorige Rangbedingung erfüllt, dann kann man wieder mit Hilfe des Satzes über implizite Funktionen eine Aufspaltung von (7) der Art

$$\begin{aligned} h^i(t, x, x_t) &= 0 \\ g^{\alpha_u}(t, x, x_t, u) &= 0 \end{aligned}$$

mit $i = 1, \ldots, n - m$ erreichen. Möchte man dieses Verfahren fortsetzen, dann muss man folglich mit impliziten Systemen operieren. Für die weiteren Untersuchungen sind also Systeme der Art

$$f^i(t, z, z_t) = 0 \qquad (8)$$

mit $i = 1, \ldots, n$ und den Koordinaten $z^1, \ldots, z^{n_z}$ von Interesse. Um auszuschließen, dass das System (8) algebraische Gleichungen enthält, wird von jetzt an angenommen, dass gilt $\operatorname{rank}\left(\left[\partial^t_{\alpha_z} f^i\right]\right) = n$.

Überträgt man nun die Überlegungen des vorigen Abschnitts auf den Fall (8), dann muss man eine Zerlegung der Art

$$\bar{f}^{\bar{i}}\left(t,\bar{z},\bar{z}_t\right) = 0 \tag{9}$$

$$g^{\alpha_s}\left(t,\bar{z},\bar{z}_t,y,y_t,s\right) = 0\,. \tag{10}$$

konstruieren, wobei gilt $n_y + n_s + n_{\bar{z}} = n_z$ sowie $\bar{i} = 1,\ldots,n-n_s$. Das System (10) erfülle noch die Bedingung $\operatorname{rank}\left(\left[\partial_{\alpha_s} g^{\beta_s}\right]\right) = n_s$ in der Umgebung eines Punktes $t,\bar{z},\bar{z}_t,y,y_t,s$ und ist somit lokal nach s auflösbar. Offensichtlich kann man wieder die Lösung von (10) sehr einfach bestimmen. Dazu muss man nur die von (9) kennen und die Funktionen $y^{\alpha_y}(t)$ vorgeben. Um jedoch diese Zerlegung zu finden muss man nun folgende Probleme für Systeme des Typs (8) lösen: 1) Wie stellt man fest, ob die Anzahl der gewählten Koordinaten von (8) minimal ist? Wenn dies nicht zutrifft, wie findet man die redundanten Koordinaten? 2) Enthält das System (8) ein Untersystem, das nach Variablen auflösbar ist, deren zugehörigen Ableitungsvariablen nicht im System vorkommen? Wie findet man es? Nach Kenntnis der Autoren, sind keine Verfahren bekannt, mit denen man die Lösbarkeit dieser Probleme testen kann, abgesehen davon, es gelingt sie mit ad hoc Methoden zu lösen. Abschließend sei noch festgehalten, dass die Konstruktion einer expliziten Form (7) zum impliziten System (8) eine durchaus anspruchsvolle Aufgabe ist, wenn die Anzahl der Koordinaten minimal sein soll.

Betrachtet man die Analogie zum linearen Fall nochmals, dann stellt man fest, dass das System (7) affin in den Ableitungskoordinanten ist. Gelingt es diese Eigenschaft zu erhalten, dann kann man zu Methoden, die für die Klasse der Pfaff'schen Systeme entwickelt worden sind, greifen. Dieser Ansatz wird im folgenden Abschnitt diskutiert.

4 Pfaff'sche und AD-Systeme

Da sich die Untersuchungen für die Klasse der allgemeinen impliziten Systeme (8) äußerst komplex gestalten, schränkt man sich im Weiteren auf Systeme des Typs

$$a^i_{\alpha_z}\left(t,z\right) z^{\alpha_z}_t = b^i\left(t,z\right) \tag{11}$$

ein, also auf Systeme, bei denen die Ableitungskoordinaten affin eingehen (affine derivatives systems). Da Systeme dieses Typs auch als Pfaff'sche Systeme der Art

$$\omega^i = a^i_{\alpha_z}\left(t,z\right)\mathrm{d}z^{\alpha_z} - b^i\left(t,z\right)\mathrm{d}t \tag{12}$$

mit $P^* = \operatorname{span}\left(B^*\right)$ und Basis $B^* = \left\{\omega^1,\ldots,\omega^n\right\}$ interpretiert werden können, werden im Folgenden einige Grundlagen für diese Systemklassen zusammengefasst.

Im Weiteren werden Pfaff'sche Systeme auf einem Bündel $\mathcal{E} \overset{\pi}{\to} \mathcal{B}$ betrachtet, wobei t die Koordinate der Basismannigfaltigkeit $\mathcal{B}$ und $t, z^{1_z},\ldots,z^{\alpha_z},\ldots,z^{n_z}$ die Koordinaten der totalen Mannigfaltigkeit $\mathcal{E}$ bezeichnen. Die Menge der Schnitte von $\mathcal{E}$ wird mit $\Gamma\left(\mathcal{E}\right)$ abgekürzt, und $C^\infty\left(\mathcal{E}\right)$ ist die Menge der glatten reellen Funktionen auf $\mathcal{E}$. Die Koordinaten des Tangentialbündels $\mathcal{T}\left(\mathcal{E}\right)$ sind $\left(t,z^{1_z},\ldots,z^{\alpha_z},\ldots,z^{n_z},\dot{t}^1,\dot{z}^{1_z},\ldots,\dot{z}^{\alpha_z},\ldots,\dot{z}^{n_z}\right)$ und $\left\{\partial_t,\partial_{1_z},\ldots,\partial_{\alpha_z},\ldots,\partial_{n_z}\right\}$ ist die zugehörige kanonische Basis. Die Koordinaten des Kotangentialbündels $\mathcal{T}^*\left(\mathcal{E}\right)$ sind $\left(t,z^{1_z},\ldots,z^{\alpha_z},\ldots,z^{n_z},\dot{t}_1,\dot{z}_{1_z},\ldots,\dot{z}_{\alpha_z},\ldots,\dot{z}_{n_z}\right)$ und $\left\{\mathrm{d}t,\mathrm{d}z^1,\ldots,\mathrm{d}z^{\alpha_z},\ldots,\mathrm{d}z^{n_z}\right\}$ ist wieder die kanonische Basis. $\mathcal{T}\left(\mathcal{E}\right)$ und $\mathcal{T}^*\left(\mathcal{E}\right)$ sind Vektorbündel. Das Vektorbündel der k-Formen auf $\mathcal{E}$ wird mit $\bigwedge^k\left(\mathcal{T}^*\left(\mathcal{E}\right)\right)$ bezeichnet, wobei die Konventionen $\mathcal{T}^*\left(\mathcal{E}\right) = \bigwedge^1\left(\mathcal{T}^*\left(\mathcal{E}\right)\right)$, $C^\infty\left(\mathcal{E}\right) = \bigwedge^0\left(\mathcal{T}^*\left(\mathcal{E}\right)\right)$ gelten. Auf einem Bündel gibt es bereits eine natürliche Struktur. Das vertikal Bündel $\mathcal{V}\left(\mathcal{E}\right) \subset \mathcal{T}\left(\mathcal{E}\right)$ ist ein Vektorunterbündel, für das gilt $\dot{t}^1 = 0$. Das horizontale Bündel $\mathcal{H}^*\left(\mathcal{E}\right) \subset \mathcal{T}^*\left(\mathcal{E}\right)$ ist ebenfalls ein Vektorunterbündel, auf das $\dot{z}_{\alpha_z} = 0$ zutrifft. Mit der Kontraktion $\rfloor : \mathcal{T}\left(\mathcal{E}\right) \times \bigwedge^k\left(\mathcal{T}^*\left(\mathcal{E}\right)\right) \to \bigwedge^{k-1}\left(\mathcal{T}^*\left(\mathcal{E}\right)\right)$, $k = 1,\ldots,n_z+1$ sieht man sofort, dass $\mathcal{V}\left(\mathcal{E}\right)\rfloor\mathcal{H}^*\left(\mathcal{E}\right) = \operatorname{span}\left\{0\right\}$ gilt.

Hier wird als *Pfaff'sches System* $P^* \subset \mathcal{T}^*\left(\mathcal{E}\right)$ ein Vektorunterbündel von $\mathcal{T}^*\left(\mathcal{E}\right)$ definiert. Eine Menge $B^* = \left\{\omega^1,\ldots,\omega^i,\ldots,\omega^n\right\}$ von Schnitten $\omega^i \in \Gamma\left(\mathcal{T}^*\left(\mathcal{E}\right)\right)$ heißt *Generator* von P^*, wenn $P^* = \operatorname{span}\left(B^*\right)$ gilt, ist die Anzahl n minimal, dann heißt B^* eine *Basis*. Das Vektorunterbündel $\mathrm{d}P^* \subset \bigwedge^2\left(\mathcal{T}^*\left(\mathcal{E}\right)\right)$ ist durch $\mathrm{d}P^* = \operatorname{span}\left(\left\{\mathrm{d}\omega^1,\ldots,\mathrm{d}\omega^i,\ldots,\mathrm{d}\omega^n\right\}\right)$ gegeben. Ein Pfaff'sches System, für

das $P^* \cap \mathcal{H}^*(\mathcal{E}) = \operatorname{span}(\{0\})$ gilt, heißt *nicht degeneriert*. Nun bezeichne $J(\mathcal{E})$ die erste Jetmannigfaltigkeit mit Koordinaten $\left(t^1, z^1, \ldots, z^{n_z}, z_1^1, \ldots, z_1^{n_z}\right)$ und $J_0^1(\mathcal{E})$ das Bündel $J(\mathcal{E}) \overset{\pi_0^1}{\to} \mathcal{E}$, dann gilt für zurückgezogene Bündel $\pi_0^{1,*}(\mathcal{T}^*(\mathcal{E})) = \mathcal{H}^*(\mathcal{E}) \oplus \hat{\mathcal{V}}^*(\mathcal{E})$ mit $\hat{\mathcal{V}}^* = \left\{\varpi^{1_z}, \ldots, \varpi^{i_z}, \ldots, \varpi^{n_z}\right\}$, $\varpi^{i_z} = \mathrm{d}z^{\alpha_z} - z_1^{\alpha_z}\mathrm{d}t$. Formt man das Pfaff'sche System (12) folgendermaßen um

$$\omega^i = a^i_{\alpha_z}\mathrm{d}z^{\alpha_z} - b^i\mathrm{d}t = a^i_{\alpha_z}\varpi^{i_z} + \left(a^i_{\alpha_z}z_1^{\alpha_z} - b^i\right)\mathrm{d}t\ ,$$

dann gilt $\omega^i \in \hat{\mathcal{V}}^*(\mathcal{E})$ genau dann, wenn die Gleichungen (11) erfüllt sind. Für Pfaff'sche Systeme existiert folglich die Abbildung $\operatorname{adsys} : \mathcal{T}^*(\mathcal{E}) \to C^\infty(J(\mathcal{E}))$, die mit $\operatorname{adsys}\left(\omega^i\right) = \left(a^i_{\alpha_z}z_1^{\alpha_z} - b^i\right)$ abgekürzt wird.

Gewisse Pfaff'sche Systeme haben die Eigenschaft, dass sie eine *Aufblätterung* der Mannigfaltigkeit erzeugen. Man sagt dann, ein Pfaff'sches System P^* ist *integrabel*, wenn es eine Basis $B^* = \left\{\mathrm{d}f^1, \ldots, \mathrm{d}f^i, \ldots, \mathrm{d}f^n\right\}$, $f^i \in C(\mathcal{E})$ besitzt, die von vollständigen Differentialen aufgespannt wird. Eine notwendige und lokal hinreichende Bedingung für Integrabilität ist, dass $\mathrm{d}P^* \subset P^* \wedge \mathcal{T}^*(\mathcal{E})$ gilt. Verallgemeinert man diese Bedingung ein wenig, gelangt man zum *abgeleiteten System* $P^{*(1)} \subset P^*$. Es ist das bezüglich der Dimension größte System, dass der Bedingung $\mathrm{d}P^{*(1)} \subset P^* \wedge \mathcal{T}^*(\mathcal{E})$ genügt. Offensichtlich erfüllen integrierbare Systeme die Beziehung $P^{*(1)} = P^*$.

Bereits mit dem Obigen kann man testen, ob ein Pfaff"sches System Q^* eine Zustandsraumdarstellung zulässt, es gilt also $Q^* = \operatorname{span}\left(\left\{\mathrm{d}x^1 - f^1(t,x,u)\,\mathrm{d}t, \ldots, \mathrm{d}x^{n_x} - f^{n_x}(t,x,u)\,\mathrm{d}t\right\}\right)$. Offensichtlich ist $Q^* \oplus \mathcal{H}^*(\mathcal{E})$ integrierbar. Erfüllt nun das System (12) diese Integrabilitätsbedingung, also $P^* \oplus \mathcal{H}^*(\mathcal{E})$ ist integrierbar, dann gilt $P^* = \operatorname{span}\left(\left\{\mathrm{d}x^1(t,z) - f^1(t,z)\,\mathrm{d}t, \ldots, \mathrm{d}x^{n_x}(t,z) - f^{n_x}(t,z)\,\mathrm{d}t\right\}\right)$ für gewisse Funkionen $x^{\alpha_x}(t,z), f^{\alpha_x}(t,z)$. Mit Hilfe der regulären Koordinatentransformation $x^{\alpha_x} = x^{\alpha_x}(t,z)$, $u^{\alpha_u} = u^{\alpha_u}(t,z)$, die immer lokal existiert, erhält man dann das gewünschte Ergebnis.

5 Eine Normalform für Pfaff'sche Systeme

Überträgt man nun die Überlegungen des allgemeinen nichtlinearen Falls, siehe (9, 10), auf Pfaff'sche Systeme, dann muss man Koordinaten $t, s, y, \bar{z}$ so finden, dass die Basis des Systems die Form

$$\bar{\omega}^{\bar{i}} = \bar{a}^{\bar{i}}_{\alpha_{\bar{z}}}(t,\bar{z})\,\mathrm{d}\bar{z}^{\alpha_{\bar{z}}} - \bar{b}^{\bar{i}}(t,\bar{z})\,\mathrm{d}t \tag{13}$$

$$\eta^{\alpha_s} = e^{\alpha_s}_{\alpha_{\bar{z}}}(t,\bar{z},y,s)\,\mathrm{d}\bar{z}^{\alpha_{\bar{z}}} + f^{\alpha_s}_{\alpha_{\bar{y}}}(t,\bar{z},y,s)\,\mathrm{d}y^{\alpha_y} - g^{\alpha_s}(t,\bar{z},y,s)\,\mathrm{d}t \tag{14}$$

hat, wobei das System $\operatorname{adsys}(\eta^{\alpha_s})$ noch lokal nach s auflösbar ist. Auch hier gehen wir wieder davon aus, dass die Anzahl der Koordinaten von (12) minimal ist. Dieses System hat nun folgende Eigenschaften: Es existiert eine Distribution $D = \operatorname{span}(\{\partial_{\alpha_s}\})$, für die gilt $\partial_{\alpha_s}\rfloor\bar{\omega}^{\bar{i}} = 0$, $\partial_{\alpha_s}\rfloor\eta^{\alpha_s} = 0$, $\partial_{\alpha_s}\left(\bar{\omega}^{\bar{i}}\right) = 0$, $\operatorname{rank}\left(\left[\partial_{\alpha_s}\left(\eta^{\beta_s}\right)\right]\right) = n_s$, sowie eine weitere Distribution $Y = \operatorname{span}\left(\left\{\partial_{\alpha_y}\right\}\right)$, die noch die Bedingungen $\partial_{\alpha_y}\rfloor\bar{\omega}^{\bar{i}} = 0$, $\partial_{\alpha_y}\left(\bar{\omega}^{\bar{i}}\right) = 0$ erfüllt. Im Folgenden werden nun diese Bedingungen des speziellen Falls (13, 14) auf den allgemeinen Fall (12) angepasst.

Als erstes soll untersucht werden, wann ein Pfaff'sches System die Form $\bar{P} = \operatorname{span}\left(\left\{\bar{\omega}^1, \ldots, \bar{\omega}^n\right\}\right)$,

$$\bar{\omega}^i = \bar{a}^i_{\alpha_x}(t,x,y)\,\mathrm{d}x^{\alpha_x} - \bar{b}^i(t,x,y)\,\mathrm{d}t \tag{15}$$

zulässt. Offensichtlich besitzt dieses System eine involutive Distribution $\bar{D} = \operatorname{span}\left(\left\{\partial_{1_y}, \ldots, \partial_{n_y}\right\}\right)$, für die $\bar{D}\rfloor\bar{P}^* \oplus \mathcal{H}^*\left(\bar{\mathcal{E}}\right) = \operatorname{span}(\{0\})$ gilt. Besitzt (12) ebenfalls so eine involutive Distribution $D = \operatorname{span}\left(\left\{v_1, \ldots, v_{n_y}\right\}\right)$ für die $D\rfloor P^* \oplus \mathcal{H}^*\left(\bar{\mathcal{E}}\right) = \operatorname{span}(\{0\})$ gilt, dann ist deren Annihilator $D^\perp$ integrierbar, es gilt $D^\perp = \operatorname{span}\left(\left\{\mathrm{d}t, \mathrm{d}x^1(t,z), \ldots, \mathrm{d}x^{n_x}(t,z)\right\}\right)$. Folglich gilt damit aber auch $P = \operatorname{span}\left(\left\{\hat{a}^1_{i_x}(t,z)\,\mathrm{d}x^{i_x}(t,z) - \hat{b}^1(t,z)\,\mathrm{d}t, \ldots, \hat{a}^n_{i_x}(t,z)\,\mathrm{d}x^{i_x} - \hat{b}^n(t,z)\,\mathrm{d}t\right\}\right)$. Mit Hilfe einer Koordinatentransformation, auch diese existiert immer, gelangt man dann zur gewünschten Form.

Als nächstes soll geklärt werden, wann das System (15) lokal eindeutig nach y auflösbar ist. Offensichtlich muss gelten $\dim(D) = n_y = n_{P^*}$. Man beachte, dass die Bedingung $\operatorname{rank}\left(\left[\partial_{i_y}\left(\operatorname{adsys}\left(\bar{\omega}^{\bar{i}}\right)\right)\right]\right) = n_y$ in den ursprünglichen Koordinaten aber die Form $\operatorname{rank}\left(\left[j\left(\partial_{i_y}\right)\left(\operatorname{adsys}\left(\omega^i\right)\right)\right]\right) = n_y$ annimmt, wobei

$j\left(v_{i_y}\right) = v_{i_y}^{\alpha_z}\partial_{\alpha_z} + d_t\left(v_{i_y}^{\alpha_z}\right)\partial_{\alpha_z}^t$, $d_t = \partial_t + z_1^{\alpha_z}\partial_{\alpha_z}$ die Erweiterung von $v_{i_y} = v_{i_y}^{\alpha_z}\partial_{\alpha_z}$ auf $\mathcal{T}\left(J\left(\mathcal{E}\right)\right)$ ist. Der Grund hiefür ist, dass die Funktionen adsys$\left(\omega^i\right)$ auch von den Variablen $z_1^{\alpha_z}$ abhängen.

Zuletzt soll geklärt werden, ob tatsächlich alle Variablen y^{α_y} zur Beschreibung des Systems (15) benötigt werden. Wenn man die Variablen $y = \bar{y}, \hat{y}$ so aufspalten kann, dass das System die Basis

$$\bar{\omega}^i \quad = \quad \bar{a}_{i_x}^i\left(t, x, \bar{y}\right)\mathrm{d}x^{i_x} - \bar{b}^i\left(t, x, \bar{y}\right)\mathrm{d}t \tag{16}$$

besitzt, ist es unabhängig von $\hat{y}$. Offensichtlich gilt $\partial_{\hat{i}}\rfloor\bar{\omega}^i = 0$, $\partial_{\hat{i}}\left(\bar{\omega}^i\right) = 0$. In den ursprünglichen Koordinaten wird nun gefordert, es gibt eine involutive Distribution $D = \mathrm{span}\left(\left\{v_1, \ldots, v_{n_{\hat{y}}}\right\}\right)$, für die gilt $D\rfloor P^* \oplus \mathcal{H}^*\left(\mathcal{E}\right) = \mathrm{span}\left(\{0\}\right)$, $D\rfloor\left(\mathrm{d}P^*\right) \subset P^*$ mit $\dim\left(D\right) = n_{\hat{y}}$. Diese Bedingungen werden offensichtlich vom System (16) erfüllt. Wegen der ersten Bedingung kann (12) auf die Form (15) gebracht werden. Als nächstes bilde man eine neue Basis mit Formen $\bar{\bar{\omega}}^i$ so, dass $\partial_{i_x}\rfloor\bar{\bar{\omega}}^i = \delta_{i_x}^i$, wobei eventuell die Reihenfolge der Variablen x^{α_x} modifiziert werden muss. Für diese spezielle Basis gilt dann $\partial_{\hat{i}}\left(\bar{\bar{\omega}}^i\right) = 0$.

Mit Hilfe des Obigen kann man nun ein Verfahren so finden, dass das System (12) auf die Form (13,14) gebracht werden kann. Im ersten Schritt konstruiere man eine Auftrennung der Art $P^* = P_C^{*,1} \oplus P^{*,1}$ und eine involutive Distribution D so, dass gilt $D\rfloor P^* \oplus \mathcal{H}^*\left(\mathcal{E}\right) = \mathrm{span}\left(\{0\}\right)$, $D\rfloor\left(\mathrm{d}P^{*,1}\right) \subset P^{*,1}$ und das Paar D, $P_C^{*,1}$ erfüllt die Bedingungen der Auflösbarkeit von $P^{*,1}$ nach den durch D gegebenen Variablen. Man beachte, dass dies im Allgemeinen ein nichtlineares Problem ist, bei dem auch lineare partielle Differentialgleichungen zu lösen sind. Im zweiten Schritt bestimme man eine Distribution Y für die $Y\rfloor P^{*,1} \oplus \mathcal{H}^*\left(\mathcal{E}\right) = \mathrm{span}\left(\{0\}\right)$, $Y\rfloor\left(\mathrm{d}P^{*,1}\right) \subset P^{*,1}$ gilt. Dies ist ein lineares Problem, und seine Lösung liefert die flachen Ausgänge. Die wiederholte Anwendung dieses Verfahrens liefert dann weitere flache Ausgänge. Falls das Verfahren jedoch nicht mehr fortgesetzt werden kann, folgt daraus nicht, dass das System (8) nicht lokal erreichbar ist. Bemerkenswert ist auch, dass für Systeme mit einem Eingang, die eine Zustandsraumdarstellung zulassen, diese Eigenschaft während des Reduktionprozesses nicht verlieren. Dies reflektiert die Tatsache, dass Eingrößensysteme genau dann flach sind, wenn sie eingangszustandslinearisierbar sind.

6 Ein einfaches Beispiel

Man betrachte das Pfaff'sche System P^* mit Basis B^*,

$$B^* = \left\{\mathrm{d}x^1 - u^1\mathrm{d}t, \mathrm{d}x^2 - u^2\mathrm{d}t, \mathrm{d}x^3 - u^1u^2\mathrm{d}t\right\} .$$

nach einigen Umformungen erhält man das System $P^{*,1}$ mit Basis $B^{*,1}$

$$B^{*,1} \quad = \quad \left\{\mathrm{d}x^2 - u^2\mathrm{d}t, \mathrm{d}x^3 - u^2\mathrm{d}x^1\right\} .$$

Offensichtlich ist das Komplement $P_C^{*,1}$ mit Basis $\left\{\mathrm{d}x^1 - u^1\mathrm{d}t\right\}$ nach u^1 auflösbar. Im zweiten und letzten Schritt berechnet man $P^{*,2}$ mit Basis $B^{*,2}$,

$$B^{*,2} = \left\{\mathrm{d}x^2 - u^2\mathrm{d}t\right\} .$$

Das Komplement $P_C^{*,2}$ ist wieder einfach auflösbar, die entsprechenden Größen werden aber erst nach einer Koordinatentransformation sichtbar. Dieses System ist folglich flach. Als nächstes führen wir gleich alle Koordinatentransformation in einem durch und erhalten mit

$$\begin{aligned} y^{1,1} &= x^2 \\ \hat{z}^{2,1} &= u^2 \\ y^{2,1} &= x^3 - x^1u^2 \\ \hat{z}^{3,1} &= x^1 \\ \hat{z}^{4,1} &= u^1 \end{aligned}$$

das spezielle Pfaff'sche System

$$\begin{matrix} \mathrm{d}y^{1,1} & & & -\hat{z}^{2,1}\mathrm{d}t \\ & \mathrm{d}y^{2,1} + \hat{z}^{3,1}\mathrm{d}\hat{z}^{2,1} & & \\ & & \mathrm{d}\hat{z}^{3,1} & -\hat{z}^{4,1}\mathrm{d}t . \end{matrix}$$

Im letzten Schritt versuchen wir noch das System in den flachen Ausgängen und ihren Ableitungen darzustellen. Die Koordinatentransformation

$$\begin{aligned} y^{1,1} &= y_0^1 \\ \hat{z}^{2,1} &= y_1^1 \\ y^{2,1} &= y_0^2 \\ \hat{z}^{3,1} &= -\frac{y_1^2}{y_2^1} \\ \hat{z}^{4,1} &= \frac{y_3^1 y_1^2 - y_2^1 y_2^2}{\left(y_2^1\right)^2} \end{aligned}$$

führt dann zu

$$\begin{aligned} & \mathrm{d}y_0^1 - y_1^1 \mathrm{d}t \\ & \left(\mathrm{d}y_0^2 - y_1^2 \mathrm{d}t\right) - \frac{y_1^2}{y_2^1}\left(\mathrm{d}y_1^1 - y_2^1 \mathrm{d}t\right) \\ & \frac{y_1^2}{\left(y_2^1\right)^2}\left(\mathrm{d}y_2^1 - y_3^1 \mathrm{d}t\right) - \frac{1}{y_2^1}\left(\mathrm{d}y_1^2 - y_2^2 \mathrm{d}t\right). \end{aligned}$$

Man beachte, dass dieses System keinen horizontalen Anteil hat.

7 Schlussbemerkungen

In diesem Beitrag wurde eine Methode vorgestellt, mit deren Hilfe man Pfaff'sche Systeme so schrittweise vereinfachen kann, dass man die flachen Ausgänge bestimmen kann, sofern diese existieren. Den Ausgangspunkt der Betrachtungen bildeten lineare und zeitinvariante Systeme, für die es natürlich einfach ist festzustellen, ob sie flach sind. Dabei wurde der Prozess der Systemreduktion so gestaltet, dass er auf nichtlineare Systeme erweiterbar wurde. Allgemeine nichtlineare Systeme erwiesen sich als zu kompliziert, für die Unterklasse der Pfaff'schen Systeme konnte jedoch ein Verfahren entwickelt werden, dass die konsequente Weiterentwicklung des linearen Falls ist. Denn wendet man es auf die Klasse der linearen Systeme an, dann vereinfacht es sich genau zu dem anfangs vorgestellten Verfahren.

8 Danksagung

Markus Schöberl ist APART Stipendiat der Österreichischen Akademie der Wissenschaften. Kurt Schlacher dankt dem Austrian center of competence in Mechatronics (ACCM) für die Unterstützung der Arbeiten zu diesen Beitrag.

Literatur

[1] M. Fliess, J. Levine, P. Martin, and P. Rouchon. Flatness and defect of nonlinear systems: introductory theory and examples. *Int. Journal of Control*, 61:1327–1361, 1995.

[2] J. Levine. *Analysis and Control of Nonlinear Systems: A Flatness-based Approach.* Springer, Berlin, 2009.

[3] K. Schlacher and M. Schöberl. Construction of flat outputs by reduction and elimination. *Proceedings 7th IFAC Symposium on Nonlinear Control Systems (NOLCOS)*, 2007.

[4] M. Schöberl, K. Rieger, and K. Schlacher. System parametrization using affine derivative systems. In *Proceedings 19th International Symposium on Mathematical Theory of Networks & Systems (MTNS)*, pages 1737–1743, 2010.

Eine minimale Schaltung für mehrdimensional passive Systeme

Karlheinz Ochs
Lehrstuhl für Nachrichtentechnik,
Ruhr-Universität Bochum
ochs@nt.rub.de*

Zusammenfassung

In diesem Aufsatz wird die Synthese einer mehrdimensional passiven Schaltung vorgestellt, die von einem passiven physikalischen System ausgeht, das durch ein gewöhnliches bzw. partielles Differentialgleichungssystem in Zustandsraumdarstellung beschrieben wird. Die zugehörigen Systemmatrizen dürfen in gewisser Weise variabel sein, so dass das System zeit- und ortsvariant sowie auch nichtlinear sein kann. Es resultiert eine regelmäßig aufgebaute elektrische Schaltung mit einer minimalen Anzahl von passiven Bauelementen. Das zugehörige Schaltbild ist eine grafische Darstellung der Differentialgleichung und begünstigt auf diese Weise eine funktionale Analyse unter Berücksichtigung energetischer Eigenschaften. Dieser Aspekt ist insofern bedeutsam, als dass eine mental beherrschbare Analyse des Algorithmus zur digitalen Nachbildung möglich wird, obgleich keine analytische Lösung für die Differentialgleichung existiert.

1 Einleitung

Eine spezielle Anwendung der Digitalen Signalverarbeitung ist die digitale Nachbildung elektrischer Schaltungen, die als Programme auf einem digitalen Signalprozessor oder als integrierte Schaltungen implementiert werden [1]. Die Nachbildung kann auch auf physikalische Systeme ausgedehnt werden, wenn zu diesen eine elektrische Repräsentation angegeben werden kann. Die digitale Nachbildung unterliegt strengen Anforderungen, da sie meistens in rückgekoppelte Systeme eingebaut wird und somit relativ leicht ein instabiles Gesamtsystem verursachen kann. Um ein gutartiges Stabilitätsverhalten zu erreichen, verwendet man für die digitale Nachbildung vornehmlich passive Systeme.

Als besonders schwierig erweist sich die digitale Nachbildung, wenn das gewünschte Systemverhalten lediglich durch ein gewöhnliches oder partielles Differentialgleichungssystem spezifiziert ist. Für ein passives System ist hierzu eine korrespondierende elektrische Schaltung zu finden, die ausschließlich aus Quellen und passiven Elementen besteht, die über ein Kichhoff'sches Netz miteinander verbunden sind. Je nachdem, ob das System durch eine gewöhnliche oder partielle Differentialgleichung spezifiziert ist, ist dabei zwischen gewöhnlicher und mehrdimensionaler Passivität zu unterscheiden. Die Synthese einer derartigen Schaltung gelingt oftmals nur, wenn der Entwickler fundierte Kenntnisse des nachzubildenden Systems besitzt und zudem eine langjährige Erfahrung in der Schaltungssynthese hat.

*Korrespondenz bitte an diese Adresse

Zumindest in Form eines Programms ist eine digitale Nachbildung gegenüber einer elektrischen Schaltung flexibler, denn das Systemverhalten kann durch den Austausch von Software problemlos modifiziert, erweitert oder ausgewechselt werden. Bedingt durch Alterung, Temperaturdrift, Leistungsschwankungen usw. unterliegt eine elektrische Schaltung stets unerwünschten Veränderungen, die mit schaltungstechnischen Maßnahmen zu kompensieren sind. Die dafür vorgesehenen Teile der elektrischen Schaltung, wie zum Beispiel die Stabilisierung eines Arbeitspunktes, sind nicht nachzubilden, da der für die digitale Nachbildung zu programmierende Algorithmus von derartigen Veränderungen unberührt bleibt. Für eine aufwandsarme digitale Nachbildung ist es somit erforderlich, den funktionalen Teil der Schaltung zu identifizieren und zu extrahieren. Eine digitale Implementierung hat zudem den Vorteil, dass man ideale Bauelemente mit exakt vorgegebenen Werten realisiert. Damit wird einerseits eine identische Reproduktion möglich und andererseits können technologische Grenzen analoger elektrischer Schaltungen überwunden werden. Auf diese Weise gelingt es beispielsweise, eine perfekt abgeglichene Brückenschaltung mit idealen Kondensatoren und Spulen zur Frequenzselektion als Digitalfilter zu implementieren [2], die wegen unvermeidbarer Fertigungstoleranzen als Analogfilter kaum nutzbar ist.

Aus der bisherigen Argumentation ist schon zu schließen, dass es für die digitale Nachbildung nur auf eine Spezifikation des geforderten Systemverhaltens ankommt, weshalb sich nicht nur reale, sondern auch ideale elektrische Schaltungen digital nachbilden lassen. Eine mögliche Anforderungsspezifikation des Systemverhaltens kann die Differentialgleichung einer elektrischen Schaltung sein. Aus dieser Perspektive erscheint die elektrische Schaltung als Repräsentant einer Differentialgleichung, zu der ihre digitale Nachbildung eine numerische Lösung generiert. Dieses Vorgehen eignet sich zwar für eine Simulation, eine digitale Nachbildung unterliegt aber ungleich strengeren Anforderungen. Denn als Teil eines technischen Systems muss sie das gewünschte Verhalten zuverlässig über lange Zeitspannen hinweg nachbilden und für eine echtzeitfähige Signalverarbeitung mit möglichst geringen Wortlängen auskommen. Erfolgt die Implementierung auf einem digitalen Signalprozessor mit Festkomma-Arithmetik oder als integrierte Schaltung, so werden die Signale und Multipliziererkoeffizienten besonders grob quantisiert. Um Instabilitäten zu vermeiden, sind diese Quantisierungen bereits beim Entwurf und der Synthese einzukalkulieren. Allerdings ist eine isolierte Betrachtung der digitalen Nachbildung nicht ausreichend, denn sie wird in der Regel nicht separat, sondern als Baustein eines Gesamtsystems betrieben, das sowohl analoge als auch digitale Bausteine enthalten kann. Selbst wenn die digitale Nachbildung stabil ist, kann es bei einem Einbau in ein rückgekoppeltes System zur Instabilität und somit zur Unbrauchbarkeit des Gesamtsystems kommen. Eine Berücksichtigung des Gesamtsystems beim Entwurf der digitalen Nachbildung hat zwei wesentliche Nachteile: Einerseits wird der Entwurf drastisch erschwert, wenn nicht sogar unmöglich und andererseits ist der Baustein durch die Spezialisierung nicht ohne Weiteres für eine andere Anwendung wiederverwendbar. Um diesem Problem zu begegnen, beschränkt sich das *Wellendigitalkonzept* bei der digitalen Nachbildung nicht auf Funktionalität, sondern berücksichtigt physikalische Eigenschaften [2]. Der wichtigste Aspekt ist hierbei, die Passivität einer nachzubildenden elektrischen Schaltung auf das *Wellendigitalmodell* zu übertragen, wobei die Passivität selbst bei einer endlichen Wortlänge für die Darstellung der Signale und Multipliziererkoeffizienten auf einfachste Weise sichergestellt werden kann [3]. Die passiven *Wellendigitalbausteine* sind universell mit passiven analogen oder digitalen Bausteinen kombinierbar, da die Kombination ein passives Gesamtsystem mit gutartigem Stabilitätsverhalten bildet.

Ursprünglich ist das Wellendigitalkonzept zum Entwurf und zur Synthese von Digitalfiltern eingesetzt worden, die auf einer Nachbildung von passiven Analogfiltern basieren [1]. Die resul-

tierenden *Wellendigitalfilter* profitieren dabei von der geringen Empfindlichkeit der Analogfilter gegenüber Toleranzen der Bauelementwerte, die sich in einer geringen Koeffizientenempfindlichkeit des Digitalfilters niederschlägt. Wellendigitalmodelle weisen eine Reihe wünschenswerter Eigenschaften auf, die zu robusten Algorithmen führen [4]. Neben einer geringen Sensitivität gegenüber Parameterschwankungen und äußeren Störeinflüssen gehören unter anderem ein geringes Rundungsrauschen, inhärente Stabilität und die Freiheit von parasitären Oszillationen dazu [2].

Die genannten Vorteile machen das Wellendigitalkonzept ebenfalls für die numerische Lösung von Differentialgleichungen, die durch eine elektrische Schaltung repräsentiert werden, attraktiv [5]. Mit dieser Methodik lassen sich nichtlineare gewöhnliche Differentialgleichungen von physikalischen Systemen erfolgreich simulieren [6], [7], [8], [9]. Parallel zu dieser Entwicklung ist der Begriff einer elektrischen Schaltung abstrakt mathematisch erweitert worden, wodurch die korrespondierenden Wellendigitalmodelle auch partielle Differentialgleichungen numerisch lösen können [10], [11], [12]. Weil das aus der Physik bekannte *Nahwirkungsprinzip* auf die Wellendigitalmodelle übertragen wird, sind die entstehenden Algorithmen für eine massive Parallelverarbeitung prädestiniert [4], deren effiziente Programmierung auf Mehrkern-Prozessoren mit gemeinsam genutztem Speicher nachgewiesen worden ist [13]. Aus mathematischer Sicht ist das unterstellte Nahwirkungsprinzip eine Einschränkung auf solche partiellen Differentialgleichungen, bei denen sich alle physikalischen Phänomene mit endlicher Geschwindigkeit ausbreiten. Für physikalische Systeme formulierte partielle Differentialgleichungen, die eine unendlich hohe Ausbreitungsgeschwindigkeit gestatten, beruhen auf weitgehenden Idealisierungen. Solche können durch eine tiefere Modellierung des physikalischen Systems, die das Nahwirkungsprinzip mit einbezieht, behoben werden [4]. Im *Wellendigitalbereich* führen unendlich hohe Ausbreitungsgeschwindigkeiten zu verzögerungsfreien gerichteten Schleifen, die nicht realisierbare Algorithmen zur Folge haben. Insofern verwundert es nicht, dass sich in gesonderten Fällen die tiefere physikalische Modellierung als ein Iterationsverfahren zum Aufbrechen von verzögerungsfreien gerichteten Schleifen herausstellt [14].

Um passive Wellendigitalmodelle zu erhalten, eignen sich für die Diskretisierung der Differentialgleichungen nur spezielle numerische Integrationsverfahren, wie zum Beispiel A-stabile lineare Mehrschritt-Verfahren, von denen die Trapez-Regel die höchste Genauigkeit erreicht [15]. Da die Trapez-Regel zudem verlustfrei ist, ist die digitale Nachbildung, wenn von Rundungseffekten einmal abgesehen wird, energetisch exakt. Diese positiven Eigenschaften und eine zudem sehr einfache Implementierung im Wellendigitalbereich machen die Trapez-Regel für eine Wellendigitalmodellierung besonders attraktiv. Bei diversen Anwendungen führt die Verlustfreiheit allerdings zu Problemen, denen man durch einen Entwurf geeigneter passiver linearer Mehrschritt-Verfahren begegnen kann [5], [16], [17]. Die resultierenden Wellendigitalmodelle benötigen generell eine Startphase [18] und bei partiellen Differentialgleichungen zudem eine gesonderte Randbehandlung [19]. Während Passivität und A-Stabilität synonym bei linearen Mehrschritt-Verfahren sind, trifft dies für Runge-Kutta-Verfahren nicht zu [20]. Hier ist Passivität eine neue Eigenschaft, die intrinsisch zu einer Reihe wünschenswerter numerischer Stabilitätseigenschaften führt [21]. Im Besonderen lassen sich passive Runge-Kutta-Verfahren entwerfen, die eine höhere numerische Genauigkeit als die Trapez-Regel aufweisen [20], [22]. Ihr Entwurf ist mathematisch höchst anspruchsvoll, weil gekoppelte Systeme von polynomialen Gleichungen und Ungleichungen in mehreren Variablen zu lösen sind [23], [24]. Beträchtlich schwieriger gestaltet sich der Entwurf passiver Mehrschritt-Runge-Kutta-Verfahren, von denen bislang nur wenige Verfahren entwickelt werden konnten [25], [26]. Ohne die Passivität numerischer Integrationsverfahren zu opfern, kann man ihre Genauigkeit durch Extrapolation

steigern [27] und den Aufwand mit einer geeigneten Schrittweitensteuerung reduzieren [28].

An dieser Stelle ist erneut zu betonen, dass bei einer Wellendigitalmodellierung die digitale Nachbildung und nicht die numerische Lösung von Differentialgleichungen im Vordergrund steht. Die Simulation eines physikalischen Systems dient vielmehr der Verifikation der digitalen Nachbildung gegenüber den spezifizierten Anforderungen. Hierfür werden Startwerte und Parameter variiert, um Modellgrenzen zu eruieren, Besonderheiten auszuwerten, Ergebnisse zu interpretieren und gegebenenfalls das Wellendigitalmodell zu vervollkommnen. Erst wenn die Simulation den Anforderungsspezifikationen genügt, wird das Wellendigitalmodell als digitale Nachbildung akzeptiert und implementiert. Für die Wellendigitalmodellierung physikalischer Systeme, die durch Differentialgleichungen beschrieben werden, existieren zahlreiche Anwendungen, von denen stellvertretend [16], [12], [29], [30] mit Beispielen aus der Elektrotechnik, der Fluiddynamik sowie der Kernphysik genannt werden. Eine digitale Nachbildung eignet sich nicht nur als Ersatz, sondern auch zur Beobachtung, um beispielsweise mit einem Wellendigitalmodell eine Näherung für unzugängliche innere Zustände eines physikalischen Systems zu gewinnen [31], [32]. Beeinflusst diese Näherung über eine Rückkopplung Stellgrößen des Systems, so liegt eine digitale Regelung vor. Eine Implementierung auf der Basis des Wellendigitalkonzepts erscheint selbst für sicherheitskritische Anwendungen geeignet, weil die grafische Darstellung von Wellendigitalmodellen mit ihrer klaren Untergliederung in Bausteine den Abstraktionsgrad verringert und somit eine formale Verifikation für die Korrektheit des Programms begünstigt [33]. Erfreulicherweise existiert ein mit formalen Methoden geführter Beweis für die Korrektheit von Software, die symmetrisch hyperbolische Differentialgleichungen mit einem Wellendigitalmodell numerisch integriert, vgl. [34], [35], [36], [37].

Die größte Schwierigkeit bei der Wellendigitalmodellierung physikalischer Systeme ist die Synthese einer elektrischen Schaltung, die sich als Referenz für ein realisierbares Wellendigitalmodell eignet und kurz *Referenzschaltung* genannt wird. Die gängige Praxis besteht darin, das physikalische System durch ein mathematisches Modell zu erfassen und eine passende elektrische Schaltung zu konstruieren, wobei die Berücksichtigung energetischer Beziehungen und physikalischer Anschauung erfahrungsgemäß probate Mittel sind. Dieses Vorgehen verlangt dem Entwickler nicht nur ein tiefes physikalisches Verständnis für das nachzubildende System ab, sondern es sind zudem fundierte Kenntnisse der Schaltungstheorie erforderlich, die meist erst mit einer gewissen Intuition zum Ziel führen. Die Ursache hierfür ist der Verlust von Strukturinformation, wenn ein physikalisches System durch ein mathematisches Modell abstrahiert wird. Aus diesem Grund ist es oft günstiger, anstelle des abstrakten mathematischen Modells als Ausgangspunkt die ursprünglichen Gleichungen zu wählen, also beispielsweise bei einem mechanischen System Bilanzgleichungen für Kräfte, Impulse, Energien und dergleichen zu nutzen. Darüber hinaus ist es häufig hilfreich, nicht direkt mit dem Gesamtsystem anzufangen, sondern einen beherrschbaren Spezialfall auszuwählen, zu dem eine elektrische Schaltung angegeben werden kann. Mit einer schrittweisen Hinzunahme von zunächst vernachlässigten Effekten ist die Hoffnung verbunden, sukzessive zur gesuchten elektrischen Schaltung zu gelangen.

In diesem Aufsatz werden passive physikalische Systeme betrachtet, die durch ein System von partiellen Differentialgleichungen erster Ordnung beschrieben werden. Ausgehend von der mathematischen Beschreibung wird eine elektrische Schaltung synthetisiert, die aus nicht gekoppelten Spulen, Widerständen, Gyratoren, Übertragern, Spannungsquellen, Reihen- und Parallelverbindungen besteht. Auf diese Weise wird Strukturinformation in Form einer elektrischen Schaltung zurückgewonnen, deren Schaltbild den modularen Aufbau widerspiegelt und eine Analyse der repräsentierten Differentialgleichung erleichtert. Die elektrische Schaltung leistet somit einen Transfer von einer Domäne der Physik hin zur Elektrotechnik und eröffnet somit die

Möglichkeit zur Beurteilung eines physikalischen Systems aus elektrotechnischer Sicht. Anhand der Bauelementwerte der elektrischen Schaltung kann direkt und zuverlässig über die Eigenschaft der Passivität entschieden werden. Selbstverständlich ist das Ziel der Schaltungssynthese eine wohldefinierte Schaltung, die per definitionem eine eindeutige Lösung für die Spannungen und Ströme besitzt.

2 Mehrdimensional passive Systeme

2.1 Mehrdimensionale Kausalität

Um die Begriffe der Kausalität und Passivität für mehrdimensionale Schaltungen einzuführen, wird das Nahwirkungsprinzip zugrunde gelegt. Letzteres heißt auch *Lokalitätsprinzip* und postuliert ein Ausbreiten aller Phänomene mit endlicher Geschwindigkeit. Für eine Erläuterung wird das Bild 1 aus [38] herangezogen, in dem verschiedene Szenarien für die zeitliche und örtliche Abfolge einer Reaktion auf eine Aktion in einem durch • markierten Referenzpunkt illustriert sind [16]. Die Gesamtheit aller Punkte, von denen eine Aktion eine Reaktion an einem Punkt hervorrufen kann, ist das Abhängigkeitsgebiet $\mathcal{A}$. Dagegen umfasst das Wirkungsgebiet $\mathcal{R}$ alle Punkte, an denen eine Aktion in einem Punkt die Reaktion beeinflussen kann.

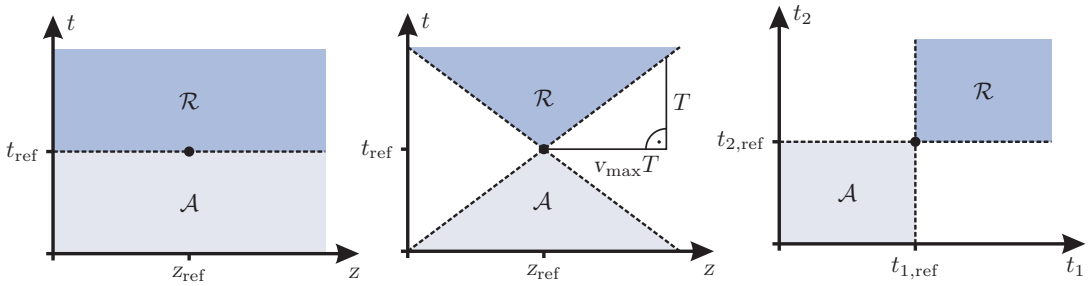


Abbildung 1: Erläuterung zur Kausalität und Lokalität bei nur einer Ortskoordinate: Kausal, aber nicht lokal (links), kausal und lokal (Mitte), mehrdimensional kausal (rechts).

Links in diesem Bild ist die Situation für ein kausales, aber nicht lokales System dargestellt, wobei im Koordinatensystem ein Referenzpunkt zum Zeitpunkt t_{ref} am Ort z_{ref} markiert ist. Kausalität ist hier gegeben, da alle Reaktionen nicht durch zukünftige Aktionen beeinflusst werden. Das Nahwirkungsprinzip ist jedoch verletzt, weil eine Aktion in diesem Punkt zum selben Zeitpunkt t_{ref} eine Reaktion in einem entfernten Punkt bewirken kann. Dies impliziert eine unendlich hohe Ausbreitungsgeschwindigkeit und steht somit im Widerspruch zur Tatsache, dass die Lichtgeschwindigkeit eine oberste Grenze für alle Ausbreitungsvorgänge ist. Im weiteren Verlauf werden Kausalität und das Nahwirkungsprinzip stillschweigend vorausgesetzt, womit die in der Mitte des Bildes 1 gezeigte Situation vorliegt, bei der die Beträge aller Ausbreitungsgeschwindigkeiten durch eine maximale Geschwindigkeit v_{max} beschränkt sind. Bei einem kausalen und lokalen System muss folglich immer eine gewisse Zeit verstreichen, bis eine Aktion in einem Punkt eine Reaktion an einem anderen Ort verursacht. Diese Aussage ist offenbar auf ein Koordinatensystem bezogen, bei dem die Zeit und die Ortskoordinaten als unabhängige Variablen auftreten. Wählt man dagegen neue Koordinaten t_1 und t_2, indem man die alten Koordinaten skaliert und anschließend das Koordinatensystem dreht, so kann die rechts im Bild 1 dargestellte Situation erreicht werden. Sie beschreibt ein *mehrdimensional kausales* System [11],

in dem die neuen Variablen die Bedeutung einer Zeit haben. Auf eine Aktion erfolgt hier erst eine Reaktion bei einem nichtnegativen Zuwachs in den unabhängigen Variablen, von denen sich mindestens eine ändern muss.

2.2 Transformation der Koordinaten

Für den weiteren Verlauf wird eine Koordinatentransformation rekapituliert, die sich im Zusammenhang mit mehrdimensional kausalen Systemen etabliert hat, vgl. [39]. Das Augenmerk ruht im Besonderen auf der Transformation partieller Ableitungen, die als Operatoren geschrieben werden:

$$\mathcal{D}_t = \frac{\partial}{\partial t}\,, \quad \mathcal{D}_{z_\varkappa} = \frac{\partial}{\partial z_\varkappa} \quad \text{für} \quad \varkappa = 1, \ldots, k-1\,. \tag{1}$$

Für die Koordinatentransformation wird die Zeitvariable mit einer positiven Geschwindigkeit v_k skaliert,

$$z_k = v_k t \quad \text{bzw.} \quad \mathcal{D}_{z_k} = \frac{1}{v_k}\mathcal{D}_t\,, \tag{2}$$

wobei v_k der Einfachheit halber konstant ist und als geeignet zu wählender Designparameter fungiert. Die skalierte Zeitvariable hat die physikalische Einheit eines Ortes und ergänzt die bisherigen Ortskoordinaten im Vektor

$$\boldsymbol{z} = \left[\, z_1, \ldots, z_k \,\right]^{\mathrm{T}} \qquad \text{bzw.} \qquad \boldsymbol{z} = \sum_{\varkappa=1}^{k} z_\varkappa\, \mathbf{e}_{z_\varkappa} \qquad \text{mit} \qquad z_\varkappa = \mathbf{e}_{z_\varkappa}^{\mathrm{T}} \boldsymbol{z}\,. \tag{3}$$

Diesen Koordinaten werden k' neue Koordinaten zugeordnet, die im Vektor

$$\boldsymbol{t} = \left[\, t_1, \ldots, t_{k'} \,\right]^{\mathrm{T}} \qquad \text{bzw.} \qquad \boldsymbol{t} = \sum_{\varkappa'=1}^{k'} t_{\varkappa'} \mathbf{e}_{t_{\varkappa'}} \qquad \text{mit} \qquad t_{\varkappa'} = \mathbf{e}_{t_{\varkappa'}}^{\mathrm{T}} \boldsymbol{t} \tag{4}$$

enthalten sind, wobei das neue Koordinatensystem mindestens so viele Koordinaten wie das alte Koordinatensystem hat [11]:

$$k' \geq k\,. \tag{5}$$

Da die Koordinatensysteme von unterschiedlicher Dimension sein können, ist es erforderlich, zwischen den Einheitsvektoren im alten und im neuen Koordinatensystem $\mathbf{e}_{z_\varkappa}$ bzw. $\mathbf{e}_{t_{\varkappa'}}$ zu unterscheiden. Für die Koordinatentransformation wird eine konstante, zeilenreguläre Transformationsmatrix $\boldsymbol{V}$ der Dimension $k \times k'$ verwendet, zu der $\boldsymbol{\Lambda}$ eine rechtsinverse Matrix ist:

$$\boldsymbol{z} = \boldsymbol{V}\boldsymbol{t} \quad \text{bzw.} \quad \boldsymbol{t} = \boldsymbol{\Lambda}\boldsymbol{z} \quad \text{mit} \quad \boldsymbol{V}\boldsymbol{\Lambda} = \mathbf{1}\,. \tag{6}$$

Die Transformationsmatrix $\boldsymbol{V}$ hat die Einheit einer Geschwindigkeit, womit der Tatsache Rechnung getragen wird, dass den neuen Koordinaten die Bedeutung von Zeitvariablen zukommt. Die vektorwertige Notation der Differentialoperatoren,

$$\boldsymbol{\mathcal{D}}_{\boldsymbol{z}} = \left[\, \mathcal{D}_{z_1}, \ldots, \mathcal{D}_{z_k} \,\right]^{\mathrm{T}} \quad \text{bzw.} \quad \boldsymbol{\mathcal{D}}_{\boldsymbol{t}} = \left[\, \mathcal{D}_{t_1}, \ldots, \mathcal{D}_{t_{k'}} \,\right]^{\mathrm{T}}\,, \tag{7}$$

ermöglicht eine kompakte Schreibweise ihrer Transformation,

$$\boldsymbol{\mathcal{D}_t} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{\mathcal{D}_z} \quad \text{bzw.} \quad \boldsymbol{\mathcal{D}_z} = \boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\mathcal{D}_t}\,. \tag{8}$$

Aufgrund der linearen Koordinatentransformation ergeben sich die Differentiale in den alten Koordinaten als Linearkombinationen der Differentiale in den neuen Koordinaten

$$\mathcal{D}_{z_\varkappa} = \sum_{\varkappa'=1}^{k'} \lambda_{\varkappa'\varkappa}\mathcal{D}_{t_{\varkappa'}} \quad \text{mit} \quad \lambda_{\varkappa'\varkappa} = \mathbf{e}_{t_{\varkappa'}}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{e}_{z_\varkappa}\,. \tag{9}$$

In den weiteren Diskussionen zur Wahl einer geeigneten Transformation für mehrdimensionale Kausalität wird grundsätzlich $k' = k$ angenommen, womit die zeilenreguläre Transformationsmatrix $\boldsymbol{V}$ nun regulär ist und für ihre jetzt eindeutige Rechtsinverse $\boldsymbol{\Lambda} = \boldsymbol{V}^{-1}$ gilt. Bei identischen Dimensionen der beiden Koordinatensysteme ist für die Einheitsvektoren keine Unterscheidung mehr notwendig und sie werden daher mit $\mathbf{e}_{z_\varkappa} = \mathbf{e}_\varkappa$ bzw. $\mathbf{e}_{t_{\varkappa'}} = \mathbf{e}_{\varkappa'}$ bezeichnet.

Welchen Bedingungen eine Koordinatentransformation genügen muss, damit aus einem kausalen und lokalen System ein mehrdimensional kausales System entsteht, wird in [39] bzw. [40] anhand der folgenden Ungleichungen definiert:

$$\boldsymbol{\mathcal{D}}_{\boldsymbol{t}}^{\mathrm{T}}\{t\} > \mathbf{0}^{\mathrm{T}} \tag{10a}$$

$$\mathcal{D}_t\{\boldsymbol{t}\} \geq \mathbf{0} \quad \text{mit} \quad \mathcal{D}_t\{\boldsymbol{t}\} \neq \mathbf{0}\,. \tag{10b}$$

Hierbei sind die Zeichen für positiv bzw. nichtnegativ koordinatenweise zu verstehen, dagegen ist das Zeichen für ungleich auf den vollständigen Vektor bezogen.

Anschaulich bedeutet mehrdimensionale Kausalität, dass eine Zunahme in einer neuen Koordinate ein Fortschreiten in der Zeit hervorruft und umgekehrt jedes Fortschreiten in der Zeit zu keiner Abnahme in einer der neuen Koordinaten führt, siehe Bild 1. Ein Beispiel für eine geeignete Koordinatentransformation ist in [4] zu finden. Dort wird eine Transformationsmatrix vorgeschlagen, die bis auf einen konstanten Faktor orthogonal ist:

$$k' = k\,, \quad \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \mathrm{v}_0^2\mathbf{1}\,, \quad \mathbf{e}_k^{\mathrm{T}}\boldsymbol{V} = \frac{\mathrm{v}_0}{\sqrt{k}}\mathbb{1}^{\mathrm{T}}\,. \tag{11}$$

Hierbei ist $\mathbb{1}$ ein Spaltenvektor mit k Einsen und v_0 bezeichnet eine konstante positive Geschwindigkeit. Für diese Koordinatentransformation ergibt sich

$$\mathrm{v}_{\mathrm{mdk}} = \mathrm{v}_{\max}\sqrt{k-1} \quad \text{mit} \quad \mathrm{v}_k \geq \mathrm{v}_{\mathrm{mdk}} \tag{12}$$

als Mindestgeschwindigkeit für mehrdimensionale Kausalität [4].

2.3 Mehrdimensionale Passivität

Zustandsgrößen, Torspannungen und -ströme eines mehrdimensionalen Mehrtores sind im Gegensatz zu einem gewöhnlichen Mehrtor nicht nur von der Zeit, sondern auch vom Ort abhängig. Für eine konkrete Beschreibung sollen die Zustandsgrößen, Torspannungen und Torströme Funktionen in den Koordinaten $\boldsymbol{z}$ sein und werden in den Vektoren $\boldsymbol{w}$, $\boldsymbol{u}$ bzw. $\boldsymbol{i}$ zusammengefasst. Das mehrdimensionale Mehrtor soll darüber hinaus *physikalisch passiv* sein: Es ist kausal und lokal und die Zunahme der gespeicherten Energiedichten $E_{z_\varkappa}$ ist von der durch die Tore des Mehrtores übertragenen Momentanleistungsdichte p beschränkt:

$$\boldsymbol{\mathcal{D}}_{\boldsymbol{z}}^{\mathrm{T}}\{\boldsymbol{E}_{\boldsymbol{z}}(\boldsymbol{w}(\boldsymbol{z}))\} \leq p(\boldsymbol{z}) \quad \text{mit} \quad \boldsymbol{E}_{\boldsymbol{z}} = \left[\, E_{z_1}, \ldots, E_{z_k} \,\right]^{\mathrm{T}} = \boldsymbol{E}_{\boldsymbol{z}}^* \quad \text{und} \quad p = \boldsymbol{i}^{\mathrm{T}}\boldsymbol{u}\,. \tag{13a}$$

Von den gespeicherten Energiedichten wird angenommen, dass sie endlich, stetig und im erforderlichen Maße differenzierbar sind. Im Besonderen ist E_{z_k} die mit v_k skalierte zeitliche Energiedichte E_t mit der Eigenschaft:

$$E_{z_k}(\boldsymbol{w}) > 0 \quad \text{für} \quad \boldsymbol{w} \neq \mathbf{0} \quad \text{und} \quad E_{z_k}(\mathbf{0}) = 0\,. \tag{13b}$$

Die verbleibenden $k-1$ Energiedichten berücksichtigen den örtlichen Austausch von Energien und können auch negativ werden, vgl. [36].

Die Koordinaten $\boldsymbol{z}$ werden wie im vorangegangenen Abschnitt so transformiert, dass das kausale und lokale mehrdimensionale Mehrtor in den neuen Koordinaten $\boldsymbol{t}$ mehrdimensional kausal wird. Eine elementweise Auswertung der Gleichung (13) und eine Substitution der Differentiale gemäß der Gleichung (8) liefert nach dem Zusammenfassen der Terme die Ungleichung für die Passivität in den neuen Koordinaten:

$$\boldsymbol{\mathcal{D}}_{\boldsymbol{t}}^{\mathrm{T}}\{\boldsymbol{E_t}(\boldsymbol{w}(\boldsymbol{t}))\} \leq p(\boldsymbol{t}) \quad \text{mit} \quad \boldsymbol{E_t} = \left[\, E_{t_1}, \ldots, E_{t_{k'}} \,\right]^{\mathrm{T}} = \boldsymbol{\Lambda} \boldsymbol{E_z}\,. \tag{14a}$$

Das mehrdimensional kausale Mehrtor heißt *mehrdimensional passiv* [4], wenn dabei jede Energiefunktion $E_{t_{\varkappa'}}$ nichtnegativ ist:

$$\boldsymbol{E_t}(\boldsymbol{w}(\boldsymbol{t})) \geq \mathbf{0}\,. \tag{14b}$$

Es wird nun untersucht, welche Anforderungen an die Koordinatentransformation zu stellen sind, damit das physikalisch passive Mehrtor mehrdimensional passiv wird, wobei wieder von $k' = k$ ausgegangen wird. Für diese Untersuchung ist es zielführend, in der Gleichung (14b) $\boldsymbol{E_t}$ durch $\boldsymbol{E_z}$ auszudrücken:

$$\boldsymbol{\Lambda} \boldsymbol{E_z} \geq \mathbf{0}\,. \tag{15}$$

Eine Aufteilung nach zeitlichen und örtlichen Energiedichten sowie eine Normierung auf die positive Energiedichte E_t liefert die Ungleichungen

$$-\mathbf{e}_{\varkappa}^{\mathrm{T}}\, \boldsymbol{\Lambda}_{k-1}\, \boldsymbol{\varepsilon}_{k-1} \leq \lambda_{\varkappa k} v_k \quad \text{für} \quad \varkappa = 1, \ldots, k\,. \tag{16}$$

Hierbei erscheint $\boldsymbol{\varepsilon}_{k-1}$ als Geschwindigkeitsvektor, dessen Koordinaten die örtlichen Energiedichten in Relation zur zeitlichen Energiedichte setzen:

$$\boldsymbol{\varepsilon}_{k-1} = \frac{1}{E_t} \left[\, E_{z_1}, \ldots, E_{z_{k-1}} \,\right]^{\mathrm{T}}\,. \tag{17}$$

Da die Kette der Ungleichungen

$$-\mathbf{e}_{\varkappa}^{\mathrm{T}}\, \boldsymbol{\Lambda}_{k-1}\, \boldsymbol{\varepsilon}_{k-1} \overset{(a)}{\leq} \left|\mathbf{e}_{\varkappa}^{\mathrm{T}}\, \boldsymbol{\Lambda}_{k-1}\, \boldsymbol{\varepsilon}_{k-1}\right| \overset{(b)}{\leq} \left\|\boldsymbol{\Lambda}_{k-1}^{\mathrm{T}}\, \mathbf{e}_{\varkappa}\right\|_2 \|\boldsymbol{\varepsilon}_{k-1}\|_2 \overset{(c)}{\leq} \left\|\boldsymbol{\Lambda}_{k-1}^{\mathrm{T}}\, \mathbf{e}_{\varkappa}\right\|_2 \varepsilon_{\max,\varkappa} \tag{18}$$

aufgestellt werden kann, ist es für die Gleichung (16) offenbar hinreichend, wenn

$$v_k \geq \frac{\varepsilon_{\max,\varkappa}}{\lambda_{\varkappa k}} \left\|\boldsymbol{\Lambda}_{k-1}^{\mathrm{T}}\, \mathbf{e}_{\varkappa}\right\|_2 \quad \text{für} \quad \varkappa = 1, \ldots, k \tag{19}$$

gilt. Um zu belegen, dass diese Bedingung auch notwendig ist, kann wie folgt argumentiert werden. Der Vektor

$$\boldsymbol{\varepsilon}_{k-1} = \boldsymbol{\varepsilon}_{\varkappa} = \beta_{\varkappa}\, \boldsymbol{\Lambda}_{k-1}^{\mathrm{T}}\, \mathbf{e}_{\varkappa} \quad \text{mit} \quad \beta_{\varkappa} < 0 \tag{20}$$

erfüllt (a) und (b) in (18) mit dem Gleichheitszeichen. Wählt man von allen möglichen Vektoren

$$\boldsymbol{\varepsilon}_{\varkappa} = \boldsymbol{\varepsilon}_{\max,\varkappa} \quad \text{mit} \quad \max_{z_1,\ldots,z_k} \{\|\boldsymbol{\varepsilon}_{\varkappa}\|_2\} = \|\boldsymbol{\varepsilon}_{\max,\varkappa}\|_2 = \varepsilon_{\max,\varkappa}\,, \tag{21}$$

so liegt auch für (c) Gleichheit vor. Zur Einhaltung der Gleichung (16) ist es daher notwendig und hinreichend, wenn v_k die Mindestgeschwindigkeit

$$v_{\mathrm{mdp}} = \max_{\varkappa=1,\ldots,k} \left\{ \frac{\varepsilon_{\max,\varkappa}}{\lambda_{\varkappa k}} \left\| \boldsymbol{\Lambda}_{k-1}^{\mathrm{T}}\, \mathbf{e}_{\varkappa} \right\|_2 \right\} \tag{22}$$

annimmt. Da ein mehrdimensional passives System per definitionem mehrdimensional kausal ist, liegt mehrdimensionale Passivität für

$$v_k \geq v_{\mathrm{mdk}} \quad \text{und} \quad v_k \geq v_{\mathrm{mdp}} \tag{23}$$

vor. Ob die Forderung nach mehrdimensionaler Passivität für v_k im Vergleich zur Forderung nach mehrdimensionaler Kausalität tatsächlich eine Verschärfung bedeutet, ist im Einzelfall zu untersuchen.

3 Synthese einer intern mehrdimensional passiven Schaltung

Ausgangspunkt für die Synthese einer intern mehrdimensional passiven Schaltung ist das lineare partielle Differentialgleichungssystem

$$\sum_{\varkappa=1}^{k-1} \boldsymbol{L}_{z_\varkappa} \mathcal{D}_{z_\varkappa}\{\boldsymbol{i}\} + \boldsymbol{L}_t \mathcal{D}_t\{\boldsymbol{i}\} + \boldsymbol{R}\boldsymbol{i} = \boldsymbol{v}_x\,, \tag{24}$$

bei dem alle auftretenden Größen reell sind. Neben der Zeit t treten $k-1$ Ortskoordinaten $z_\varkappa$ als unabhängige Variablen auf, deren zugehörige partielle Ableitungen die Operatoren $\mathcal{D}_t$ bzw. $\mathcal{D}_{z_\varkappa}$ sind. Für eine elektrische Interpretation dient die Schaltung des Bildes 2, bei der die jeweils n Spulenströme und Quellenspannungen in den Vektoren $\boldsymbol{i}$ und $\boldsymbol{v}_x$ als Koordinaten eingetragen sind. Die Dimension der Matrizen ist $n \times n$, wobei $\boldsymbol{R}$ die Widerstandsmatrix des Mehrtor-Widerstandes ist, während $\boldsymbol{L}_t$ und $\boldsymbol{L}_{z_\varkappa}$ Induktivitätsmatrizen der entsprechenden Mehrtor-Spulen sind. Damit ein extern physikalisch passives System mit endlichen Ausbreitungsgeschwindigkeiten vorliegt, ist es notwendig und hinreichend, dass die Matrizen die folgenden Eigenschaften aufweisen [36]:

$$\boldsymbol{L}_{z_\varkappa} = \boldsymbol{L}_{z_\varkappa}^{\mathrm{T}}\,, \quad \boldsymbol{L}_t = \boldsymbol{L}_t^{\mathrm{T}} > \mathbf{0} \quad \text{und} \quad \boldsymbol{R} + \boldsymbol{R}^{\mathrm{T}} \geq \mathbf{0}\,. \tag{25}$$

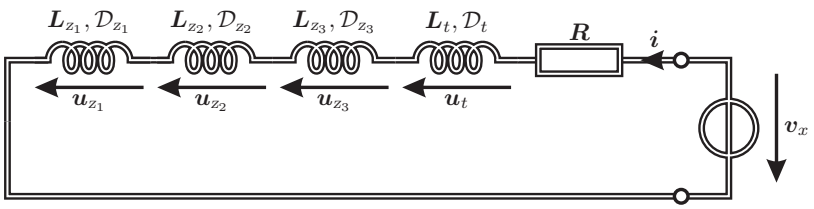


Abbildung 2: Schaltung zu einem konstanten partiellen Differentialgleichungssystem

Inspiriert von diesem Ergebnis wird die Problemstellung der Gleichung (24) auf ein partielles Differentialgleichungssystem mit variablen Koeffizienten

$$\sum_{\varkappa=1}^{k-1} \hat{\boldsymbol{L}}_{z_\varkappa} \mathcal{D}_{z_\varkappa}\{\boldsymbol{i}\} + \hat{\boldsymbol{L}}_t \mathcal{D}_t\{\boldsymbol{i}\} + \hat{\boldsymbol{R}}_{\mathrm{diss}}\, \boldsymbol{i} = \boldsymbol{v}_x \tag{26}$$

erweitert, bei dem sich alle Vorgänge höchstens mit der maximalen Geschwindigkeit $v_{\max}$ ausbreiten sollen. Die Parameter sind nun variable Induktivitätsmatrizen $\hat{\boldsymbol{L}}_{z_\varkappa}$ und $\hat{\boldsymbol{L}}_t$ von differentiellen Mehrtor-Spulen sowie eine variable Widerstandsmatrix $\hat{\boldsymbol{R}}_{\mathrm{diss}}$, die den Einschränkungen der Gleichung (25) unterliegen:

$$\hat{\boldsymbol{L}}_{z_\varkappa} = \hat{\boldsymbol{L}}_{z_\varkappa}^{\mathrm{T}}\,, \quad \hat{\boldsymbol{L}}_t = \hat{\boldsymbol{L}}_t^{\mathrm{T}} > \boldsymbol{0} \quad \text{und} \quad \hat{\boldsymbol{R}}_{\mathrm{diss}} + \hat{\boldsymbol{R}}_{\mathrm{diss}}^{\mathrm{T}} \geq \boldsymbol{0}\,. \tag{27}$$

Die positive Definitheit der zeitlichen Induktivitätsmatrix garantiert die Existenz ihrer Inversen, so dass es immer möglich ist, die Gleichung (26) von links mit $L\hat{\boldsymbol{L}}_t^{-1}$ zu multiplizieren. Offenbar bleibt die algebraische Struktur der Differentialgleichung dabei unverändert und man darf ohne Einschränkung der Allgemeinheit $\hat{\boldsymbol{L}}_t = L\boldsymbol{1}$ mit $L > 0$ wählen:

$$\sum_{\varkappa=1}^{k-1} \hat{\boldsymbol{L}}_{z_\varkappa} \mathcal{D}_{z_\varkappa}\{\boldsymbol{i}\} + L\mathcal{D}_t\{\boldsymbol{i}\} + \hat{\boldsymbol{R}}_{\mathrm{diss}}\, \boldsymbol{i} = \boldsymbol{v}_x\,. \tag{28}$$

Für die Synthese einer intern mehrdimensional passiven Schaltung wird die Zeitvariable wie in der Gleichung (2) mit einer noch geeignet zu wählenden positiven konstanten Geschwindigkeit v_k skaliert und als Koordinate z_k eingeführt. Das Differentialgleichungssystem lautet somit

$$\sum_{\varkappa=1}^{k} \hat{\boldsymbol{L}}_{z_\varkappa} \mathcal{D}_{z_\varkappa}\{\boldsymbol{i}\} + \hat{\boldsymbol{R}}_{\mathrm{diss}}\, \boldsymbol{i} = \boldsymbol{v}_x \quad \text{mit} \quad \hat{\boldsymbol{L}}_{z_k} = \mathrm{v}_k \hat{\boldsymbol{L}}_t = \hat{\boldsymbol{L}}_{z_k}^{\mathrm{T}} > \boldsymbol{0} \tag{29}$$

und tritt als Maschengleichung der elektrischen Schaltung des Bildes 3 in Erscheinung.

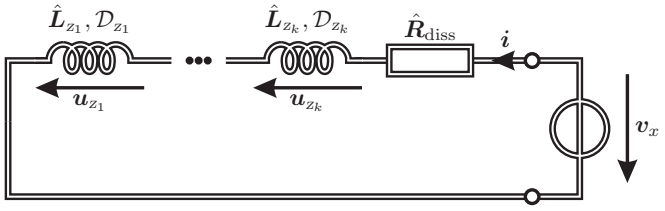


Abbildung 3: Schaltung zu einem variablen partiellen Differentialgleichungssystem

Um eine mehrdimensional passive Schaltung zu konstruieren, werden entsprechend zur Gleichung (4) neue Koordinaten eingeführt, wobei v_k größer oder gleich der Mindestgeschwindigkeit für mehrdimensionale Kausalität ist. Eine Transformation der Differentialoperatoren gemäß der Gleichung (9) führt auf das transformierte partielle Differentialgleichungssystem

$$\sum_{\varkappa'=1}^{k'} \hat{\boldsymbol{L}}_{t_{\varkappa'}} \mathcal{D}_{t_{\varkappa'}}\{\boldsymbol{i}\} + \hat{\boldsymbol{R}}_{\mathrm{diss}}\, \boldsymbol{i} = \boldsymbol{v}_x\,, \tag{30}$$

das entsprechend dem Bild 3 interpretiert wird. Durch die Transformation entstehen die neuen Induktivitätsmatrizen

$$\hat{\boldsymbol{L}}_{t_{\varkappa'}} = \sum_{\varkappa=1}^{k} \lambda_{\varkappa'\varkappa}\, \hat{\boldsymbol{L}}_{z_\varkappa} = \hat{\boldsymbol{L}}_{t_{\varkappa'}}^{\mathrm{T}} \,, \tag{31}$$

die als Linearkombinationen der ursprünglichen Induktivitätsmatrizen ebenfalls symmetrisch sind. Es ist zu beachten, dass die Spannungen und Ströme nun von den neuen Koordinaten abhängig sind. Genau genommen müssten sie daher umbenannt werden, aus Gründen der Übersichtlichkeit wird jedoch hierauf verzichtet.

Da L und $\lambda_{\varkappa'k}$ positiv sind, kann immer ein hinreichend großes v_k gewählt werden, so dass die symmetrischen Induktivitätsmatrizen

$$\hat{\boldsymbol{L}}_{t_{\varkappa'}} = \sum_{\varkappa=1}^{k-1} \lambda_{\varkappa'\varkappa}\, \hat{\boldsymbol{L}}_{z_\varkappa} + \mathrm{v}_k \lambda_{\varkappa'k}\, L\mathbf{1} \tag{32}$$

durch den positiven Term $v_k \lambda_{\varkappa'k} L$ diagonaldominant werden. Unter dieser Annahme sind die Induktivitätsmatrizen positiv semidefinit und das partielle Differentialgleichungssystem der Gleichung (30) lässt sich mit Hilfe der Bauelementgleichungen energetischer Mehrtor-Spulen formulieren:

$$\sum_{\varkappa'=1}^{k'} \hat{\mathcal{D}}_{t_{\varkappa'}}\{\hat{\boldsymbol{L}}_{t_{\varkappa'}}, \boldsymbol{i}\} - \hat{\boldsymbol{R}}_{[\varkappa']}^{\mathrm{T}} \boldsymbol{i} + \hat{\boldsymbol{R}}_{\mathrm{diss}}\, \boldsymbol{i} = \boldsymbol{v}_x \tag{33}$$

mit

$$\hat{\mathcal{D}}_{t_{\varkappa'}}\{\hat{\boldsymbol{L}}_{t_{\varkappa'}}, \boldsymbol{i}\} = \sqrt{\hat{\boldsymbol{L}}_{t_{\varkappa'}}}^{\mathrm{T}}\, \mathcal{D}_{t_{\varkappa'}} \left\{ \sqrt{\hat{\boldsymbol{L}}_{t_{\varkappa'}}}\, \boldsymbol{i} \right\} \,, \tag{34}$$

vgl. [41]. In dieser Gleichung werden die von den variablen Induktivitätsmatrizen verursachten parametrischen Effekte durch die Widerstandsmatrizen

$$\hat{\boldsymbol{R}}_{[\varkappa']} = \mathcal{D}_{t_{\varkappa'}} \left\{ \sqrt{\hat{\boldsymbol{L}}_{t_{\varkappa'}}} \right\}^{\mathrm{T}} \sqrt{\hat{\boldsymbol{L}}_{t_{\varkappa'}}} \tag{35}$$

berücksichtigt. Zur Vermeidung eines unnötigen Implementierungsaufwands werden die einzelnen Widerstandsmatrizen $\hat{\boldsymbol{R}}_{[\varkappa']}$ in der Widerstandsmatrix

$$\hat{\boldsymbol{R}} = \hat{\boldsymbol{R}}_{\mathrm{diss}} - \sum_{\varkappa'=1}^{k'} \hat{\boldsymbol{R}}_{[\varkappa']}^{\mathrm{T}} \tag{36}$$

zusammengefasst. Hierdurch entsteht das partielle Differentialgleichungssystem

$$\sum_{\varkappa'=1}^{k'} \hat{\mathcal{D}}_{t_{\varkappa'}}\{\hat{\boldsymbol{L}}_{t_{\varkappa'}}, \boldsymbol{i}\} + \hat{\boldsymbol{R}}\boldsymbol{i} = \boldsymbol{v}_x \quad \text{mit} \quad \hat{\boldsymbol{L}}_{t_{\varkappa'}} = \hat{\boldsymbol{L}}_{t_{\varkappa'}}^{\mathrm{T}} \geq \mathbf{0} \,, \tag{37}$$

welches nur dann ein extern mehrdimensional passives System beschreibt, wenn der symmetrische Teil von $\hat{\boldsymbol{R}}$ positiv semidefinit ist. Das partielle Differentialgleichungssystem der Gleichung (37) beschreibt daher genau dann ein extern mehrdimensional passives System, wenn

mehrdimensionale Kausalität vorliegt und der symmetrische Teil der Widerstandsmatrix $\hat{\boldsymbol{R}}$ positiv semidefinit ist. Unter dieser Voraussetzung findet man für das partielle Differentialgleichungssystem die Darstellung

$$L\sum_{\varkappa'=1}^{k'}\hat{\boldsymbol{N}}_{[\varkappa']}^{\mathrm{T}}\mathcal{D}_{t_{\varkappa'}}\{\hat{\boldsymbol{N}}_{[\varkappa']}\,\boldsymbol{i}\}+\left[\hat{\boldsymbol{R}}_{\mathrm{G}}-\hat{\boldsymbol{R}}_{\mathrm{G}}^{\mathrm{T}}+\hat{\boldsymbol{N}}_{\mathrm{s}}^{\mathrm{T}}\hat{\boldsymbol{R}}_{x}\hat{\boldsymbol{N}}_{\mathrm{s}}\right]\boldsymbol{i}=\hat{\boldsymbol{N}}_{\mathrm{s}}^{\mathrm{T}}\boldsymbol{u}_{x}\,. \tag{38}$$

Zu ihr gehören die in den Bildern 4 und 5 dargestellten Schaltungen, wobei $\hat{n}_{\mu\nu}^{[\varkappa']}$ der entsprechende Eintrag in der unteren Dreiecksmatrix $\hat{\boldsymbol{N}}_{[\varkappa']}$ ist.

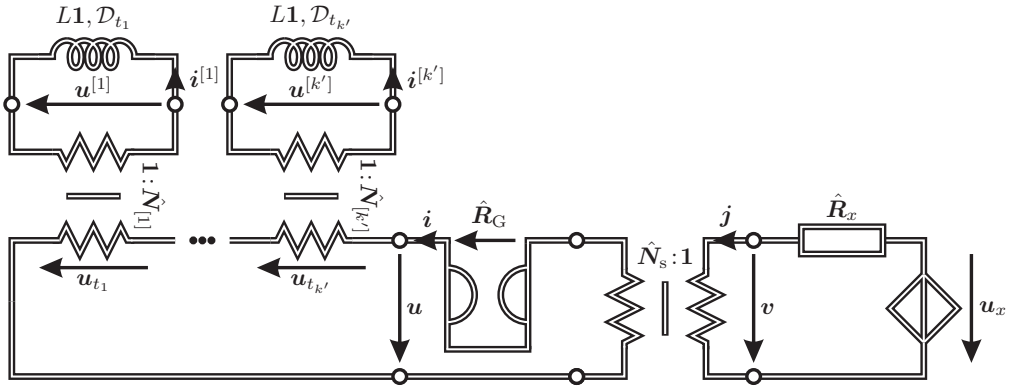


Abbildung 4: Schaltung zu einem mehrdimensional passiven System

Damit ist es zwar unter gewissen Voraussetzungen gelungen, für die Gleichung (28) eine intern mehrdimensional passive KIRCHHOFF'sche Schaltung zu synthetisieren, sie ist aber nicht minimal, wie man durch einen Vergleich der Zahl der Parameter des partiellen Differentialgleichungssystems und der Zahl der Bauelementwerte der Schaltung feststellt. Für eine minimale Realisierung ist offenbar eine Koordinatentransformation mit $k'=k$ erforderlich, so dass die Zahl der Koordinaten unverändert bleibt:

$$L\sum_{\varkappa'=1}^{k}\hat{\boldsymbol{N}}_{[\varkappa']}^{\mathrm{T}}\mathcal{D}_{t_{\varkappa'}}\{\hat{\boldsymbol{N}}_{[\varkappa']}\boldsymbol{i}\}+\left[\hat{\boldsymbol{R}}_{\mathrm{G}}-\hat{\boldsymbol{R}}_{\mathrm{G}}^{\mathrm{T}}+\hat{\boldsymbol{N}}_{\mathrm{s}}^{\mathrm{T}}\hat{\boldsymbol{R}}_{x}\hat{\boldsymbol{N}}_{\mathrm{s}}\right]\boldsymbol{i}=\hat{\boldsymbol{N}}_{\mathrm{s}}^{\mathrm{T}}\boldsymbol{u}_{x}\,.$$

Des Weiteren kann man v_k auch so groß wählen, dass mindestens eine der Induktivitätsmatrizen in der Gleichung (32) positiv definit wird. Ist dies beispielsweise für die Matrix $\hat{\boldsymbol{N}}_{[1]}$ gegeben, so lässt sich das partielle Differentialgleichungssystem von links mit der transponierten Inversen $\hat{\boldsymbol{N}}_{[1]}^{-\mathrm{T}}$ von $\hat{\boldsymbol{N}}_{[1]}$ multiplizieren und der Vektor $\boldsymbol{i}_{[1]}=\hat{\boldsymbol{N}}_{[1]}\boldsymbol{i}$ der Ströme einführen. Das Differentialgleichungssystem hat damit die Gestalt

$$L\mathcal{D}_{t_{\varkappa'}}\{\boldsymbol{i}_{[1]}\}+L\sum_{\varkappa'=2}^{k}\hat{\boldsymbol{N}}_{[1][\varkappa']}^{\mathrm{T}}\mathcal{D}_{t_{\varkappa'}}\{\hat{\boldsymbol{N}}_{[1][\varkappa']}\boldsymbol{i}_{[1]}\}+\left[\hat{\boldsymbol{R}}_{[1]\mathrm{G}}-\hat{\boldsymbol{R}}_{[1]\mathrm{G}}^{\mathrm{T}}+\hat{\boldsymbol{N}}_{[1]\mathrm{s}}^{\mathrm{T}}\hat{\boldsymbol{R}}_{x}\hat{\boldsymbol{N}}_{[1]\mathrm{s}}\right]\boldsymbol{i}_{[1]}=\hat{\boldsymbol{N}}_{[1]\mathrm{s}}^{\mathrm{T}}\boldsymbol{u}_{x}\,,$$

wobei anstelle der unteren Dreiecksmatrizen $\hat{\boldsymbol{R}}_{\mathrm{G}}$, $\hat{\boldsymbol{N}}_{\mathrm{s}}$ und $\hat{\boldsymbol{N}}_{[\varkappa']}$ die Matrizen

$$\hat{\boldsymbol{R}}_{[1]\mathrm{G}}=\hat{\boldsymbol{N}}_{[1]}^{-\mathrm{T}}\hat{\boldsymbol{R}}_{\mathrm{G}}\hat{\boldsymbol{N}}_{[1]}^{-1}\,,\quad \hat{\boldsymbol{N}}_{[1]\mathrm{s}}=\hat{\boldsymbol{N}}_{\mathrm{s}}\hat{\boldsymbol{N}}_{[1]}^{-1}\quad\text{bzw.}\quad \hat{\boldsymbol{N}}_{[1][\varkappa']}=\hat{\boldsymbol{N}}_{[\varkappa']}\hat{\boldsymbol{N}}_{[1]}^{-1}$$

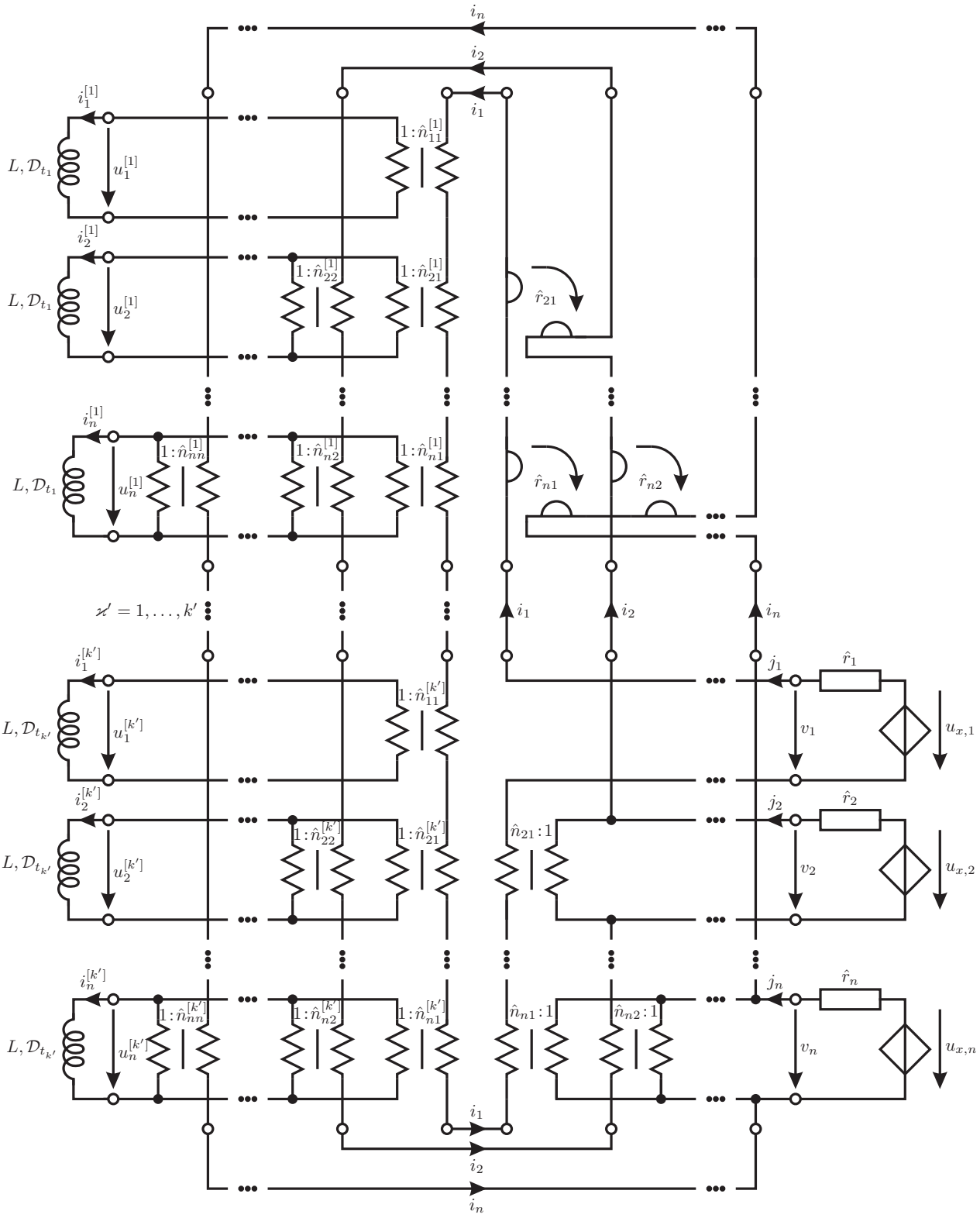


Abbildung 5: Explizite Darstellung der Schaltung des Bildes 4

vorliegen. Da ein Produkt von unteren (oberen) Dreiecksmatrizen wieder eine untere (obere) Dreiecksmatrix ist, bleibt die Struktur der ursprünglichen Dreiecksmatrizen erhalten, denn $\hat{\boldsymbol{N}}_{[1]}$ ist wie ihre Inverse eine untere Dreiecksmatrix, die durch eine Transposition zu einer oberen Dreiecksmatrix wird. Zum resultierenden Differentialgleichungssystem lässt sich daher weiterhin die Schaltung des Bildes 4 nutzen, jedoch kann der erste Mehrtor-Übertrager mit dem Übersetzungsverhältnis $\mathbf{1} : \mathbf{1}$ nun entfallen. Von der Möglichkeit zur Reduktion eines Mehrtor-Übertragers ist in den Bildern 4 und 5 abgesehen worden, weil a priori nicht bekannt ist, welcher der idealen Mehrtor-Übertrager eingespart wird.

Das Ergebnis ist in der Tat eine minimale intern mehrdimensional passive Schaltung für das partielle Differentialgleichungssystem der Gleichung (28). Für eine minimale Realisierung werden bei k' Koordinaten somit

$$n^2 + [k' - 1]\frac{n[n+1]}{2} \quad \text{und} \quad nk' \tag{39}$$

Bauelementwerte als Parameter bzw. Spulen als Energiespeicher benötigt. Die Zahl der Bauelementwerte setzt sich aus n Widerständen, $n[n-1]/2$ Gyratoren und $k'n[n+1]/2 - n$ idealen Übertragern zusammen.

3.1 Ein Beispiel

Die einzelnen Schritte der Schaltungssynthese werden nun für eine partielle Differentialgleichung gezeigt. Um die Analogie zu gewöhnlichen Differentialgleichungen aufzuzeigen, wird die Gleichung eines nichtlinearen Oszillators um eine örtliche Ableitung zweiter Ordnung erweitert,

$$\mathcal{D}_{tt}\{\vartheta\} + \Omega_\mathrm{s}\hat{f}\,\mathcal{D}_t\{\vartheta\} + \Omega_\mathrm{a}^2\,\hat{g}(\vartheta) - \mathrm{v}_z^2\,\mathcal{D}_{zz}\{\vartheta\} = \Omega_x^2\vartheta_x\,, \tag{40}$$

wobei $\Omega_\mathrm{a} > 0$ und $\Omega_\mathrm{s}, \Omega_x \geq 0$ gilt, vgl. [41]. In dieser partiellen Differentialgleichung sind $\vartheta = \vartheta(z,t)$ sowie $\vartheta_x = \vartheta_x(z,t)$ von der Zeit und der Ortskoordinate z abhängig. Die Argumente der Funktion $\hat{f}$ haben keinen Einfluss auf die Schaltungssynthese und werden aus Gründen der Übersichtlichkeit weggelassen. Es wird ferner eine hinreichend glatte Funktion für ϑ angenommen, so dass alle partiellen Ableitungen bis zur zweiten Ordnung existieren und die Reihenfolge bei der Differentiation unerheblich ist. Eine Vorgabe von Anfangs- oder Randwerten wird bewusst vermieden, weil sie ohnehin keine Rolle für die beabsichtigte Schaltungssynthese spielen und daher nicht diskutiert werden.

Für die Schaltungssynthese ist die partielle Differentialgleichung (40) mit ihren Ableitungen zweiter Ordnung wie in der Gleichung (28) als ein System partieller Differentialgleichungen erster Ordnung zu formulieren. Durch Einsetzen aller bislang bekannten Größen in die Gleichung (28) ermittelt man den Stromvektor mit seinen partiellen Ableitungen

$$\boldsymbol{i} = \frac{1}{\sqrt{L}}\begin{bmatrix} \Omega_\mathrm{a}\hat{h}(\vartheta) \\ \mathcal{D}_t\{\vartheta\} \\ \mathrm{v}_z\mathcal{D}_z\{\vartheta\} \end{bmatrix}, \quad \mathcal{D}_t\{\boldsymbol{i}\} = \frac{1}{\sqrt{L}}\begin{bmatrix} \Omega_\mathrm{a}\mathcal{D}_t\{\hat{h}(\vartheta)\} \\ \mathcal{D}_{tt}\{\vartheta\} \\ \mathrm{v}_z\mathcal{D}_{tz}\{\vartheta\} \end{bmatrix}, \quad \mathcal{D}_z\{\boldsymbol{i}\} = \frac{1}{\sqrt{L}}\begin{bmatrix} \Omega_\mathrm{a}\mathcal{D}_z\{\hat{h}(\vartheta)\} \\ \mathcal{D}_{zt}\{\vartheta\} \\ \mathrm{v}_z\mathcal{D}_{zz}\{\vartheta\} \end{bmatrix}$$

sowie die Widerstandsmatrix und den Vektor der Quellenspannungen

$$\hat{\boldsymbol{R}}_\mathrm{diss} = L\begin{bmatrix} 0 & -\Omega_\mathrm{a}\mathcal{D}_\vartheta\{\hat{h}(\vartheta)\} & 0 \\ \Omega_\mathrm{a}\mathcal{D}_\vartheta\{\hat{h}(\vartheta)\} & \Omega_\mathrm{s}\hat{f} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{bzw.} \quad \boldsymbol{v}_x = \boldsymbol{u}_x = \sqrt{L}\begin{bmatrix} 0 \\ \Omega_x^2\vartheta_x \\ 0 \end{bmatrix}.$$

Ein Vergleich mit der partiellen Differentialgleichung (40) liefert die Induktivitätsmatrix

$$\hat{\boldsymbol{L}}_z = -v_z L \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} ,$$

Während in der ersten und dritten Zeile der Gleichung (28) Identitäten stehen, die auf der Kettenregel bzw. auf der Vertauschbarkeit der Ableitungen beruhen, stellt die zweite Zeile die partielle Differentialgleichung (40) dar.

Ausgehend von dieser Darstellung werden nun die Bauelementwerte der Schaltung des Bildes 5 gesucht, mit denen die Schaltung intern mehrdimensional passiv wird. Für diesen Zweck werden die Koordinaten des partiellen Differentialgleichungssystems transformiert, um eine extern mehrdimensional passive Schaltung zu erhalten:

$$\begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = \boldsymbol{\Lambda} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \quad \text{mit} \quad \boldsymbol{\Lambda} = \frac{1}{2v_0} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \frac{1}{2v_0^2} \boldsymbol{V}^{\mathrm{T}} \quad \text{und} \quad \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} z \\ v_2 t \end{bmatrix} .$$

Für die Transformation muss notwendigerweise $v_2 \geq v_{\max}$ gelten, damit die Schaltung zumindest mehrdimensional kausal ist. Die Koordinatentransformation führt zu neuen Differentialoperatoren und Induktivitätsmatrizen, siehe Gleichungen (9) und (32):

$$\mathcal{D}_{t_{1,2}} = \mathrm{v}_0 \left[\frac{1}{v_2} \mathcal{D}_t \pm \mathcal{D}_z \right] \quad \text{bzw.} \quad \hat{\boldsymbol{L}}_{t_{1,2}} = \frac{1}{2v_0} \left[v_2 \hat{\boldsymbol{L}}_t \mp \hat{\boldsymbol{L}}_z \right] = \frac{L}{2v_0} \begin{bmatrix} v_2 & 0 & 0 \\ 0 & v_2 & \mp v_z \\ 0 & \mp v_z & v_2 \end{bmatrix} .$$

Für die mehrdimensionale Passivität der Schaltung ist mehrdimensionale Kausalität nur notwendig und es sind zusätzlich positiv semidefinite Induktivitätsmatrizen erforderlich, die durch die Vorgabe $v_2 \geq |v_z|$ erreicht werden. Bei dieser Vorgabe haben die Induktivitätsmatrizen die Cholesky-Zerlegung

$$\hat{\boldsymbol{L}}_{t_{1,2}} = \boldsymbol{N}_{[1,2]}^{\mathrm{T}} L \boldsymbol{N}_{[1,2]} \quad \text{mit} \quad \boldsymbol{N}_{[1,2]} = \sqrt{\frac{v_2}{2v_0}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{1-n_z^2} & 0 \\ 0 & \mp n_z & 1 \end{bmatrix} ,$$

die vom Verhältnis der Geschwindigkeiten v_z und v_2 abhängig sind:

$$n_z = \frac{v_z}{v_2} \quad \text{mit} \quad n_z^2 \leq 1 . \tag{41}$$

Von den aus der Cholesky-Zerlegung gewonnenen Matrizen können die Übersetzungsverhältnisse der Übertrager für die Mehrtor-Spulen entnommen werden. Es resultiert die Schaltung des Bildes 6, die eine intern mehrdimensional passive Implementierung der partiellen Differentialgleichung (40) ist.

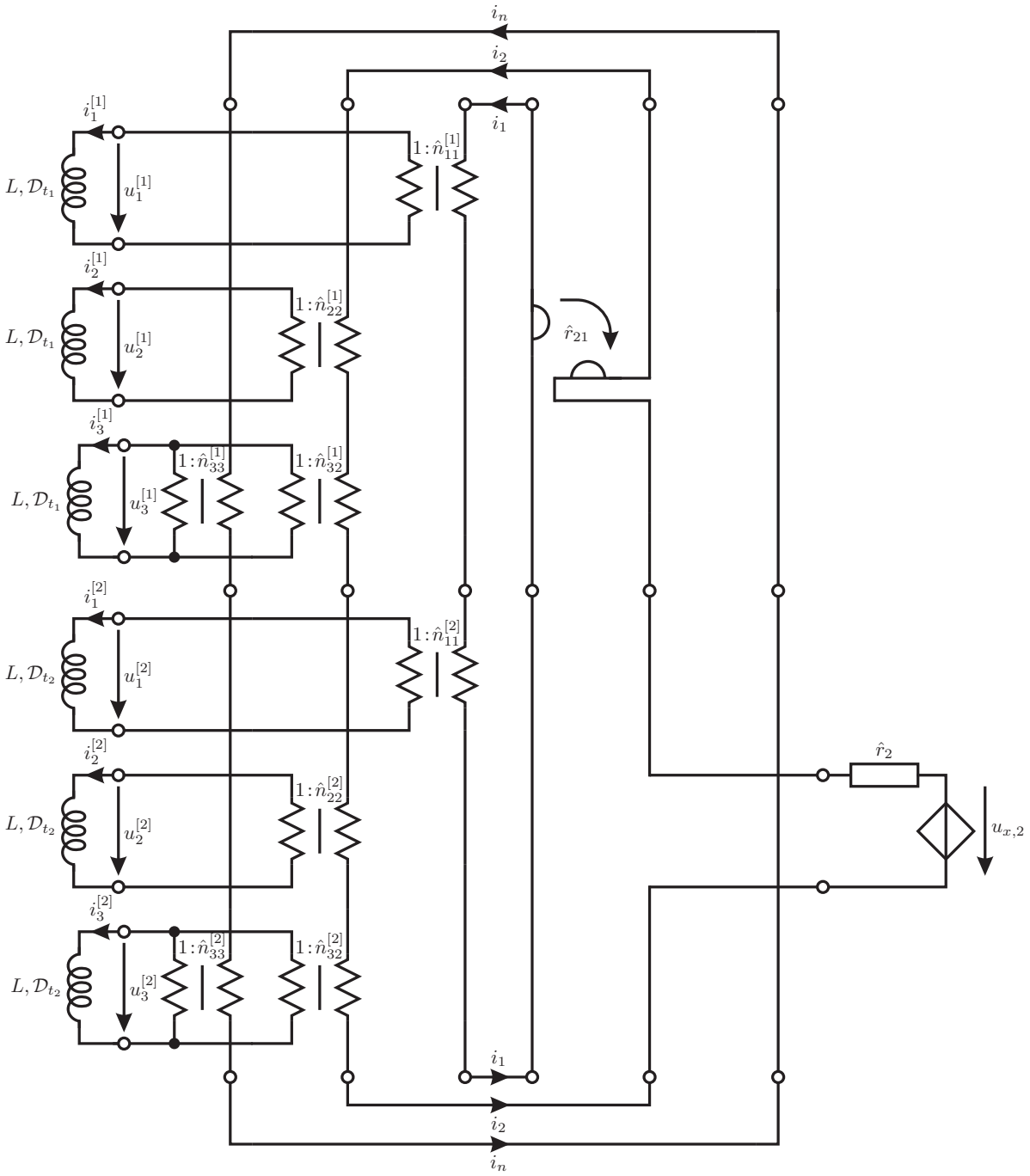


Abbildung 6: Schaltung des Bildes 5 für eine partielle Differentialgleichung zweiter Ordnung

Zur besseren Übersichtlichkeit werden in einem ersten Schritt die Übertrager mit den Spulen der Induktivität L zusammengefasst. Diese Zusammenfassung liefert Spulen mit konstanter Induktivität

$$L_1 = \frac{v_2}{2v_0} L \quad \text{bzw.} \quad L_2 = [1 - n_z^2] L_1 \,,$$

womit die Schaltung des Bildes 6 auf die im Bild 7 reduziert werden kann.

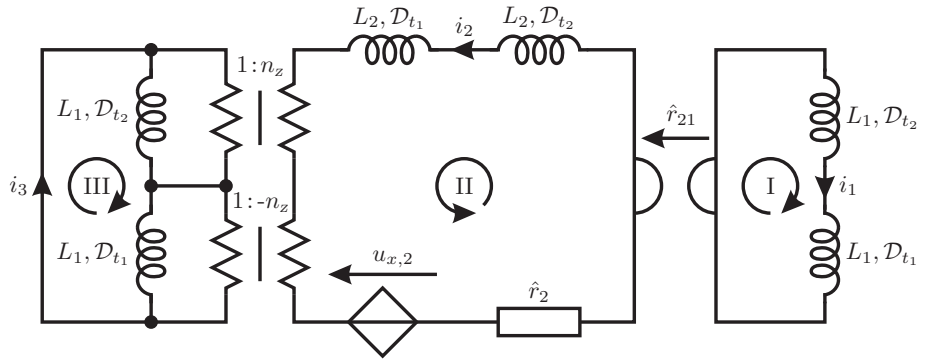


Abbildung 7: Schaltung des Bildes 6 nach modifizierter CHOLESKY-Zerlegung

Nun wird berücksichtigt, dass die Summe der Differentialoperatoren $\mathcal{D}_{t_1}$ und $\mathcal{D}_{t_2}$ bis auf eine Skalierung gleich der zeitlichen Ableitung ist:

$$\mathcal{D}_{t_1} + \mathcal{D}_{t_2} = \frac{2v_0}{v_2}\mathcal{D}_t \, .$$

Diese Beziehung kann ausgenutzt werden, um die Reihenschaltungen von Spulen gleicher Induktivität, die in Richtung der Koordinate t_1 bzw. t_2 wirken, wie im Bild 8 zu gewöhnlichen Spulen zusammenzufassen:

$$\begin{aligned} L_1\mathcal{D}_{t_1}\{i_1\} + L_1\mathcal{D}_{t_2}\{i_1\} &= L\mathcal{D}_t\{i_1\} \, , \\ L_2\mathcal{D}_{t_1}\{i_2\} + L_2\mathcal{D}_{t_2}\{i_2\} &= [1 - n_z^2]L\mathcal{D}_t\{i_2\} \, . \end{aligned}$$

Insbesondere ist zu bemerken, dass für $v_z = 0$ bzw. $n_z = 0$ der Einfluss der örtlichen Ableitungen verschwindet und die aus [41] bekannte Implementierung eines nichtlinearen Oszillators resultiert.

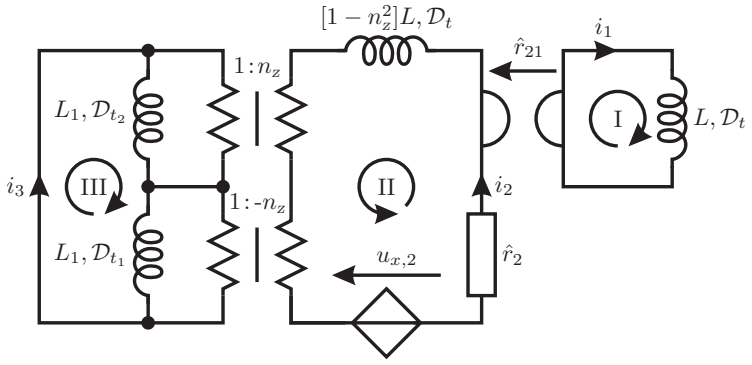


Abbildung 8: Schaltung des Bildes 7 nach einer Koordinatentransformation

In einem weiteren Schritt wird ausgenutzt, dass in der Schaltung des Bildes 8 zwei Übertrager mit dem Übersetzungsverhältnissen $1 : n_z$ und $1 : -n_z$ auftreten. Der Übertrager mit dem Übersetzungsverhältnis $1 : -n_z$ kann durch eine Kaskade zweier Übertrager mit den Übersetzungsverhältnissen $1 : n_z$ und $1 : -1$ ersetzt werden. Es entstehen zwei gleiche Übertrager mit dem Übersetzungsverhältnis $1 : n_z$, die man durch eine Transformation des links davon befindlichen Teils der Schaltung berücksichtigen kann, siehe Bild 9.

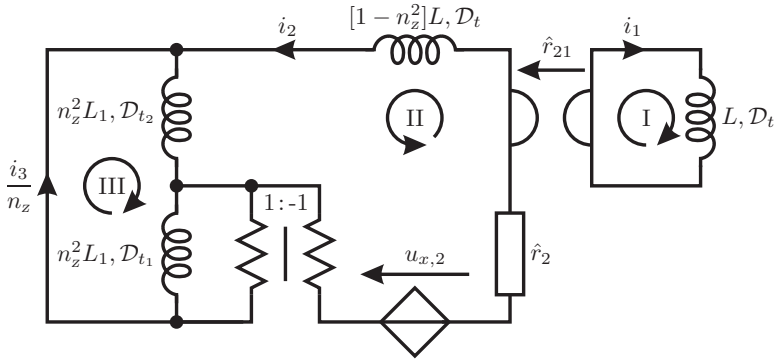


Abbildung 9: Schaltung des Bildes 8 nach einer Zustandstransformation

Im Einzelnen gelten für die Maschen I, II und III die Gleichungen

$$L\mathcal{D}_t\{i_1\} = \hat{r}_{21} i_2 \,, \qquad L\mathcal{D}_t\{i_2\} + \hat{r}_2 i_2 + \hat{r}_{21} i_1 - v_z L\mathcal{D}_z\{i_3\} = u_{x,2} \,, \qquad \mathcal{D}_t\{i_3\} = v_z \mathcal{D}_z\{i_2\} \,,$$

welche die Kettenregel, die Differentialgleichung und die Vertauschbarkeit der Ableitungen wieder zum Vorschein bringen. Aus der Schaltung des Bildes 9 ist ersichtlich, dass für $n_z^2 = 1$ eine Spule eingespart werden kann. Da n_z in der Gleichung (41) als Verhältnis der Geschwindigkeiten v_k und v_z definiert worden ist, müsste $v_k = |v_z|$ gelten und $|v_z|$ größer oder gleich der maximalen Ausbreitungsgeschwindigkeit sein. Ob dieser Umstand für alle möglichen Nichtlinearitäten gegeben ist, ist nicht überschaubar, zumal die Lösung der partiellen Differentialgleichung noch von Randwertvorgaben abhängig ist, die hier nicht behandelt werden. Für eine konkrete Nachbildung eines physikalischen Systems besteht jedoch die berechtigte Hoffnung, die maximale Ausbreitungsgeschwindigkeit mit physikalischen Überlegungen herauszufinden.

4 Zusammenfassung

In diesem Aufsatz ist die digitale Nachbildung mit dem Wellendigitalkonzept zwar in den Vordergrund gestellt worden, der Fokus liegt aber auf der Synthese einer intern passiven Schaltung zu einer vorgegebenen Differentialgleichung. Die Rekapitulation des Wellendigitalkonzepts ist daher knapp gehalten und auf das Notwendigste beschränkt. Die hier vorgestellte Synthese schließt zwar nicht die Lücke in der Synthese eines Wellendigitalmodells zu einer vorgelegten Differentialgleichung, sie erleichtert jedoch das Auffinden einer Referenzschaltung. Letztere zeichnet sich durch die Möglichkeit aus, die Quellen, Bausteine und elementaren Verbindungen einzeln im Wellendigitalbereich nachzubilden, sodass ihre torweise Verbindung ein Signalflussdiagramm ergibt, das frei von verzögerungsfreien gerichteten Schleifen ist. Die hier synthetisierte elektrische Repräsentation ist in diesem Sinne zwar keine Referenzschaltung, sie kann aber dennoch für die Synthese eines Wellendigitalmodells genutzt werden. Zu diesem Zweck wird die elektrische Schaltung torweise in Spulen, resistive Spannungsquellen und ein verbleibendes energieneutrales Netz zerlegt. Die Spulen und resistiven Quellen können direkt im Wellendigitalbereich modelliert werden, während die Synthese des energieneutralen Netzes durch eine Faktorisierung der Streumatrix nach Householder bzw. Givens gelingt. Dieses ursprünglich für Kirchhoff'sche Verbindungsnetze eingeführte Verfahren [42] lässt sich auch auf energieneutrale Netze erweitern [43]. Da die Faktorisierung auf algebraischen Umformungen basiert, ist diese Synthese selbst bei energieneutralen Netzen anwendbar, die Übertrager und Gyratoren mit variablen Übersetzungsverhältnissen bzw. Gyrationswiderständen enthalten [44]. Diese Vorgehensweise hat aber

keinen großen Nutzen, weil die symbolischen Rechnungen lediglich bei kleinen energieneutralen Netzen zu handhabbaren Ausdrücken führen. Zudem steigt der Implementierungsaufwand für das energieneutrale Netz im Wellendigitalbereich quadratisch mit der Zahl der Tore [43] und führt schnell zu einer ineffizienten Lösung. Der Aufwand für die Implementierung kann durch eine geeignete Zerlegung des energieneutralen Netzes reduziert werden, wenn eine Abspaltung energieneutraler Mehrtore gelingt, die im Wellendigitalbereich keine verzögerungsfreien gerichteten Schleifen verursachen. Für die automatische Zerlegung des energieneutralen Netzes gibt es einen effizienten Algorithmus, der zunächst für KIRCHHOFF'sche Schaltungen, die allein aus Eintoren aufgebaut sind, entwickelt worden ist [45]. Mit einer Modifikation wird dieser Algorithmus auch zur Zerlegung von KIRCHHOFF'schen Schaltungen, die zudem Übertrager [46] oder generell Mehrtore [47] enthalten, anwendbar. Ausgehend von der elektrischen Repräsentation der Differentialgleichung führt diese Methode systematisch zu einer Referenzschaltung bzw. zu einem Wellendigitalmodell. Da gerade mehrdimensionale Wellendigitalmodelle einen hohen Rechenaufwand benötigen, besteht hier ein erhöhtes Interesse, den Aufwand zu reduzieren.

Das Beispiel des vorherigen Abschnitts zeigt, dass eine Aufwandsreduktion erreicht werden kann, indem die elektrische Repräsentation der partiellen Differentialgleichung mit den Methoden der Schaltungstheorie vereinfacht wird. In [38] wird eine praktische Anleitung für die manuelle Synthese einer mehrdimensionalen Referenzschaltung angegeben, die von der übersichtlichen Schreibweise des Differentialoperators einer energetischen Spule profitiert. Die Synthese wird am Beispiel der NAVIER-STOKES-Gleichungen demonstriert, zu denen bereits Wellendigitalmodelle existieren [48] bzw. [30], die eine gesteuerte Spannungsquelle enthalten, deren zugeführte Energie von einem Widerstand dissipiert wird. Diese Anordnung konnte in der neuen Referenzschaltung durch einen energieneutralen gesteuerten Gyrator ersetzt werden. Die besondere Herausforderung der NAVIER-STOKES-Gleichungen ist der unstetige Verlauf der Lösungen, da die Unstetigkeiten bei einer Verwendung der Trapez-Regel zu erheblichen numerischen Schwierigkeiten führen [30]. Diese durch Idealisierungen verursachten Schwierigkeiten ließen sich ohne eine tiefere Modellierung mit einem Wechsel auf ein anderes numerisches Integrationsverfahren bei einer Wellendigitalsimulation vermeiden [38], [49]. Die Aufwandsreduktion einer mehrdimensionalen KIRCHHOFF'schen Schaltung zielt vornehmlich darauf ab, die Zahl der Bauelemente zu minimieren, ohne die Passivität der Schaltung zu opfern. In [44] wird diese Minimierung für passive physikalische Systeme, die durch eine lineare partielle Differentialgleichung beschrieben werden, systematisch am Beispiel der MAXWELL-Gleichungen dargelegt. Die Güte dieser Schaltungssynthese zeigt die Einsparung einer Spule bei einem Vergleich mit einer etablierten Lösung [50], wobei allerdings eine Einschränkung auf den linearen Fall und zwei Ortskoordinaten vorliegt. In einer umfangreicheren Arbeit [51] ist die Anwendbarkeit der vorgeschlagenen Schaltungssynthese an partiellen Differentialgleichungen aus der Mechanik, Elektrodynamik, Festkörper- und Kernphysik, zu denen bekannte Referenzschaltungen vorliegen [16], [52], [32], gezeigt worden. Zur Effizienzsteigerung der digitalen Nachbildung wird für manuelle Umformungen ein Kalkül angegeben, mit dem sich sowohl die Zahl der Bauelemente als auch die Geschwindigkeit v_k reduzieren lässt. Zumindest bei konstanten Parametern ist die Effizienz der resultierenden Wellendigitalmodelle im Vergleich zu den etablierten Wellendigitalmodellen gleich oder gesteigert. Gegenwärtig erscheint es aussichtsreich, eine hyperbolische partielle Differentialgleichung nach Ortskoordinaten aufzuteilen, so dass eine Summe hyperbolischer partieller Differentialgleichungen entsteht, für deren Summanden es leichter ist, eine Referenzschaltung zu synthetisieren [53]. Ungeachtet der erzielten Erfolge gelingt die Synthese bislang nur dann systematisch, wenn schwach besetzte Induktivitätsmatrizen vorliegen. Für die Zukunft ist die Entwicklung eines rechnergestützten Verfahrens erstrebenswert, das ausgehend von der Differentialgleichung eines

physikalischen Systems automatisiert einen Algorithmus zur digitalen Nachbildung generiert.

Zu guter Letzt werden ein paar Vorteile der elektrischen Repräsentation erwähnt, die nicht im unmittelbaren Zusammenhang mit der digitalen Nachbildung stehen. Die elektrische Repräsentation ist so universell wie eine Differentialgleichung und transferiert die Beurteilung eines physikalischen Systems auf eine elektrotechnische Ebene. Unter vorausgesetzter Kausalität kann Passivität auf einfachste Weise anhand der Bauelementwerte der minimalen Schaltung analysiert werden. In diesem Sinne ist die hier vorgestellte elektrische Repräsentation eine *Passivitätsnormalform* für die Differentialgleichung eines passiven physikalischen Systems. Im Gegensatz zu anderen Normalformen, die auf einer speziellen mathematischen Formulierung beruhen, liefert die Passivitätsnormalform mit der elektrischen Repräsentation Strukturinformation. Das zugehörige Schaltbild ist eine grafische Darstellung der Differentialgleichung und begünstigt auf diese Weise eine funktionale Analyse unter Berücksichtigung energetischer Eigenschaften. Dieser Aspekt ist insofern bedeutsam, als dass eine mental beherrschbare Analyse des Algorithmus zur digitalen Nachbildung möglich wird, obgleich keine analytische Lösung für die Differentialgleichung existiert.

Literatur

[1] Alfred Fettweis. Digital Filters related to Classical Filter Networks. *Intern. Journal of Electronics and Communications (AEÜ)*, 25:79–89, 1971.

[2] Alfred Fettweis. Wave digital filters: Theory and practice. *Proceedings of the IEEE*, 74:270–327, February 1986.

[3] Klaus Meerkötter. *Beiträge zur Theorie der Wellendigitalfilter*. Ruhr-Universität Bochum, 1979.

[4] Alfred Fettweis. Robust numerical integration using wave-digital concepts. *Multidimensional Systems and Signal Processing*, 17:7–25, January 2006.

[5] Hans Dieter Fischer. Wave digital filters for numerical integration. *NTZ Archiv*, 6:37–40, 1984.

[6] Klaus Meerkötter and Reinhard Scholz. Digital Simulation of Nonlinear Circuits by Wave Digital Filter Principles. *Proceedings IEEE International Symposium on Circuits and Systems*, 1:720–723, 1989.

[7] Thomas Felderhoff. Simulation of nonlinear circuits with periodic doubling and chaotic behavior by wave digital filter principles. *IEEE Transactions on Circuits and Systems*, 41(1):485–491, 1994.

[8] Thomas Felderhoff. A new wave description for nonlinear elements. *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, 3:221–224, 1996.

[9] S. Petrausch and R. Rabenstein. Wave Digital Filters with Multiple Nonlinearities. *European Signal Processing Conference (EUSIPCO)*, 47(6):77–80, 2004.

[10] Alfred Fettweis and Gunnar Nitsche. Numerical integration of partial differential equations by means of multidimensional wave digital filters. *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, 2:954–957, 1990.

[11] Alfred Fettweis and Gunnar Nitsche. Numerical integration of partial differential equations using principles of multidimensional wave digital filters. *Journal of VLSI Signal Processing*, 3:7–24, 1991.

[12] Alfred Fettweis. Discrete modelling of lossless fluid dynamic systems. *Int. J. Electron. Commun. (AEÜ)*, 46:209–218, 1992.

[13] Georg Hetmanczyk. Exploiting the parallelism of multidimensional wave digital algorithms on multicore computers. *Multidimensional Systems and Signal Processing*, 21(1):45–58, March 2010.

[14] Katrin Luhmann and Karlheinz Ochs. A novel interpretation of the hyperbolization method used to solve the parabolic neutron diffusion equations by means of the wave digital concept. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 19(4):345–364, July 2006.

[15] Germund Dahlquist. A special stability problem for linear multistep methods. *BIT*, 3(1):27–43, 1963.

[16] Gunnar Nitsche. *Numerische Lösung partieller Differentialgleichungen mit Hilfe von Wellendigitalfiltern*. Fortschritt-Berichte VDI, 1993.

[17] Thomas Felderhoff. *Digitale Simulation nichtlinearer Systeme mit Methoden der Netzwerktheorie*. Shaker Verlag, Aachen, 1995.

[18] Karlheinz Ochs. *Passive Integrationsmethoden*. Shaker Verlag, Aachen, 2001.

[19] Georg Hetmanczyk and Karlheinz Ochs. Initialization of linear multistep methods in multidimensional wave digital models. *Proceedings of the 52nd Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 786–789, August 2009.

[20] Karlheinz Ochs. Passive integration methods: Fundamental theory. *Intern. Journal of Electronics and Communications (AEÜ)*, 55(3):153–163, May 2001.

[21] Dietrich Fränken and Karlheinz Ochs. Numerical stability properties of passive Runge-Kutta methods. *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, 3:473–476, May 2001.

[22] Dietrich Fränken. *Digitale Simulation physikalischer Systeme mit Methoden der Netzwerktheorie*. VDI-Verlag, Düsseldorf, 2003.

[23] Dietrich Fränken and Karlheinz Ochs. Synthesis and design of passive Runge-Kutta methods. *Intern. Journal of Electronics and Communications (AEÜ)*, 55(6):417–425, November 2001.

[24] Dietrich Fränken and Karlheinz Ochs. Passive Runge-Kutta methods - properties, parametric representation, and order conditions. *BIT Numerical Mathematics*, 43(2):339–361, June 2003.

[25] Dietrich Fränken, Klaus Meerkötter, and Karlheinz Ochs. Eine besondere Klasse 2-stufiger passiver Runge-Kutta-Verfahren. *15. Symposium Simulationstechnik (ASIM)*, pp. 503–508, September 2001.

[26] Dietrich Fränken, Karlheinz Ochs, and Markus Schmidt. A passive two-step Runge-Kutta method for the simulation of nonlinear electrical networks. *International Symposium on Nonlinear Theory and its Applications (NOLTA)*, pp. 347–350, Oktober 2002.

[27] Dietrich Fränken and Karlheinz Ochs. Improving wave digital simulation by extrapolation techniques. *Intern. Journal of Electronics and Communications (AEÜ)*, 56(5):327–336, September 2002.

[28] Dietrich Fränken and Karlheinz Ochs. Automatic step-size control in wave digital simulation using passive numerical integration methods. *Intern. Journal of Electronics and Communications (AEÜ)*, 58(6):391–401, November 2004.

[29] Hans Dieter Fischer. Anwendung von Wellen-Digitalfiltern auf parabolische PDEs: Neutronendiffusion in 2 Energiegruppen. *Siemens Arbeitsbericht*, KWU/U8 241-89/76:1–23, 1989.

[30] André Mengel. *Untersuchung zur numerischen Lösung der Navier-Stokes Gleichungen mit Wellendigital-Prinzipien.* Cuvillier, Göttingen, 2007.

[31] Dietrich Fränken, Klaus Meerkötter, and Joachim Waßmuth. Observer-based Feedback Linearization of Dynamic Loudspeakers with AC Amplifiers. *IEEE Transactions on Speech and Audio Processing*, 13(2):233–242, 2005.

[32] Katrin Luhmann. *Die numerische Lösung der Neutronendiffusionsgleichungen in zwei Energiegruppen mit dem Wellendigital-Konzept.* Cuvillier, Göttingen, 2004.

[33] Hans Dieter Fischer. Technische Zuverlässigkeit. Monographie zur Vorlesung, Ruhr-Universität Bochum, 2009.

[34] Rüdiger Pott. *Numerische Integration von Neutronendiffusionsgleichungen unter Verwendung mehrdimensionaler Wellendigitalfilter auf parallelen Rechnerarchitekturen.* Shaker Verlag, Aachen, 1998.

[35] Elmar Rummert. *Methodik eines formalen Korrektheitsbeweises bei graphisch spezifizierter Software am Beispiel von Wellendigitalfiltern.* Shaker Verlag, Aachen, 1998.

[36] Michael Vollmer. *Automatische Code-Erzeugung zur numerischen Integration partieller Differentialgleichungen für sicherheitskritische Anwendungen.* Cuvillier, Göttingen, 2004.

[37] Thomas Lutter. *Eine schnelle numerische Lösung der Neutronendiffusionsgleichungen nach dem Wellendigitalprinzip.* Cuvillier, Göttingen, 2009.

[38] Georg Hetmanczyk and Karlheinz Ochs. A practical guide to multidimensional wave digital algorithms using an example of fluid dynamics. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 2010. Zur Veröffentlichung angenommen.

[39] Alfred Fettweis and Gunnar Nitsche. Transformation approach to numerically integrating PDEs by means of WDF principles. *Multidimensional Systems and Signal Processing*, 2:127–159, May 1991.

[40] Alfred Fettweis. Wellendigitalfilter. Folien zur Vorlesung, Ruhr-Universität Bochum, 2009.

[41] Karlheinz Ochs. Wave digital simulation of passive systems in linear state-space form. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 23(1):42–61, Januar/Februar 2010.

[42] Dietrich Fränken and Klaus Meerkötter. Digital realization of connection networks by voltage-wave two-port adaptors. *Intern. Journal of Electronics and Communications (AEÜ)*, 50(6):362–367, 1996.

[43] Dietrich Fränken. *Passive Systeme zur Verarbeitung komplexer zeitdiskreter Signale*. Shaker Verlag, Aachen, 1997.

[44] Christiane Leuer and Karlheinz Ochs. Systematic derivation of reference circuits for wave digital modeling of passive linear partial differential equations. *Proceedings of the 52nd Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 782–785, August 2009.

[45] Karlheinz Ochs and Benno Stein. Systematischer Entwurf von Wellendigitalstrukturen. *15. Symposium Simulationstechnik (ASIM)*, 45:61–66, September 2001.

[46] Dietrich Fränken, Jörg Ochs, and Karlheinz Ochs. Generation of wave digital structures for connection networks containing ideal transformers. *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, 3:240–243, May 2003.

[47] Dietrich Fränken, Jörg Ochs, and Karlheinz Ochs. Generation of wave digital structures for networks containing multi-port elements. *IEEE Transactions on Circuits and Systems-I: Regular Papers*, 52(3):586–596, March 2005.

[48] Alfred Fettweis. Improved wave-digital approach to numerically integrating the PDEs of fluid dynamics. *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, 3:361–364, 2002.

[49] Georg Hetmanczyk and Karlheinz Ochs. Wave digital simulation of Burgers' equation using Gear's method. *Proceedings of the 51st Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 161–164, August 2008.

[50] Alfred Fettweis. Multidimensional wave digital filters for discrete-time modelling of Maxwell's equations. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 5(3):183–201, 1992.

[51] Christiane Leuer and Karlheinz Ochs. Systematische Wellendigital-Synthese für eine Klasse passiver verteilter Systeme. *16. Steirisches Seminar über Regelungstechnik und Prozessautomatisierung (SSRP)*, pp. 117–149, September 2009.

[52] Alfred Fettweis. Numerical integration of partial differential equations using wave-digital principles. Collection of detailed viewgraphs (at present 357 individual viewgraphs), 2000-2002.

[53] Christiane Leuer and Karlheinz Ochs. On systematic wave digital modeling of passive hyperbolic partial differential equations. *International Journal of Circuit Theory and Applications*, pp. 1–23, February 2011. doi:10.1002/cta.752.

SSRP2011 ISBN: 978-3-901439-09-4