

Peer Review: 01.04.2019

Ein neues Verfahren zur Modellierung linearer Systeme

Experimentelle Modellbildung mit Ein- und Ausgangsdaten

Peter Zentgraf, Technische Hochschule Rosenheim

Es wird ein Verfahren zum Simulieren von Vorgängen vorgestellt, das auf der Messung des Eingriffs in ein reales System und dessen gemessener Reaktion beruht. Das Verfahren ist vollständig transparent, denn die notwendigen theoretischen Voraussetzungen zum Verständnis des Verfahrens bestehen nur aus Lehrinhalten aus dem Bachelorstudium für Ingenieure: das Lösen linearer DGLs mit konstanten Koeffizienten, die Laplace-Transformation und die Methode der kleinsten Quadrate. Es kommt ohne nichtlineare Optimierer oder Iterationen aus, weil die Lösung in zwei analytischen Gleichungen ausgedrückt werden kann. Das Verfahren hat in der Praxis ein großes Einsatzspektrum, denn es ist auf stabile und instabile Systeme mit und ohne Dämpfung anwendbar. Für die Wahl des Eingangssignals gibt es keine Einschränkung und die Anfangsbedingungen können mitgeschätzt werden. Das Verfahren kann im offenen und im geschlossenen Regelkreis eingesetzt werden.

#Experimentelle Modellbildung #Simulation #Lineare DGL #Einfachheit

A new method for modelling linear systems

Experimental modelling using input/output data

A novel method is presented for simulating technical processes on the basis of measured reaction of a system excited by a known input signal. The method is fully transparent and involves only linear differential equations, Laplace transformations, and the least squares method. There is no need for a nonlinear optimizer or an algorithm requiring iterations, and the solutions can be expressed by two linear matrix equations. The method can be used for a wide range of practical applications because the initial conditions can also be estimated it is applicable for stable or unstable systems, for open or closed loop systems and for dynamic processes in progress.

#experiment-based modelling #simulation #linear differential equations #simplicity

Zur Lösung einer regelungstechnischen Entwurfsaufgabe benötigt man ein Simulationsmodell der Regelstrecke. Entweder verwendet man es, weil es die meisten Entwurfsverfahren zur Bestimmung der Reglerparameter als gegeben voraussetzt oder weil man das Verhalten des Regelkreises zur Verifikation des gewünschten Verhaltens gefahrlos simulieren möchte, bevor man den Regler in dem realen geschlossenen Regelkreis verwendet.

Die Haupt-Motivation zur Entwicklung des hier vorgestellten Verfahrens ist es, Bachelor-Studenten der Ingenieurwissenschaften aus dem höheren Semester, die zum ersten Mal mit Regelungstechnik in Berührung kommen, ein möglichst allgemein anwendbares Verfahren zur Bildung von Streckenübertragungsfunktionen an die Hand zu geben, das von der Theorie her vollständig transparent und allein mit dem aus der Mathematik bisher Gelernten für sie verstehbar ist.

Grundsätzlich gibt es zwei verschiedene Arten, ein Simulationsmodell der Regelstrecke zu gewinnen [1]: Bei der *physikalischen*

oder *theoretischen* Modellbildung werden physikalische Erhaltungssätze auf den idealisierten, sprich vereinfachten technischen Vorgang angewendet und nach den gewählten Ein- beziehungsweise Ausgangssignalen umgeformt. Diese Vorgehensweise kann sehr zeitaufwendig sein. Es kommt auch oft vor, dass Parameter des Modells meist unbekannt sind. Bei der *experimentellen* Modellbildung [1, 2] geht man von einem gemessenen Ein- beziehungsweise Ausgangssignal aus und gleicht über mehr oder weniger aufwendige Optimierungsverfahren die simulierte Antwort an die gemessene Antwort an. Den *bislang* bekannten experimentellen Verfahren [1, 2] ist gemein, dass sie eines oder mehrere der folgenden Merkmale aufweisen:

- » Sie sind nur auf *bestimmte Streckentypen* anwendbar.
- » Sie sind nur im *offenen* Kreis und mit nur *speziellen*, meist sprungförmigen Testsignalen anwendbar.

- » Sie gehen von einer Regelstrecke in der *Ruhelage* aus.
- » Sie erfordern den Einsatz komplizierter *mathematischer Optimierungsverfahren*.
- » Sie identifizieren die Koeffizienten *zeitdiskreter z-Übertragungsfunktionen*.

Bei dem letzten Merkmal könnte man sich fragen, was denn der Vorteil sei, direkt im zeitkontinuierlichen Laplace- beziehungsweise s-Bereich zu identifizieren. Denn eine Identifikation im z-Bereich mit anschließender Rücktransformation in den s-Bereich ergäbe ja auch eine s-Übertragungsfunktion. Der Unterschied liegt darin, dass sich abhängig von den Rücktransformationsverfahren – zum Beispiel ausgehend von dem Euler-Vorwärts-, Euler-Rückwärts- oder dem Tustin-Verfahren [1] – verschiedene s-Übertragungsfunktionen teilweise mit unterschiedlichen Zähler-Ordnungen ergeben. Somit verlieren auch die rückgerechneten Koeffizienten der s-Übertragungsfunktion ihre physikalische Bedeutung. Mit dem hier vorgestellten Verfahren hingegen lassen sich in dem RC-Netzwerk-Beispiel in Abschnitt 2.1 aus dem Identifikationsergebnis über die Gleichungen (36), (37) wieder die Werte des elektrischen Widerstandes und die der Kondensatoren errechnen.

In [3] wird ein Verfahren [4] zitiert, das dem hier vorgestellten Verfahren ähnelt. Allerdings muss sich dort als Voraussetzung für die Anwendbarkeit das System in der Ruhelage befinden. In dem hier vorgestellten Verfahren können auch die Anfangsbedingungen mitgeschätzt werden. Ferner werden mehr Freiheitsgrade in der Verbesserung der Modellgüte verwendet. Die Herleitung hier mit Hilfe der Laplace-Transformation ist auch deutlich transparenter.

1. Beschreibung des Identifikationsverfahrens

Die Aufgabenstellung ist die Folgende: Aus einem einzigen Datensatz mit fehlerhaft gemessenen Eingangssignal $u(t)$ und Ausgangssignal $y(t)$ soll das komplette, allgemeine Systemverhalten des Modells bestimmt werden – dabei beschränkt sich der vorliegende Beitrag der Einfachheit halber und ohne Einschränkung der Allgemeingültigkeit auf die linearen zeitinvarianten DGLs und nimmt die Messfehler als mittelwertfreie Zufallszahlen an. Das Eingangssignal sollte genug Dynamik besitzen, damit die Ableitungen von Ein- und Ausgangssignal möglichst in einem großen Zeitabschnitt von Null verschieden sind. Denn nur in diesem Fall lassen sich die zugehörigen Koeffizienten numerisch sicher berechnen. Hat man im anderen Extremfall nur das stationäre Verhalten eines Systems zur Verfügung, so ließe sich nur ein reines Proportionalglied identifizieren. Das Ziel der Identifikation soll sein, auch für vollkommen andere Eingangssignale das zugehörige Ausgangsverhalten vorherzusagen, ohne erneut messen zu müssen.

1.1 Beschreibung der Regelstrecke im Zeitbereich

Das dynamische Verhalten eines linearen – oder um einen Arbeitspunkt linearisierten – zeitinvarianten Systems mit Eingangsgröße $u(t)$ und Ausgangsgröße $y(t)$ lässt sich in Form einer DGL (DGL) folgendermaßen beschreiben:

$$\begin{aligned} b_0 \cdot y(t) + b_1 \cdot \dot{y}(t) + b_2 \cdot \ddot{y}(t) + \dots + b_n \cdot y^{(n)}(t) = \\ a_0 \cdot u(t) + a_1 \cdot \dot{u}(t) + a_2 \cdot \ddot{u}(t) + \dots + a_m \cdot u^{(m)}(t) \end{aligned} \quad (1)$$

Dabei ist die Methode aber nicht auf lineare Systeme beschränkt. Auch nichtlineare Systeme lassen sich häufig in eine passende Form umformen. Entscheidend ist viel mehr, ob die zu identifizierenden Systemparameter linear in der Systembeschreibung vorkommen oder, wenn das nicht der Fall ist, ob sie wie im vorliegenden Fall durch Ersetzen von Zwischengrößen in die passende Form gebracht werden, um anschließend die Zwischengrößen wieder nach den eigentlich gesuchten Systemparametern aufzulösen.

Willkürlich kann man die gesamte Gleichung (1) durch einen der Koeffizienten (der nicht Null sein darf) teilen. Die DGL wird so auf einen Koeffizienten normiert. Auf eine entsprechende ausführlichere Nomenklatur wie beispielsweise $\hat{b}_1 = b_1/b_3$ wird hier verzichtet, da sich die DGL nicht verändert hat. Die Folge ist, dass sich die Anzahl der unbekanntenen Koeffizienten um eins verringert und ein Koeffizient auf Eins gesetzt wird. Alternativ hätte man auch durch $-b_3$ oder durch $-10 \cdot b_3$ teilen können: Dann würde ein Koeffizient auf den Wert -1 beziehungsweise -0.1 gesetzt.

Im Grunde genommen kann man den Wert *eines* Koeffizienten frei wählen. Wäre die Messung perfekt und handelte es sich exakt um ein wie in Gleichung (3) beschriebenes System, dann hätte dieses Vorgehen keinen Einfluss auf die zu bestimmende DGL. Weil aber das nicht der Fall ist, ändern sich die Ergebnisse und man erhält weitere mögliche Lösungen: Hier ist ein Freiheitsgrad versteckt, mit dem sich die Modellgüte verbessern lässt. In den folgenden mathematischen Beschreibungen wurde willkürlich $b_3 = 1$ gesetzt.

Das Verfahren wird ohne Einschränkung der Allgemeingültigkeit anhand einer DGL dritter Ordnung (in Gleichung (1) mit $m=0, n=3$) vorgestellt:

$$b_0 \cdot y(t) + b_1 \cdot \dot{y}(t) + b_2 \cdot \ddot{y}(t) + 1 \cdot \ddot{y}(t) = a_0 \cdot u(t) \quad (2)$$

Diese Gleichung lässt sich für die Korrespondenz mit dem allgemeinen Laplace'schen Differentiationssatz

$$f^{(n)}(t) = s^n \cdot F(s) - \sum_{k=0}^{n-1} f^{(k)}(0) \cdot s^{n-k-1} \quad (3)$$

für Ein- und Ausgangssignale in den Laplace-Bereich transformieren.

1.2 Beschreibung der Regelstrecke im Laplace-Bereich

In der Praxis hat man es mit kontinuierlich betriebenen Prozessen in Produktionsstätten zu tun, die man nur in Ausnahmefällen aus einer Ruhelage heraus starten kann. Es ist kaum davon auszugehen, dass ein Prozess nur für eine Modellbildung gestoppt und dann aus der Ruhelage hochgefahren wird. In einem laufenden Prozess aber haben die Ableitungen des Ausgangssignals nicht gleichzeitig alle den Wert Null, das heißt

$y_0 = \dot{y}_0 = \ddot{y}_0 = 0$ gilt nicht. Für die Laplace-Transformierten der einzelnen Summanden aus Gleichung (2), (3) ergibt sich für nicht verschwindende Anfangsbedingungen:

$$b_0 \cdot y(t) \rightsquigarrow b_0 \cdot Y(s) \quad (4)$$

$$b_1 \cdot \dot{y}(t) \rightsquigarrow b_1 \cdot (s \cdot Y(s) - y_0) \quad (5)$$

$$b_2 \cdot \ddot{y}(t) \rightsquigarrow b_2 \cdot (s^2 \cdot Y(s) - s \cdot y_0 - \dot{y}_0) \quad (6)$$

$$1 \cdot \ddot{y}(t) \rightsquigarrow s^3 \cdot Y(s) - s^2 \cdot y_0 - s \cdot \dot{y}_0 - \ddot{y}_0 \quad (7)$$

$$a_0 \cdot u(t) \rightsquigarrow a_0 \cdot U(s) \quad (8)$$

Es werden ohne Einschränkung der Allgemeingültigkeit nur Eingangssignale $u(t)$ betrachtet, die für $t < 0$ den Wert Null haben, das gilt dann auch für alle Anfangswerte der Ableitungen. Die Anfangswerte des Ausgangssignals $y(t)$ können für $t < 0$ beliebige Werte ungleich Null annehmen, sie reflektieren den allgemeinen Zustand des Systems aus der Vergangenheit bis zum Zeitpunkt $t = 0$, der nicht unbedingt einer Ruhelage entsprechen muss.

Im Allgemeinen sind die Aufnahme der Messdaten und die Auswahl des Messbereichs für die Schätzung unabhängig voneinander. Es ist sogar wichtig, nicht alle Messdaten für die Schätzung zu verwenden, damit man das ermittelte Modell auch anhand von unabhängigen Messdaten überprüfen kann. Das Gleiche gilt für den Zeitpunkt, den man als *Anfang* festlegt, ab dem geschätzt wird. Dieser kann, aber muss nicht mit dem ersten aufgenommenen Datensatz übereinstimmen. Für Sensitivitätsanalysen empfiehlt es sich außerdem, den Bereich der Messdaten, der für die Schätzung verwendet wird, zu verändern.

Als Wert der Anfangsbedingung y_0 wird der Messwert an dem Zeitpunkt verwendet, der als *Anfang* festgelegt wurde; das hat auch den Vorteil, dass die Simulation exakt mit demselben Wert der Messung beginnt. Die Anfangsableitungen \dot{y}_0, \ddot{y}_0 sollte man genau wie die unbekannt Koeffizienten b_0, b_1, b_2, a_0 mitschätzen. Denn bekanntermaßen ist die numerische Differentiation von verrauschten Signalen fehleranfällig; das heißt, selbst ein im Prinzip konstantes Signal wie beispielsweise $y(t) = 5,00 \text{ V} = \text{const.}$ aus einer elektrischen Gleichspannungsquelle hat nie einen absolut glatten Verlauf, sondern ist immer von zufälligen Anteilen (Rauschen) überlagert. Man erhält beliebig steile Flanken zwischen den einzelnen Messpunkten. Der Differenzenquotient $\Delta y / \Delta t$, mit dem sich die Ableitung bei rauscharmen Signalen annähern lässt, wird auf diese Weise bei sehr kleinem Δt beliebig groß. Dieser Effekt verstärkt sich mit höheren Ableitungen. Beim Integrieren hingegen verkleinert sich der Einfluss. Die Rauschanteile werden dann *herausgemittelt*. Diese Überlegung motiviert das vorgestellte Verfahren in der Art, dass die in einer klassischen DGL vorkommenden Ableitungen durch Integration derselben vermieden werden; dieser Aspekt ist dem in [4] beschriebenen Verfahren ähnlich, denn dort wird unter der Annahme, dass die Anfangsbedingung verschwindet, ebenfalls die DGL integriert.

Laplace-Transformation von Gleichung (2) mit den Gleichungen (4) – (8) ergibt

$$b_0 \cdot Y(s) + b_1 \cdot (s \cdot Y(s) - y_0) + b_2 \cdot (s^2 \cdot Y(s) - s \cdot y_0 - \dot{y}_0) + s^3 \cdot Y(s) - s^2 \cdot y_0 - s \cdot \dot{y}_0 - \ddot{y}_0 = a_0 \cdot U(s) \quad (9)$$

Anschließendes Auflösen nach $Y(s)$ löst die DGL im Laplace-Bereich unter Berücksichtigung der Anfangsbedingungen $y_0, \dot{y}_0, \ddot{y}_0$ und Auflösen nach dem Ausgangssignal:

$$Y(s)(b_0 + b_1 \cdot s + b_2 \cdot s^2 + s^3) = a_0 \cdot U(s) + b_1 \cdot y_0 + b_2 \dot{y}_0 + \ddot{y}_0 + s \cdot (b_2 \cdot y_0 + \dot{y}_0) + s^2 \cdot y_0 \quad (10)$$

$$Y(s) = \underbrace{\frac{a_0}{(b_0 + b_1 \cdot s + b_2 \cdot s^2 + s^3)} \cdot U(s)}_{\text{Teil 1}} + \underbrace{\frac{b_1 \cdot y_0 + b_2 \dot{y}_0 + \ddot{y}_0 + s \cdot (b_2 \cdot y_0 + \dot{y}_0) + s^2 \cdot y_0}{(b_0 + b_1 \cdot s + b_2 \cdot s^2 + s^3)}}_{\text{Teil 2}} \quad (11)$$

Es wird deutlich, dass sich das Ausgangssignal $Y(s)$ in Gleichung (11) aus zwei Teilen zusammensetzt:

1. In Teil 1 wird das Eingangssignal $U(s)$ ganz klassisch nur über die Übertragungsfunktion $G(s)$ auf den Ausgang übertragen. Das ist ausreichend, wenn die Anfangsbedingungen Null sind.
2. Der Teil 2 enthält nicht das Eingangssignal $U(s)$ und macht den Einfluss der Anfangsbedingungen deutlich. Die Pole von Teil 2 sind die Pole der Übertragungsfunktion $G(s)$. Die Nullstellen werden aus den Anfangsbedingungen $y_0, \dot{y}_0, \ddot{y}_0$ gebildet. Sind diese Null, dann ist auch der ganze Ausdruck von Teil 2 Null. Mit Gleichung (11) lässt sich das System mit Anfangsbedingungen ungleich Null simulieren.

Im Laplace-Bereich entspricht einer Integration die Multiplikation mit $1/s$. Aus diesem Grunde wird im nächsten Schritt Gleichung (9) mit $1/s^3$ multipliziert. Dadurch werden alle Multiplikationen mit „s“, welche im Zeitbereich den Ableitungen entsprechen, vermieden. Aber auch eine Multiplikation mit $1/s^4$ oder höherer Ordnung ist denkbar. Dieses *Überintegrieren* stellt somit einen weiteren Freiheitsgrad der Modellbildung der hier vorgestellten Methode dar. Dadurch erhält man weitere Kandidaten zur Lösung des Problems, aus denen dann das beste Modell zum Beispiel im Sinne der kleinsten Abweichung oder in Form einer gewünschten Eigenschaft wie Stabilität oder Instabilität ausgewählt wird.

Somit ist aus der klassischen DGL, die Ableitungen der Signale enthält, eine Integralbeziehung geworden. Dieses technisch einfachere zu lösende Ersatzproblem lautet sortiert nach den Unbekannten wie folgt:

$$\begin{aligned} & \left(Y(s) - \frac{1}{s} \cdot y_0 - \frac{1}{s^2} \cdot \dot{y}_0 - \frac{1}{s^3} \cdot \ddot{y}_0 \right) + b_2 \cdot \left(\frac{1}{s} \cdot Y(s) - \frac{1}{s^2} \cdot y_0 \right) \\ & + b_1 \cdot \dot{y}_0 \cdot \left(-\frac{1}{s^3} \right) + b_1 \cdot \left(\frac{1}{s^2} \cdot Y(s) - \frac{1}{s^3} \cdot y_0 \right) + b_0 \cdot \frac{1}{s^3} \cdot Y(s) = \\ & a_0 \cdot \frac{1}{s^3} \cdot U(s) \end{aligned} \tag{12}$$

In Gleichung (12) taucht das Produkt zweier Unbekannter auf, nämlich $b_2 \cdot \dot{y}_0$, das ebenfalls unbekannt ist. In dieser Form lässt sich die Methode der kleinsten Quadrate [6] nicht direkt anwenden, denn mit dem Ansatz des Verfahrens ist es nicht möglich, dass sich zwei Unbekannte miteinander multiplizieren. Es werden daher zwei Zwischengrößen, y_{20} und y_{10} , eingeführt, die letztendlich doch eine Anwendung der Methode erlauben:

$$\begin{aligned} & Y(s) - y_0 \cdot \frac{1}{s} + b_2 \cdot \left(\frac{1}{s} \cdot Y(s) - \frac{1}{s^2} \cdot y_0 \right) + \\ & b_1 \cdot \left(\frac{1}{s^2} \cdot Y(s) - \frac{1}{s^3} \cdot y_0 \right) + b_0 \cdot \frac{1}{s^3} \cdot Y(s) - \\ & \left(\frac{b_2 \cdot \dot{y}_0 + \ddot{y}_0}{=:y_{20}} \right) \cdot \frac{1}{s^3} - \frac{\dot{y}_0}{=:y_{10}} \cdot \frac{1}{s^2} = a_0 \cdot U(s) \cdot \frac{1}{s^3} \end{aligned} \tag{13}$$

Anschließend kann nach Berechnung von b_2 von den errechneten Werten der Zwischengrößen auf \dot{y}_0, \ddot{y}_0 zurückgerechnet werden.

In Gleichung (13) entgeht man dadurch, dass y_0 als bekannt angenommen wird, zwei zusätzlichen unabhängigen Zwischengrößen für $b_1 \cdot y_0$ und $b_2 \cdot y_0$. Weil aber die Unbekannten b_1 und b_2 ebenfalls unabhängig voneinander zu ermitteln sind, käme es bei der Auflösung der Zwischengrößen für $b_1 \cdot y_0$ und $b_2 \cdot y_0$ zu Widersprüchen.

Gleichung (13) lässt sich mit dem Laplace'schen Integrationssatz

$$\frac{1}{s} \cdot F(s) \Leftrightarrow \int_{t_0}^t f(\tau) d\tau =: f_i(t) \tag{14}$$

in den Zeitbereich zurück transformieren. Das einfache Integral im Zeitbereich wird abkürzend als $f_i(t)$ bezeichnet, das zweifache Integral mit $f_{ii}(t)$ und so weiter.

Die Laplace-Korrespondenzen der Zeitfunktionen für $t \geq 0$ lassen sich wie folgt angeben [1]:

$$y_0 \cdot \frac{1}{s} \Leftrightarrow y_0 \tag{15}$$

$$y_0 \cdot \frac{1}{s^2} \Leftrightarrow y_0 \cdot t \tag{16}$$

$$y_0 \cdot \frac{1}{s^3} \Leftrightarrow y_0 \cdot \frac{t^2}{2} \tag{17}$$

Den Zeitpunkt t_i , ab dem die Messwerte für die Schätzung verwendet werden, legt man danach fest, wie geeignet der Messbereich für die Identifikation ist. Es ist sinnvoll, mit zwei Zeitkoordinaten zu arbeiten: Die eine Zeitkoordinate t_i bezieht sich auf die Auswahl der Messdaten, und die andere Zeitkoordinate \tilde{t} stellt den Beginn für die Schätzung dar. Dann ist der Beitrag der unteren Integrationsgrenze bei der Auflösung der bestimmten Integrale Null, und das reduziert die Anzahl der Terme. Die spätere Systemsimulation soll aber immer zum Zeitpunkt „0 s“ beginnen, Aus diesen Gründen wird daher die Variable \tilde{t} eingeführt, welche diese Bedingung erfüllt, also an der Stelle $t = t_i$ den Wert $\tilde{t} = 0$ s hat. Es gilt also $\tilde{t} = t - t_i$. Der Ausgang $y(t)$ wird dadurch zu $y(\tilde{t})$ und ebenso $u(t)$ zu $u(\tilde{t})$ zu.

Entsprechend den Ausdrücken $1/s, 1/s^2$ und $1/s^3$ werden die zugehörigen Signale ein-, zwei- oder dreimal im Zeitbereich integriert. Aus Gleichung (13) wird

$$\begin{aligned} & y(\tilde{t}) - y_0 + b_2 \cdot (y_i(\tilde{t}) - \tilde{t} \cdot y_0) + b_1 \cdot \left(y_{ii}(\tilde{t}) - \frac{1}{2} \cdot \tilde{t}^2 \cdot y_0 \right) + \\ & b_0 \cdot y_{iii}(\tilde{t}) = \frac{1}{2} \cdot \tilde{t}^2 \cdot y_{20} + \tilde{t} \cdot y_{10} + a_0 \cdot u_{iii}(\tilde{t}). \end{aligned} \tag{18}$$

Diese Gleichung wird nun an den Abtastzeitpunkten $\tilde{t} = [\tilde{t}_1, \tilde{t}_2, \dots, \tilde{t}_n]$ ausgewertet und pro Abtastzeitpunkt in Matrixschreibweise als eigene Zeile zusammengefasst:

$$\begin{aligned} & \begin{bmatrix} y(\tilde{t}_1) - y_0 \\ \vdots \\ y(\tilde{t}_n) - y_0 \end{bmatrix} + \begin{bmatrix} y_i(\tilde{t}_1) - \tilde{t}_1 \cdot y_0 \\ \vdots \\ y_i(\tilde{t}_n) - \tilde{t}_n \cdot y_0 \end{bmatrix} \cdot b_2 + \\ & \begin{bmatrix} y_{ii}(\tilde{t}_1) - \frac{1}{2} \cdot \tilde{t}_1^2 \cdot y_0 \\ \vdots \\ y_{ii}(\tilde{t}_n) - \frac{1}{2} \cdot \tilde{t}_n^2 \cdot y_0 \end{bmatrix} \cdot b_1 + \begin{bmatrix} y_{iii}(\tilde{t}_1) \\ \vdots \\ y_{iii}(\tilde{t}_n) \end{bmatrix} \cdot b_0 = \\ & \frac{1}{2} \cdot \begin{bmatrix} \tilde{t}_1^2 \\ \vdots \\ \tilde{t}_n^2 \end{bmatrix} \cdot y_{20} + \begin{bmatrix} \tilde{t}_1 \\ \vdots \\ \tilde{t}_n \end{bmatrix} \cdot y_{10} + \begin{bmatrix} u_{iii}(\tilde{t}_1) \\ \vdots \\ u_{iii}(\tilde{t}_n) \end{bmatrix} \cdot a_0 \end{aligned} \tag{19}$$

Nach Umstellen der obigen Gleichung entsteht die passende Form für die Bestimmung der Koeffizienten durch das Verfahren der kleinsten Quadrate [6]:

$$\begin{aligned} & \underbrace{\begin{bmatrix} y(\tilde{t}_1) - y_0 \\ \vdots \\ y(\tilde{t}_n) - y_0 \end{bmatrix}}_{\tilde{y}} = \dots \\ & = \underbrace{\begin{bmatrix} \tilde{t}_1 \cdot y_0 - y_i(\tilde{t}_1) & \frac{1}{2} \cdot \tilde{t}_1^2 \cdot y_0 - y_{ii}(\tilde{t}_1) & -y_{iii}(\tilde{t}_1) & u_{iii}(\tilde{t}_1) & \frac{1}{2} \cdot \tilde{t}_1^2 & \tilde{t}_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{t}_n \cdot y_0 - y_i(\tilde{t}_n) & \frac{1}{2} \cdot \tilde{t}_n^2 \cdot y_0 - y_{ii}(\tilde{t}_n) & -y_{iii}(\tilde{t}_n) & u_{iii}(\tilde{t}_n) & \frac{1}{2} \cdot \tilde{t}_n^2 & \tilde{t}_n \end{bmatrix}}_H \cdot \underbrace{\begin{bmatrix} b_2 \\ b_1 \\ b_0 \\ a_0 \\ y_{20} \\ y_{10} \\ x \end{bmatrix}}_x \end{aligned} \tag{20}$$

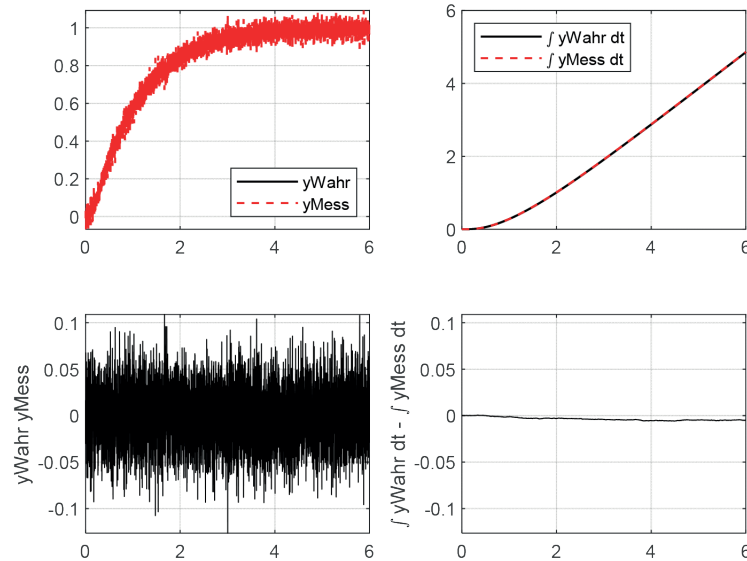


Bild 1: Direkte und integrierte Sprungantworten eines PT3-Systems

Gesucht wird derjenige Vektor x , der in Gleichung (20) die Vektoren \tilde{y} und $H \cdot x$ einander annähert. Man benötigt dazu ein passendes Gütekriterium.

Bei Gleichung (20) handelt es sich um eine Auswertung der dreifach integrierten DGL von Gleichung (2). Dies bedeutet, bei perfekter Messung des Ausgangs y als Reaktion eines wirklichen PT3-Systems würde man nur sechs Messungen benötigen, um in Gleichung (20) nach den sechs unbekannt Elementen des x -Vektors aufzulösen. Da Gleichung (20) aber zur Bestimmung von x auch bei fehlerbehafteten Messungen verwendet werden soll, werden viel mehr als nur sechs Messungen benötigt, damit sich die Messfehler ausgleichen können. Für die Verwendung von Gleichung (20) mit realen Messsignalen wird zu jedem Abtastzeitpunkt ein unbekannter Fehler-Vektor e hinzugefügt, der sowohl die nicht modellierbaren, zufälligen Messfehler als auch die Modellierungsfehler beinhaltet:

$$\tilde{y} = H \cdot x + e. \tag{21}$$

Für das Modell in Gleichung (21) lässt sich dann ein erwartungstreuer Schätzwert \hat{x} für den wahren x -Vektor x_{Wahr} bestimmen, das heißt der Erwartungswert $E(\hat{x}) = x_{Wahr}$, unter der Voraussetzung, dass die H -Matrix linear unabhängige Spalten hat und genau wie der \tilde{y} -Vektor deterministisch ist [2]. Aber das sind nicht alle Spalten der H -Matrix. Auch der \tilde{y} -Vektor in Gleichung (21) beinhaltet die verrauschten Messungen des Ausgangssignals.

Dies ist in Bild 1 anhand der Sprungantwort des PT3-Systems von Gleichung (44) veranschaulicht, die im linken oberen Subplot die unverrauschte Sprungantwort mit y_{Wahr} und als verrauschtes Messsignal y_{Mess} (modelliert als normalverteilte Zufallsgröße mit der Standardabweichung 0,03) zu sehen ist. Darunter ist ihre Differenz abgebildet. Diese Differenz wird erheblich kleiner, wenn man diese beiden Signale numerisch integriert (Subplot rechts oben) und wieder die Differenz bildet (Subplot rechts unten). Das so entstandene Signal wird durch Integration also „deterministischer“. Nun enthält die H -Matrix in Gleichung (20) aber nur integrierte Ein-/ und Ausgangs-

signale und ihre Spalten sind weniger empfindlich bezüglich linearer Abhängigkeit. Diesen Effekt kann man sogar noch verstärken, wenn man die DGL noch öfter integriert, als es die Ordnung verlangt. In Gleichung (12) würde das einer Multiplikation mit $1/s^4$ oder $1/s^5$ entsprechen. Auf diese Weise kann man sich der Erwartungstreue weiter annähern und sich gleichzeitig noch mehr mögliche Modelle generieren lassen.

Die Größe des Fehler-Vektors e lässt sich in Abhängigkeit vom x -Vektor ausdrücken,

$$e(x) = \tilde{y} - H \cdot x. \tag{22}$$

Als Maß für seine Länge dient die skalare Fehlerquadratsumme $J(x)$:

$$J(x) = \frac{1}{2} \cdot e^T(x) \cdot e(x). \tag{23}$$

Geht die Zahl $J(x)$ gegen Null, gehen auch alle Elemente des Fehler-Vektors $e(x)$ gegen Null. Es wird nun derjenige x -Vektor gesucht, der die Fehlerquadratsumme $J(x)$ so klein wie möglich macht. Die Lösung wurde als *Methode der kleinsten Quadrate* von Gauss [6] hergeleitet und kann analytisch wie folgt angegeben werden:

$$\hat{x} = (H^T \cdot H)^{-1} \cdot H^T \cdot \tilde{y}. \tag{24}$$

In Gleichung (24) ist eine Matrix-Invertierung des Produktes aus der transponierten H -Matrix mit sich selbst enthalten, die nur dann möglich ist, wenn die Spalten der H -Matrix unabhängig voneinander sind: Dass der zu erwartende Einfluss der Messfehler darauf gering ist, wurde bereits ausführlich diskutiert. Es kann noch mehr Gründe dafür geben, dass die Spalten *nicht* unabhängig voneinander sind, und somit keine Lösung existiert, zum Beispiel wenn:

- a. die H -Matrix so viele enthält Messdaten enthält, dass numerische Fehler eine Matrix-Invertierung verhindern. In dem Fall sollte man die Messdaten *ausdünnen*, bis eine

Lösung errechnet werden kann. Der so erhaltene Informationsverlust lässt sich durch Vergleich der Messdaten mit der Simulation des Ausgangssignals überprüfen.

- b. das Ein- oder das Ausgangssignal oder ihre Ableitungen stationäre Werte annehmen. Dann gehen die Werte in den entsprechenden Spalten gegen Null und somit werden sie immer ähnlicher. Dies sieht man auch direkt an der DGL (1), denn offensichtlich können nur dann Koeffizienten errechnet werden, wenn die zugehörigen Ableitungen ungleich Null sind. In dem Fall sollte natürlich ein anderer Messbereich verwendet werden.
- c. die Ordnung des zu identifizierenden Systems höher geschätzt wird, als sie ist. Dies soll an einem zu identifizierenden System erster Ordnung exemplarisch gezeigt werden, dessen Systemverhalten nach Gleichung (2) wie folgt beschrieben wird:

$$y(t) + b_1 \cdot \dot{y}(t) = a_0 \cdot u(t). \tag{25}$$

Das System werde wie in Bild 2 dargestellt durch einen Einheitssprung angeregt. Die Reaktion am Ausgang $y(t)$ wird exakt gemessen, um daraus die unbekannt Koeffizienten b_1 und a_0 bestimmen zu können. Bekanntermaßen [1] lässt sich die Antwort mathematisch berechnen zu:

$$y(t) = a_0 \cdot \left(1 - e^{-\frac{t}{b_1}}\right). \tag{26}$$

Der Einfachheit halber wird nun angenommen, dass auch die Ableitung des Ausgangssignals, $\dot{y}(t)$, genau bestimmt werden kann, sodass aus Gleichung (26) folgt

$$\dot{y}(t) = \frac{a_0}{b_1} \cdot e^{-\frac{t}{b_1}}. \tag{27}$$

In dem Fall wird das Integrieren der DGL nach Gleichung (12) überflüssig. So gelangt man zu einem ähnlichen und ebenfalls überbestimmten Gleichungssystem wie in Gleichung (20),

$$\underbrace{\begin{bmatrix} y(t_1) \\ \vdots \\ y(t_n) \end{bmatrix}}_{\hat{y}} = \underbrace{\begin{bmatrix} \dot{y}(t_1) & u(t_1) \\ \vdots & \vdots \\ \dot{y}(t_n) & u(t_n) \end{bmatrix}}_H \cdot \underbrace{\begin{bmatrix} b_1 \\ a_0 \end{bmatrix}}_x, \tag{28}$$

aus dem sich dann die unbekannt Koeffizienten b_1 und a_0 bestimmen lassen. Nimmt man irrtümlich an, die Messungen gehörten zu einem System zweiter Ordnung, so führt das auf das Ersatzproblem

$$\underbrace{\begin{bmatrix} y(t_1) \\ \vdots \\ y(t_n) \end{bmatrix}}_{\hat{y}} = \underbrace{\begin{bmatrix} \dot{y}(t_1) & \dot{y}(t_1) & u(t_1) \\ \vdots & \vdots & \vdots \\ \dot{y}(t_n) & \dot{y}(t_n) & u(t_n) \end{bmatrix}}_H \cdot \underbrace{\begin{bmatrix} b_2 \\ b_1 \\ a_0 \end{bmatrix}}_x, \tag{29}$$

mit der zweiten Ableitung

$$\dot{y}(t) = -\frac{a_0}{(b_1)^2} \cdot e^{-\frac{t}{b_1}} = -\frac{1}{b_1} \dot{y}(t). \tag{30}$$

Wegen Gleichung (30) sind nun die ersten beiden Spalten der

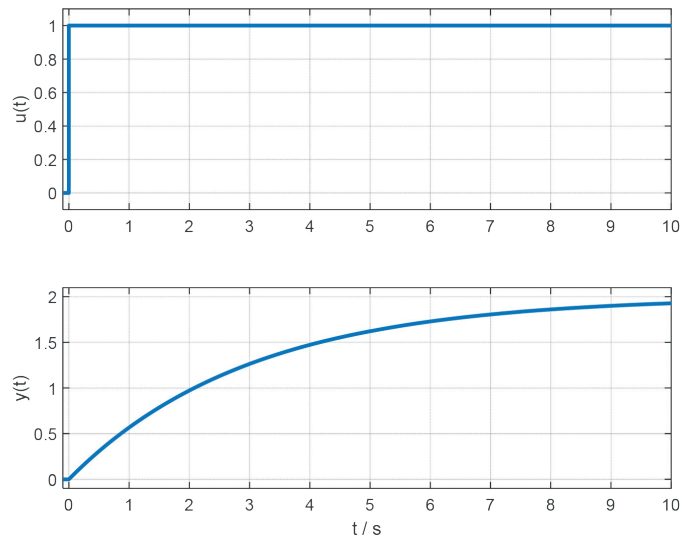


Bild 2: Eingangssprung und Ausgangssignal an einem System erster Ordnung

H -Matrix in Gleichung (29) linear abhängig und es existiert keine Lösung. Dieser Effekt taucht auch bei linearen DGLs beliebiger Ordnung auf: Wird die Ordnung zu hoch angenommen, treten immer lineare Abhängigkeiten in den Spalten auf. Tritt dies bei einem Modellierungsvorgang auf, sollte man die angenommene Ordnung reduzieren und die Berechnung wiederholen.

In der Praxis können aber nicht berücksichtigte *Nichtlinearitäten* dazu führen, dass die Spalten linear unabhängig voneinander bleiben und eine gute Lösung mit einer *linearen* DGL für den zu Grunde liegenden Ein- und Ausgangsdatensatz nur vortäuschen. Daher benötigt man immer Messdaten, die nicht für die Schätzung verwendet wurden, mit denen man dann aber zur unabhängigen Verifikation Simulation und Messung miteinander vergleichen und damit das Modellierungsergebnis bewerten kann. Dies wird noch an den Anwendungsbeispielen von Abschnitt 2 verdeutlicht.

Zusammenfassend gibt es für das **praktische Vorgehen** beim Modellierungsprozess folgende Möglichkeiten:

1. Es gibt keine Lösung \hat{x} in Gleichung (24):

Die Maßnahmen sind:

- » Die Ordnung der Übertragungsfunktion reduzieren
- » Ein anderes Ein- und Ausgangsdaten-Intervall verwenden
- » Die Anzahl der Daten ausdünnen

2. Es gibt eine Lösung \hat{x} in Gleichung (24):

Die Prüfmaßnahme besteht darin, eine Simulation mit dem identifizierten Modell durchzuführen und die Güte mit unabhängigen Messdaten zu bewerten: Ist das Ergebnis gut, dann hat man ein Modell gefunden, ist es schlecht, handelt es sich um ein nichtlineares System.

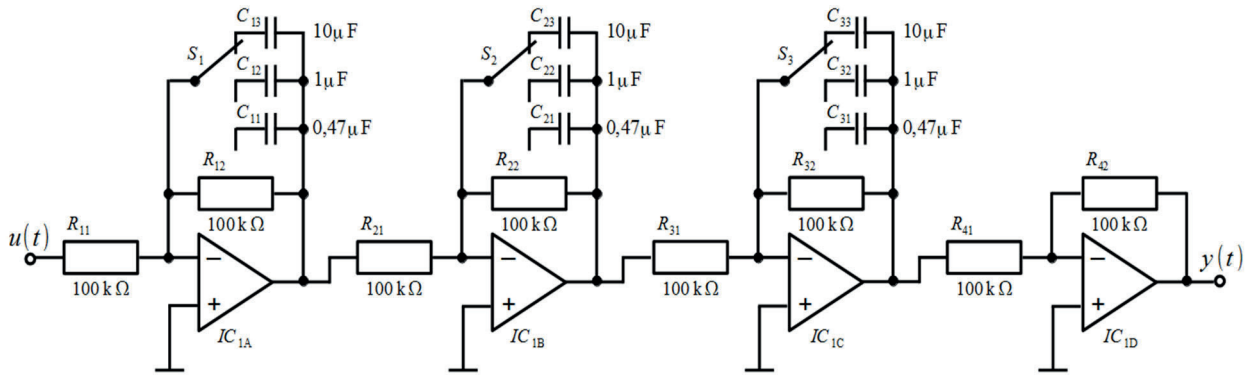


Bild 3: Schaltplan der elektronischen Regelstrecke [10]

Soll ausgewählten Zeitabschnitten bei den Messdaten besondere Bedeutung in der Berechnung zugewiesen werden – zum Beispiel einen Bereich, in dem das Eingangssignal eine besonders starke Dynamik aufweist und somit auch am Ausgang ein ausgeprägtes Einschwingverhalten – kann man einzelne Elemente des Fehler-Vektors $e(x)$ durch einen entsprechend großen Wert an der entsprechenden Stelle einer diagonalen Wichtungsmatrix W wichten, sodass nun die gewichtete Fehlerquadratsumme

$$J(x) = \frac{1}{2} \cdot e^T(x) \cdot W \cdot e(x), \quad (31)$$

minimiert wird mit dem zugehörigen Ergebnis

$$\hat{x} = (H^T \cdot W \cdot H)^{-1} \cdot H^T \cdot W \cdot \tilde{y}. \quad (32)$$

Somit sind nun die Koeffizienten der Übertragungsfunktion b_0, b_1, b_2, a_0 sowie die Zwischengrößen y_{10} und y_{20} berechenbar. Die zwei sich aus Gleichung (13) ergebenden Bestimmungsgleichungen werden in ein lineares Gleichungssystem der Form

$$\begin{bmatrix} y_{20} \\ y_{10} \end{bmatrix} = \begin{bmatrix} b_2 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \dot{y}_0 \\ y_0 \end{bmatrix} \quad (33)$$

gebracht. So können auch die restlichen Anfangsbedingungen \dot{y}_0, y_0 durch Invertierung der Matrizengleichung berechnet werden.

Erweitert man nun die Identifikation für das angegebene Beispiel dritter Ordnung in Gleichung (20) auf eine DGL beliebiger Ordnung wie in Gleichung (1), so ergibt sich statt des Gleichungssystems von Gleichung (20) und (33) das neue Gleichungssystem

$$\underbrace{\begin{bmatrix} \frac{1}{1!} \tilde{t}_1 y_0 - y_i(\tilde{t}_1) & \dots & \frac{1}{(p-1)!} \tilde{t}_1^{p-1} y_0 - y_{ii}(\tilde{t}_1) & -y_{ii}(\tilde{t}_1) & u_{ii}(\tilde{t}_1) \dots u_{ii}(\tilde{t}_1) & \frac{1}{(p-1)!} \tilde{t}_1^2 & \dots & \frac{1}{1!} \tilde{t}_1 \\ \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{1!} \tilde{t}_n y_0 - y_i(\tilde{t}_n) & \dots & \frac{1}{(p-1)!} \tilde{t}_n^{p-1} y_0 - y_{ii}(\tilde{t}_n) & -y_{ii}(\tilde{t}_n) & u_{ii}(\tilde{t}_n) \dots u_{ii}(\tilde{t}_n) & \frac{1}{(p-1)!} \tilde{t}_n^2 & \dots & \frac{1}{1!} \tilde{t}_n \end{bmatrix}}_H \cdot \begin{bmatrix} b_{p-1} \\ \vdots \\ b_0 \\ a_q \\ \vdots \\ a_0 \\ y_{(p-1)0} \\ \vdots \\ y_{10} \\ x \end{bmatrix} = \underbrace{\begin{bmatrix} y(\tilde{t}_1) - y_0 \\ \vdots \\ y(\tilde{t}_n) - y_0 \end{bmatrix}}_{\tilde{y}} \quad (34)$$

$$\begin{bmatrix} y_{(n-1)0} \\ \vdots \\ y_{20} \\ y_{10} \end{bmatrix} = \begin{bmatrix} b_2 & \dots & b_{n-1} & b_n \\ \vdots & b_{n-1} & b_n & 0 \\ b_{n-1} & 0 & 0 & 0 \\ b_n & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \dot{y}_0 \\ y_0 \\ \vdots \\ y_0 \end{bmatrix} \quad (35)$$

2. Das Verfahren in der praktischen Anwendung

2.1 Modellierung eines seriellen RC-Netzwerkes

Als zu modellierende Regelstrecke werden drei in Serie geschaltete RC-Netzwerke verwendet, deren Schaltung in Bild 3 dargestellt ist. Die Schaltung ist deshalb so variabel in der Wahl der Kondensatoren realisiert, damit in dem zugehörigen Praktikumsversuch [10] schnell unterschiedliche Regelstrecken realisiert werden können. Idealiert betrachtet handelt es sich um drei in Serie geschaltete PT_1 -Glieder. Jedes einzelne PT_1 -Glieder ist als elektronische Schaltung mit einem RC-Glied und einem OPV realisiert. Für die Kapazitäten können bei jedem RC-Glied drei mögliche Werte getrennt gewählt werden, sodass verschiedene Regelstrecken mit schneller oder langsamer Dynamik realisiert werden können.

Der Widerstand ist bei jedem RC-Glied $R = 100 \text{ k}\Omega$. Die Werte für die Kondensatoren $C_1 = 0,47 \mu\text{F}$; $C_2 = 1,0 \mu\text{F}$; $C_3 = 10 \mu\text{F}$ werden willkürlich eingestellt. Das physikalische Modell der Regelstrecke ist eine Serienschaltung von drei PT_1 -Gliedern und ergibt die Übertragungsfunktion [7]

$$G_{RC}(s) = \frac{1}{1 + T_1 \cdot s} \cdot \frac{1}{1 + T_2 \cdot s} \cdot \frac{1}{1 + T_3 \cdot s} \quad (36)$$

mit den Zeitkonstanten $T_j = RC_j$ für $j = 1, 2, 3$ und den Werten

$$T_1 = 0,047; T_2 = 0,1; T_3 = 1,0 \quad (37)$$

und ergibt dann die Übertragungsfunktion

$$G_{RC}(s) = \frac{a_0}{s^3 + b_2 s^2 + b_1 s + b_0} \quad (38)$$

mit

$$a_0 = 212,76, b_2 = 32,28, b_1 = 244,07, b_0 = 212,77. \quad (39)$$

2.1.1 Rückrechnung einer vorgegebenen Übertragungsfunktion aus simulierten Ein- und Ausgangsdaten

Zunächst wird das Verfahren selbst wie in Bild 4 dargestellt getestet. Es werden Messdaten künstlich durch Simulation erzeugt. Die Testidee ist also, die bekannte Übertragungsfunktion G_{RC} als das zu bestimmende reale System zu nehmen und die simulierte Antwort des Systems auf den Sägezahn-Eingang künstlich durch stochastische Anteile als ungenaue Messung dem Verfahren zur Verfügung zu stellen. Das heißt, das optimale Ergebnis ist vor der Schätzung bekannt und dieses wird von dem Verfahren rückgerechnet. Im Einzelnen heißt das:

1. Auf die Übertragungsfunktion in Gleichung (38) wird als Eingangssignal die in Bild 5 im obersten Subplot gezeigte Sägezahnspannung angelegt – es hätte auch ein beliebiges anderes anregendes Eingangssignal sein können – und das Ausgangsverhalten wird mit den willkürlich gewählten Anfangsbedingungen

$$y_0^T := [y(t=0) \quad \dot{y}(t=0) \quad \ddot{y}(t=0)] \\ = [-1,79 \quad -0,07 \quad -0,25] \quad (40)$$

durch Matlab-funktion `lsim` [8] simuliert. Diese Anfangsbedingungen sollen durch das Verfahren wieder errechnet werden. Auf das Ausgangssignal werden dann zum Simulieren von Messfehlern normalverteilte Zufallszahlen mit vorgegebener Standardabweichung σ gemäß Tabelle 1 hinzuaddiert. Das Ergebnis ist für das Signal y_{Mess} für $\sigma_5 = 0,05$ im mittleren Subplot in Bild 5 zu sehen.

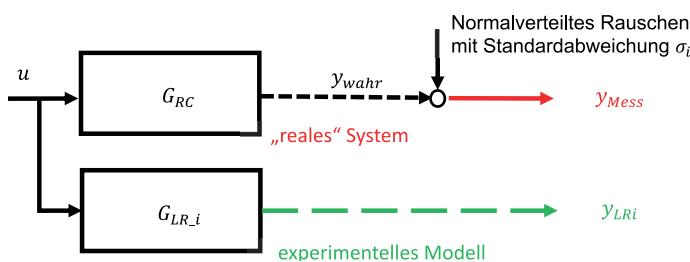


Bild 4: Signale und Systeme beim Testen durch Rückrechnung auf ein bekanntes „reales“ System

2. Es wird nun die Identifikation nur mit den blau gekreuzten Messdaten in den oberen beiden Subplots in Bild 5 durchgeführt und anhand von zwei Kriterien bewertet:

- a. Mit dem *Fit-Wert*:

$$fit := 100\% \cdot \left(1 - \frac{norm(y_{Wahr} - y_{LR5})}{norm(y_{Wahr} - mean(y_{Wahr}))} \right), \quad (41)$$

wobei unter *norm* die Länge beziehungsweise der Betrag des Vektors (bekannt auch als die sogenannte Euklidische Norm oder auch 2-Norm) verstanden und durch die gleichnamige Matlab-funktion errechnet wird. *Mean* bezeichnet den Mittelwert und die zugehörige Matlab-funktion. Der Fit-Wert wird auch in der Matlab-Identifikation-Toolbox in der function *goodnessoffit* als *normalized root mean square error* als Kriterium für die Güte einer Identifikation verwendet [8].

- b. Mit der *Rückrechnungsgenauigkeit*, dem RR-Wert:

Dieser ist ein Maß für die Abweichungen der geschätzten Koeffizienten von den wahren Koeffizienten sowie zwischen den geschätzten Anfangsbedingungen und den wahren an:

$$RR := 100\% \cdot \left(1 - \frac{norm(c - c_e)}{norm(c)} \right), \quad (42)$$

wobei in dem Vektor c die *wahren* Koeffizienten aus Gleichung (39) und die *wahren* Anfangsbedingungen aus Gleichung (40) enthalten sind:

$$c = [a_0 \quad b_2 \quad b_1 \quad b_0 \quad \dot{y}_0 \quad \ddot{y}_0]^T, \quad (43)$$

Dabei umfasst der Vektor c_e die Ergebnisse der Identifikation in Tabelle 1.

In Bild 5 sind das wahre beziehungsweise fehlerfrei gemessene Signal y_{Wahr} , das gemessene Signal y_{Mess} und das simulierte Signal y_{LR5} dargestellt. In dem vergrößerten Ausschnitt und an dem kleinen Differenzsignal im unteren Subplot ist deutlich zu erkennen, wie gut das wahre Signal y_{Wahr} durch das simulierte Signal y_{LR5} angenähert wird, obwohl nur die verrauschte Messung y_{Mess} für die Schätzung zur Verfügung stand. Die Güte der Rückrechnung wird in Bild 7 verdeutlicht. Eine exakte Rückrechnung der Koeffizienten ist demnach nur für kleine Rauschwerte σ möglich, auch wenn das Simulationsergebnis sogar für höhere Rauschwerte als sehr gut erscheint. Der Fit-Wert, der die Abweichung von Messung und Simulation beschreibt, bleibt unabhängig davon bei nahezu 100 %. Das bedeutet, dass Übertragungsfunktionen, die von der *wahren* numerisch deutlich abweichen, dennoch gute Simulationsergebnisse mit geringer Abweichung zum gemessenen Signal erzielen können. Ein guter Fit-Wert bedeutet demnach nicht zwingend, dass die wahre Übertragungsfunktion numerisch genau gefunden wurde. Der Test mit weiteren, andersartigen Ein-/Ausgangssignalen demonstriert in Bild 6 und Bild 8 durch die kleinen Abweichungen, dass die Identifikation erfolgreich war. Im Ergebnis bedeutet

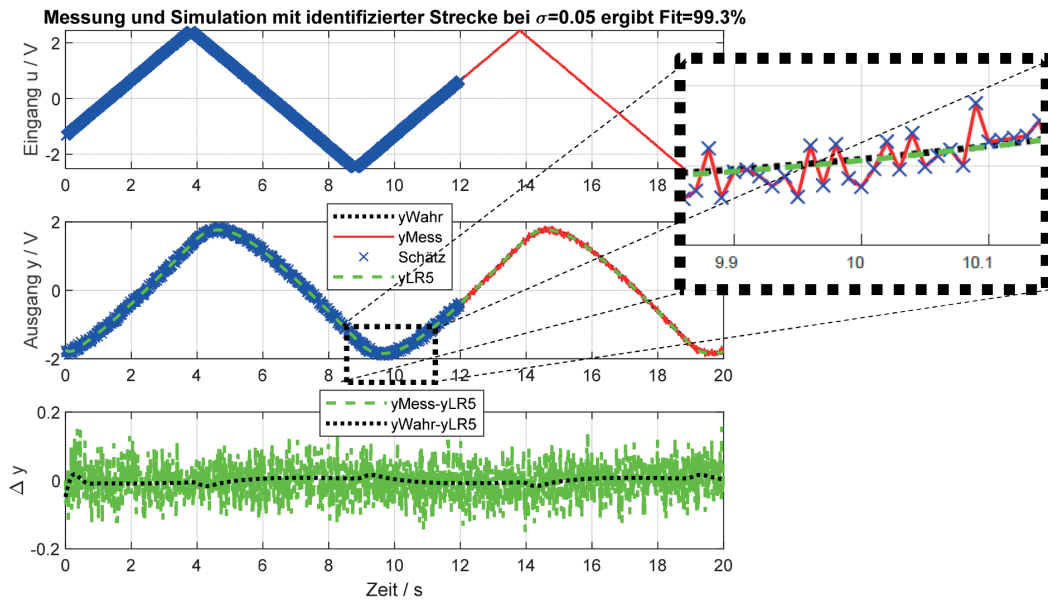


Bild 5: Messung und Simulation mit der identifizierten Übertragungsfunktion mit Modellierungsfehler

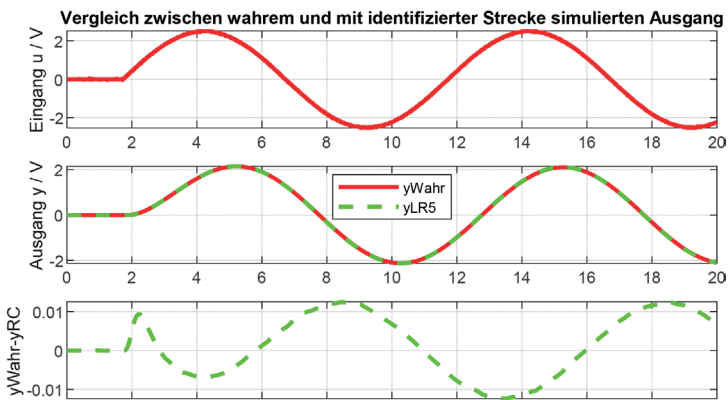


Bild 6: Simulation des mit verrauschten Ausgangssignals mit dem wahren Signal mit Sinuseingang

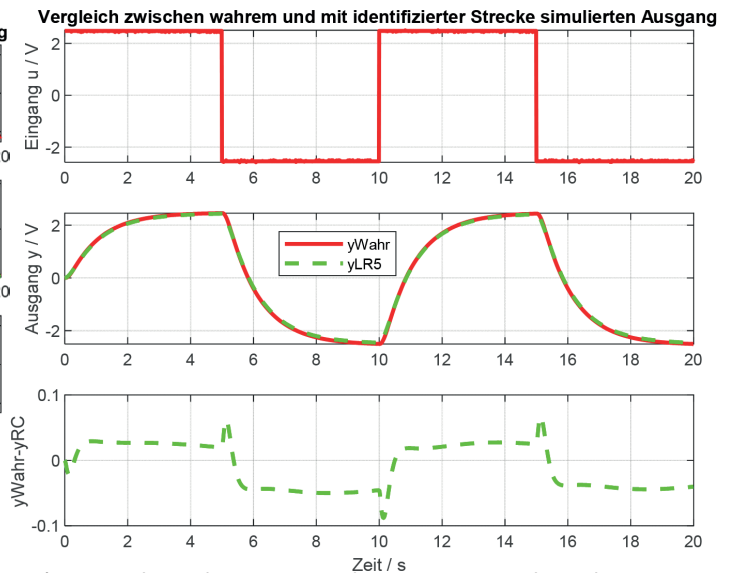


Bild 8: Simulation des mit verrauschten Ausgangssignals mit dem wahren Signal mit Sprungeingang

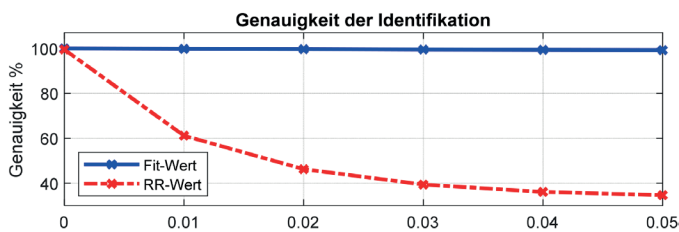


Bild 7: Ergebnisse der Identifikation für verschiedene Rausch-Level in der Messung

dies, dass die Methode funktioniert. Im Anschluss in Abschnitt 2.1.2 wird sie an einem realen System ausprobiert, bei dem natürlich die wahren Werte unbekannt sind.

2.1.2 Berechnen der Übertragungsfunktion aus gemessenen Ein- und Ausgangsdaten

Die Methode wird nun an einem realen System getestet, das wie in Bild 3 dargestellt aus drei RC-Netzwerken in Serie besteht. Die Werte für die Kapazitäten werden zum besseren Vergleich mit Abschnitt 2.1 gleich gewählt zu $C_1 = 0,47 \mu F$; $C_2 = 1,0 \mu F$; $C_3 = 10 \mu F$.

Zunächst werden mit dem in Gleichung (20) beschriebenen Verfahren aus den in Bild 10 in den oberen beiden Subplots nur mit den blauen Kreuzen gekennzeichneten Messsignalen die Übertragungsfunktion mit Zählergrad Null und Nennergrad Drei und die Anfangsbedingungen berechnet. In Gleichung (2) wurde nur der Koeffizient bei der höchsten Ableitung zu Eins gesetzt und die restlichen Koeffizienten wurden bestimmt. Wäre ein anderer Koeffizient gewählt worden, so hätte sich ein anderes überbestimmtes Gleichungssystem wie in Gleichung (20) ergeben. Darin hätte sich die H -Matrix aus anderen Spalten gebildet und der \tilde{y} -Vektor wäre ein anderer geworden.

Selbst bei fehlerfreier Messung und einem zu identifizierenden System, das sich exakt wie ein lineares System verhält, werden sich die Identifikationsergebnisse mehr oder weniger unterscheiden, wenn man verschiedene Koeffizienten normiert und somit immer andere Koeffizienten berechnet. Der Grund dafür liegt darin, dass die nicht vermeidbaren

Tabelle 1: Testergebnisse der Rekonstruktion der wahren Werte mit verschiedenen verrauschten Messsignalen

Parameter	Wahre Werte	Geschätzte Werte mit Rauschanteilen $\sigma = i \cdot 0,05$					
		$\sigma = i \cdot 0,00$	$\sigma = i \cdot 0,01$	$\sigma = i \cdot 0,02$	$\sigma = i \cdot 0,03$	$\sigma = i \cdot 0,04$	$\sigma = i \cdot 0,05$
a_0	212,80	213,65	130,16	98,44	83,85	76,88	73,87
b_2	32,28	32,42	19,90	15,15	12,98	11,95	11,51
b_1	244,00	245,06	149,16	112,70	95,89	87,78	84,20
b_0	212,80	213,66	130,26	98,59	84,03	77,08	74,10
y_0	-0,07	-0,07	0,21	0,28	0,34	-0,42	-0,51
\dot{y}_0	-0,25	-0,23	1,52	1,39	1,24	1,28	1,53

numerischen Integrationsfehler bei den Ein- und Ausgangssignalen je nach Normierung ein etwas anderes überbestimmtes Gleichungssystem für Gleichung (20) ergeben, in dem sich diese Fehler bei der Berechnung der Lösung einschließlich Matrizen-Invertierung und -Multiplikation (Gleichung (24)) unterschiedlich auswirken. Sind auch noch Modellierungsfehler dabei, können sich die Ergebnisse noch mehr in ihrer Güte unterscheiden. Auch das *Ausdünnen* der Messdaten in einem gewählten Bereich kann positive oder negative Effekte für die Modellgüte mit sich bringen: Natürlich gehen beim Weglassen von Messdaten immer Informationen verloren, auf der anderen Seite verkleinern sich aber auch die numerischen Fehler bei der Matrizeninvertierung in Gleichung (24). Diese Effekte lassen sich im konkreten Fall nur durch Ausprobieren gegeneinander abwägen.

Für die Wahl von Zähler- und Nennergrad gibt es zwei Möglichkeiten: Zum einen können physikalische Überlegungen eine Mindestordnung für Zähler und Nenner ergeben. Zum anderen analysiert man das Differenzsignal zwischen Messung und Simulation, zum Beispiel das im unteren Subplot von Bild 10: Je unkorrelierter dieses Signal ist, desto besser ist die Identifikation, denn die Rauschanteile durch die Messung sollen ja herausgefiltert und nicht mitmodelliert werden. Die Ordnungen werden dann solange vom Anwender erhöht, bis sich das Differenzsignal nicht mehr signifikant verbessert. Erhöht man die Ordnung immer weiter, so werden bei einem linearen System die Spalten der H -Matrix in Gleichung (20) – wie oberhalb schon diskutiert – und es existiert keine Lösung mehr. Entstammen die Messdaten aus einem nichtlinearen System, so können die Spalten zwar linear unabhängig bleiben, aber die anschließende Simulation offenbart dann, dass es sich um kein brauchbares Simulationsmodell handelt.

Da Stabilität, Instabilität oder Minimalphasigkeit nicht in die Minimierung des Gütekriteriums eingehen, können auch Lösungen generiert werden, bei denen diese Eigenschaften verloren gehen. Solche Lösungen werden dann ausgefiltert. Integratoren oder Differenzglieder bedeuten in der Übertragungsfunktion, dass ein Koeffizient oder mehrere den Wert Null annehmen. Dies lässt sich hingegen leicht dadurch berücksichtigen, dass in der H -Matrix in Gleichung (20) die entsprechenden Spalten gelöscht werden und die Unbekannte im x -Vektor entfernt wird.

Weil die Berechnungsdauer für die Erstellung der Lösung (Gleichung (24)) kurz ist, können sehr schnell viele Modelle aus dem gleichen Datensatz generiert werden. Die Güte der Modelle kann anhand des Fit-Wertes bewertet werden und

das beste Ergebnis wird dann als das Endergebnis ausgewählt. Diese Tatsache kann auch bei der Berechnung von Totzeiten im System verwendet werden: Über einen Algorithmus wird das Ausgangssignal stufenweise Zeitschritt für Zeitschritt zurück in die Vergangenheit geschoben. Am Ende dieses Prozesses ist die gesuchte Lösung diejenige mit dem kleinsten Gütekriterium. Genau dieser Zeitversatz entspricht dann der gesuchten Totzeit.

Die oben beschriebenen Einflussmöglichkeiten in der Gestaltung der Lösung sind in dem kostenlos herunterladbaren Tool pzMove unter Registerkarte *Identification* [5] programmiert und mit einer Bedienungsanleitung hinterlegt; in Bild 11 ist die Schaltfläche dargestellt. Hier lassen sich die Einflüsse auf

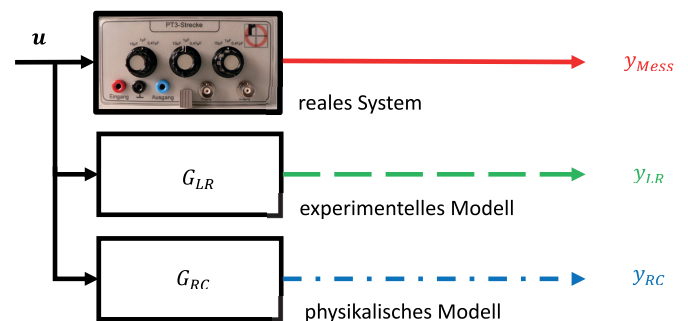


Bild 9: Bezeichnung der Signale und Systeme beim Test mit realen Messsignalen

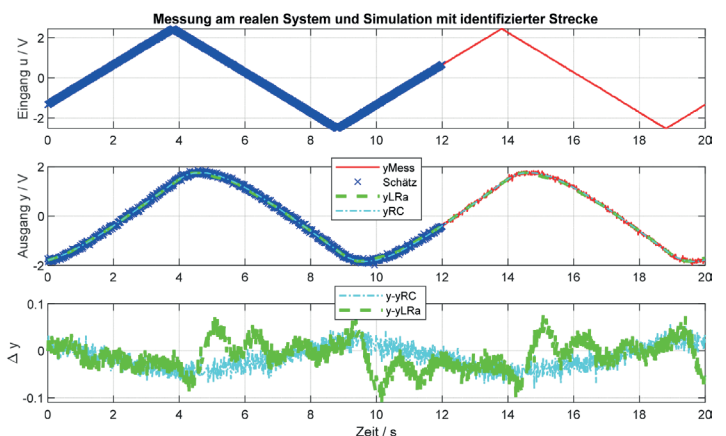


Bild 10: Ergebnis der Simulation mit dem physikalischen (Fit = 97,5%) und dem aus den Messdaten gewonnenen experimentellen Modell (Fit = 97,3%)

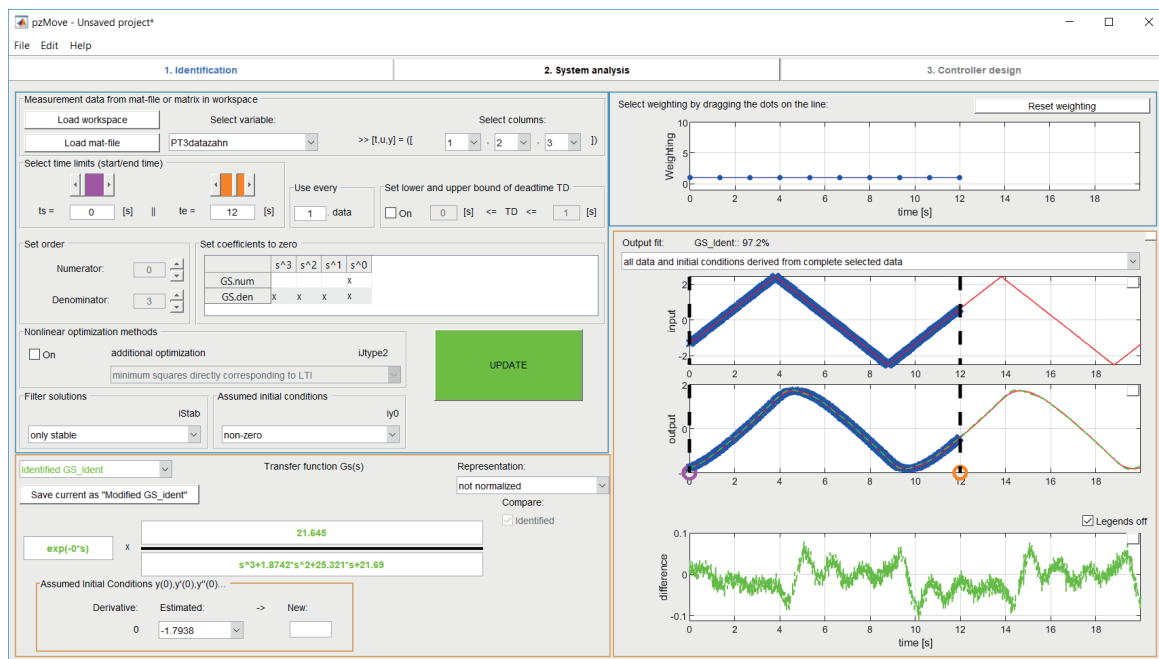


Bild 11: Schaltfläche des Identifikation-Tools zur experimentellen Modellbildung nach dem hier vorgestellten Verfahren

das Ergebnis schnell verändern und ausprobieren:

1. Vorgeben des Zähler- und des Nennergrades der Übertragungsfunktion
2. Vorgabe System ist in Ruhelage oder System ist nicht in Ruhelage
3. Auswahl des Messbereiches, der für die Schätzung verwendet wird
4. Die Wichtungen ausgedrückt durch die Matrix W der Messungen nach Gleichung (32)
5. Nullsetzen von einzelnen Koeffizienten.

6. Sortieren der mit allen Koeffizienten erzielten Ergebnisse nach stabilen oder instabilen Lösungen
7. Berücksichtigung von Totzeiten

Das so erzielte Ergebnis ist die Übertragungsfunktion

$$G_{LRa}(s) = \frac{22,22}{s^3 + 1,894 s^2 + 25,98 s + 22,25} \quad (44)$$

Der Vergleich von Simulation und Messung in Bild 10 zeigt, dass die Lösung im Vergleich zum RC-Modell eine höhere Differenz zur Messung und den etwas geringeren Fit-Wert von 97,3 % statt 97,5 % hat. Dies ändert sich aber deutlich, wenn nun mit einem Zählergrad Eins statt Null experimentiert wird:

$$G_{LRb}(s) = \frac{6,26 s + 0,3471}{s^3 + 7,331 s^2 + 6,764 s + 0,3337} \quad (45)$$

Mit dieser Übertragungsfunktion wird ein höherer Fit-Wert von 98,8 % erreicht. Das Differenzsignal im unteren Subplot in Bild 12 sieht nahezu wie unkorreliertes weißes Rauschen aus. Das ist das bestmögliche Ergebnis.

Dies macht das konstruktive Zusammenspiel von physikalischer und experimenteller Modellierung besonders deutlich: Weil das Hinzufügen einer weiteren Nullstelle das Modellierungsergebnis noch einmal verbessert und weitere Erhöhungen der Modellordnungen keine Lösung mehr zur Folge haben, ist das ein Hinweis darauf, dass sich dahinter ein bislang nicht modellierter physikalischer Effekt verbirgt. Die Vorstellung vom idealen RC-Netzwerk stößt also an ihre Grenzen und lädt zu einer höchstwahrscheinlich zeitraubenden Iterationsschleife bei der physikalischen Modellierung ein, sollte die Genauigkeit des physikalischen Modells nicht ausreichen.

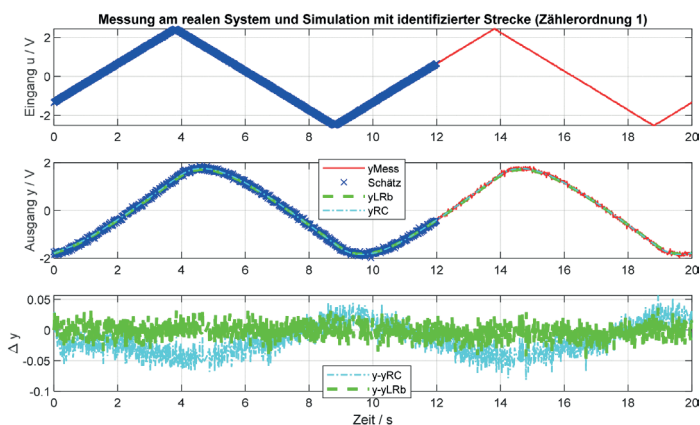


Bild 12: Ergebnis der Simulation mit dem physikalischen (Fit 97,5 %) und dem aus den Messdaten gewonnenen experimentellen Modell mit Zählerordnung Eins und Nennerordnung drei mit erhöhtem Fit =98,8 %

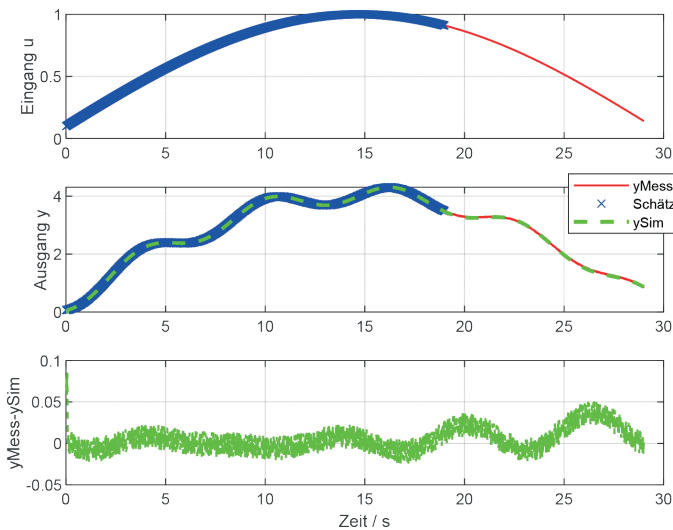


Bild 13: Identifikationsergebnis eines schwach gedämpften Systems mit Totzeit

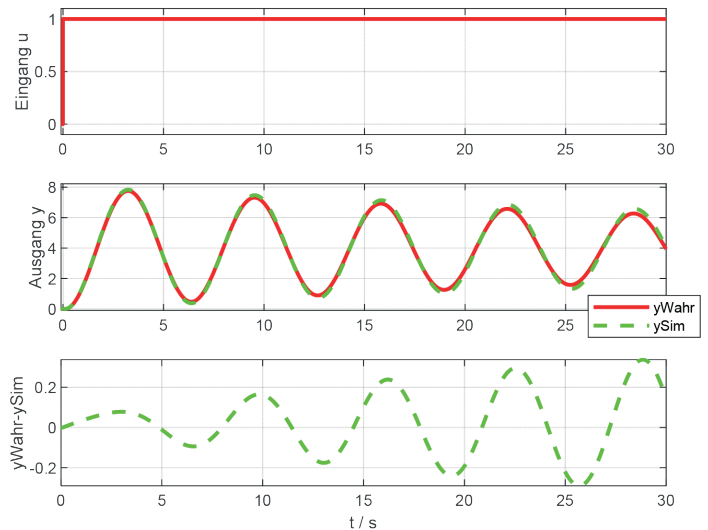


Bild 14: Verifikation des identifizierten Modells anhand der Sprungantwort

Allerdings sieht man es der Übertragungsfunktion nicht an, auf welche Art und mit wieviel Aufwand sie gebildet wurde, so dass man es auch einfach dabei belassen und mit der experimentell gewonnenen Übertragungsfunktion weiterarbeiten kann. Der Name *experimentelle Übertragungsfunktion* ist also Programm: Man darf an den Messdaten experimentieren, um das Identifikationsergebnis zu verbessern.

2.2 Weitere Modellierungsbeispiele

Um das weite Anwendungsspektrum der Methode zu demonstrieren werden zwei weitere Modellierungen von Systemen mit ausgeprägten Eigenschaften durchgeführt.

2.2.1 Modellierung eines schwachgedämpften Systems mit Totzeit

Ähnlich wie in Abschnitt 2.1.1 wird eine Regelstrecke $G_{wahr}(s)$ vorgegeben. Diese wird mit einem zur Eigenfrequenz des Systems bewusst sehr langsam gewählten sinusförmigen Eingangssignal $u(t)$ (oberer Subplot in Bild 13) angeregt, um die Identifikation zu erschweren. Das Simulationsergebnis ist mit künstlichem Messrauschen (normalverteilt mit Standardabweichung 0,03) versehen, welches das Messsignal Ausgang $y_{Mess}(t)$ im mittleren Subplot in Bild 13 darstellt. Es wird eine schwach gedämpfte Regelstrecke zweiter Ordnung mit Dämpfungsfaktor $d = 0,02$ und der Zeitkonstanten $T = 1s$ einschließlich Totzeit $T_t = 0,10s$ und den Anfangsbedingungen ungleich Null identifiziert:

$$G_{wahr}(s) = e^{-0,10 \cdot s} \frac{4}{1 + 2 \cdot 0,02 \cdot 1,0 \cdot s + 1^2 \cdot s^2} \quad (46)$$

Die Identifikation der Regelstrecke nach dem hier beschriebenen Verfahren ergibt eine Übertragungsfunktion $G_{mod}(s)$, die in guter Übereinstimmung mit den vorgegebenen Werten von $G_{wahr}(s)$ ist:

$$G_{mod}(s) = e^{-0,09 \cdot s} \frac{4,009}{1 + 2 \cdot 0,016 \cdot 1,002 \cdot s + 1,002^2 \cdot s^2} \quad (47)$$

Zur Verifikation werden zusätzlich die Sprungantworten aus der Ruhelage heraus miteinander in Bild 14 verglichen. Das gute Ergebnis der Identifikation wird noch einmal bestätigt.

2.2.2 Modellierung eines instabilen Systems

Als Beispiel für die Identifikation eines *realen* instabilen Systems wird ein Magnetschweberversuch [9] vorgestellt, der in den Ingenieurstudiengängen der TH Rosenheim zum Praktikum der Grundlagenvorlesung Regelungstechnik gehört. Dabei geht es darum, einen magnetischen Schwebekörper so zu regeln, dass er konstant in der Luft schwebt, was bei den Studenten nach dreistündiger Bearbeitungszeit immer wieder sichtliche Faszination auslöst (Bild 15).

Ein wichtiger Teil der Regelungsaufgabe ist die Modellbildung, die mit dem hier vorgestellten Verfahren durchgeführt wird. Dabei wird der Spulenstrom, der das externe Magnetfeld beeinflusst und die Eingangsgröße darstellt, sprungartig

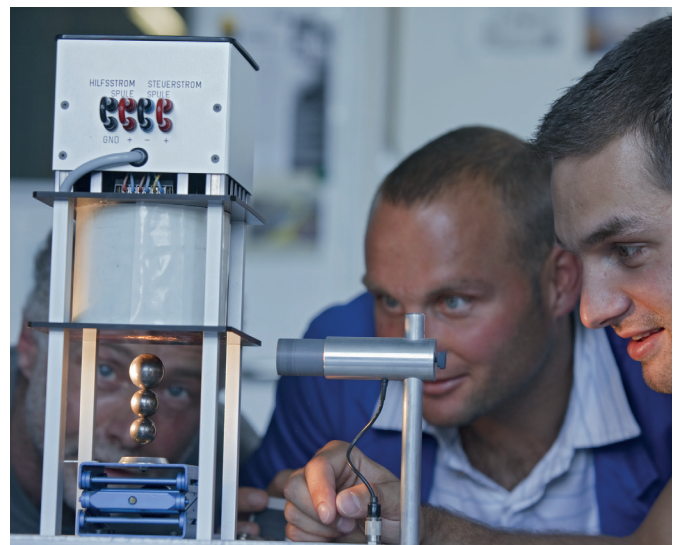


Bild 15: Studenten nach erfolgreicher Stabilisierung eines Schwebekörpers (©TH Rosenheim)

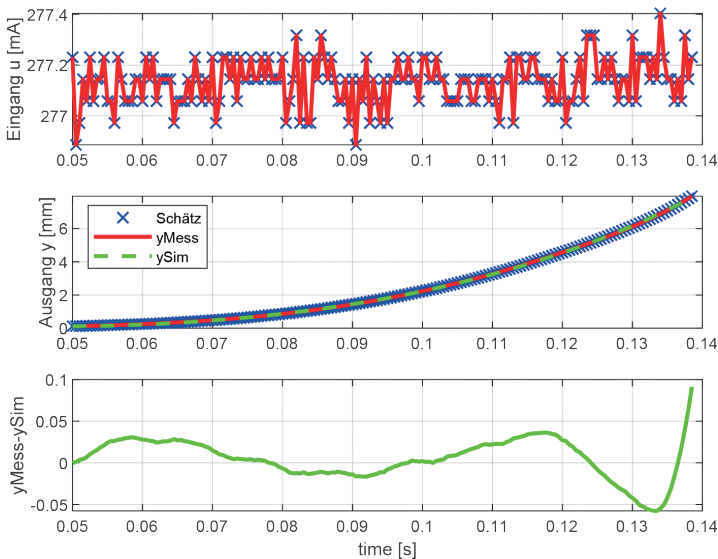


Bild 16: Gemessene und simulierte Sprungantwort des Magnetschwebekörpers

erhöht, und die mit einer Fotodiode gemessene Lichtmenge, die ein gutes Höhensignal für den Schwebekörper liefert, wie in Bild 16 dargestellt gemessen. Ebenfalls in Bild 16 ist das Simulationsergebnis gezeigt, das mit Hilfe des Modells der Regelstrecke,

$$G_{mod}(s) = \frac{4,946}{s^2 - 431,3'} \quad (48)$$

erzeugt wurde. Man erkennt in Bild 16 an dem Knick zwischen Messung und Simulation zum Zeitpunkt 0,13 s, dass hier zwischen Modell und Messung eine große Abweichung entsteht. Weil aus der Physik bekannt ist, dass Magnetfeldkräfte die Position eines magnetischen Körpers nichtlinear zum Abstand beeinflussen, wird wahrscheinlich diese Nichtlinearität den Knick verantworten. Für die Regelung hat die geringere Modellierungsgenauigkeit ab 6 mm keine Bedeutung, weil diese Höhe außerhalb des Arbeitsbereiches der Regelung liegt.

Das geregelte System wird mit einem Sollwert beaufschlagt, dessen Wert sich über einen Schieberegler verändern lässt. Dieser Wert wird dann dem realen System und dem simulierten System zugeführt. Als Ergebnis sind in Bild 17 der Sollwert, die gemessene und die simulierte Position des Magnetschwebekörpers dargestellt. Die Übereinstimmung zwischen der gemessenen und der simulierten Position ist sehr gut.

3. Zusammenfassung und Ausblick

Es wurde ein Verfahren zur experimentellen Modellbildung für Systeme vorgestellt, die sich als lineare DGLs mit konstanten Koeffizienten beschreiben lassen. Das Verfahren hat ein großes Einsatzspektrum, weil es keine Einschränkung des Streckentyps benötigt. Es ist anwendbar für den offenen und

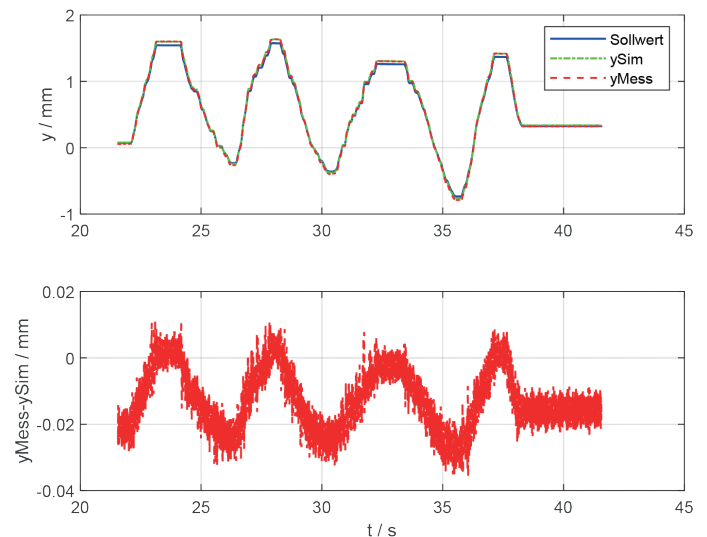


Bild 17: Übereinstimmung bei der gemessenen und simulierten Position eines Magnetschwebekörpers

den geschlossenen Regelkreis und für Vorgänge, die nicht in der Ruhelage beginnen – die Anfangswerte können nämlich mitgeschätzt werden. Wie gezeigt wurde, können stabile oder instabile Systeme, gedämpfte, schwach gedämpfte oder Systeme mit und ohne Totzeit identifiziert werden.

Die Richtigkeit der Methode wurde an zwei *simulierten* Beispielen demonstriert, bei denen das optimale Schätzergebnis bekannt war und die geringe Abweichung davon festgestellt wurde. Anschließend wurde es auch an zwei *realen* Systemen mit echten Messungen erfolgreich angewendet. Das Verfahren ist in der Theorie völlig transparent und baut allein auf dem Vorwissen von Bachelorstudenten der Ingenieurwissenschaften auf. Dies sollte im Bachelorstudium die Vernetzung zwischen der Mathematik und den ingenieurwissenschaftlichen Fächern weiter verstärken.

Das Verfahren einschließlich des in Abschnitt 2.1.2 vorgestellten Beispiels ist für jedermann zugänglich [5] und als Matlab-unabhängiges *.exe-Programm herunterladbar. Das vorgestellte Verfahren wenden Studenten der Ingenieurwissenschaften im Rahmen von Praktikumsversuchen [9, 10] an. Dozenten können auf Nachfrage die Praktikumsanleitung und die Musterlösung erhalten.

Es wird bei dem Verfahren an einer Erweiterung gearbeitet, um garantiert das bestmögliche Identifikationsergebnis im Sinne der kleinsten Fehlerquadratsumme bei vorgegebenen Streckeneigenschaften wie Stabilität, Instabilität oder Minimalphasigkeit zu erhalten. Auch die Erweiterung für Systeme mit mehreren Ein- und Ausgängen ist von Interesse.

Eine weiteres interessantes Projekt ist eine komfortable Variante des hier vorgestellten *manuellen* Verfahrens, in dem eine übergeordnete Optimierung über alle beschriebenen Einflussparameter den Grad der Autokorrelation [1] des Differenzsignals zwischen Messung und Simulation minimiert.

Referenzen

- [1] Lutz, J., Wendt, W. (2014). *Taschenbuch der Regelungstechnik*. Harri Deutsch
- [2] Bohn, C., Unbehauen, H. (2016). *Identifikation dynamischer Systeme-Methoden zur experimentellen Modellbildung aus Messdaten*. Springer, 2016
- [3] Kahlert, J. (2004). *Simulation technischer Systeme*. Vieweg Praxiswissen, 2004.
- [4] Golubev, B., Horowitz, I. (1982). Plant rational transfer approximation from input-output data. *International Journal of Control*, 36(4), (pp. 711-723).
- [5] Technische Hochschule Rosenheim. (2019). *Lerntool für die Regelungstechnik - pzMove*. Abgerufen von: www.th-rosenheim.de/rtpzMove.html
- [6] Crassidi, J., Junkins, J. (2004). *Optimal Estimation of Dynamic Systems*, (p. 551). Chapman and Hall/CRC, 2004
- [7] Tieste, K. D., Romberg, O. (2015). *Keine Panik vor Regelungstechnik!*. Springer Fachmedien Wiesbaden.
- [8] The MathWorks, Inc. (2019). *Goodness of fit between test and reference data*. Abgerufen von: <https://de.mathworks.com/help/ident/ref/goodnessoffit.html>
- [9] Zentgraf, P., Huber, M., Zeitler, A., Bernhardt, A. (2018). *Praktikum Regelungstechnik Magnetischer Schwebekörper, Praktikumsanleitung*. Technische Hochschule Rosenheim, 2018.
- [10] Zentgraf, P., Huber, M., Zeitler, A., Bernhardt, A. (2018). *Praktikum Regelungstechnik Streckenidentifikation, Praktikumsanleitung*. Technische Hochschule Rosenheim, 2018.

AUTOR

Prof. Dr. Peter Zentgraf, M.Sc., (geb. 1965) arbeitet seit 2010 als Professor für Simulation und Regelungstechnik an der Technischen Hochschule Rosenheim. Nach seiner Promotion 1994 am Zentrum für Luft und Raumfahrt (DLR) war er bei Airbus in der Satelliten-Lageregelung und anschließend bei der Europäischen Raumfahrtagentur (ESA) tätig. Seine aktuellen Forschungsschwerpunkte sind die Entwicklung von Verfahren zur robusten Stabilisierung und Verfahren zur Störgrößenunterdrückung, die mit dem Grundwissen eines Bachelor-Ingenieurs verstanden werden können.



Peter Zentgraf

Technische Hochschule Rosenheim
Hochschulstraße 1
83024 Rosenheim
☎ +49 (0) 8031 805 26 60
@ peter.zentgraf@th-rosenheim.de