

Generieren einer Streckenübertragungsfunktion aus gemessenen Ein-/Ausgangsdaten

Von Peter Zentgraf¹

1 Einleitung und Motivation

Zur Lösung einer regelungstechnischen Entwurfsaufgabe benötigt man ein Simulationsmodell der Regelstrecke. Entweder benötigt man es, weil es die meisten Entwurfsverfahren zur Bestimmung der Reglerparameter als gegeben voraussetzen 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 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 „verstehbar“ für sie 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- bzw. Ausgangssignalen umgeformt. Diese Vorgehensweise kann sehr zeitaufwendig sein und hinzukommt, dass Parameter des Modells meist unbekannt sind. Bei der *experimentellen* Modellbildung [1], [2] geht man von einem gemessenen Ein- bzw. Ausgangssignal aus und gleicht über mehr oder weniger aufwändige Optimierungsverfahren die simulierte Antwort an die gemessene Antwort an.

Den bislang bekannten experimentellen Verfahren [1], [2] ist gemein, dass sie entweder

- nur für *bestimmte Streckentypen* gelten,
- nur im *offenen* Kreis mit speziellen meist sprungförmigen Testsignalen anwendbar sind,
- nur für *diskrete* Systeme geeignet sind,
- von einer Regelstrecke in der *Ruhelage* ausgehen,
- oder die Anwendung komplizierter *mathematischer Optimierungsverfahren* erfordern.

Das hier vorgestellte Verfahren ähnelt dem in [3] zitierten Verfahren von [4], allerdings ist die Herleitung hier mit Hilfe der Laplace-Transformation deutlich transparenter. Das System muss sich nicht in der Ruhelage befinden. Zusätzlich können in dem hier vorgestellten Verfahren die Anfangsbedingungen mitgeschätzt werden, und außerdem werden mehr Freiheitsgrade in der Verbesserung der Modellgüte verwendet.

¹ Prof. Dr. Peter Zentgraf, M.Sc., Technische Hochschule Rosenheim, Hochschulstr. 1, 83024 Rosenheim

2 Beschreibung des Identifikationsverfahrens

Die Aufgabenstellung ist die Folgende: Aus einem einzigen Datensatz mit Eingangssignal $u(t)$ und Ausgangssignal $y(t)$ soll das komplette, allgemeine Systemverhalten des Modells bestimmt werden, so dass auch für vollkommen andere Eingangssignale das zugehörige Ausgangsverhalten vorhersagt werden kann, ohne erneut messen zu müssen.

2.1 Beschreibung der Regelstrecke

Das dynamische Verhalten eines Systems mit Eingangsgröße $u(t)$ und Ausgangsgröße $y(t)$ lässt sich in Form einer Differentialgleichung folgendermaßen beschreiben:

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

Willkürlich kann man die gesamte Gleichung Gl. (1) durch einen der Koeffizienten (welcher nicht null sein darf) teilen. Die DGL wird also auf einen Koeffizienten normiert. Auf eine entsprechende ausführlichere Nomenklatur wie beispielsweise $\widetilde{b}_1 = \frac{b_1}{b_3}$ wird hier verzichtet, da sich die DGL nicht verändert hat. Die Folge ist, dass sich die Anzahl der unbekanntenen Koeffizienten um einen 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 eben auf den Wert auf -1 bzw. -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 Gl. (3) beschriebenes System, dann hätte dieses Vorgehen keinen Einfluss auf die zu bestimmende Differentialgleichung. Weil aber das eben nicht der Fall ist ändern sich die Ergebnisse eben doch und man erhält andere Modelle. Das ist aber keine schlechte sondern eine gute Nachricht: Hier ist ein Freiheitsgrad versteckt, mit dem man die Modellgüte verbessern kann. In den folgenden mathematischen Beschreibungen wurde willkürlich $b_3 = 1$ gesetzt.

Das Verfahren wird anhand einer DGL dritter Ordnung (in Gl. (1) mit $m=0, n=3$) vorgestellt:

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

Diese Gleichung lässt sich für die Korrespondenz $f(t) \leftrightarrow F(s)$ 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 Ausgangssignal in den Laplace-Bereich transformieren. Für verschwindende Anfangsbedingungen, d.h. wenn $y(t=0) = y_0$ und die Anfangsbedingungen der Ableitungen \dot{y}_0 und \ddot{y}_0 zu Null gesetzt werden können, ergibt sich die Laplace-Transformierte von Gl. (2):

$$b_0 \cdot Y(s) + b_1 \cdot s \cdot Y(s) + b_2 \cdot s^2 \cdot Y(s) + s^3 \cdot Y(s) = a_0 \cdot U(s) \quad (4)$$

Nach Umstellen dieser Gleichung kommt man auf die bekannte Form der Übertragungsfunktion des PT_3 -Systems:

$$G_S(s) = \frac{Y(s)}{U(s)} = \frac{a_0}{s^3 + b_2 \cdot s^2 + b_1 \cdot s + b_0} \quad (5)$$

In dieser Darstellung wird noch einmal deutlich, dass sich die Übertragungsfunktion nicht verändert, wenn man sowohl Zähler als auch Nenner durch denselben Wert teilt.

In der Praxis hat man es mit kontinuierlich betriebenen Prozessen zu tun, die man nur in Ausnahmefällen aus einer Ruhelage heraus starten kann. Man kann nicht davon ausgehen, 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 Gl. (2), (3) erhält man mit Gl. für nicht verschwindende Anfangsbedingungen:

$$b_0 \cdot y(t) \rightsquigarrow b_0 \cdot Y(s) \quad (6)$$

$$b_1 \cdot \dot{y}(t) \rightsquigarrow b_1 \cdot (s \cdot Y(s) - y_0) \quad (7)$$

$$b_2 \cdot \ddot{y}(t) \rightsquigarrow b_2 \cdot (s^2 \cdot Y(s) - s \cdot y_0 - \dot{y}_0) \quad (8)$$

$$1 \cdot \ddot{y}(t) \rightsquigarrow s^3 \cdot Y(s) - s^2 \cdot y_0 - s \cdot \dot{y}_0 - \ddot{y}_0 \quad (9)$$

$$a_0 \cdot u(t) \rightsquigarrow a_0 \cdot U(s) \quad (10)$$

Hinweis: 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.

Als Wert der Anfangsbedingung y_0 kann man den Messwert an dem Zeitpunkt nehmen, den man als den "Anfang" festgelegt hat. Dies geht aber für die Anfangsableitungen \dot{y}_0, \ddot{y}_0 nicht so einfach, diese Werte sollte man genauso wie die unbekanntenen Koeffizienten b_0, b_1, b_2, a_0 mitschätzen: Denn die numerische Differentiation von verrauschten Signalen ist fehleranfällig; d.h., selbst ein im Prinzip konstantes Signal wie beispielsweise $u(t) = 5,00V = const.$ aus einer elektrischen Gleichspannungsquelle hat nie einen absolut glatten Verlauf, sondern ist immer von zufälligen Anteilen (Rauschen) überlagert. Man erhält dann beliebig steile Flanken zwischen den einzelnen Messpunkten. Die Steigung $\frac{\Delta y}{\Delta t}$ und damit die Ableitung am jeweiligen Punkt wird so bei sehr kleinem Δt beliebig groß. Beim Ableiten vergrößert sich der Einfluss des Rauschens demnach. Beim Integrieren hingegen verkleinert sich der Einfluss. Die Rauschanteile werden dann "herausgemittelt".

Laplace-Transformation von Gl. (2) mit Gln. (6) - (10) ergibt

$$b_0 \cdot Y(s) + b_1 \cdot (s \cdot Y(s) - y_0) + b_2 \cdot (s^2 \cdot Y(s) - s \cdot y_0 - \dot{y}_0) + s^3 \cdot Y(s) - s^2 \cdot y_0 - s \cdot \dot{y}_0 - \ddot{y}_0 = a_0 \cdot U(s) \quad (11)$$

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

$$Y(s)(b_0 + b_1 \cdot s + b_2 \cdot s^2 + s^3) = a_0 \cdot U(s) + b_1 \cdot y_0 + b_2 \dot{y}_0 + \ddot{y}_0 + s \cdot (b_2 \cdot y_0 + \dot{y}_0) + s^2 \cdot y_0 \quad (12)$$

$$Y(s) = \underbrace{\frac{a_0}{(b_0 + b_1 \cdot s + b_2 \cdot s^2 + s^3)}}_{\text{Teil 1}} \cdot U(s) + \underbrace{\frac{b_1 \cdot y_0 + b_2 \dot{y}_0 + \ddot{y}_0 + s \cdot (b_2 \cdot y_0 + \dot{y}_0) + s^2 \cdot y_0}{(b_0 + b_1 \cdot s + b_2 \cdot s^2 + s^3)}}_{\text{Teil 2}} \quad (13)$$

Man erkennt, dass sich das Ausgangssignal $Y(s)$ in Gl.(13) aus zwei Teilen zusammensetzt:

1. In Teil 1 wird das Eingangssignal $U(s)$ „ganz klassisch“ 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 $G(s)$, und die Nullstellen werden gebildet aus den Anfangsbedingungen $y_0, \dot{y}_0, \ddot{y}_0$. Sind diese null, dann ist auch der ganze Ausdruck von Teil 2 null. Mit Gl. (13) lässt sich also das System mit Anfangsbedingungen ungleich null simulieren.

Im Laplace-Bereich entspricht einer Integration die Multiplikation mit $\frac{1}{s}$. Aus diesem Grunde wird im nächsten Schritt Gl. (11) mit $\frac{1}{s^3}$ multipliziert. Dadurch werden alle Multiplikationen mit "s", welche im Zeitbereich Ableitungen entsprechen, vermieden; eine Multiplikation mit $\frac{1}{s^4}$ oder höherer Ordnung ist denkbar: Dieses „Überintegrieren“ stellt somit einen weiteren Freiheitsgrad der Modellbildung der hier vorgestellten Methode dar.

Somit ist eine „DGL mit Ableitungen“ durch mehrfaches Integrieren in eine „DGL mit Integralen“ umgewandelt worden. Dieses technisch einfacher zu lösende Ersatzproblem lautet damit sortiert nach den Unbekannten:

$$\left(Y(s) - \frac{1}{s} \cdot y_0 - \frac{1}{s^2} \cdot \dot{y}_0 - \frac{1}{s^3} \cdot \ddot{y}_0 \right) + b_2 \cdot \left(\frac{1}{s} \cdot Y(s) - \frac{1}{s^2} \cdot y_0 \right) + b_2 \cdot \dot{y}_0 \cdot \left(-\frac{1}{s^3} \right) + \dots \quad (14)$$

$$+ b_1 \cdot \left(\frac{1}{s^2} \cdot Y(s) - \frac{1}{s^3} \cdot y_0 \right) + b_0 \cdot \frac{1}{s^3} \cdot Y(s) = a_0 \cdot \frac{1}{s^3} \cdot U(s)$$

In Gl. (14) taucht das Produkt zweier Unbekannter auf, nämlich $b_2 \cdot \dot{y}_0$, welches natürlich ebenfalls unbekannt ist. In dieser Form lässt sich aber 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, welche dann doch eine Anwendung der Methode erlauben:

$$Y(s) - y_0 \cdot \frac{1}{s} + b_2 \cdot \left(\frac{1}{s} \cdot Y(s) - \frac{1}{s^2} \cdot y_0 \right) + b_1 \cdot \left(\frac{1}{s^2} \cdot Y(s) - \frac{1}{s^3} \cdot y_0 \right) + \dots \quad (15)$$

$$+ b_0 \cdot \frac{1}{s^3} \cdot Y(s) - \underbrace{\left(b_2 \cdot \dot{y}_0 + \ddot{y}_0 \right)}_{=:y_{20}} \cdot \frac{1}{s^3} - \underbrace{\dot{y}_0}_{=:y_{10}} \cdot \frac{1}{s^2} = a_0 \cdot U(s) \cdot \frac{1}{s^3}$$

Anschließend kann nach Berechnung von b_2 von den errechneten Werten der Zwischengrößen auf \dot{y}_0, \ddot{y}_0 zurückgerechnet werden. Gl. (15) lässt sich mit dem Laplace'schen Integrationssatz

$$\frac{1}{s} \cdot F(s) \Leftrightarrow \int_{t_0}^t f(\tau) d\tau := f_i(t) \quad (16)$$

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

Zur Erinnerung werden noch folgende Laplace-Korrespondenzen für $t \geq 0$ angegeben:

$$y_0 \cdot \frac{1}{s} \Leftrightarrow y_0 \quad (17)$$

$$y_0 \cdot \frac{1}{s^2} \Leftrightarrow y_0 \cdot t \quad (18)$$

$$y_0 \cdot \frac{1}{s^3} \Leftrightarrow y_0 \cdot \frac{t^2}{2} \quad (19)$$

Den Zeitpunkt t_1 , ab dem die Messwerte für die Schätzung verwendet werden, legt man danach fest, wie geeignet der Messbereich für die Identifikation ist. Der Signalbereich für die Schätzung und auch für die spätere Systemsimulation soll zum Zeitpunkt „0 s“ beginnen, denn dann ist der Beitrag der unteren Integrationsgrenze bei der Auflösung der bestimmten Integrale null, und das reduziert die Anzahl der Terme. Aus diesem Grunde wird die Variable \tilde{t} eingeführt, welche diese Bedingung erfüllt, also an der Stelle $t = t_1$ den Wert $\tilde{t} = 0$ s hat. Es gilt also $\tilde{t} = t - t_1$. Der Ausgang $y(t)$ wird dadurch zu $y(\tilde{t})$ und ebenso $u(t)$ zu $u(\tilde{t})$.

Entsprechend den Ausdrücken $\frac{1}{s}$, $\frac{1}{s^2}$ und $\frac{1}{s^3}$ werden die zugehörigen Signale ein-, zwei- oder dreimal im Zeitbereich integriert. Aus Gl. (15) wird dann

$$y(\tilde{t}) - y_0 + b_2 \cdot (y_i(\tilde{t}) - \tilde{t} \cdot y_0) + b_1 \cdot \left(y_{ii}(\tilde{t}) - \frac{1}{2} \cdot \tilde{t}^2 \cdot y_0 \right) + b_0 \cdot y_{iii}(\tilde{t}) = \dots \quad (20)$$

$$\frac{1}{2} \cdot \tilde{t}^2 \cdot y_{20} + \tilde{t} \cdot y_{10} + a_0 \cdot u_{iii}(\tilde{t}).$$

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

$$\begin{bmatrix} y(\tilde{t}_1) - y_0 \\ \vdots \\ y(\tilde{t}_n) - y_0 \end{bmatrix} + \begin{bmatrix} y_i(\tilde{t}_1) - \tilde{t}_1 \cdot y_0 \\ \vdots \\ y_i(\tilde{t}_n) - \tilde{t}_n \cdot y_0 \end{bmatrix} \cdot b_2 + \begin{bmatrix} y_{ii}(\tilde{t}_1) - \frac{1}{2} \cdot \tilde{t}_1^2 \cdot y_0 \\ \vdots \\ y_{ii}(\tilde{t}_n) - \frac{1}{2} \cdot \tilde{t}_n^2 \cdot y_0 \end{bmatrix} \cdot b_1 + \begin{bmatrix} y_{iii}(\tilde{t}_1) \\ \vdots \\ y_{iii}(\tilde{t}_n) \end{bmatrix} \cdot b_0 = \dots \quad (21)$$

$$= \frac{1}{2} \cdot \begin{bmatrix} \tilde{t}_1^2 \\ \vdots \\ \tilde{t}_n^2 \end{bmatrix} \cdot y_{20} + \begin{bmatrix} \tilde{t}_1 \\ \vdots \\ \tilde{t}_n \end{bmatrix} \cdot y_{10} + \begin{bmatrix} u_{iii}(\tilde{t}_1) \\ \vdots \\ u_{iii}(\tilde{t}_n) \end{bmatrix} \cdot a_0$$

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

$$\underbrace{\begin{bmatrix} y(\tilde{t}_1) - y_0 \\ \vdots \\ y(\tilde{t}_n) - y_0 \end{bmatrix}}_{\tilde{y}} = \dots = \underbrace{\begin{bmatrix} \tilde{t}_1 \cdot y_0 - y_i(\tilde{t}_1) & \frac{1}{2} \cdot \tilde{t}_1^2 \cdot y_0 - y_{ii}(\tilde{t}_1) & -y_{iii}(\tilde{t}_1) & u_{iii}(\tilde{t}_1) & \frac{1}{2} \cdot \tilde{t}_1^2 & \tilde{t}_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{t}_n \cdot y_0 - y_i(\tilde{t}_n) & \frac{1}{2} \cdot \tilde{t}_n^2 \cdot y_0 - y_{ii}(\tilde{t}_n) & -y_{iii}(\tilde{t}_n) & u_{iii}(\tilde{t}_n) & \frac{1}{2} \cdot \tilde{t}_n^2 & \tilde{t}_n \end{bmatrix}}_H \cdot \underbrace{\begin{bmatrix} b_2 \\ b_1 \\ b_0 \\ a_0 \\ y_{20} \\ y_{10} \end{bmatrix}}_x \quad (22)$$

Gesucht wird derjenige Vektor x , der in Gl. (22) die Vektoren \tilde{y} und $H \cdot x$ so gleich wie möglich macht.

Hinweis: Bei Gl. (22) handelt es sich nur um eine Auswertung der dreifach integrierten DGL von Gl. (2). Dies bedeutet, bei perfekter Messung des Ausgangs y als Reaktion eines wirklichen PT3-Systems würde man nur 6 Messungen benötigen, um in Gl. (22) nach den sechs unbekannt Elementen des x -Vektors aufzulösen. Da Gl. (22) aber zur Bestimmung von x bei realen und bei fehlerbehafteten Messungen verwendet werden soll, werden viel mehr als nur sechs Messungen benötigt, damit sich die Messfehler ausgleichen können.

Für die Verwendung von Gl. (22) mit realen Messsignalen wird zu jedem Abtastzeitpunkt ein unbekannter Fehler-Vektor $e(t)$ hinzugefügt, der sowohl die nicht modellierbaren, zufälligen Messfehler als auch die Modellierungsfehler beinhaltet:

$$\tilde{y}(t) = H(t) \cdot x + e(t). \quad (23)$$

Die „Größe“ des Fehler-Vektors $e(t)$,

$$e(t) = \tilde{y}(t) - H(t) \cdot x, \quad (24)$$

kann durch die (skalare) Fehlerquadratsumme $J(x)$ beschrieben werden:

$$J(x) = \frac{1}{2} \cdot e^T(x) \cdot e(x), \quad (25)$$

denn geht die Zahl $J(x)$ gegen Null geht gehen auch alle Elemente des Fehler-Vektors $e(t)$ gegen Null. Es wird nun derjenige x -Vektor gesucht, der den Fehler-Vektor $e(t)$ so klein wie möglich macht. Die Lösung ist wird mit der Methode der kleinsten Quadrate nach Gauss [6] hergeleitet und kann analytisch angegeben werden mit

$$\hat{x} = (H^T \cdot H)^{-1} \cdot H^T \cdot \tilde{y}; \quad (26)$$

will man ausgewählte Zeitabschnitte bei den Messdaten besondere Bedeutung in der Berechnung zuweisen kann man einzelne Elemente des Fehler-Vektors $e(t)$ durch einen entsprechend großen Wert an der entsprechenden Stelle einer diagonalen Wichtungsmatrix W die wichten, so dass nun die gewichtete Fehlerquadratsumme

$$J(x) = \frac{1}{2} \cdot e^T(x) \cdot W \cdot e(x), \quad (27)$$

minimiert wird mit dem zugehörigen Ergebnis

$$\hat{x} = (H^T \cdot W \cdot H)^{-1} \cdot H^T \cdot W \cdot \tilde{y}. \quad (28)$$

Somit sind nun die Koeffizienten der Übertragungsfunktion, b_0, b_1, b_2, a_0 sowie die Zwischengrößen y_{10} und y_{20} berechenbar: Bringt man die zwei sich aus Gl. (15) ergebenden Bestimmungsgleichungen in ein lineares Gleichungssystem der Form

$$\begin{bmatrix} y_{20} \\ y_{10} \end{bmatrix} = \begin{bmatrix} b_2 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \dot{y}_0 \\ \ddot{y}_0 \end{bmatrix}, \quad (29)$$

so können auch die restlichen Anfangsbedingungen \dot{y}_0, \ddot{y}_0 durch Invertierung der Matrixgleichung berechnet werden. Erweitert man nun die Identifikation für das angegebene Beispiel dritter Ordnung in Gl. (4) auf eine DGL beliebiger Ordnung wie in Gl. (1), so ergibt sich statt des Gleichungssystems von Gl. (22),(29) das neue Gleichungssystem

$$\underbrace{\begin{bmatrix} y(\tilde{t}_1) - y_0 \\ \vdots \\ y(\tilde{t}_n) - y_0 \end{bmatrix}}_{\tilde{y}} = \dots \quad (30)$$

$$\underbrace{\begin{bmatrix} \frac{1}{1!} \tilde{t}_1 y_0 - y_i(\tilde{t}_1) & \dots & \frac{1}{(p-1)!} \tilde{t}_1^{p-1} y_0 - y_{i \dots i}(\tilde{t}_1) & -y_{i \dots i}(\tilde{t}_1) & u_{i \dots i}(\tilde{t}_1) \dots u_{i \dots i}(\tilde{t}_1) & \frac{1}{(p-1)!} \tilde{t}_1^2 & \dots & \frac{1}{1!} \tilde{t}_1 \\ \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{1!} \tilde{t}_n y_0 - y_i(\tilde{t}_n) & \dots & \frac{1}{(p-1)!} \tilde{t}_n^{p-1} y_0 - y_{i \dots i}(\tilde{t}_n) & -y_{i \dots i}(\tilde{t}_n) & u_{i \dots i}(\tilde{t}_n) \dots u_{i \dots i}(\tilde{t}_n) & \frac{1}{(p-1)!} \tilde{t}_n^2 & \dots & \frac{1}{1!} \tilde{t}_n \end{bmatrix}}_H \cdot \underbrace{\begin{bmatrix} b_{p-1} \\ \vdots \\ b_0 \\ a_q \\ \vdots \\ a_0 \\ y_{(p-1)0} \\ \vdots \\ y_{10} \\ x \end{bmatrix}}_x \quad (30)$$

$$\begin{bmatrix} y_{(n-1)0} \\ \vdots \\ y_{20} \\ y_{10} \end{bmatrix} = \begin{bmatrix} b_2 & \dots & b_{n-1} & b_n \\ \vdots & & b_n & 0 \\ b_{n-1} & & 0 & 0 \\ b_n & & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \dot{y}_0 \\ \ddot{y}_0 \\ \vdots \\ (n-1) \\ y_0 \end{bmatrix} \quad (31)$$

3 Das Verfahren in der praktischen Anwendung

Als zu modellierende Regelstrecke werden drei in Serie geschaltete RC-Netzwerke verwendet, deren Schaltung in Bild 1 dargestellt ist. Idealiert betrachtet handelt es sich um drei in Serie geschaltete PT_1 -Glieder. Jedes einzelne PT_1 -Glied ist in diesem Praktikumsversuch 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.

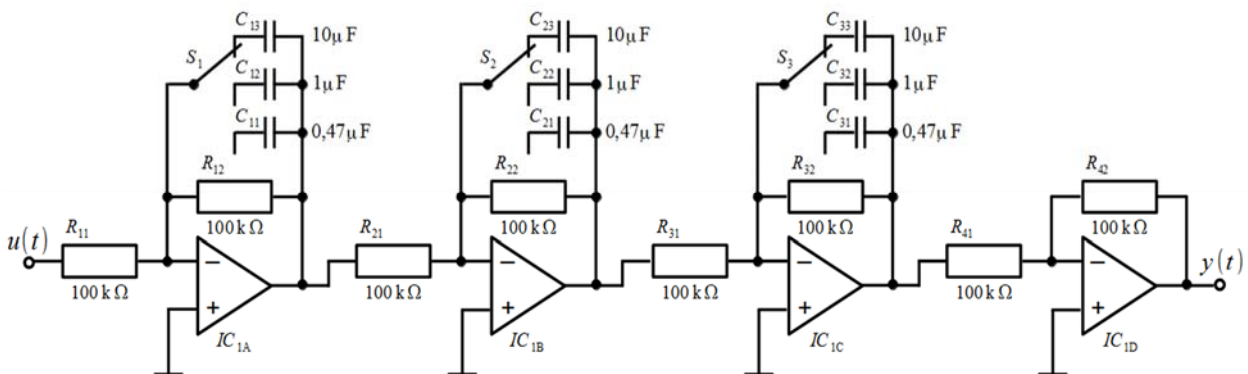


Bild 1: Schaltplan der elektronischen Regelstrecke

Der Widerstand ist bei jedem RC-Glied $R = 100 \text{ k}\Omega$ und die Werte für die Kondensatoren werden gewählt zu $C_1 = 0,47 \text{ }\mu\text{F}$; $C_2 = 1,0 \text{ }\mu\text{F}$; $C_3 = 10\text{ }\mu\text{F}$. Das physikalische Modell der Regelstrecke ist eine Serienschaltung von drei PT1-Gliedern und ergibt die Übertragungsfunktion [7]

$$G_{RC}(s) = \frac{1}{1 + T_1 \cdot s} \cdot \frac{1}{1 + T_2 \cdot s} \cdot \frac{1}{1 + T_3 \cdot s} \quad (32)$$

mit den Zeitkonstanten $T_i = RC_i$ für $i = 1,2,3$ und den Werten

$$T_1 = 0,47; T_2 = 1,0; T_3 = 10,0 \quad (33)$$

und ergibt dann die Übertragungsfunktion

$$G_{RC}(s) = \frac{a_0}{s^3 + b_2 s^2 + b_1 s + b_0} \quad (34)$$

mit

$$a_0 = 212,8, b_2 = 32,28, b_1 = 244,0, b_0 = 212,8. \quad (35)$$

3.1 Rückrechnung einer vorgegebenen Übertragungsfunktion aus *simulierten* Ein-/Ausgangsdaten

Das Verfahren wird nun wie in Bild 2 getestet. Die Testidee ist die, die bekannte Übertragungsfunktion G_{RC} als das zu bestimmende reale System zu nehmen und die simulierte Antwort des

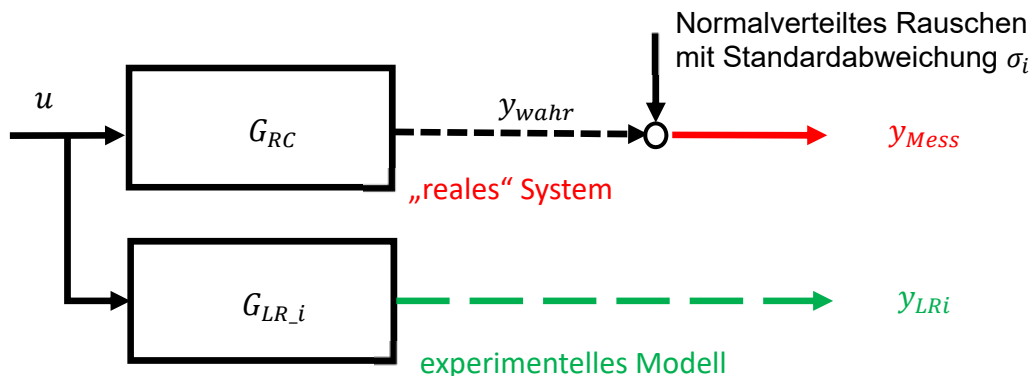


Bild 2: Signale und Systeme beim Testen durch Rückrechnung auf ein bekanntes „reales“ System

Systems auf den Sägezahn-Eingang künstlich durch stochastische Anteile als ungenaue Messung dem Verfahren zur Verfügung zu stellen. Im Einzelnen heißt das:

1. Auf die Übertragungsfunktion in Gl. (34) wird als Eingangssignal die in Bild 3 im obersten Subplot gezeigte Sägezahnspannung angelegt und das Ausgangsverhalten wird mit den Anfangsbedingungen

$$y_0^T := [y(t=0) \quad \dot{y}(t=0) \quad \ddot{y}(t=0)] = [-1,7938 \quad -0,0689 \quad -0,2544] \quad (36)$$

durch MATLAB-funktion *lsim* [8] simuliert.

2. Auf das Ausgangssignal wird dann zum Simulieren von Messfehlern normalverteilte Zufallszahlen mit vorgegebener Standardabweichung σ gemäß Tabelle 1 hinzuaddiert. Das Ergebnis ist für das Signal y_{Mess} für $\sigma = 0,05$ im mittleren Subplot in Bild 3 zu sehen.

3. Es wird nun die Identifikation mit den blau gekreuzten Messdaten durchgeführt (die oberen beiden Subplots in Bild 3) und anhand von zwei Kriterien bewertet:

A) Mit dem *Fit-Wert*:

$$\text{fit} := 100\% \cdot \left(1 - \frac{\text{norm}(y_{\text{Wahr}} - y_{\text{LR5}})}{\text{norm}(y_{\text{Wahr}} - \text{mean}(y_{\text{Wahr}}))} \right) \quad (37)$$

wobei unter „norm“ die Länge bzw. der Betrag des Vektors verstanden wird, und „mean“ bezeichnet den Mittelwert. Der Fit-Wert wird auch in MATLAB als Kriterium für Identifikationen verwendet.

B) Mit der „Rückrechnungsgenauigkeit“, dem RR-Wert, er gibt die Abweichungen der geschätzten Koeffizienten von den wahren Koeffizienten sowie zwischen den geschätzten Anfangsbedingungen und den wahren an:

$$\text{RR} := 100\% \cdot \left(1 - \frac{\text{norm}(c - c_e)}{\text{norm}(c)} \right), \quad (38)$$

wobei in dem Vektor c die „wahren“ Koeffizienten aus Gl. (35) und die „wahren“ Anfangsbedingungen aus Gl. (36) enthalten sind:

$$c = [a_0 \quad b_2 \quad b_1 \quad b_0 \quad \dot{y}_0 \quad \ddot{y}_0]^T, \quad (39)$$

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

In Bild 4 wird deutlich, dass eine exakte Rückrechnung der Koeffizienten nur für kleine Werte von σ möglich ist. Der Fit-Wert bleibt unabhängig davon bei nahezu 100 %, und das bedeutet, dass Übertragungsfunktionen, die von der „wahren“ numerisch deutlich abweichen, dennoch

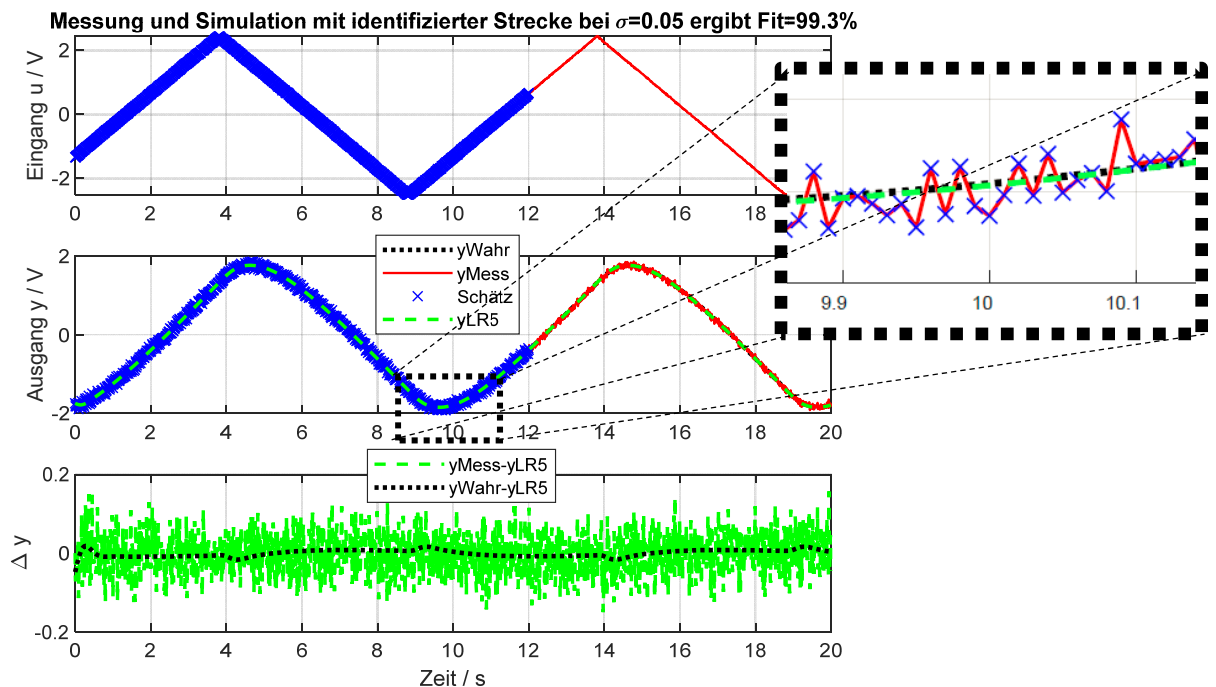


Bild 3: Messung und Simulation mit der identifizierten Übertragungsfunktion mit Modellierungsfehler

gute Simulationsergebnisse mit geringer Abweichung zum gemessenen Signal erzielen können. Ein guter Fit-Wert bedeutet also noch nicht zwingend, dass die wahre Übertragungsfunktion numerisch genau gefunden wurde.

Parameter	Wahre Werte	Geschätzte Werte mit Rauschanteilen σ					
		$\sigma = 0,00$	$\sigma = 0,01$	$\sigma = 0,02$	$\sigma = 0,03$	$\sigma = 0,04$	$\sigma = 0,05$
a_0	212,8000	213,6520	130,1557	98,4428	83,8548	76,8812	73,8661
b_2	32,2800	32,4219	19,9028	15,1521	12,9756	11,9463	11,5150
b_1	244,000	245,0567	149,1647	112,7071	95,8856	87,7809	84,1970
b_0	212,800	213,6570	130,2577	98,5880	84,0300	77,0834	74,0957
\dot{y}_0	-0,0689	-0,0696	0,2103	0,2814	0,3432	-0,4156	-0,5050
\ddot{y}_0	-0,2544	-0,2343	1,5195	1,3861	1,2410	1,2837	1,5347

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

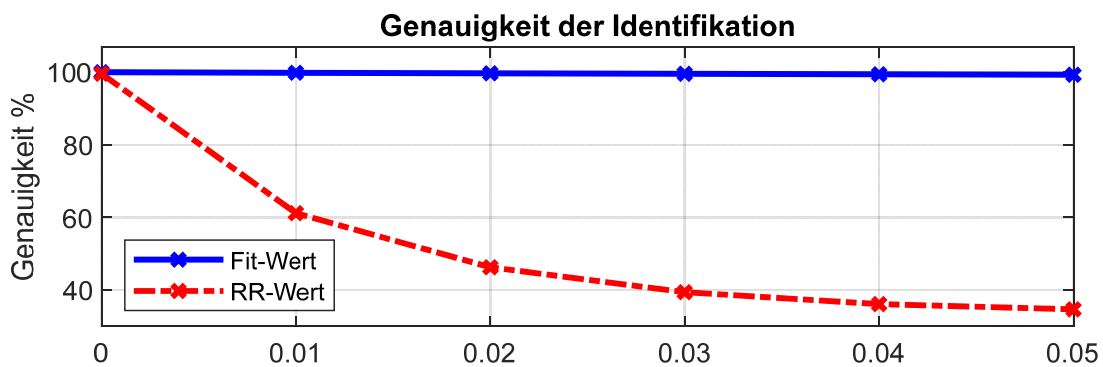


Bild 4: Ergebnisse der Identifikation für verschiedene Rausch-Level σ in der Messung

Der Test mit einem weiteren, andersartigen Ein-/Ausgangsgangssignal demonstriert in Bild 5 aber durch den kleinen Fehler, dass die Identifikation erfolgreich war.

Im Ergebnis heißt das, dass die Methode funktioniert. Im Anschluss wird sie an einem realen System ausprobiert, bei dem natürlich die „wahren Werte“ immer unbekannt bleiben.

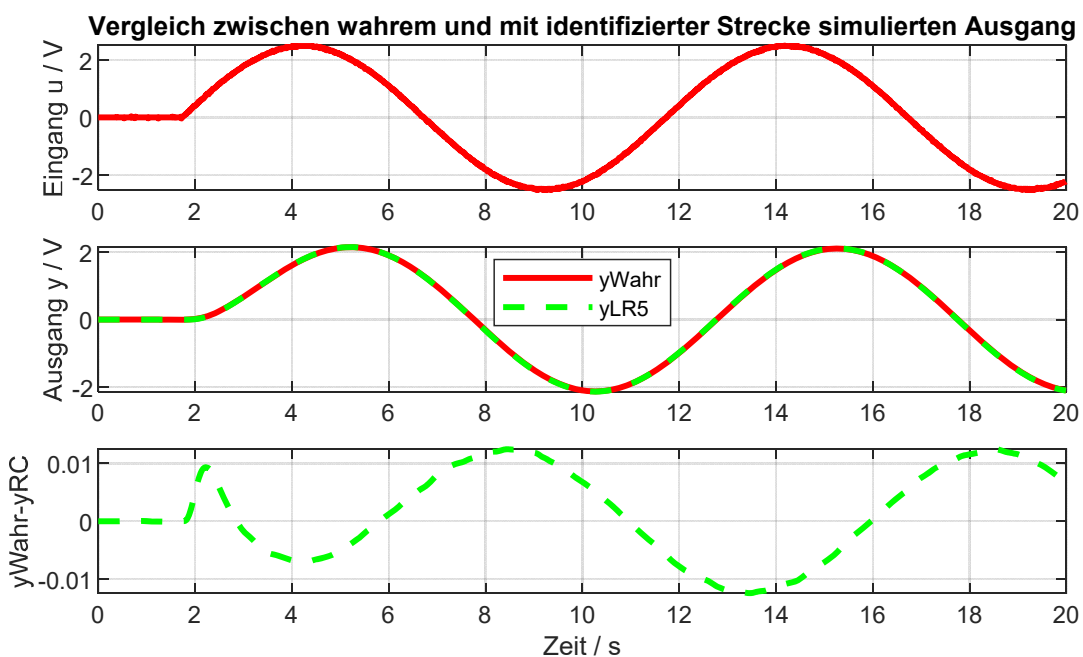


Bild 5: Simulation des mit $\sigma = 0,05$ verrauschten Ausgangssignals mit dem wahren Signal

3.2 Berechnen der Übertragungsfunktion aus gemessenen Ein-/Ausgangsdaten

Die Methode wird nun an einem realen System getestet, das wie in Bild 1 aus drei RC Netzwerken in Serie besteht. Die Werte für die Kapazitäten können an den drei Drehknöpfen eingestellt werden.

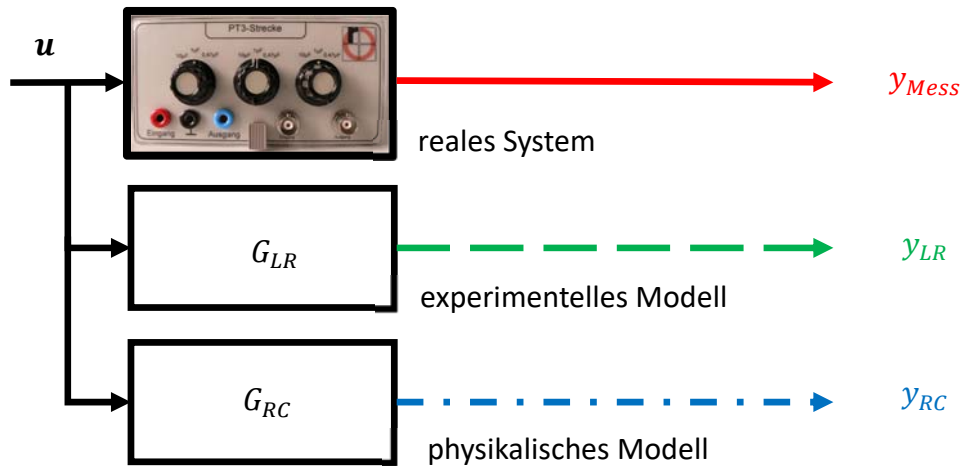


Bild 6: Bezeichnung der Signale und Systeme beim Test mit realen Messsignalen

Zunächst werden mit dem nach Gl. (22) beschriebenen Verfahren aus den in Bild 8 in den oberen beiden Subplots mit blauen Kreuzen gekennzeichneten Messsignalen die Übertragungsfunktion mit Zählergrad null und Nennergrad drei und die Anfangsbedingungen berechnet. Dabei werden anders als in Gl. (2), bei der nur der Koeffizient bei der höchsten Ableitung zu eins gesetzt wurde, nacheinander alle anderen Koeffizienten ebenfalls zu eins gesetzt und das Ergebnis anhand des Fit-Wertes bewertet. Dieser Algorithmus ist in dem kostenlos herunterladbaren Tool pzMove unter Registerkarte „identification“ [5] programmiert. Hier lassen sich die Einflüsse auf das Ergebnis schnell verändern und ausprobieren:

1. Auswahl des *Messbereiches*, der für die Schätzung verwendet wird
2. Die *Wichtungen* der Messungen nach Gl. (22)
3. Sortieren der mit allen Koeffizienten erzielten Ergebnisse nach „stabilen“ oder „instabilen“ Lösungen
4. Berücksichtigung von *Totzeiten* durch zeitliche Verschiebung von Ein- und Ausgangssignal und Bewertung der Simulationsgüte mit dem Fit-Wert Gl. (37)
5. Vorgeben des *Zähler- und des Nennergrades* der Übertragungsfunktion

Das Ergebnis ist die Übertragungsfunktion

$$G_{LRa}(s) = \frac{22,22}{s^3 + 1,894 s^2 + 25,98 s + 22,25'} \quad (40)$$

und man erkennt nach Simulation und Vergleich mit der Messung in Bild 8, 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:

$$G_{LRb}(s) = \frac{6,26 s + 0,3471}{s^3 + 7,331 s^2 + 6,764 s + 0,3337'} \quad (41)$$

Denn mit dieser Übertragungsfunktion wird ein höherer Fit-Wert von 98,8% erreicht und das Differenzsignal im unteren Subplot in Bild 9 sieht nun wie unkorreliertes weißes Rauschen aus – das bestmögliche Ergebnis also.

Dies macht das konstruktive Zusammenspiel von physikalischer und experimenteller Modellierung besonders deutlich: Weil das Hinzufügen einer weiteren Nullstelle das Modellierungsergebnis noch einmal deutlich verbessert ist das ein starker Hinweis, dass sich dahinter ein bislang ignoriertes 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. 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.

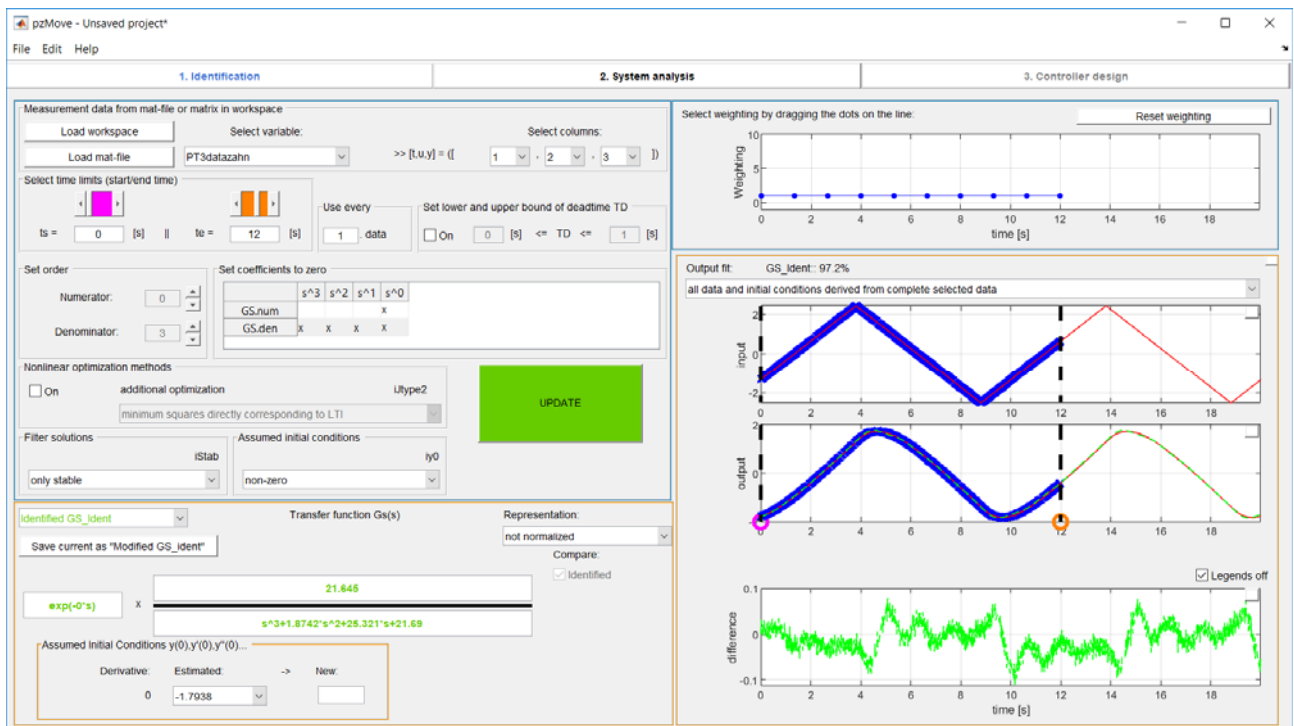


Bild 7: Schaltfläche des Identifikation-Tools zur experimentellen Modellbildung nach dem hier vorgestellten Verfahren

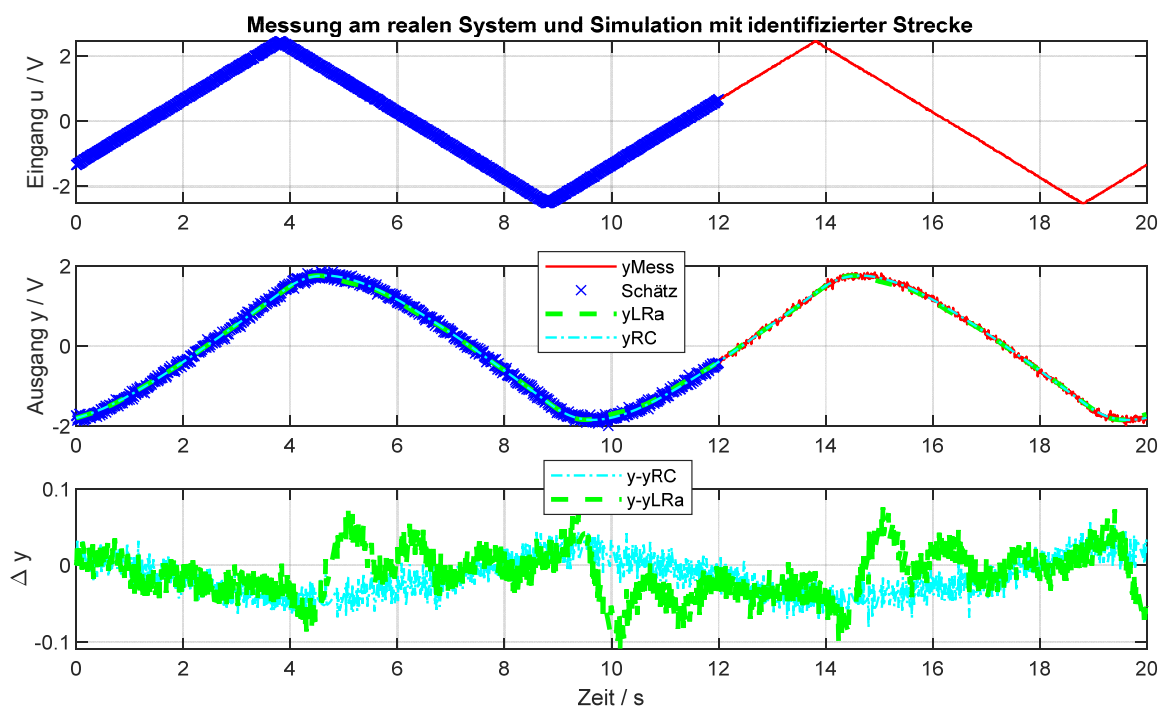


Bild 8: Ergebnis der Simulation mit dem physikalischen (Fit 97,5%) und dem aus den Messdaten gewonnenen experimentellen Modell (Fit=97,3%)

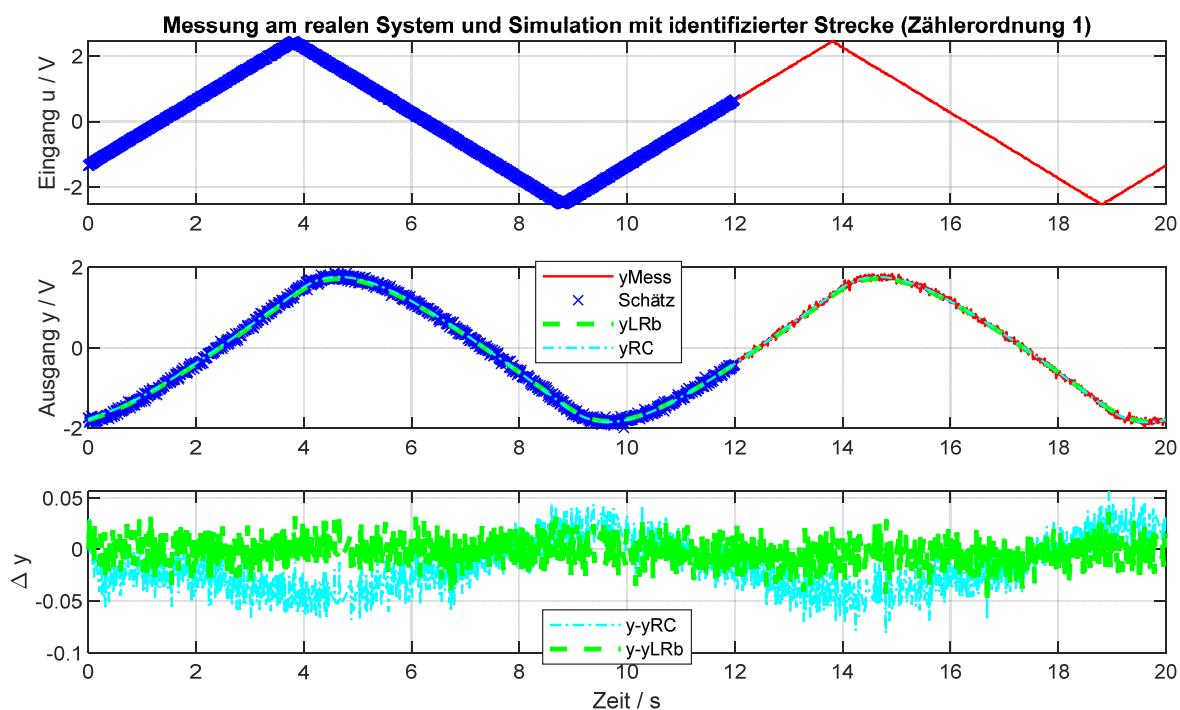


Bild 9: Ergebnis der Simulation mit dem physikalischen (Fit 97,5%) und dem aus den Messdaten gewonnenen experimentellen Modell mit Zählerordnung eins und Nennerordnung drei mit erhöhtem Fit =98,8%

4 Zusammenfassung und Ausblick

Es wurde ein Verfahren zur experimentellen Modellbildung für Systeme vorgestellt, die sich als lineare Differentialgleichungen mit konstanten Koeffizienten beschreiben lassen. Das Verfahren hat ein großes Einsatzspektrum, weil es keine Einschränkung im Streckentyp benötigt, für den offenen und geschlossenen Regelkreis geeignet ist und für Vorgänge geeignet ist, die sich nicht in der Ruhelage beginnen – die Anfangswerte können mitgeschätzt werden.

Das Verfahren ist völlig transparent und baut allein auf dem Vorwissen von Bachelorstudenten der Ingenieurwissenschaft auf. Die Richtigkeit der Methode wurde demonstriert und sie wurde an einem realen Beispiel angewendet. Das Zusammenspiel von physikalischer und experimenteller Modellbildung wurde verdeutlicht.

Das Verfahren einschließlich des hier vorgestellten numerischen Beispiels ist für jedermann zugänglich [5] und als MATLAB-unabhängiges *.exe-Programm herunterladbar. Dozenten können auf Nachfrage die Praktikumsanleitung und die Musterlösung für einen regelungstechnischen Versuch 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.

5 Literatur

- [1] Lutz, J., und Wendt, W., *Taschenbuch der Regelungstechnik*, Harri Deutsch, 2014
- [2] Bohn, C., und Unbehauen, H., *Identifikation dynamischer Systeme- Methoden zur experimentellen Modellbildung aus Messdaten*, Springer, 2016
- [3] Kahlert, Jörg, *Simulation technischer Systeme*, Vieweg, 2004
- [4] Golubev, B.; Horowitz, I.: *Plant rational transfer approximation from input-output data*, International Journal on Control 4/1982
- [5] www.fh-rosenheim.de/~rt-pzMove.html
- [6] Crassidi, J., and Junkins, J., *Optimal Estimation of Dynamic Systems*, p, 551, Chapman and Hall/CRC, 2004
- [7] Karl-Dieter Tieste, Oliver Romberg: *Keine Panik vor Regelungstechnik!*, Springer-Verlag, 2014
- [8] www.mathworks.de