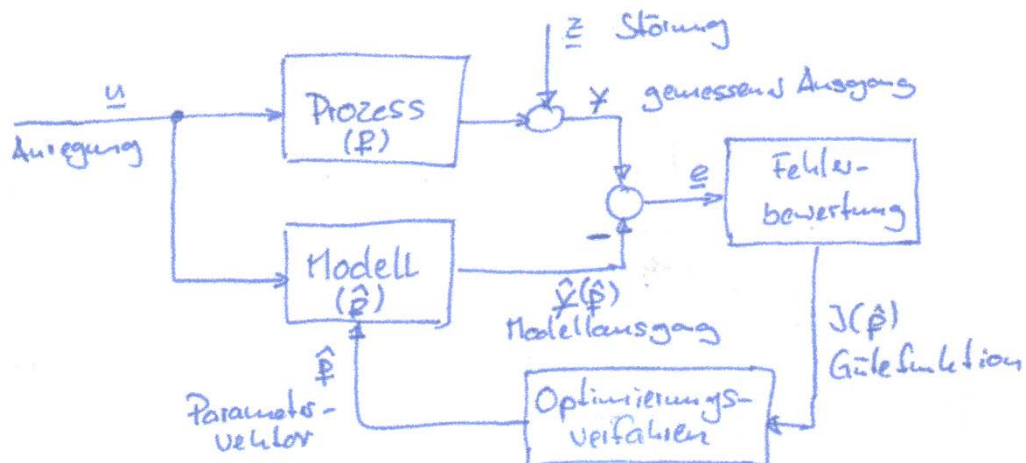


Skript zur Vorlesung

Modellbildung, Identifikation und Simulation dynamischer Systeme

Prof. Dr.-Ing Christoph Ament

Lehrstuhl Regelungstechnik in der Ingenieurinformatik



Stand: Februar 2017

Gliederung

Literatur	0-3
1 Einführung.....	1-1
1.1 Motivation.....	1-1
1.2 Modellbegriff	1-1
1.3 Modellbildung.....	1-2
Teil A: Modellbildung	1-1
2 Whitebox-Modelle	2-1
2.1 Bilanzgleichungen	2-1
2.2 Mechanische Systeme	2-2
2.3 Elektrische Systeme.....	2-6
2.4 Analogiebetrachtungen.....	2-8
2.5 Methode von Lagrange.....	2-13
2.6 Modellierung im Blockschaltbild	2-16
2.7 Zustandsraumdarstellung	2-18
3 Modellvereinfachung	3-1
3.1 Linearisierung	3-1
3.2 Zeitliche Diskretisierung	3-3
3.3 Diskretisierung der Zustände	3-4
3.4 Ortsdiskretisierung	3-5
3.5 Ordnungsreduktion	3-5
4 Blackbox-Modelle	4-1
4.1 Look-up Table Modell	4-2
4.2 Polynomiale Modelle.....	4-3
4.3 Modelle mittels radialer Basisfunktionen.....	4-3
4.4 Neuronale Netze.....	4-5
4.5 Erweiterung auf Signale	4-7
4.6 Erweiterung auf dynamische Systeme.....	4-8
Teil B: Identifikation	4-1
5 Identifikation statischer Modelle.....	5-1
5.1 Aufgabenstellung.....	5-1
5.2 Lineare Optimierung: Direkte Lösung.....	5-2
5.3 Lineare Optimierung: Rekursive Lösung	5-5
5.4 Nichtlineare Optimierung	5-6

6	Versuchsplanung	6-1
6.1	Grundlagen	6-1
6.2	Versuchsplanung	6-4
6.3	Versuchsauswertung	6-9
7	Identifikation von Systemantworten und Signalen	7-1
7.1	Systemantworten	7-1
7.2	Signale	7-3
8	Identifikation dynamischer Systeme	8-1
Teil C: Simulation		8-1
9	Explizite Verfahren	9-1
9.1	Grundlagen	9-1
9.2	Lokaler Diskretisierungsfehler	9-3
9.3	Globaler Diskretisierungsfehler und Schrittweitensteuerung	9-4
9.4	Stabilität	9-7
10	Implizite Verfahren	10-1
10.1	Grundlagen	10-1
10.2	Stabilität	10-3
10.3	Steife Systeme	10-3
Anhang		1
A. Statistische Tests		1

Literatur

- R. Isermann, M. Münchhof: Identification of Dynamic Systems – An Introduction with Applications, Springer Verlag, 2011 (93,80 €)
- C. Bohn, H. Unbehauen: Identifikation dynamischer Systeme – Methoden zur experimentellen Modellbildung aus Messdaten, Springer, 2016 (39,99 €)
- J. Wernstedt: Experimentelle Prozessanalyse, VEB Verlag Technik, 1989
- B. Fabien: Analytical System Dynamics – Modeling and Simulation, Springer, 2009
- L. Ljung: System Identification: Theory for the User, 2nd Edition, Prentice Hall, 1999
- O. Nelles: Nonlinear System Identification, Springer, Berlin, 2001
- K. Janschek: Systementwurf mechatronischer Systeme, Methoden – Modelle – Konzepte, Springer, 2010 (59,99 €)
- R. Nollau: Modellierung und Simulation technischer Systeme, Eine praxisnahe Einführung, Springer, 2009 (59,95 €)
- H. Scherf: Modellbildung und Simulation dynamischer Systeme, Eine Sammlung von Simulink-Beispielen, Oldenbourg, 2009 (27,95 €)
- H.-J. Bungartz, St. Zimmer, M. Buchholz, D. Pflüger: Modellbildung und Simulation, Eine anwendungsorientierte Einführung, Springer, 2009 (29,99 €)

Zur Statistischen Versuchsplanung:

- B. Klein: Versuchsplanung – DoE, Einführung in die Taguchi/Shainin-Methodik, Oldenbourg, 3. Auflage, 2011 (44,80 €)
- W. Kleppmann: Taschenbuch Versuchsplanung, Produkte und Prozesse optimieren, 7. Auflage, Hanser, 2011 (34,99 €)
- E. Scheffler, Einführung in die Praxis der statistischen Versuchsplanung, 1986
- R. Schmitt, T. Pfeifer: Qualitätsmanagement – Strategien, Methoden, Techniken, 4. Auflage, Hanser, 2010 (49,90 €)

1 Einführung

1.1 Motivation

Bei der Konzeption technischer Systeme oder Prozesse gehört eine begleitende Modellbildung heute in der Regel zum „guten“ ingenieurwissenschaftlichen Vorgehen. Das so entstehende mathematische Modell des Systems oder Prozesses kann in unterschiedlicher Weise genutzt werden:

- **Analyse des Systemverhaltens:** Auf Basis eines Modells können Einsichten in das Systemverhalten gewonnen werden: Unter welchen Voraussetzungen ist das System z. B. instabil oder schwingungsfähig? Welche Zeitkonstanten besitzt es?
- **Vorhersage des Systemverhaltens:** Die Dynamik des Systems kann numerisch simuliert werden. Eine Simulation kann bereits in der Entwurfsphase durchgeführt werden und ist oft kostengünstiger und ungefährlicher als reale Experimente.
- **Reglerentwurf:** Das Modell ist Grundlage des Reglerentwurfs (siehe Vorlesung „Mess- und Regelungstechnik“).
- **Diagnose und Monitoring:** Ein Modell kann genutzt werden, um im Betrieb des Prozesses online Fehler zu erkennen und ggf. eine Fehlermeldung auszulösen. So könnte beispielsweise der Strom eines elektrischen Antriebs überwacht werden: Wird die Differenz zwischen Messung und modellbasierter Berechnung zu groß, kann ein Alarm ausgelöst werden.
- **Optimierung:** Auf Basis des Modells können z.B. Parameter des Systems optimiert werden, um etwa einen ressourceneffizienten Betrieb zu ermöglichen.

1.2 Modellbegriff

Modell (in Wissenschaft und Technik)

„Darstellung, die nur die als wichtig angesehenen Eigenschaften des Vorbilds ausdrückt, um durch diese Vereinfachung zu einem überschaubaren oder mathematische berechenbaren oder zu experimentellen Untersuchungen geeigneten Modell zu kommen.“ [aus: dtv Brockhaus, 1998]

Mathematisches Modell

„Ein mathematisches Modell verwendet die mathematische Sprache für die Beschreibung eines Systems, z. B. aus der Physik, der Biologie oder den Sozialwissenschaften. Der Prozess zur Erstellung wird als Modellierung bezeichnet. Durch Vergleich der Berechnungsergebnisse mit Beobachtungen kann die Korrektheit eines Modells verifiziert werden. Anwendungsbeispiele mathematischer Modelle sind etwa Prognosen des Klimawandels, des Wetters oder der Standsicherheitsnachweis (Statik) eines Gebäudes.“ [aus: Wikipedia, Abruf 2011]

Insbesondere die erste Definition macht deutlich, dass Modelle immer nur einen Teil der Realität (in vereinfachender Weise) wiedergeben. Diese Sichtweise auf die Realität ist geprägt durch die spezielle Fragestellung, die mit Hilfe des Modells untersucht werden soll.

Klassifikation nach der Abbildungsart

Materielle Modelle

sind verkleinerte oder vereinfachte materielle Realisierungen, wie etwa ein Architekturmodell, ein Rapid-Prototyping-Bauteil im Rahmen einer Designstudie oder auch eine Modell-Eisenbahn.

Ideelle Modelle

sind demgegenüber „gedankliche“ Modelle z.B. in Form mathematischer Gleichungen. Der Schwerpunkt dieser Vorlesung liegt in der Entwicklung solcher ideeller Modelle.

Klassifikation nach Art der Beschreibung

Strukturmodelle („White Box Modelle“)

stellen die kausalen Zusammenhänge innerhalb eines Systems dar. Diese analytische Beschreibung kann z.B. durch physikalische Gesetze oder durch Petri-Netze gegeben sein. Diese Modelle sind i.d.R. das Ergebnis einer theoretischen Modellbildung.

Verhaltensmodell („Black Box Modelle“)

hingegen beschreiben das Ein-Ausgangsverhalten in deskriptiver Weise. Sie verwenden einen allgemeinen Modellansatz zur Approximation des Verhaltens oder eine nicht-parametrische Darstellung und sind meist das Ergebnis einer experimentellen Modellbildung.

Auch Mischformen („Grey Box Modelle“) beider Formen sind möglich.

Parametrische Modelle

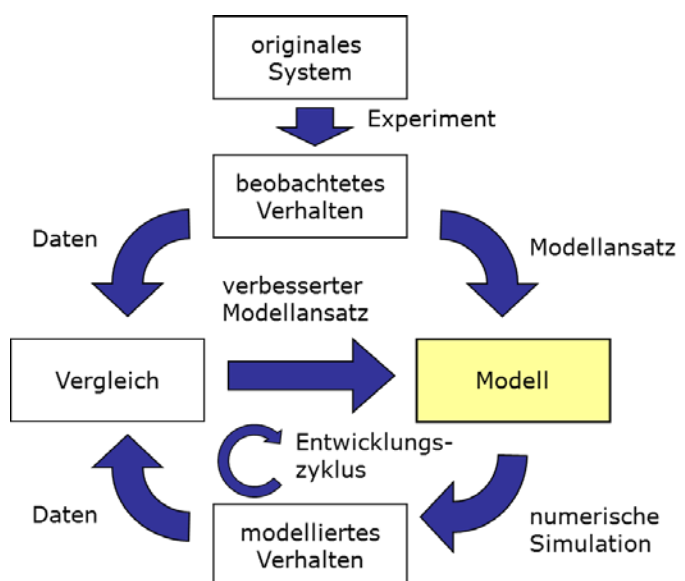
bestehen aus einem (physikalischen oder allgemeinen) Modellansatz und einer endlichen Zahl von Parametern, welche diesen Ansatz quantifizieren. Im Vergleich zu nicht-parametrischen Modellen bieten sie eine kompaktere Modelldarstellung.

Nicht-parametrische Modelle

besitzen keine spezielle Struktur und beschreiben das System mittelbar durch Daten, die häufig durch Messung ermittelt werden (z.B. eine Sprungantwort). Sie besitzen prinzipiell eine unendliche Anzahl von Parametern.

1.3 Modellbildung

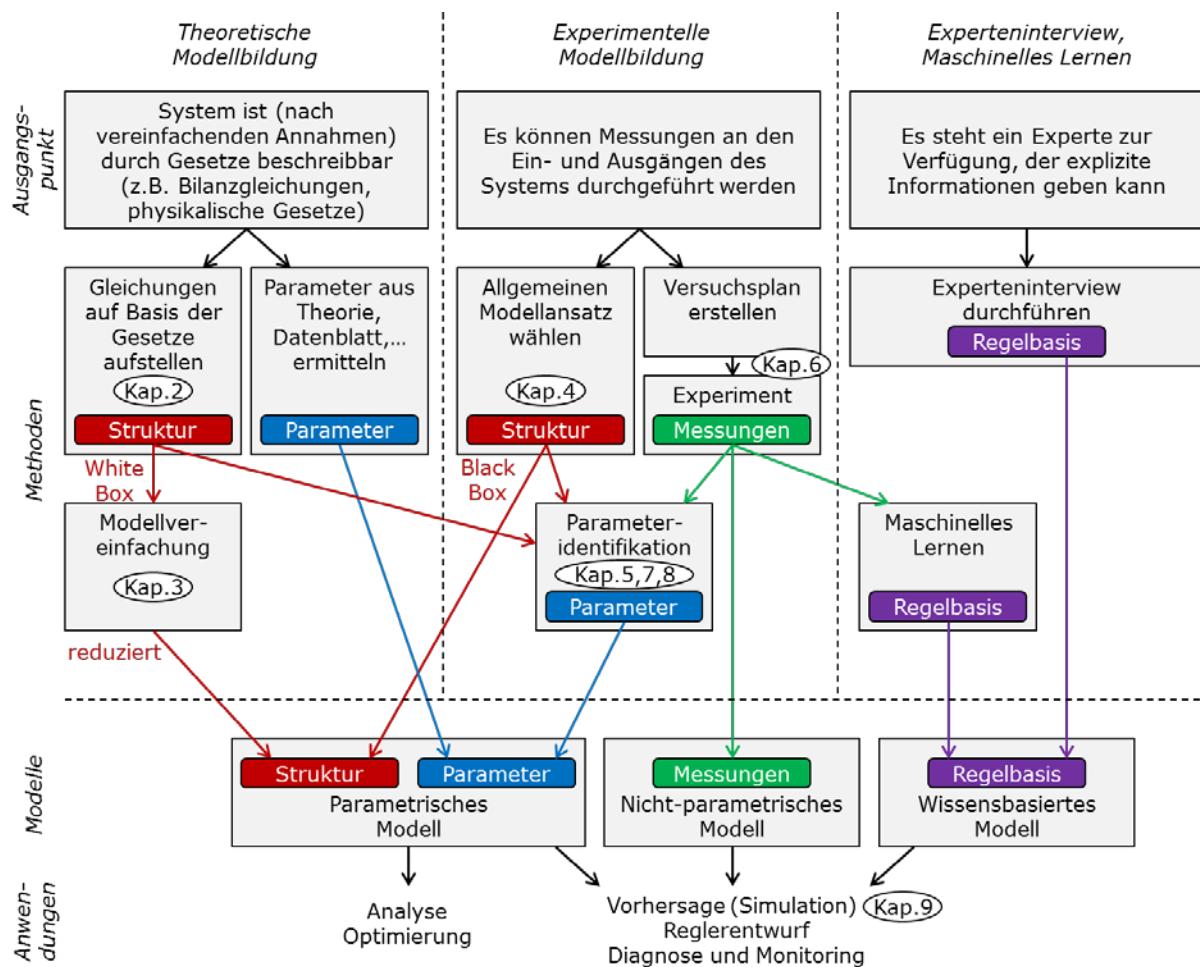
Als *Modellbildung* (synonym: *Modellierung*) wird der Prozess zur Entwicklung eines Modells bezeichnet. Das originale System und das Modell lassen sich nicht unmittelbar vergleichen wie nebenstehendes Bild zeigt. Auf Basis der Beobachtung des realen Prozesses muss (in systematischer und evtl. auch kreativer Weise) ein Modellansatz gefunden werden. Dann kann das modellierte und das beobachtete Verhalten miteinander verglichen und auf dieser Basis der Modellansatz verbessert werden. Hierdurch entsteht ein (mitunter langwieriger!) Entwurfszyklus der Modellbildung.



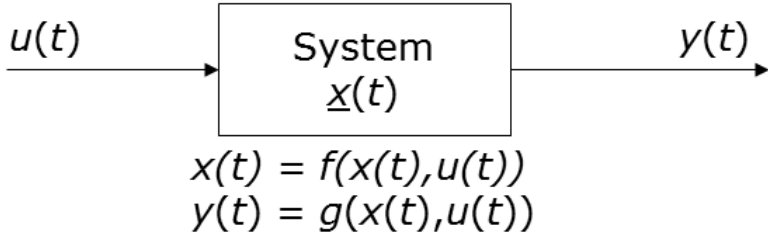
Je nachdem, welche Informationen über den realen Prozess zur Verfügung stehen, werden unterschiedliche Wege (oder Strategien) der Modellbildung unterschieden:

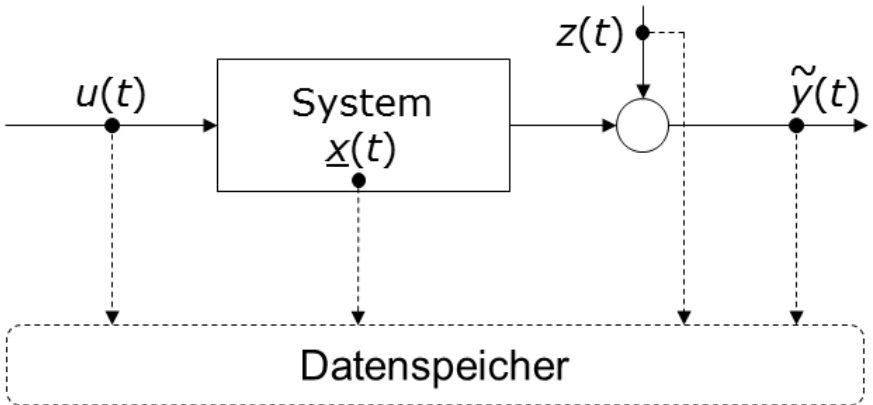
Ausgangspunkt (vorhandene Information)	Modellbildungsstrategie	Typische Modellstruktur
Theoretische Kenntnisse, z.B. Gleichungen	Theoretische Modellbildung	Strukturmodell („White Box“)
Experimentelle Ergebnisse, Messungen	Experimentelle Modellbildung	Verhaltensmodell („Black Box“)
Erfahrungs- oder Expertenwissen eines Menschen	Experten-Interview, Maschinelles Lernen	Regelbasis

Nachfolgend ist das Zusammenspiel dieser Wege der Modellbildung dargestellt. Man erkennt, dass häufig eine Kombination von Strategien sinnvoll ist, abhängig davon, für welche Teile eines Prozesses welche Informationen am einfachsten gewonnen werden können. In der Praxis werden diese Wege im Zuge des Entwicklungszyklus der Modellbildung (Bild auf Seite 1-2) evtl. mehrfach durchlaufen.

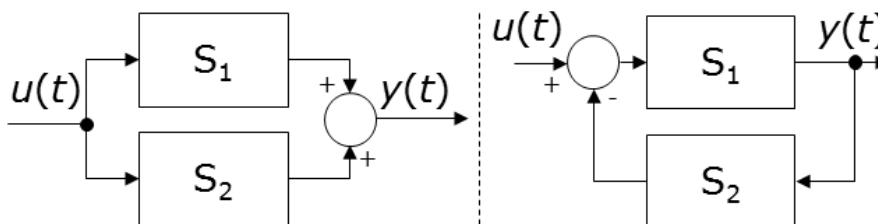



Zusammenfassung der Strategien nach Wernstedt

Theoretische Modellbildung	
Grundlage	Ergebnis
Bilanzgleichungen von u.a. Massen, Energien, Kräften	Physikalische Modelle, häufig im Zustandsraum
<div style="text-align: center;">  <p> $x(t) = f(x(t), u(t))$ $y(t) = g(x(t), u(t))$ </p> </div>	
Vorteile	Nachteile
<ul style="list-style-type: none"> • Beliebig genaue Modellierung des realen Verhaltens (je nach Anforderungen / vorhandenen Ressourcen) • Parameter sind oft Systemkomponenten zugeordnet (z.B. Masse, Widerstand), daher einfache Bestimmung (Wiegen, Datenblatt) 	<ul style="list-style-type: none"> • Bestimmung der nicht zuordenbaren Parametern • Einfluss von (Mess-)Störungen schwierig erfassbar • Zeitvariantes Verhalten schwierig erfassbar (z.B. Abnutzung)

Experimentelle Modellbildung	
Grundlage	Ergebnis
Experimente am System und Aufnahme der Systemgrößen $u(t)$, $y(t)$, $x(t)$, $z(t)$	Ein-Ausgangsmodelle oder Zustandsmodelle
	
Vorteile	Nachteile
<ul style="list-style-type: none"> Modellparameterbestimmung gut möglich (Bestimmung/ Schätzung) Zeitvarianz durch adaptive Strategien erfassbar System wird in realer Umwelt abgebildet 	<ul style="list-style-type: none"> Strukturerkennung schwer bzw. nicht möglich* Modell nur im Bereich der Messdaten gültig Parameterzuordnung zu Systemkomponenten schwer möglich Experimente am System erforderlich: <ul style="list-style-type: none"> System muss existieren Betrieb nur im stabilen Bereich möglich (Modellierung des instabilen Bereichs schwierig) Aufwand (Zeit, Ausfall (z.B. Maschine), Abnutzung, Auswertung)

Systeme, die in begrenzten Bereichen ähnliche Messwerte generieren, können stark abweichende Strukturen haben. Die Gefahr instabilen Verhaltens wird dabei möglicherweise übersehen:



Experteninterview/ Maschinelles Lernen	
Grundlage	Ergebnis
Befragen/ Handlungsanalysen von Experten	Eine der vorgestellten Modellbeschreibungen; Schwerpunkt linguistische/ qualitative Modelle
<div style="text-align: center;">  <pre> graph TD A[Experteninterview/ Handlungsanalyse] --> B[Interviews] A --> C[Handlungsanalyse] </pre> </div>	
Vorteile	Nachteile
<ul style="list-style-type: none"> • Geringer Aufwand • Abbildung des realen Verhaltens • Nachbildung der langjährigen Erfahrung des Experten 	<ul style="list-style-type: none"> • Wissen und Bereitschaft zur Wissensabgabe des „Experten“ • Qualität des Interviewers und seiner Befragungsstrategie • Subjektivität des Menschen und von Gruppierungen • Erfahrungen/Beispiele müssen existieren => ungeeignet für neue Technologien/ Produkte

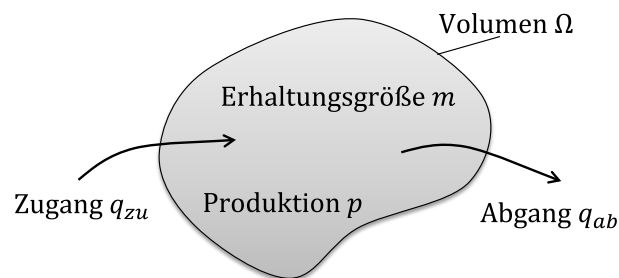
Teil A: Modellbildung

2 Whitebox-Modelle

2.1 Bilanzgleichungen

Die Bilanzierung einer Erhaltungsgröße ist ein sehr universelles Modellierungsprinzip, das z.B. bei verfahrenstechnischen oder biologischen Prozessen Anwendung findet.

In einem spezifizierten Volumen Ω wird eine Erhaltungsgröße m bilanziert.



Die Änderung dieser Größe innerhalb des Volumens in einem Zeitintervall muss dem Zugang q_{zu} abzüglich dem Abgang q_{ab} in diesem Zeitintervall entsprechen. Zusätzlich kann noch eine Produktion p im Inneren des Volumens ergänzt werden.

Die Bilanzgleichung in integraler Schreibweise lautet:

$$m(t) = m(0) + \int_0^t [q_{zu}(\tau) - q_{ab}(\tau) + p(\tau)] d\tau$$

Die entsprechende Bilanzgleichung in differentieller Schreibweise lautet:

$$\dot{m}(t) = q_{zu}(t) - q_{ab}(t) + p(t)$$

Typische Erhaltungsgrößen sind:

- Masse oder Volumen
- Energie
- Anzahl von Zellen
- Elektrische Ladung
- Impuls

Für jedes Bilanzierungsvolumen Ω erhält man so eine lineare Differenzialgleichung erster Ordnung.

2.2 Mechanische Systeme

Newton'sche Mechanik eines Teilchens

Es gilt das *Zweite Newton'sche Bewegungsgesetz*:

$$\underline{F} = \frac{d\underline{p}}{dt}$$

Die Summe aller am Massepunkt angreifenden Kräfte \underline{F} ist gleich der zeitlichen Ableitung des Impulses \underline{p} . Letzteres ist das Produkt aus Masse m und Geschwindigkeit \underline{v} des Massepunktes:

$$\underline{p} = m \cdot \underline{v}$$

Die Geschwindigkeit \underline{v} ist die zeitliche Ableitung der Trajektorie \underline{x} des Massepunktes:

$$\underline{v} = \frac{d\underline{x}}{dt}$$

Es gibt Fälle, in denen die Masse m zeitveränderlich ist, z.B. eine Rakete. Häufig kann die Masse m aber als konstant betrachtet werden, dann gilt

$$\underline{F} = m \cdot \frac{d\underline{v}}{dt} = m \cdot \underline{a}$$

wobei \underline{a} die Beschleunigung des Massenpunktes ist.

Für die *geleistete Arbeit*, um das Teilchen von Punkt 1 zu Punkt 2 zu bewegen, gilt:

$$W_{12} = \int_1^2 \underline{F} \cdot d\underline{s}$$

Die Kraft \underline{F} wird als *konservativ* bezeichnet, wenn für einen geschlossenen Weg gilt:

$$W = \oint \underline{F} \cdot d\underline{s} = 0$$

dann existiert ein *Potential* V (potentielle Energie) dessen negativer Gradient die Kraft \underline{F} ist:

$$\underline{F} = -\text{grad}(V)$$

Die *kinetische Energie* des Teilchens ist:

$$T = \frac{1}{2} m \cdot |\underline{v}|^2$$

Energieerhaltung:

Wenn die Kräfte, die auf ein Teilchen wirken, konservativ sind, dann bleibt die Gesamtenergie $T + V$ des Teilchens konstant.

Newtonsche Mechanik für Teilchensysteme

Das Teilchensystem besteht aus $i = 1, \dots, N$ Teilchen. In Teilchensystemen kann zwischen den *inneren Kräften*, die innerhalb des Systems wirken, und den *äußeren Kräften*, deren Ursache außerhalb des Systems liegt, unterschieden werden.

Die Bewegungsgleichung für das Teilchen i lautet:

$$\underline{F}_i = \frac{d\underline{p}_i}{dt} \quad \text{mit} \quad \underline{F}_i = \underline{F}_i^{(a)} + \sum_{\substack{j=1 \\ j \neq i}}^N \underline{F}_{ji}$$

$\underline{F}_i^{(a)}$: äußere Kraft auf das Teilchen i

\underline{F}_{ji} : innere Kraft des Teilchens j auf das Teilchen i

Dabei gilt für die inneren Kräfte:

- „actio = reactio“ (3. Newtonsche Bewegungsgesetz):

$$-\underline{F}_{ij} = \underline{F}_{ji}$$

- Die Kräfte \underline{F}_{ij} und \underline{F}_{ji} wirken in der Verbindungslinie der beiden Teilchen.

Die kinetische Energie eines Teilchensystems entspricht der Summe der kinetischen Energien der Einzelteilchen

$$T = \frac{1}{2} \sum_{i=1}^N m_i \cdot |\underline{v}_i|^2$$

mit der Masse m_i und der Geschwindigkeit \underline{v}_i des Teilchens i .

Zur Bestimmung der potentiellen Energie eines Teilchensystems werden die Potentiale der äußeren bzw. der inneren Kräfte unterschieden. (Der Faktor 1/2 bei den Potentialen der inneren Kräfte vermeidet, dass diese „doppelt“ gezählt werden.)

$$V = \sum_{i=1}^N V_i^{(a)} + \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N V_{ji}$$

$V_i^{(a)}$: Potential von $\underline{F}_i^{(a)}$

V_{ij} : Potential von \underline{F}_{ij}

Zwangsbedingungen

Häufig werden Freiheitsgrade der Bewegung eines mechanischen Systems durch geometrische Randbedingungen eingeschränkt. Dies kann beispielsweise das Seil eines Fadenpendels, die Führung einer Spindel oder eine Auflagefläche sein.

Man unterscheidet *holonome* und *nicht-holonome Zwangsbedingungen*: Holonome Zwangsbedingungen können als Gleichung der Ortskoordinaten der N Teilchen angegeben werden:

$$f_k(\underline{x}_1, \dots, \underline{x}_N) = 0$$

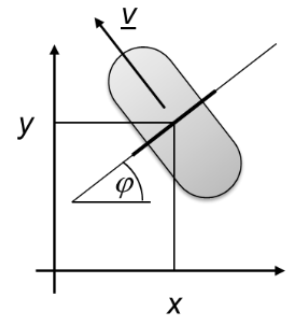
(für die Zwangsbedingung $k=1, \dots, K$)

Bei nicht-holonomen Bedingungen ist das nicht möglich. Hier hängt die Randbedingung auch von der Geschwindigkeit ab.

Ein Beispiel einer nicht-holonomen Zwangsbedingung ist die Bewegung eines Rades. Bei einer Rollbewegung (ohne Gleiten) ist die Zwangsbedingung

$$\dot{x} \cos \varphi + \dot{y} \sin \varphi = 0 \quad \text{mit} \quad \underline{\dot{v}} = \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix}$$

immer erfüllt und beschränkt die Freiheitsgrade x , y , φ geschwindigkeitsabhängig.



Weiterhin werden *skleronome Zwangsbedingung*, die nicht explizit von der Zeit abhängen, von *rheonome Zwangsbedingung*, die zeitabhängig sind, unterschieden.

Um Zwangsbedingungen im Modell zu berücksichtigen, werden *Zwangskräfte* im System eingeführt, welche die Wirkung der Zwangsbedingungen sicherstellen. Dies sind beispielsweise Seilkräfte, Reaktionskräfte einer Führung oder Normalkräfte. Die Zwangskraft am Teilchen i ist \underline{Z}_i .

Von den Zwangskräften werden die zu Beginn eingeführten *eingepprägten Kräfte* \underline{F}_i unterschieden. Letztere sind durch Kraftgesetze gegeben, das kann z.B. eine Gravitations-, Feder- oder Lorenzkraft sein. Die Bewegungsgleichung mit Zwangskräften für das Teilchen i lautet:

$$\underline{F}_i + \underline{Z}_i = \frac{d\underline{p}_i}{dt}$$

Aufstellen der Bewegungsgleichungen für Teilchensysteme unter Zwangsbedingungen:

1. Auswahl geeigneter *Koordinaten* zur Beschreibung der Bewegungen. (Koordinaten in eine Skizze eintragen!)
2. Bestimmung aller *Kräfte*: Neben den eingepprägten Kräften

$$\underline{F}_i = \underline{F}_i^{(a)} + \sum_{\substack{j=1 \\ j \neq i}}^N \underline{F}_{ji}$$

sind die notwendigen Zwangskräfte \underline{Z}_i anzusetzen, um Zwangsbedingungen zu berücksichtigen. (Kräfte in Skizze eintragen!)

3. Aus den Trajektorien \underline{x}_i der Teilchen werden durch zweifache zeitliche Differentiation die zugehörige *Beschleunigungen* $\underline{\ddot{x}}_i$ berechnet.
4. Aufstellen der *Newtonschen Bewegungsgleichung*

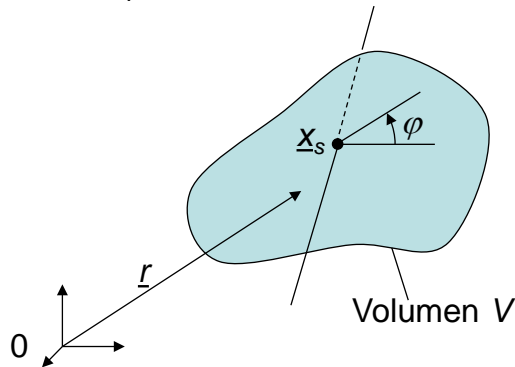
$$\underline{F}_i + \underline{Z}_i = m \cdot \underline{\ddot{x}}_i$$

für alle Teilchen $i = 1, \dots, N$.

5. Man erhält somit ein System von gekoppelten, i.A. nichtlinearen Differentialgleichungen. Aus deren *Lösung* können der Zeitverlauf der „freien“ Bewegungen im System sowie der Zeitverlauf der Zwangskräfte bestimmt werden.

Erweiterung auf Starrkörper

Ein *starrer Körper* ist ein physikalisches Modell eines nicht verformbaren Körpers. Hierfür erfolgt eine Idealisierung durch die Annahme, dass zwei beliebige Punkte des Körpers unabhängig von äußeren Kräften immer den gleichen Abstand zueinander besitzen. [aus: Wikipedia]



Integriert man die Dichte $\rho(\underline{r})$ über das Volumen des Starrkörpers, erhält man seine *Gesamtmasse*

$$m = \iiint_V \rho(\underline{r}) dV$$

Für den Ortsvektor des *Massenmittelpunktes* (= Schwerpunkt bei konstanter Gravitation) gilt:

$$\underline{x}_S = \frac{1}{m} \iiint_V \rho(\underline{r}) \underline{r} dV$$

Bei einer Bewegung des Starrkörpers um eine raumfeste Achse (s. Bild), gilt im Schwerpunkt

- die translatorische Bewegungsgleichung: $m \ddot{\underline{x}}_S(t) = \underline{F}_S(t)$
- die rotatorische Bewegungsgleichung: $J_S \ddot{\phi}(t) = \underline{M}_S(t)$

Darin ist \underline{F}_S die Summe der im Schwerpunkt angreifenden Kräfte und \underline{M}_S die Summe der im Schwerpunkt angreifenden Momente.

In der rotatorischen Bewegungsgleichung tritt das *Trägheitsmoment* J anstelle der Masse. Im Schwerpunkt S gilt

$$J_S = \iiint_V \rho(\underline{x}) x^2 dV$$

wobei \underline{x} den Abstand zum Schwerpunkt S bezeichnet. Mit Hilfe des *Steinerschen Satzes* lässt sich das Trägheitsmoment auf einen anderen Punkt P umrechnen. Mit dem Abstand s zwischen den Punkten S und P gilt:

$$J_P = J_S + m s^2$$

Für die *kinetische Energie* des Starrkörpers gilt:

$$T = \frac{1}{2} m \dot{\underline{x}}_S^2 + \frac{1}{2} J_S \dot{\phi}^2$$

Sonderfall: Gibt es einen ruhenden Punkt Q im Körper (z.B. fester Drehpunkt), so ist nach dem Steinerschen Satz:

$$T = \frac{1}{2} J_Q \dot{\phi}^2$$

Für die *potentielle Energie* z.B. eines Starrkörpers im Schwerfeld gilt mit der Höhe h_s des Schwerpunktes:

$$V = m g h_s$$

2.3 Elektrische Systeme

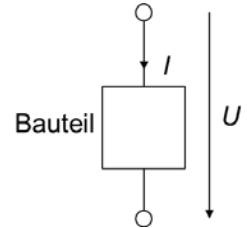
Gleichungen zur Beschreibung von linearen elektrischen Netzwerken

Strom-Spannungs-Beziehung elektrischer Bauteile

Ein elektrisches Bauteil wird über zwei Klemmen angeschlossen (Bild rechts). Es können der Strom I , der durch das Bauteil fließt, und die Spannung U , die zwischen den Klemmen auftritt, bestimmt werden.

Im Bildbereich wird die Übertragungsfunktion vom Strom $I(s)$ auf die Spannung $U(s)$ als *Impedanz* $Z(s)$ bezeichnet:

$$U(s) = Z(s) \cdot I(s)$$



Der Vergleich mit dem Ohmschen Gesetz zeigt, dass diese Übertragungsfunktion einen verallgemeinerten elektrischen Widerstand darstellt. Nachfolgende Tabelle fasst die Strom-Spannungs-Beziehungen für Widerstände, Kondensatoren und Spulen zusammen.

Element	Strom-Spannungs-Beziehung im Zeitbereich	Strom-Spannungs-Beziehung im Bildbereich
Widerstand R	$u(t) = R \cdot i(t)$	$U(s) = R \cdot I(s)$
Kondensator C	$u(t) = \frac{1}{C} \int_0^t i(\tau) d\tau$	$U(s) = \frac{1}{s C} \cdot I(s)$
Spule L	$u(t) = L \cdot \frac{di(t)}{dt}$	$U(s) = s L \cdot I(s)$

Die Struktur eines elektrischen Netzwerks kann durch einen *Graphen* beschrieben werden. Darin sind die elektrischen Bauteile die *Kanten*, die *Knoten* entstehen durch die Verbindung der Bauteilklemmen durch Leitungen und besitzen gleiches Potential. Innerhalb des elektrischen Netzwerks gelten die beiden folgenden Gesetze:

Kirchhoffsche Spannungsregel (Schleifenregel)

Die Summe der Spannungen entlang eines beliebigen geschlossenen Weges (Schleife) ist Null. Dabei sind die durch willkürliche Zählrichtungen festzulegenden Spannungen mit unterschiedlichem Vorzeichen einzusetzen, je nachdem, ob sie in Richtung oder in Gegenrichtung zum willkürlichen Umlaufsinn liegen.

$$\sum_j u_j(t) = 0$$

Kirchhoffsche Stromregel (Knotenregel)

Die Summe der Ströme in einem Verzweigungspunkt (Knoten) ist Null. Dabei sind die durch willkürliche Zählrichtung festzulegenden Ströme mit unterschiedlichem Vorzeichen einzusetzen, je nachdem, ob sie zu- oder abfließen.

$$\sum_j i_j(t) = 0$$

Maschenstrom-Analyse von linearen elektrischen Netzwerken

1. Man wandle ggf. *Stromquellen* in Spannungsquellen um.
2. Man zeichne den *Graphen* des Netzwerkes.
Man wähle einen *Baum* des Graphen aus.

Definitionen:

Die *Knoten* entsprechen Leitungen; die *Kanten* entsprechenden Bauelementen. Ein geschlossener Weg im Graph wird als *Schleife* bezeichnet.

Ein *Baum* verbindet alle Knoten des Graphens, so dass keine Schleifen entstehen.

Für einen Graphen kann es mehrere Bäume geben.

3. Man führe durch jede Kante, die nicht zum Baum gehört, und der dazugehörigen Schleife einen *Schleifenstrom* I_i mit $i=1, \dots, n$ beliebiger Richtung ein.
4. Aufstellen des *Gleichungssystems im Bildbereich*:

$$\underbrace{\begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{nn} \end{bmatrix}}_{\underline{Z}} \cdot \underbrace{\begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_n \end{bmatrix}}_{\underline{I}} = \underbrace{\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}}_{\underline{U}}$$

- \underline{Z} ist darin die Impedanzmatrix des Systems mit den Nebendiagonalelementen Z_{ik} und den Hauptdiagonalelementen Z_{ii} :

Z_{ik} : von Schleifenstrom I_i und I_k gemeinsam durchflossene Impedanz, bei gleicher Flussrichtung mit positivem, andernfalls mit negativem Vorzeichen einzusetzen.

Z_{ii} : Summe aller Impedanzen in Schleife i

- \underline{I} ist der Vektor der Schleifenströme:

I_i : Schleifenstrom der Schleife i (mit $i=1, \dots, n$)

- \underline{U} ist der Vektor der eingprägten Schleifenspannungen:

U_i : Summe der unabhängigen (eingprägten) Quellenspannungen in der Schleife i (gegen die Richtung des Schleifenstroms I_i positiv)

(Dual dazu kann auch die *Knotenpotential-Analyse* verwendet werden.)

2.4 Analogiebetrachtungen

Die Analyse von Systemen mit konzentrierten Elementen (Bauteilen) in einem Netzwerk, wie sie im letzten Abschnitt für elektrische System eingeführt wurde, kann verallgemeinert werden und erlaubt dann eine domänenübergreifende Analyse dynamischer Systeme. Die Methode wird in [Fabien: Analytical System Dynamics] detailliert vorgestellt.

Zur generalisierten Beschreibung sind zwei Paare von Größen erforderlich:

1. Kinematische Variablen

Sie charakterisieren die sichtbare Wirkung (die verallgemeinerte Geometrie) einer dynamischen Bewegung. Dazu führt man die *Verschiebung* $x(t)$ und ihre zeitliche Ableitung, den *Fluss* $v(t)$ ein. Es gilt also

$$v(t) = \frac{dx(t)}{dt} \quad \text{oder} \quad x(t) = \int_{\tau=0}^t v(\tau) d\tau.$$

Im elektrischen System entspricht die Verschiebung $x(t)$ der elektrischen Ladung $Q(t)$, die zeitliche Ableitung ist der Strom $i(t)$ entsprechend dem Fluss $v(t)$. In einem translatorischen mechanischen System ist die Verschiebung eine Position $x(t)$ und der Fluss entspricht der Geschwindigkeit $v(t)$.

2. Kinetische Variablen

Sie beschreiben die nicht sichtbaren Ursachen einer dynamischen Bewegung. Dazu wird der verallgemeinerte Impuls $p(t)$ und seine zeitliche Ableitung, die verallgemeinerte Kraft $F(t)$ eingeführt:

$$F(t) = \frac{dp(t)}{dt} \quad \text{oder} \quad p(t) = \int_{\tau=0}^t F(\tau) d\tau.$$

Dies kann direkt als translatorisches mechanisches System interpretiert werden. Im elektrischen System entspricht die verallgemeinerte Kraft der elektrischen Spannung $u(t)$ und der verallgemeinerte Kraftfluss dem (negativen) magnetischen Fluss $\phi(t)$, so dass sich das Induktionsgesetz ergibt.

Größen in verschiedene Domänen

Domäne	kinematische Größen		kinetische Größen	
	Verschiebung $x(t)$	Fluss $v(t)$	verallg. Impuls $p(t)$	verallg. Kraft $F(t)$
elektrisch	el. Ladung $Q(t)$ in C=As	el. Strom $i(t)$ in A	neg. magn. Fluss – $\phi(t)$ in W=Vs	el. Spannung $u(t)$ in V
mechanisch, translatorisch	Position $x(t)$ in m	Geschwindigkeit $v(t)$ in m/s	Impuls $p(t)$ in Ns	Kraft $F(t)$ in N
mechanisch, rotatorisch	Winkel $\phi(t)$ in rad	Winkelgeschw. $\omega(t)$ in rad/s	Drehimpuls $L(t)$ in Nms	Drehmoment $M(t)$ in Nm
pneumatisch	Volumen $V(t)$ in m ³	Volumenstrom $q(t)$ in m ³ /s	Druckimpuls $I(t)$ in Ns/m ²	Druck $p(t)$ in Pa= N/m ²
thermisch	Wärme $Q(t)$ in J	Wärmestrom $q(t)$ in W=J/s	–	Temperatur $T(t)$ in K, °C

Leistung und Energie

Führt eine ursächliche verallgemeinerte Kraft $F(t)$ in der Wirkung zu einer (kleinen) Verschiebung $\partial x(t)$, so wird die zugehörige Arbeit

$$\partial W(t) = F(t) \partial x(t)$$

Verrichtet. Bezieht man das auf die Zeit, erhält man eine Leistung:

$$P(t) = \frac{dW(t)}{dt} = F(t) \frac{dx(t)}{dt} = \frac{dp(t)}{dt} v(t) = F(t) v(t)$$

In allen Domänen ist die Leistung (z.B. in Nm/s) also das Produkt aus verallgemeinerter Kraft und Fluss!

Die Energie ist das zeitliche Integral über die Leistung, also:

$$E(t) = \int_{\tau=0}^t F(\tau) v(\tau) d\tau = \int_{C_q} F(t) dx(t) = \int_{C_p} v(t) dp(t)$$

Nun können verallgemeinerte Bauteile definiert werden und in den jeweiligen Domänen interpretiert werden:

1. Verallgemeinerte Induktivität

Dieses Bauteil verknüpft kinetische und kinematische Größen, in dem es eine Beziehung vom Impuls p zum Fluss v zu definiert:

$$v = \phi_I^{-1}(p)$$

Ist der Impuls p von Null verschieden, speichert das Bauteil Energie. Diese wird als *kinetische Energie* T bezeichnet, es gilt:

$$T(t) = \int_{C_p} v(t) dp(t) = \int_0^p \phi_I^{-1}(p) dp$$

In der elektrischen Domäne entspricht das Bauteil einer *Spule*, und mit der Induktivität L gilt entsprechend $p = \phi_I(v)$:

$$-\phi(t) = L i(t) \quad \text{oder zeitlich differenziert} \quad u(t) = L \frac{di(t)}{dt}$$

Für translatorische mechanische Systeme ist die Entsprechung ein *bewegtes Teilchen* mit der Masse m und es gilt:

$$p(t) = m v(t) \quad \text{oder zeitlich differenziert} \quad F(t) = m \frac{dv(t)}{dt}$$

Weitere Domänen sind in der nachfolgenden Tabelle zusammengefasst.

2. Verallgemeinerte Kapazität

Auch dieses Bauteil verknüpft kinetische und kinematische Größen, nun aber durch eine Beziehung von der verallgemeinerten Kraft F zur Verschiebung x :

$$x = \phi_C^{-1}(F)$$

Ist die verallgemeinerte Verschiebung x von Null verschieden, speichert das Bauteil Energie. Diese wird als *potentielle Energie oder Potential* V bezeichnet, es gilt:

$$V(t) = \int_{C_p} F(t) dx(t) = \int_0^p \phi_C(x) dx$$

In der elektrischen Domäne entspricht das Bauteil einem *Kondensator*, und mit der Kapazität C gilt entsprechend $F = \phi_C(x)$:

$$u(t) = \frac{1}{C} Q(t) \quad \text{oder zeitlich differenziert} \quad \frac{u(t)}{dt} = \frac{1}{C} i(t)$$

Für translatorische mechanische Systeme ist die Entsprechung eine *Feder* mit Federkonstante c und es gilt für die Federkraft

$$F(t) = c x(t) \quad \text{oder zeitlich differenziert} \quad p(t) = c \frac{dx(t)}{dt}.$$

Weitere Domänen sind in der nachfolgenden Tabelle zusammengefasst.

3. Verallgemeinerter Widerstand

Auch dieses Bauteil verknüpft kinetische und kinematische Größen, nun aber verallgemeinerte Kraft F und Fluss v . Dieses Bauteil speichert keine Energie, sondern charakterisiert verlustbehaftete (dissipative) Prozesse. Es gilt:

$$v = \phi_R^{-1}(F)$$

Die im Widerstand entstehende *Verlustleistung* (dissipative Leistung) ist mit $F = \phi_R(v)$:

$$D(v) = \int_0^v F dv = \int_0^v \phi_R(v) dv$$

In der elektrischen Domäne entspricht das Bauteil einem *ohmschen Widerstand*, und mit dem Widerstand R gilt entsprechend $F = \phi_R(v)$ das Ohmsche Gesetz:

$$u(t) = R i(t)$$

Für translatorische mechanische wird ein *Dämpfer* modelliert. Mit der Dämpfungskonstante d gilt:

$$F(t) = d v(t)$$

Weitere Domänen sind in der nachfolgenden Tabelle zusammengefasst.

4. Ideale Umsetzer und Umwandler

Umsetzer erhöhen den Fluss v um ein Übersetzungsverhältnis \ddot{u} , während die verallgemeinerte Kraft F um \ddot{u} reduziert wird:

$$v_2(t) = \ddot{u} v_1(t) \quad \text{und} \quad F_2(t) = \frac{1}{\ddot{u}} F_1(t)$$

Die Leistung ist das Produkt beider Größen und bleibt daher bei der Umsetzung mit einem idealen Umsetzer konstant.

In einem translatorisch mechanischen System kann ein Umsetzer durch einen *Hebel* realisiert werden. Im elektrischen System entspricht dies einem (verlustfreien) *Transformator*. Weitere Domänen sind in der nachfolgenden Tabelle zu finden.

Umwandler arbeiten ähnlich, aber über die Domänen hinweg, z.B.:

- Eine Spindel wandelt eine rotatorische Bewegung $\{\omega, M\}$ in eine translatorische Bewegung $\{v, F\}$ um, wobei die Spindelsteigung das Übersetzungsverhältnis \ddot{u} definiert.
- Ein (verlustfreier) Gleichstrommotor wandelt eine elektrische Bewegung $\{i, u\}$ in eine rotatorische mechanische Bewegung $\{\omega, M\}$ um, wobei die Maschinenkonstante wiederum das Übersetzungsverhältnis \ddot{u} definiert.

5. Ideale Quellen

Ideale Quellen geben eine der Größen aus $\{x, v, p, F\}$ vor. Wird eine kinetische Größe vorgegeben („eingepägt“), stellt sich als Reaktion eine kinematische Größe ein, und umgekehrt.

Neben ungesteuerten Quellen, die eine Größe konstant vorgeben, können auch gesteuerte Quellen definiert werden, mit denen insbesondere nichtlineare Zusammenhänge modelliert werden können (z.B. Diode oder Haftreibung).

Bauteile in verschiedenen Domänen

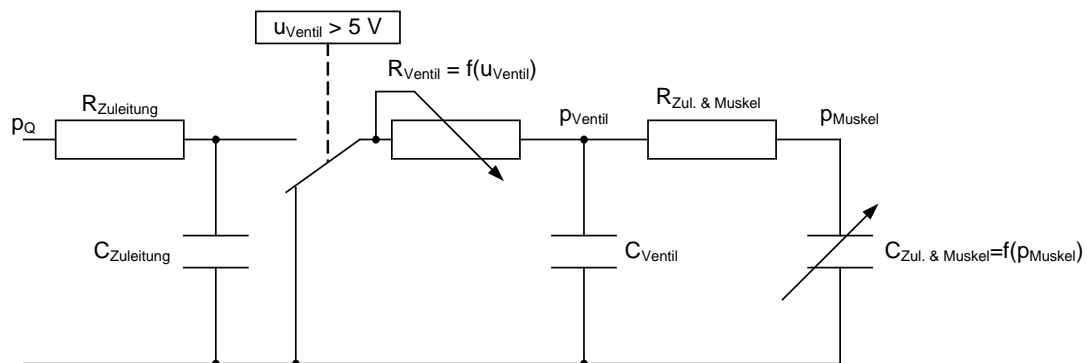
Domäne	verallg. Induktivität $p = \phi_I(v)$	verallg. Kapazität $F = \phi_C(x)$	verallg. Widerstand $F = \phi_R(v)$	Umsetzer $v_2(t) = \ddot{u} v_1(t)$ $F_2(t) = \frac{1}{\ddot{u}} F_1(t)$
elektrisch	Spule mit Induktivität L : $-\dot{\phi}(t) = L i(t)$	Kondensator mit Kapazität C : $u(t) = \frac{1}{C} Q(t)$	Ohmscher Widerstand R : $u(t) = R i(t)$	Transformator (für Wechselströme)
mechanisch, translatorisch	Teilchen mit Masse m : $p(t) = m v(t)$	transl. Feder mit Federsteifigkeit c : $F(t) = c x(t)$	transl. Dämpfer mit D.-konst. d : $F(t) = d v(t)$	Hebel (\ddot{u} aus Verhältnis der Hebellängen)
mechanisch, rotatorisch	Starrkörper mit Trägheitsmoment J : $L(t) = J \omega(t)$	rot. Feder mit Federsteifigkeit c : $M(t) = c \rho(t)$	rot. Dämpfer mit D.-konst. d : $M(t) = d \omega(t)$	Getriebe (Zahnräder mit verschiedenem Umfang)
pneumatisch	Teilchenstrom mit Fluidträgheitsmoment J_F : $\Gamma(t) = J_F q(t)$	Volumen mit Kapazität C_F : $p(t) = \frac{1}{C_F} V(t)$	pneumatischer Widerstand R_F : $p(t) = R_F q(t)$	Druckübersetzer (2 Kolben mit verschiedener Fläche)
thermisch	–	Wärmespeicher mit Wärmekapazität C_T : $T(t) = \frac{1}{C_T} Q(t)$	Wärmewiderstand R_T : $T(t) = R_T q(t)$	Wärmetauscher

Beispiel: Pneumatischer Muskel

Der im Bild gezeigte pneumatische Muskel ist ein pneumatischer Aktor: Wird im Inneren des Schlauchs der Druck erhöht, dehnt sich dieser radial aus. Dies führt – unterstützt durch die Gewebestruktur der Außenwand – zu einer Kontraktion der Enden des Muskels. Das Bauteil kann mit geringem Eigengewicht große Kontraktionskräfte (bis 1 800 N) erzeugen.

Der pneumatische Muskel wird über ein Proportionalventil an eine Druckluftversorgung angeschlossen. Um eine Druckregelung entwerfen zu können, soll ein Modell dieses Systems aufgebaut werden.

In Analogie zwischen dem pneumatischen und einem elektrischen System wird das nachstehende pneumatische Ersatzschaltbild aufgebaut. Darin entsprechen Drücke den elektrischen Spannungen und Masseströme den elektrischen Strömen; pneumatische Volumen werden durch Kondensatoren beschrieben. Unterschieden werden Muskel, Ventil und Zuleitung. Für die Entlüftung des Muskels öffnet das Ventil zur Umgebung. Im Ersatzschaltbild ist das durch den Schalter beschrieben, der Atmosphärendruck entspricht dem Massepotential:



2.5 Methode von Lagrange

Herleitung

Die *virtuelle Verrückung* eines Teilchens i wird als $\partial \underline{x}_i$ bezeichnet. Als handelt sich um eine kleine (infinitesimale) Verrückung mit beliebiger Richtung bei „festgehaltener“ Zeit. Es ist also nur eine gedachte (virtuelle) Verrückung des Teilchens.

Formal kann eine zugehörige *virtuelle Arbeit* des Teilchens i definiert werden:

$$\partial W_i = \underline{F}_i \cdot \partial \underline{x}_i$$

Diese (ebenfalls infinitesimale) virtuelle Arbeit wird also durch die virtuelle Verrückung $\partial \underline{x}_i$ verrichtet.

Das *d'Alembertsche Prinzip* ist (neben den Newtonschen Bewegungsgesetzen) ein zentrales Axiom der Mechanik. Es besagt, dass die virtuelle Arbeit der Zwangskräfte \underline{Z}_i innerhalb eines Teilchensystems mit $i=1, \dots, N$ Teilchen verschwindet:

$$\sum_{i=1}^N \underline{Z}_i \cdot \partial \underline{x}_i = 0$$

D. h. die Natur der Zwangskräfte ist derart, dass im Gesamtsystem durch sie keine virtuelle Arbeit geleistet wird.

Teilt man alle am Teilchen i angreifenden Kräfte in eingeprägte Kräfte \underline{F}_i und Zwangskräfte \underline{Z}_i auf, so kann man das Newtonsche Bewegungsgesetz schreiben als:

$$m_i \cdot \ddot{\underline{x}}_i = \underline{F}_i + \underline{Z}_i$$

Löst man das Newtonsche Bewegungsgesetz nach \underline{Z}_i auf und setzt in die vorstehende Bedingung des d'Alembertschen Prinzips ein, so erhält man als *d'Alembert-Gleichung*:

$$\sum_{i=1}^N (m_i \cdot \ddot{\underline{x}}_i - \underline{F}_i) \cdot \partial \underline{x}_i = 0$$

Die Methode von Lagrange ermöglicht es, die Bewegungsgleichungen eines Teilchensystems aufzustellen, ohne die Zwangskräfte mitzuführen. Es kann aus dem d'Alembertschen Prinzip entwickelt werden: Die Trajektorie \underline{x}_i der $i=1, \dots, N$ Teilchen soll als Funktion der generalisierten Koordinaten q_1, \dots, q_n beschrieben werden können:

$$\underline{x}_i = \underline{x}_i(q_1, \dots, q_n)$$

Für die virtuelle Verrückung gilt dann in Abhängigkeit der generalisierten Koordinaten:

$$\partial \underline{x}_i = \sum_{j=1}^n \frac{\partial \underline{x}_i}{\partial q_j} \cdot \partial q_j$$

Die Differenz der d'Alembert-Gleichung (s.o.) wird jetzt getrennt betrachtet. Zunächst der hintere Term, die virtuelle Arbeit der eingepprägten Kräfte:

$$\sum_{i=1}^N \underline{F}_i \cdot \partial \underline{x}_i = \sum_{j=1}^n \left(\sum_{i=1}^N \underline{F}_i \cdot \frac{\partial \underline{x}_i}{\partial q_j} \right) \cdot \partial q_j = \sum_{j=1}^n Q_j \cdot \partial q_j \quad \text{mit} \quad Q_j = \sum_{i=1}^N \underline{F}_i \cdot \frac{\partial \underline{x}_i}{\partial q_j}$$

Die Größen Q_j werden als *generalisierte Kräfte* bezeichnet.

Nun der vordere Term der d'Alembert-Gleichung:

$$\sum_{i=1}^N m_i \cdot \ddot{\underline{x}}_i \cdot \partial \underline{x}_i = \sum_{j=1}^n \left(\sum_{i=1}^N m_i \cdot \ddot{\underline{x}}_i \cdot \frac{\partial \underline{x}_i}{\partial \dot{q}_j} \right) \cdot \partial q_j$$

Durch die Produktregel der Differentiation lässt sich die Gleichheit mit

$$= \sum_{j=1}^n \left[\sum_{i=1}^N \frac{d}{dt} \left(m_i \cdot \dot{\underline{x}}_i \cdot \frac{\partial \underline{x}_i}{\partial \dot{q}_j} \right) - m_i \cdot \dot{\underline{x}}_i \cdot \frac{d}{dt} \frac{\partial \underline{x}_i}{\partial \dot{q}_j} \right] \cdot \partial q_j$$

zeigen. Mit $\underline{v}_i = \dot{\underline{x}}_i$ folgt:

$$\begin{aligned} &= \sum_{j=1}^n \left[\sum_{i=1}^N \frac{d}{dt} \left(m_i \cdot \underline{v}_i \cdot \frac{\partial \underline{v}_i}{\partial \dot{q}_j} \right) - m_i \cdot \underline{v}_i \cdot \frac{\partial \underline{v}_i}{\partial \dot{q}_j} \right] \cdot \partial q_j \\ &= \sum_{j=1}^n \left[\frac{d}{dt} \left(\frac{\partial}{\partial \dot{q}_j} \sum_{i=1}^N \frac{1}{2} m_i \cdot |\underline{v}_i|^2 \right) - \frac{\partial}{\partial q_j} \sum_{i=1}^N \frac{1}{2} m_i \cdot |\underline{v}_i|^2 \right] \cdot \partial q_j \end{aligned}$$

Mit der kinetischen Energie $T = \frac{1}{2} \sum_{i=1}^N m_i \cdot |\underline{v}_i|^2$ des Teilchensystems lässt sich kompakter schreiben:

$$= \sum_{j=1}^n \left[\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} \right] \cdot \partial q_j$$

Schließlich können beide Zwischenergebnisse in die d'Alembert-Gleichung eingesetzt werden:

$$= \sum_{j=1}^n \left[\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} - Q_j \right] \cdot \partial q_j = 0$$

Die virtuellen Verrückungen ∂q_j sollen unabhängig voneinander gewählt werden können, dann ist die Gleichung erfüllt, falls

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} - Q_j = 0 \quad \text{für alle } j = 1, \dots, n$$

gilt. Dies ist die *Lagrange-Gleichung* mit generalisierten Kräften. Vorausgesetzt wird, dass die Zwangsbedingungen holonom sind.

Sind in Q_j *konservative Kräfte* enthalten (z.B. Federkräfte oder Schwerkraft), so können diese aus einem Potential V bestimmt werden:

$$Q_{j,V} = - \frac{\partial V(q_1, \dots, q_n)}{\partial q_j}$$

Ebenso können in Q_j *dissipative Kräfte* (wie z.B. eine geschwindigkeitsabhängige Reibung) aus ihrer Verlustleistung D berechnet werden:

$$Q_{j,D} = - \frac{\partial D(\dot{q}_1, \dots, \dot{q}_n)}{\partial \dot{q}_j}$$

In Q_j verbleiben *eingeprägte Kräfte* (wie z.B. Stellkräfte eines Aktors), die als $Q_{j,E}$ bezeichnet werden. Es gilt:

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} + \frac{\partial V}{\partial q_j} + \frac{\partial D}{\partial \dot{q}_j} = Q_{j,E}$$

Anwendung der Methode von Lagrange

Voraussetzung: Das System besitzt nur holonome Zwangsbedingungen.

1. Wahl unabhängiger **Lagrange-Koordinaten** q_1, \dots, q_n (in Zeichnung eintragen!)

Die Zahl der Freiheitsgrade aller N Teilchen bzw. Starrkörper wird aufsummiert, dabei gilt jeweils:

	Modellierung in der Ebene	Modellierung im Raum
Teilchen	2 Freiheitsgrade	3 Freiheitsgrade
Starrkörper	3 Freiheitsgrade	6 Freiheitsgrade

Davon werden alle Zwangskräfte im System abgezogen, die durch Zwangsbedingungen entstehen (z.B. Seil, Auflage, Rollen). Die verbleibenden n Freiheitsgrade werden mit Lagrange-Koordinaten parametrisiert.

2. Bestimmung der **Ortsvektoren** $\underline{x}_1(q_1, \dots, q_n), \dots, \underline{x}_N(q_1, \dots, q_n)$ zu allen Teilchen bzw. Starrkörpern als Funktion der Lagrange-Koordinaten, sowie deren zeitliche Ableitungen $\dot{\underline{x}}_1, \dots, \dot{\underline{x}}_N$.
3. **Kinetische Energie** T des Gesamtsystems in Abhängigkeit von q_1, \dots, q_n sowie $\dot{q}_1, \dots, \dot{q}_n$ bestimmen.
4. Für alle konservativen Kräfte wird das **Potential** V in Abhängigkeit von q_1, \dots, q_n berechnet.
5. Für alle dissipativen Kräfte wird die **Verlustleistung** D in Abhängigkeit von $\dot{q}_1, \dots, \dot{q}_n$ berechnet.
6. Für die verbleibenden eingepprägten Kräfte (z.B. Stellkräfte von Aktoren) o werden hingegen die **generalisierten Kräfte**

$$\begin{aligned}
 Q_1 &= \sum_{i=1}^N \underline{E}_i \cdot \frac{\partial \underline{x}_i}{\partial q_1} \\
 &\vdots \\
 Q_n &= \sum_{i=1}^N \underline{E}_i \cdot \frac{\partial \underline{x}_i}{\partial q_n}
 \end{aligned}$$

berechnet.

7. Die n **Bewegungsgleichungen** lassen sich aus den Lagrange-Gleichungen berechnen:

$$\begin{aligned}
 \frac{d}{dt} \frac{\partial T}{\partial \dot{q}_1} - \frac{\partial T}{\partial q_1} + \frac{\partial V}{\partial q_1} + \frac{\partial D}{\partial \dot{q}_1} &= Q_{1,E} \\
 &\vdots \\
 \frac{d}{dt} \frac{\partial T}{\partial \dot{q}_n} - \frac{\partial T}{\partial q_n} + \frac{\partial V}{\partial q_n} + \frac{\partial D}{\partial \dot{q}_n} &= Q_{n,E}
 \end{aligned}$$

2.6 Modellierung im Blockschaltbild

Gerade komplexere Systeme sollten schrittweise aus Teilsystemen aufgebaut werden. Der in Abschnitt 1.3 vorgestellte Entwicklungszyklus der Modellbildung kann dann für die Teilsysteme durchlaufen werden, so dass man schrittweise verifizierte Modelle der Teilsysteme erhält, die schrittweise zum Gesamtsystem kombiniert werden. Diese Vorgehensweise wird durch die graphische Darstellung eines Blockschaltbildes unterstützt.

Vorteile einer Darstellung im Blockschaltbild:

- Anschauliche Darstellung (als Alternative zur Darstellung durch Gleichungen).
- Modelle können schrittweise implementiert und getestet werden (s.o.).
- Substrukturen unterstützen die Übersicht: Es können Teilmodelle aus mehreren Blöcken zu einem zusammengefasst und gekapselt werden. Somit können komplexe Modelle hierarchisch und auf der oberen Ebene übersichtlich gestaltet werden.
- Im Rahmen von Projekten können Blöcke schnell ausgetauscht werden.
- Mit Software-Werkzeugen ist eine numerische Simulation des Modells möglich.
- Mit Software-Werkzeugen kann z.B. C-Code automatisch generiert werden.

Klassifikation:

Objektorientierte Modellierung

Im Blockschaltbild sind die Knoten physikalische Einheiten (z.B. Motor, Getriebe, Drehzahlsensor, ...) und die Kanten stellen entsprechend physikalische Verbindungen her, über die i.d.R. eine Leistung (als Produkt aus Fluss- und Potential (s. Abschnitt 2.4) ausgetauscht wird.

Diese Form der Darstellung kommt den physikalischen Gegebenheiten nahe.

Software: z.B. Modelica/Dymola

Kausale Modellierung

Im Blockschaltbild bilden die Knoten Übertragungsfunktionen, welche aufgrund der gegebenen Eingänge die Ausgänge des Blocks bestimmen. Die gerichteten Kanten beschreiben den Signalfluss.

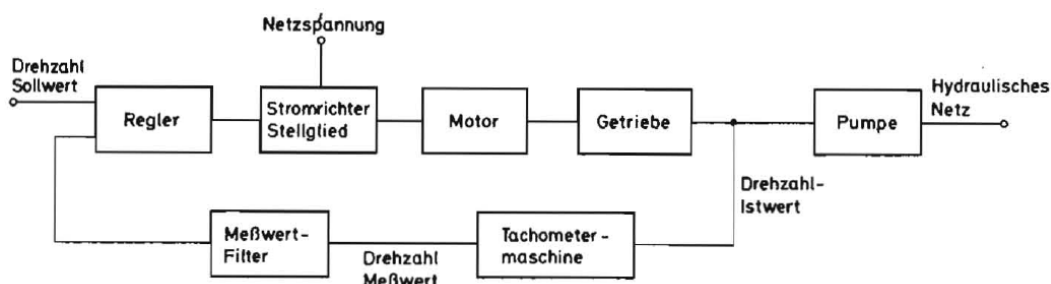
Diese Form der Darstellung eignet sich zur Analyse von Systemen.

Software: z.B. Matlab/Simulink

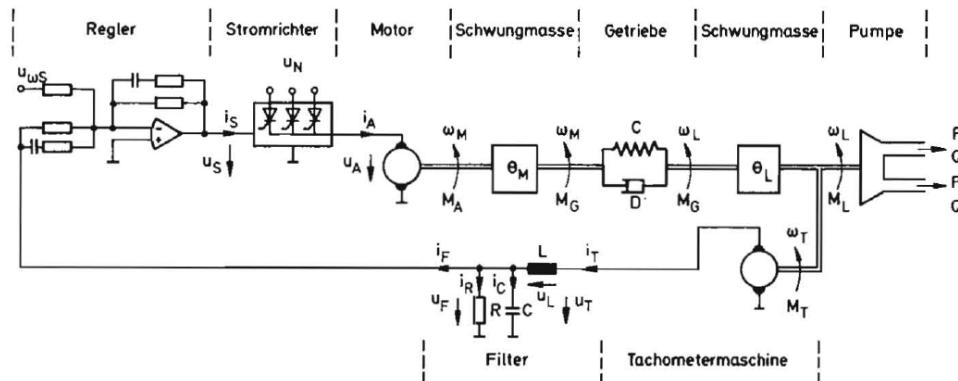
Beispiel: Modellierung eines Pumpenaggregats im Blockschaltbild

aus Föllinger: Regelungstechnik

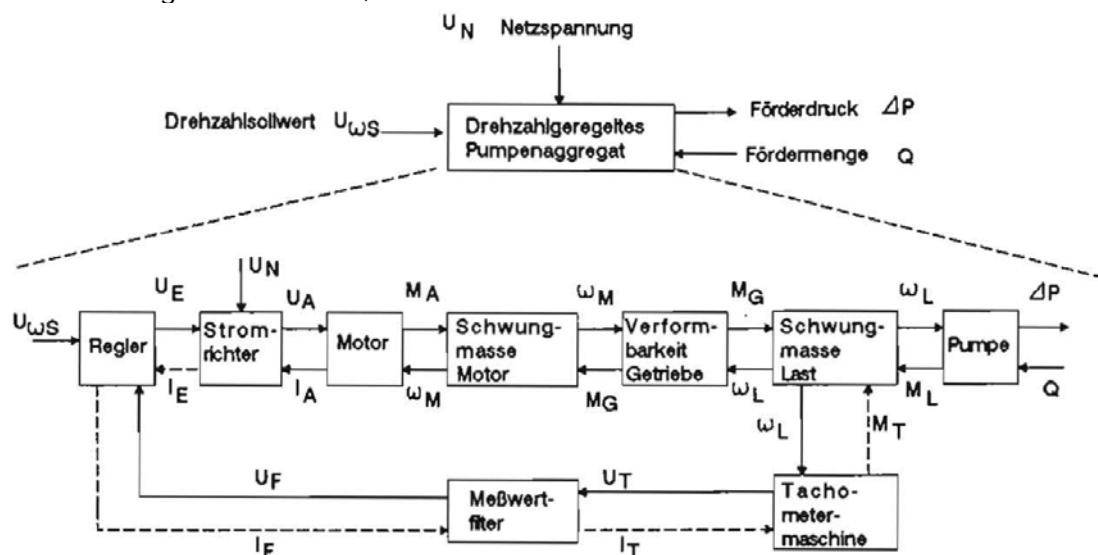
Nachfolgend dargestellt ist ein Blockschaltbild der Geräte, das eine erste Übersicht über das Gesamtsystem gibt und die Wechselwirkung zwischen den Teilsystemen beschreibt:



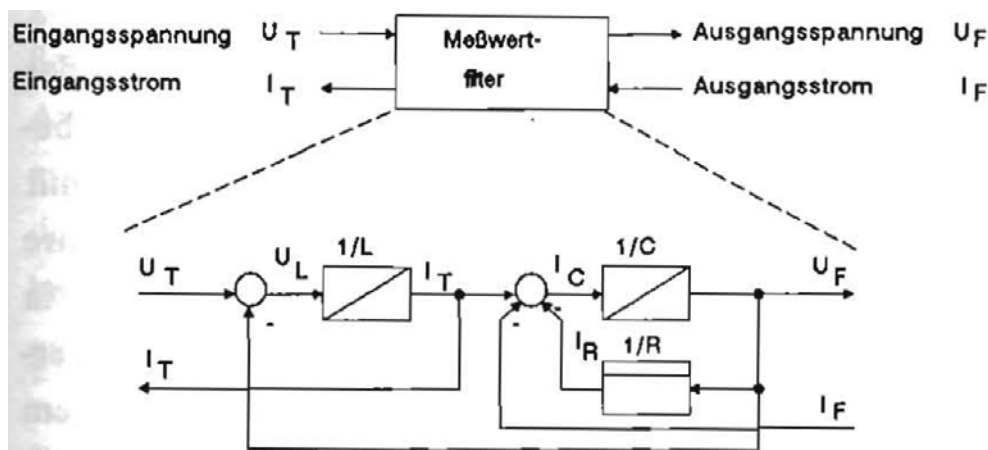
Eine Detaillierung führt zu folgendem Schaltschema, in dem die Schnittstellen bereits durch physikalische Größen spezifiziert sind. Dies ist ein Beispiel einer *objektorientierten Modellierung*.



Aus der Detaillierung des nachfolgenden Funktionsplans, in dem kausale Beziehungen durch Pfeile dargestellt werden,



gelangt man zum *Blockschaltbild* (oder *Signalflussplan*), das eine *kausale Modellierung* erlaubt:



2.7 Zustandsraumdarstellung

Der Zustand eines dynamischen Systems wird zum Zeitpunkt t durch eine Reihe von Zustandsgrößen (= Zuständen) $x_1(t), \dots, x_n(t)$ eindeutig beschrieben. Diese Zustandsgrößen werden im Zustandsvektor $\underline{x}(t)$ zusammengefasst.

Das dynamische Verhalten des Systems wird durch die Zustandsdifferentialgleichung definiert: Sie beschreibt die zeitliche Änderung des Zustandsvektors $\dot{\underline{x}}(t)$ als Funktion des aktuellen Zustands $\underline{x}(t)$ sowie eines Eingangsvektors $\underline{u}(t)$. Es handelt sich um eine vektorielle Differentialgleichung erster Ordnung.

Die (messbaren) Ausgangsgrößen werden im Ausgangsvektor $\underline{y}(t)$ zusammengefasst. Die Ausgangsgleichung bestimmt $\underline{y}(t)$ auf Basis des aktuellen Zustands $\underline{x}(t)$ sowie ggf. des Eingangsvektors $\underline{u}(t)$. Die Ausgangsgleichung ist eine algebraische Gleichung.

Systemdarstellung

Zustandsdifferentialgleichung:	$\dot{\underline{x}}(t) = \underline{f}(\underline{x}(t), \underline{u}(t))$
Ausgangsgleichung:	$\underline{y}(t) = \underline{g}(\underline{x}(t), \underline{u}(t))$

mit folgenden Größen:

Zustandsvektor	$\underline{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}$	mit n : Anzahl der Zustände
Eingangsvektor	$\underline{u}(t) = \begin{bmatrix} u_1(t) \\ \vdots \\ u_m(t) \end{bmatrix}$	mit m : Anzahl der Eingänge
Ausgangsvektor	$\underline{y}(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_p(t) \end{bmatrix}$	mit p : Anzahl der Ausgänge

Umwandlung

Erster Schritt ist die geeignete Wahl der Zustände in \underline{x} :

Bilanzgleichungen (Abschnitt 2.1)	Wählen Sie die Erhaltungsgrößen als Zustände.
Mechanische Systeme (Abschnitt 2.2)	Die Bewegungsgleichung ist eine DGL 2. Ordnung. Um daraus zwei DGLen 1. Ordnung zu gewinnen, wird neben der Position (bzw. Winkel) auch die Geschwindigkeit (bzw. Winkelgeschwindigkeit) als Zustand eingeführt.
Elektrische Systeme (Abschnitt 2.3)	Kondensator-Spannungen und Spulen-Ströme werden (als „Speicher“ eines elektrischen Netzwerkes) als Zustände eingeführt.
Pneumatische, thermische Systeme (Abschnitt 2.4)	Der Druck bzw. die Temperatur werden als Zustände eingeführt.
Blockschaltbilder (Abschnitt 2.5)	Die Ausgänge der „Speicher“ im Blockschaltbild (z.B. Integratoren) werden als Zustände eingeführt.

Anschließend sind die DGLen nach der ersten Ableitung dieser Zustände umzustellen und vektoriell darzustellen.

3 Modellvereinfachung

3.1 Linearisierung

Viele Methoden für die Systemanalyse oder für den Reglerentwurf stehen nur für lineare Systeme zur Verfügung. Daher kann die Vereinfachung eines nichtlinearen Modells zu einem linearen Modell notwendig sein. Dazu muss zuerst ein Arbeitspunkt festgelegt werden, in dem anschließend die Linearisierung durchgeführt wird.

Arbeitspunkt

Technische Anlagen und überhaupt dynamische Systeme sollen häufig in einem stationären Zustand betrieben werden, bei dem die Ausgangsgrößen des Systems auf ihrem Sollwert sind. Diesen gewünschten Betriebszustand bezeichnet man gewöhnlich als Arbeitspunkt.

Der Arbeitspunkt ist somit ein spezieller stationärer Zustand, bei dem die Ausgangsgrößen ihre Sollwerte annehmen. [Föllinger 1994]

Linearisierung

Jede zeitveränderliche Größe $x(t)$ des Systems kann ersetzt werden durch ihren entsprechenden Arbeitspunkt x_0 zuzüglich einer Abweichung $\Delta x(t)$ von diesem Arbeitspunkt:

$$x(t) = x_0 + \Delta x(t)$$

Arbeitet das System in der Nähe des gewünschten Betriebszustandes, kann erwartet werden, dass die Abweichung $\Delta x(t)$ vom Arbeitspunkt klein ist.

Wird das System im Arbeitspunkt linearisiert, sind die Abweichungen zwischen dem nichtlinearen und dem linearen Modell in der Umgebung des Arbeitspunktes klein.

Linearisierung einer nichtlinearen Kennlinie

Gegeben ist eine nichtlineare Kennlinie:

$$y(t) = F(u(t))$$

Übergang zu Abweichungen vom Arbeitspunkt:

$$y_0 + \Delta y(t) = F(u_0 + \Delta u(t))$$

Entwicklung in eine Taylorreihe bis zum linearen Glied:

$$y_0 + \Delta y(t) = F(u_0) + \left. \frac{\partial f(u)}{\partial u} \right|_{u_0} \cdot \Delta u(t) + (\text{höhere Terme})$$

Mit $y_0 = F(u_0)$ und unter Vernachlässigung höherer Terme gilt also:

$$\Delta y(t) = \left. \frac{\partial f(u)}{\partial u} \right|_{u_0} \cdot \Delta u(t)$$

Linearisierung einer nichtlinearen Zustandsraum-Darstellung

In analoger Weise kann auch eine nichtlineare Zustandsraum-Darstellung (Abschnitt 2.6) linearisiert werden. Man wählt einen durch $\underline{x}_0, \underline{u}_0$ definierten Arbeitspunkt, der eine Ruhelage mit

$$\underline{0} = \underline{f}(\underline{x}_0, \underline{u}_0)$$

darstellt. Für den Ausgang folgt dann als Arbeitspunkt $\underline{y}_0 = \underline{g}(\underline{x}_0, \underline{u}_0)$.

Mit den Abweichungen vom Arbeitspunkt

$$\underline{x}(t) = \underline{x}_0 + \underline{\Delta x}(t), \quad \underline{u}(t) = \underline{u}_0 + \underline{\Delta u}(t), \quad \underline{y}(t) = \underline{y}_0 + \underline{\Delta y}(t)$$

werden die Zustandsdifferentialgleichung und die Ausgangsgleichung in Taylor-Reihen entwickelt und die Entwicklung nach dem linearen Glied abgebrochen:

$$\begin{aligned} \underline{\Delta \dot{x}}(t) &= \underbrace{\underline{f}(\underline{x}_0, \underline{u}_0)}_{\underline{0}} + \underbrace{\left. \frac{\partial \underline{f}(\underline{x}, \underline{u})}{\partial \underline{x}} \right|_{\underline{x}_0, \underline{u}_0}}_{\underline{A}} \underline{\Delta x}(t) + \underbrace{\left. \frac{\partial \underline{f}(\underline{x}, \underline{u})}{\partial \underline{u}} \right|_{\underline{x}_0, \underline{u}_0}}_{\underline{B}} \underline{\Delta u}(t) \\ \underline{y}_0 + \underline{\Delta y}(t) &= \underbrace{\underline{g}(\underline{x}_0, \underline{u}_0)}_{\underline{y}_0} + \underbrace{\left. \frac{\partial \underline{g}(\underline{x}, \underline{u})}{\partial \underline{x}} \right|_{\underline{x}_0, \underline{u}_0}}_{\underline{C}} \underline{\Delta x}(t) + \underbrace{\left. \frac{\partial \underline{g}(\underline{x}, \underline{u})}{\partial \underline{u}} \right|_{\underline{x}_0, \underline{u}_0}}_{\underline{D}} \underline{\Delta u}(t) \end{aligned}$$

Man erhält die Zustandsraum-Darstellung für lineare Systeme:

Zustandsdifferentialgleichung:	$\underline{\Delta \dot{x}}(t) = \underline{A} \underline{\Delta x}(t) + \underline{B} \underline{\Delta u}(t)$
Ausgangsgleichung:	$\underline{\Delta y}(t) = \underline{C} \underline{\Delta x}(t) + \underline{D} \underline{\Delta u}(t)$

Mit folgenden Matrizen:

$$\begin{aligned} \text{Systemmatrix } \underline{A} &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{nn} \end{bmatrix}, & \text{Eingangsmatrix } \underline{B} &= \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ b_{21} & \cdots & b_{2m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix} \\ \text{Ausgangsmatrix } \underline{C} &= \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pn} \end{bmatrix}, & \text{Durchgangsmatrix } \underline{D} &= \begin{bmatrix} d_{11} & \cdots & d_{1m} \\ \vdots & \ddots & \vdots \\ d_{p1} & \cdots & d_{pm} \end{bmatrix} \end{aligned}$$

3.2 Zeitliche Diskretisierung

Ausgehend vom physikalischen System werden dynamische Systeme bisher durch Differentialgleichungen in der (kontinuierlichen) Zeit t beschrieben. Die Zustandsdifferentialgleichung lautet beispielsweise

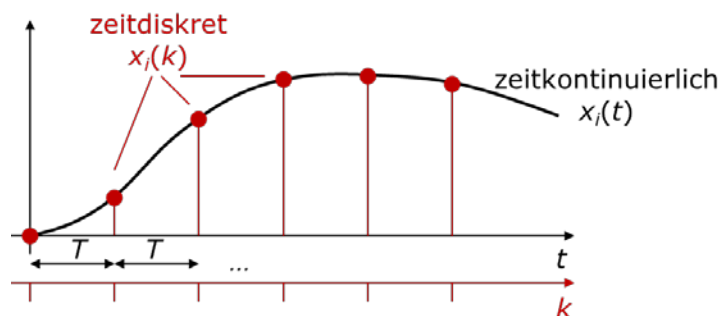
$$\dot{\underline{x}}(t) = \underline{f}(\underline{x}(t), \underline{u}(t)) \quad \text{für nichtlineare Systeme bzw.}$$

$$\dot{\underline{x}}(t) = \underline{A} \underline{x}(t) + \underline{B} \underline{u}(t) \quad \text{für lineare Systeme.}$$

Mit Blick auf die Simulation des Systems oder den Reglerentwurf kann eine *zeitdiskrete Systembeschreibung* vorteilhaft sein. Dazu wird das zeitkontinuierliche System äquidistant mit einer *Abtastzeit* T abgetastet und eine neue Notation für das zeitdiskrete System eingeführt:

$$\underline{x}(k) := \underline{x}(t = k T)$$

Darin ist $k = 0, 1, 2, \dots$ ein Zähler für die diskreten Zeitschritte.



Wünschenswert wäre nun eine zeitdiskrete Systembeschreibung in Form einer *Zustandsdifferenzengleichung*:

$$\underline{x}(k+1) = \underline{f}_D(\underline{x}(k), \underline{u}(k)) \quad \text{für nichtlineare Systeme}$$

$$\underline{x}(k+1) = \underline{A}_D \underline{x}(k) + \underline{B}_D \underline{u}(k) \quad \text{für lineare Systeme}$$

Wie kann ein zeitkontinuierliches in ein zeitdiskretes Modell umgewandelt werden?

Für nichtlineare Systeme

Für nichtlineare Systeme ist dies i.A. nur näherungsweise möglich. Die Näherung des Zustandsvektors $\underline{x}(k)$ wird mit $\tilde{\underline{x}}(k)$ bezeichnet.

Ein einfacher Ansatz ist das *Euler-Verfahren*. Ausgehend von einem bekannten Anfangszustand $\tilde{\underline{x}}(0)$ des Systems kann die Näherung des Zustands rekursiv bestimmt werden:

$$\tilde{\underline{x}}(k+1) = \tilde{\underline{x}}(k) + T \cdot \underline{f}(\tilde{\underline{x}}(k), \underline{u}(k)) \quad \text{für } k = 1, 2, 3, \dots$$

Damit ist also eine zeitdiskrete Zustandsdifferenzengleichung mit

$$\underline{f}_D(\underline{x}(k), \underline{u}(k)) = \underline{x}(k) + T \cdot \underline{f}(\underline{x}(k), \underline{u}(k))$$

gefunden. Verfahren höherer Ordnung und weiterführende Methoden werden im Teil C „Simulation“ dieser Vorlesung vorgestellt.

Für lineare Systeme

Für lineare Systeme kann die Zustandsdifferentialgleichung direkt gelöst werden. Damit kann der Übergang von Zustand $\underline{x}(k)$ zu $\underline{x}(k+1)$ exakt berechnet werden:

$$\underline{x}(k+1) = e^{\underline{A}T} \underline{x}(k) + \int_0^T e^{\underline{A}(T-\tau)} \underline{B} d\tau \cdot \underline{u}(k)$$

Somit ist eine zeitdiskrete Zustandsdifferenzengleichung mit

$$\underline{A}_D = e^{\underline{A}T} \text{ und } \underline{B}_D = \int_0^T e^{\underline{A}(T-\tau)} \underline{B} d\tau$$

gefunden.

3.3 Diskretisierung der Zustände

Bisher wurden die Zustände eines Systems als wertkontinuierlich angenommen. Im nebenstehend skizzierten Beispiel eines Fahrstuhls könnte so die Position des Fahrstuhls durch die Höhe x in Metern spezifiziert werden.

Zu einer Beschreibung durch diskrete Zustände gelangt man durch *Klassifikation*. Im Beispiel könnte mit der Klassifikation

$$\begin{aligned} z = \text{„EG“} & \quad \text{falls } 0 \text{ m} \leq x < 3 \text{ m} \\ z = \text{„1. OG“} & \quad \text{falls } 3 \text{ m} \leq x < 6 \text{ m} \\ z = \text{„2. OG“} & \quad \text{falls } 6 \text{ m} \leq x \end{aligned}$$

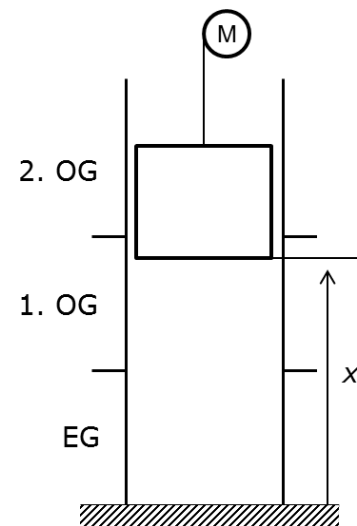
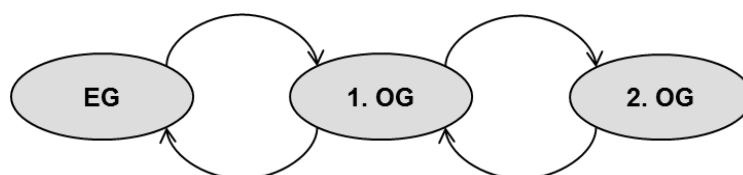
die Position des Fahrstuhls nun durch die Angabe eines diskreten Stockwerks $z = \{\text{„EG“}, \text{„1. OG“}, \text{„2. OG“}\}$ erfolgen.

Hier beschreibt w also einen diskreten Zustand. Den Wechsel von einem diskreten Zustand in einen anderen bezeichnet man als *Ereignis*, so modellierte System daher als *ereignisdiskrete Systeme*.

Eine analoge Zustandsraum-Darstellung ist auch für ereignisdiskrete Systeme möglich:

$z(k+1) = G(z(k), v(k))$	Zustandsüberföhrungsfunktion
$w(k) = H(z(k), v(k))$	Ausgabefunktion

Mit dem diskreten Zustand $z(k)$, dem diskreten Eingang $v(k)$ und dem diskreten Ausgang $w(k)$ stellt dies einen *Mealy-Automaten* dar. Eine Spezifikation eines Automaten ist z.B. durch einen Automatengraphen möglich:

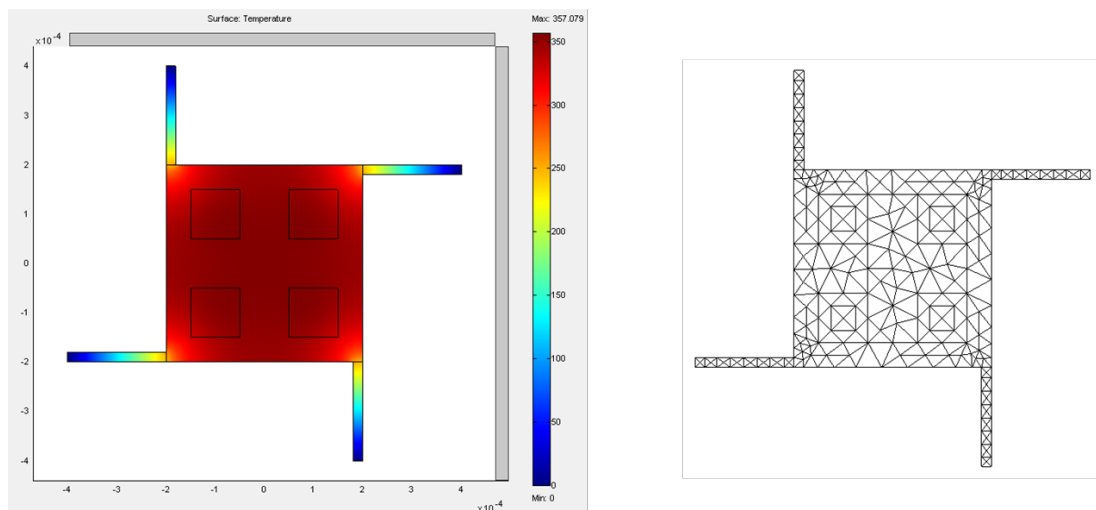


3.4 Ortsdiskretisierung

Eine weitere, bisher nicht betrachtete Klasse von Systemen umfassen Feldprobleme: So werden z.B. thermische, elastomechanische oder elektromagnetische Feldprobleme durch *partielle Differentialgleichungen* beschrieben, die neben der zeitlichen auch räumliche Ableitungen beinhalten. Man nennt diese Systeme auch als *verteilt-parametrisch*.

Solche Systeme besitzen im Prinzip unendlich viele Freiheitsgrade. Durch eine *Ortsdiskretisierung* kann eine Approximation durch ein System mit endlich vielen Freiheitsgraden (und entsprechend endlich vielen Zuständen) gefunden werden.

Nachstehendes Beispiel zeigt das Temperaturfeld eines Bauteils (links) und die Ortsdiskretisierung durch Finite Elemente (rechts). Das System besteht aus 913 Dreieckselementen. Für jedes Element wird die Temperatur als Zustand eingeführt, so dass man ein Zustandsraum-Modell der Ordnung $n = 913$ erhält.



3.5 Ordnungsreduktion

Komplexe Systeme hoher Ordnung – wie sie z.B. durch eine Ortsdiskretisierung (Abschnitt 3.4) entstehen können – sind ungeeignet für Analyse, Reglerentwurf und echtzeitfähige Implementierungen.

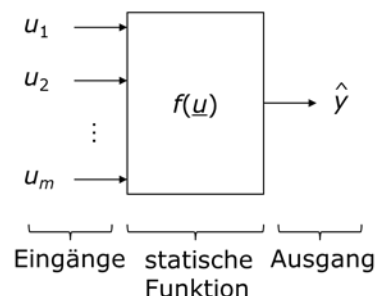
Verfahren zur *Ordnungsreduktion* haben zum Ziel, das originale System hoher Ordnung n durch ein System niedriger Ordnung $N \ll n$ zu ersetzen, ohne dass dabei wesentliche Systemeigenschaften vernachlässigt werden. Dazu wird im ersten Schritt die Relevanz aller Zustände bewertet, um dann im zweiten Schritt die nicht relevanten Zustände zu eliminieren.

4 Blackbox-Modelle

Liegen keine physikalischen Gesetzmäßigkeiten vor oder wäre eine physikalische Modellbildung (Kapitel 2) zu aufwändig, können allgemeine (nicht-physikalische) Modellansätze gewählt werden, um das Ein-Ausgangsverhalten eines realen Prozesses zu modellieren. Diese Ansätze sollen nachfolgend vorgestellt werden.

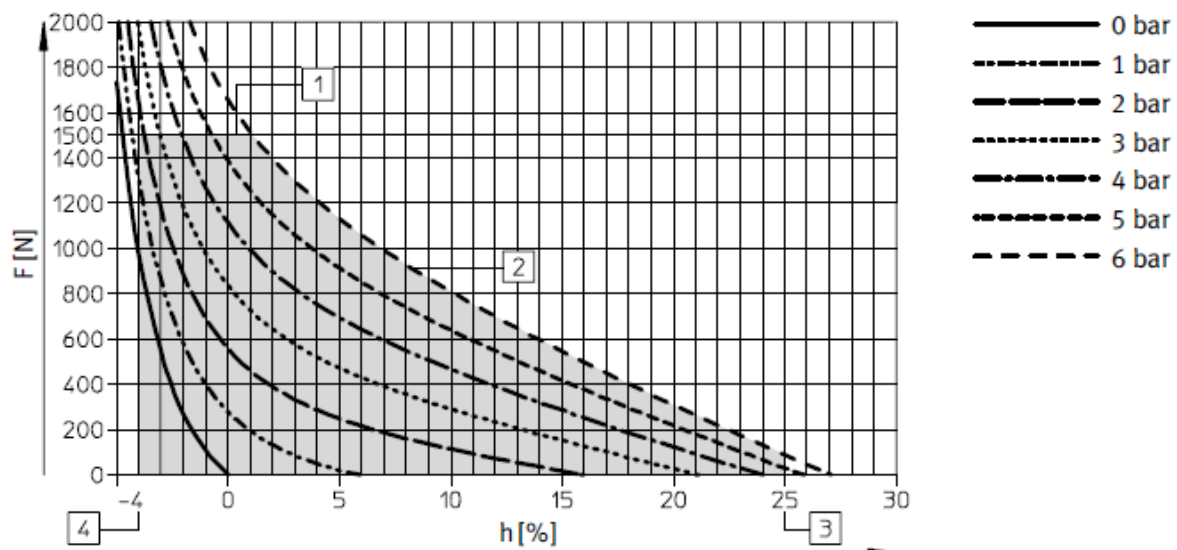
Die Parameter dieser Ansätze besitzen keine physikalische Interpretation. Sie sind zunächst unbekannt und müssen auf Basis von Messungen identifiziert werden (Kapitel 5 ff.).

In den folgenden vier Abschnitten 4.1 bis 4.4 werden Modellansätze für statische Zusammenhänge vorgestellt. Der Funktionszusammenhang kann linear, im Allgemeinen aber auch nichtlinear sein. Es sollen generell mehrere Eingänge $\underline{u} = [u_1 \dots u_m]^T$, aber (der Übersichtlichkeit halber) nur ein skalarer Modellausgang \hat{y} betrachtet werden:

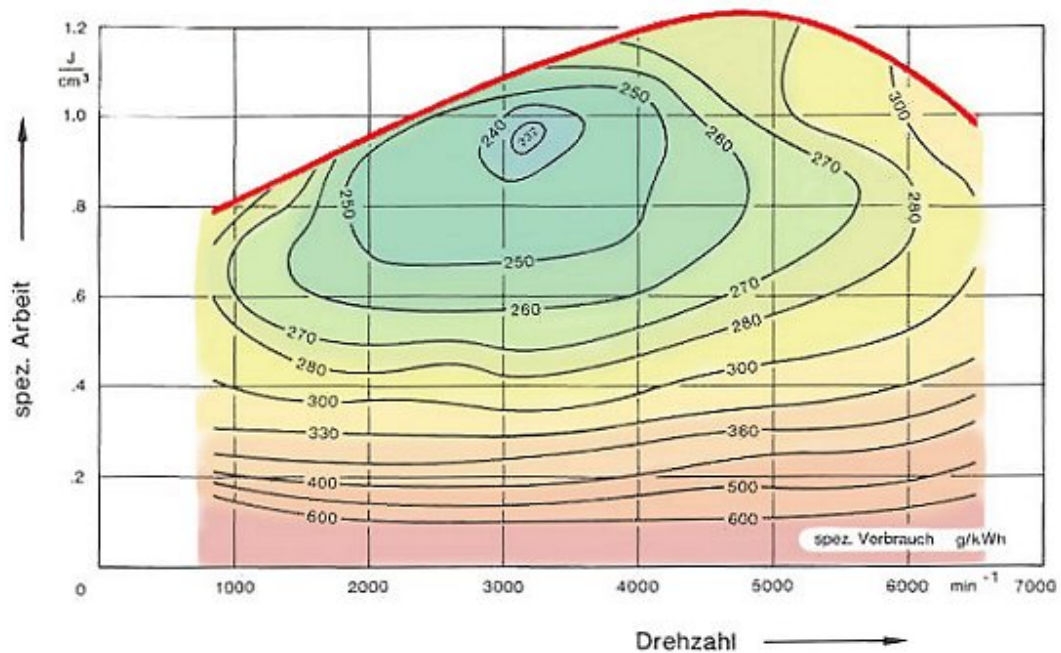


Beispiele für die graphische Darstellung von Modellzusammenhängen

1. Pneumatischer Muskel: Funktionszusammenhang von Luftdruck und Kontraktion h des Muskels zur resultierenden Kraft F bei einem Muskel mit dem Durchmesser von 20 mm [Festo Datenblatt]:



2. Muscheldiagramm eines Verbrennungsmotors: Funktionszusammenhang von Drehzahl und Arbeit zum Verbrauch des Motors:



4.1 Look-up Table Modell

Look-up Table definieren einen Funktionszusammenhang, in dem für den Ausgang \hat{y} explizite Stützstellen gespeichert werden. Mit Blick auf geringe Kosten für Speicher kann dies durchaus eine sinnvolle Methode sein, um Funktionszusammenhänge wie z.B. eine Kalibrierkurve zur Verfügung zu stellen.

Beispiel für ein Look-up Table Modell

Widerstandswerte eines PT100-Sensors in Abhängigkeit der Temperatur [aus www.pt100.de]:

	[-200...+39 °C] [+40...+289 °C] [+290...+339 °C] [+340...850 °C]									
	0	1	2	3	4	5	6	7	8	9
40	115,539	115,925	116,311	116,697	117,083	117,469	117,854	118,240	118,625	119,010
50	119,395	119,780	120,165	120,550	120,934	121,319	121,703	122,087	122,471	122,855
60	123,239	123,623	124,007	124,390	124,774	125,157	125,540	125,923	126,306	126,689
70	127,072	127,454	127,837	128,219	128,602	128,984	129,366	129,748	130,130	130,511
80	130,893	131,274	131,656	132,037	132,418	132,799	133,180	133,561	133,941	134,322
90	134,702	135,083	135,463	135,843	136,223	136,603	136,982	137,362	137,741	138,121
100	138,500	138,879	139,258	139,637	140,016	140,395	140,773	141,152	141,530	141,908

4.2 Polynomiale Modelle

Ansatz für Systeme mit einem Eingang ($m=1$), SISO-System mit der Ordnung des Polynoms n :

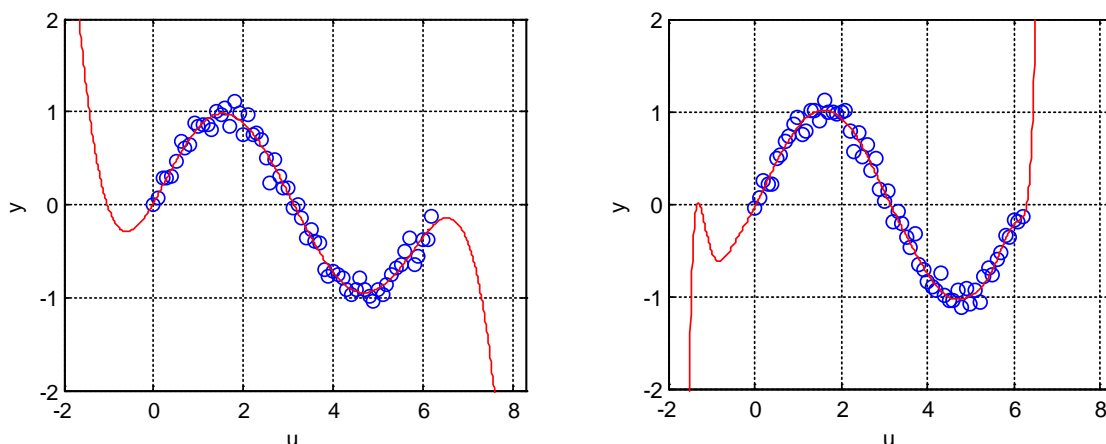
$$\hat{y} = f(u) = a_0 + a_1 u + a_2 u^2 + \dots + a_n u^n = \sum_{i=0}^n a_i u^i$$

Ansatz für System mit mehreren Eingängen ($m>1$), MISO-System:

$$\hat{y} = f(\underline{u}) = a_0 + \sum_{j=1}^m a_j u_j + \sum_{j=1}^m \sum_{k=1}^m a_{j,k} u_j u_k + \dots$$

Der Ausgang \hat{y} dieses Ansatzes ist für $n>1$ eine nichtlineare Funktion des Eingangs. D.h. es können prinzipiell nichtlineare Ein-Ausgangsfunktionen durch diesen Ansatz beschrieben werden. Demgegenüber ist der Ausgang \hat{y} aber eine lineare Funktion der Parameter a_j , dies wird sich bei der Parameteridentifikation in Kapitel 5 als großer Vorteil herausstellen.

Mit Hilfe des Polynoms wird die unbekannte Funktion über ihren gesamten Wertebereich hinweg approximiert. Lokale Abweichungen sind dadurch schlecht beschreibbar, ebenfalls ist keine gute *Extrapolation* zu erwarten, wie nachfolgendes Beispiel zur Approximation einer Sinus-Funktion für $n=5$ (links) und $n=15$ (rechts) zeigt [aus Nelles: Nonlinear System Identification]. Diese wird mit steigender Ordnung des Polynoms sogar schlechter.



4.3 Modelle mittels radialer Basisfunktionen

Radiale Basisfunktionen approximieren eine Funktion innerhalb eines lokalen Bereiches des durch die Eingänge \underline{u} aufgespannten Eingangsraums. Für die Beschreibung des gesamten Eingangsraums werden mehrere dieser lokalen Beschreibungen superponiert.

Als *radiale Basisfunktionen* θ eignen sich Funktionen,

- deren Funktionswerte nicht negativ sind,
- die in ihrem Mittelpunkt \underline{u}_0 ein Maximum besitzen und
- mit steigendem Abstand zu \underline{u}_0 monoton fallende Funktionswerte besitzen, die schließlich Null werden können.

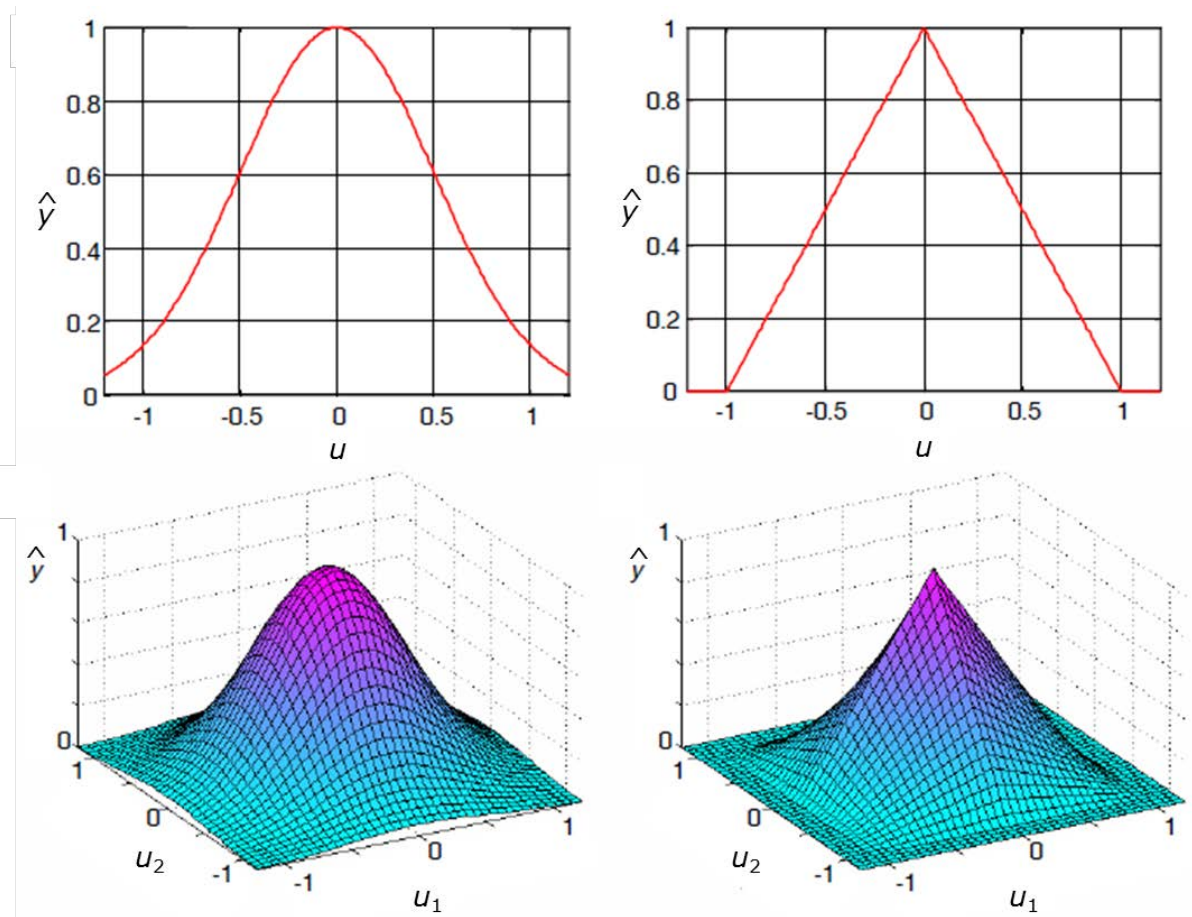
Nachfolgendes Bild zeigt zwei häufig verwendete radiale Basisfunktionen. In den Bildern oben ist der eindimensionale Fall gezeigt. Die einer Normalverteilung entsprechende, glockenförmige Gaußfunktion

$$\hat{y} = \theta(u) = e^{-\frac{1}{2}v^2} \quad \text{mit } v = \frac{1}{\sigma}(u - u_0)$$

ist stetig differenzierbar, während die Dreiecksfunktion

$$\hat{y} = \theta(u) = \begin{cases} 1 - \frac{1}{2\sigma}(u - u_0) & \text{für } u_0 \leq u \leq u_0 + 2\sigma \\ 1 + \frac{1}{2\sigma}(u - u_0) & \text{für } u_0 - 2\sigma \leq u \leq u_0 \\ 0 & \text{sonst} \end{cases}$$

in den Eckpunkten nicht differenzierbar ist. Der Parameter σ ist in beiden Funktionen ein Maß für die Breite der Basisfunktion. In den beiden unteren Bildern ist die Erweiterung auf einen zweidimensionalen Eingangsraum dargestellt.

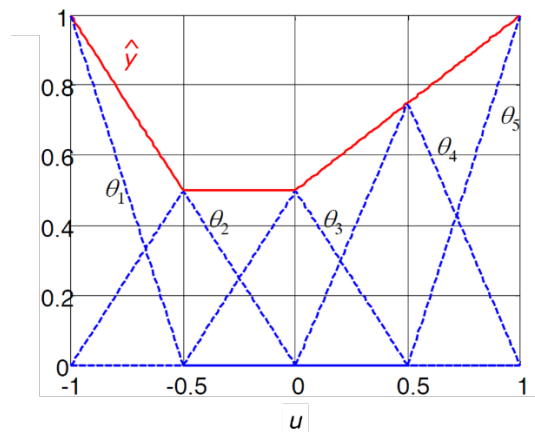


Um einen Funktionszusammenhang zu approximieren, werden mehrere radiale Basisfunktionen θ_i mit $i = 1, \dots, n$, die jeweils verschiedene Mittelpunkte $u_{0,i}$ besitzen, jeweils mit einem Gewicht a_i skaliert und dann superponiert:

$$\hat{y} = \sum_{i=1}^n a_i \cdot \theta_i(u)$$

Mit diesem Ansatz können nichtlineare Funktionszusammenhänge zwischen den Eingängen \underline{u} und dem Ausgang \hat{y} approximiert werden. Wie bei den polynomialen Ansätzen in Abschnitt 4.2 ist auch hier der Ausgang \hat{y} eine lineare Funktion der Parameter a_i , was wiederum bei der Parameteridentifikation in Kapitel 5 vorteilhaft sein wird.

Nachfolgendes Bild zeigt als Beispiel die Superposition einer Funktion (rote durchgezogene Linie) durch fünf Dreiecksfunktionen (blau gestrichelte Linie).



4.4 Neuronale Netze

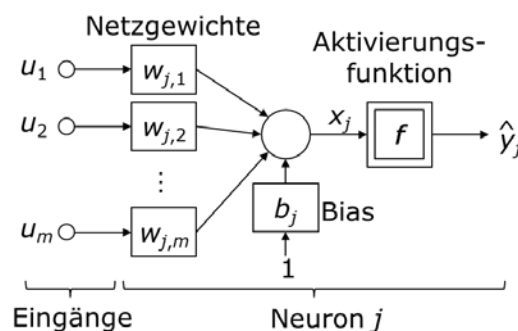
Künstliche Neuronale Netze abstrahieren biologische neuronale Netze, um einen parallelen und vernetzten Ansatz zur Modellbildung technischer Prozesse zu erhalten.

Eine parallele Architektur kann auch parallel implementiert und ausgewertet werden. Ein Vorteil ist die Robustheit gegenüber Ausfällen von Teilsystemen. Dadurch wird zwar die Gesamtleistung verschlechtert, eine Grundfunktion bleibt aber meist erhalten.

Nachfolgend wird eine einfache Netzarchitektur vorgestellt, in der *Neuronen* in *Schichten* angeordnet werden, die keine inneren Rückkopplungen besitzen. Diese werden als *Multi Layer Perceptron (MLP)* oder *Feedforward Netze* bezeichnet.

Neuron

Die Eingänge u_1, \dots, u_m am Neuron j werden durch Netzgewichte $w_{j,1}, \dots, w_{j,m}$ verstärkt und zusammen mit dem Bias b_j zum Signal x_j aufsummiert, das am Eingang der Aktivierungsfunktion f anliegt (vgl. auch Matlab, Neural Network Toolbox, User's Guide):



Die Funktion f kann eine nichtlineare Funktion sein, häufig werden Schwellwertfunktionen verwendet wie

$$f(x) = \tanh(x) \quad \text{oder} \quad f(x) = \frac{1}{1 + e^{-x}}.$$

Es kann aber auch eine lineare Funktion $f(x) = x$ gewählt werden (insbesondere in der letzten Schicht eines Netzes). Mit den Vektoren $\underline{u} = [u_1 \dots u_m]^T$ und $\underline{w}_j = [w_{j,1} \dots w_{j,m}]^T$ lässt sich die Funktionszusammenhang eines Neurons kompakt schreiben:

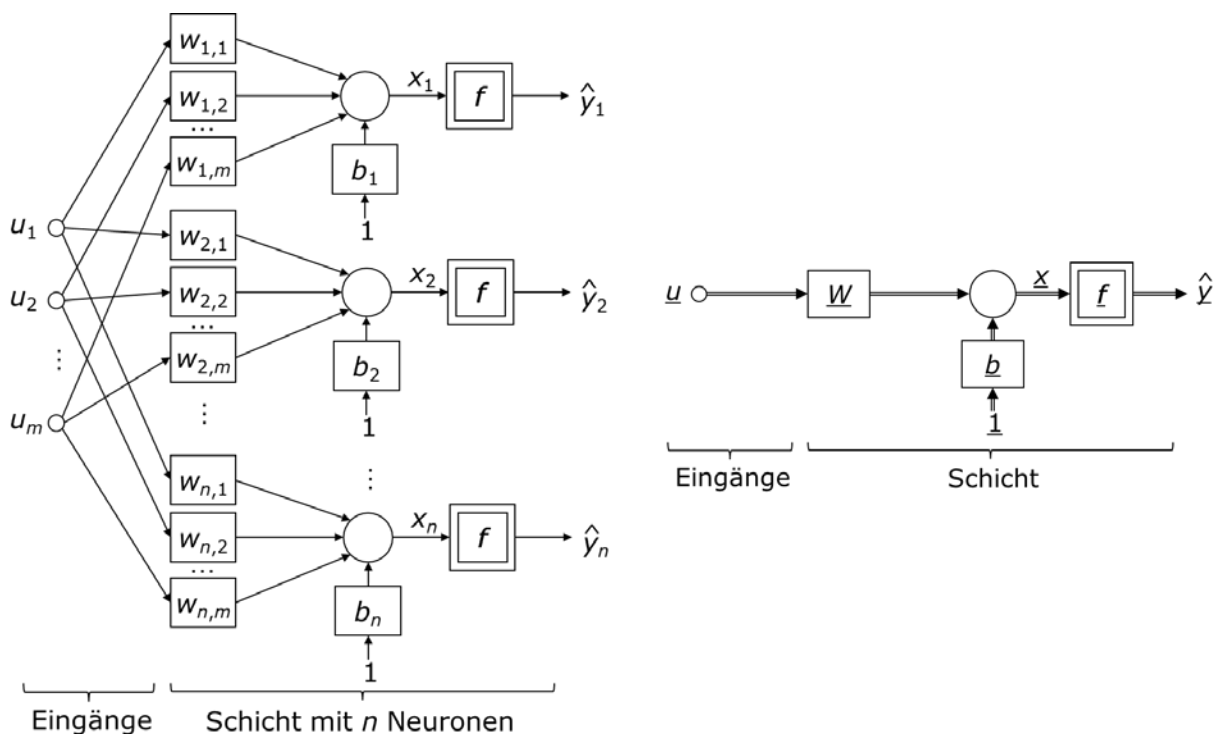
$$\hat{y}_j = f(\underline{w}_j^T \underline{u} + b_j)$$

Schichten

Eine Schicht wird aus n parallelen Neuronen realisiert. Mit dem Biasvektor $\underline{b} = [b_1 \dots b_n]^T$ und der Matrix der Netzgewichte $\underline{W} = [\underline{w}_1^T \dots \underline{w}_n^T]^T$ lässt sich der Funktionszusammenhang für die Ausgänge $\underline{\hat{y}} = [\hat{y}_1 \dots \hat{y}_n]^T$ der gesamten Schicht wie folgt schreiben:

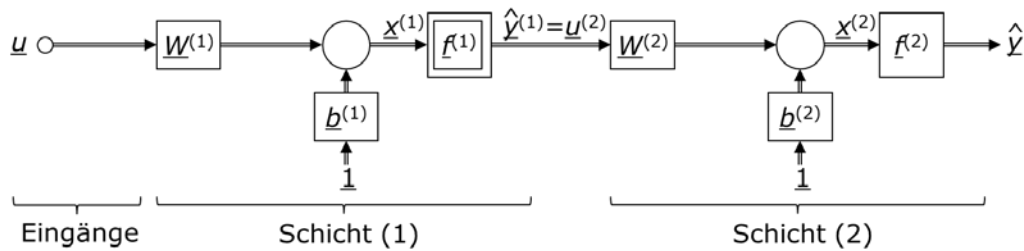
$$\underline{\hat{y}} = f(\underline{W}\underline{u} + \underline{b})$$

Nachstehend ist das Blockschaltbild einer Schicht dargestellt. Links auf Basis von n Neuronen, rechts dasselbe in vektorieller Darstellung:



Multi Layer Perceptron (MLP)

Ein typisches MLP ist aus zwei Schichten aufgebaut, für die Aktivierungsfunktion $f^{(1)}$ wird eine nichtlineare, für $f^{(2)}$ eine lineare Funktion verwendet.



Der entsprechende Funktionszusammenhang lautet:

$$\hat{y} = \underline{f}^{(2)} \left[\underline{W}^{(2)} \underline{f}^{(1)} (\underline{W}^{(1)} \underline{u} + \underline{b}^{(1)}) + \underline{b}^{(2)} \right]$$

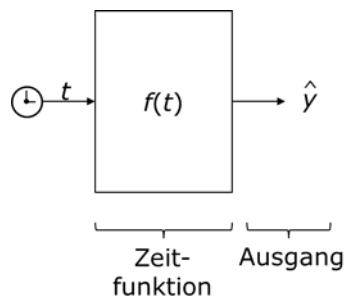
Die Parameter dieses Modellansatzes sind die Netzgewichte $\underline{W}^{(\cdot)}$ und die Biaswerte $\underline{b}^{(\cdot)}$. Diese können durch eine nichtlineare Parameteroptimierung bestimmt werden, wie sie in Abschnitt 5.4 vorgestellt wird.

4.5 Erweiterung auf Signale

Auch zur Modellierung des zeitlichen Verlaufs von Signalen können Blackbox-Modelle hilfreich sein. Das gilt vor allem dann, wenn die Entstehung des Signals recht komplex ist bzw. keine physikalische Beschreibung dazu zur Verfügung steht. Dies könnte z.B. der Verlauf der Außentemperatur $T(t)$ über den Zeitraum mehrerer Jahre hinweg sein oder der zeitliche Verlauf einer Leistungsabforderung $P(t)$ in einem elektrischen Netz. Bei einfachen Zusammenhängen wie z.B. einer Sprungantwort $h(t)$ eines gegebenen Übertragungssystems sollte hingegen die Systemantwort des Whitebox-Modells herangezogen werden.

Das Modell eines Signals charakterisiert die Signaleigenschaften und ermöglicht eine im Vergleich zu den Rohdaten kompaktere Darstellung. Es kann u.A. zur Vorhersage des zukünftigen Signalverlaufs, zur Signalanalyse oder zur Korrektur eines Trends genutzt werden.

Prinzipiell können alle Modellansätze der Abschnitte 4.1 bis 4.4 übernommen werden, wenn anstelle der Eingänge \underline{u} die Zeit t tritt:



Dabei kann das gemessene Signal als Gesamtsignal aus der Superposition mehrerer Signalanteile formuliert werden. Nachfolgend sollen zwei häufige Fälle näher betrachtet werden:

Polynomiales Signalmodell

Es wird angenommen, das ungestörte Signal entspräche einer polynomialen Zeitfunktion, dann lautet der Modellansatz analog zu Abschnitt 4.2:

$$\hat{y} = f(t) = \sum_{i=0}^n a_i t^i$$

Periodisches Signalmodell

Für harmonische Signalverläufe kann folgender Modellansatz gewählt werden:

$$\hat{y} = f(t) = \sum_{i=1}^n [a_i \cos(\omega_i t) + b_i \sin(\omega_i t)]$$

Darin sind ω_i die für das Signal wesentlichen Kreisfrequenzen. Diese müssen bekannt sein! Häufig kann man diese aus zusätzlichem Wissen über den Prozess ableiten. Bei einer Außentemperaturmessung kann z.B. vermutet werden, dass der Zeitverlauf die Periode T_1 eines Tages und die Periode T_2 eines Jahres besitzen wird. Entsprechend können $\omega_1 = 2\pi/T_1$ und $\omega_2 = 2\pi/T_2$ gewählt werden. Alternativ könnte man auch vorab eine Signalanalyse im Frequenzbereich durchführen, um relevante Kreisfrequenzen ω_i zu finden.

4.6 Erweiterung auf dynamische Systeme

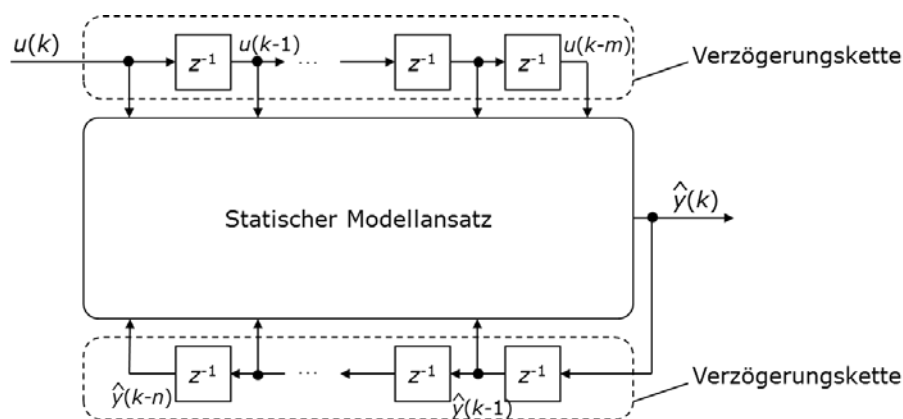
Dynamische Systeme haben im Unterschied zu statischen Systemen ein „Gedächtnis“, d.h. ihre Antwort ist nicht nur eine Funktion der aktuellen Eingänge, sondern hängt auch von zeitlich zurückliegenden Ein- und Ausgangswerten ab. Auch hier sind Blackbox-Modelle hilfreich, wenn keine Whitebox-Modelle gefunden werden können.

Nachfolgend beschränken wir uns auf *zeitdiskrete* dynamische Systeme. Diese können aus zeitkontinuierlichen Systemen durch Abtastung gewonnen werden. Bei äquidistanter Abtastung mit konstanter Abtastzeit T erhält man aus $y(t)$ das zeitdiskrete Signal

$$y(k) = y(t = k T)$$

Wobei k der Zähler über die Zeitschritte ist.

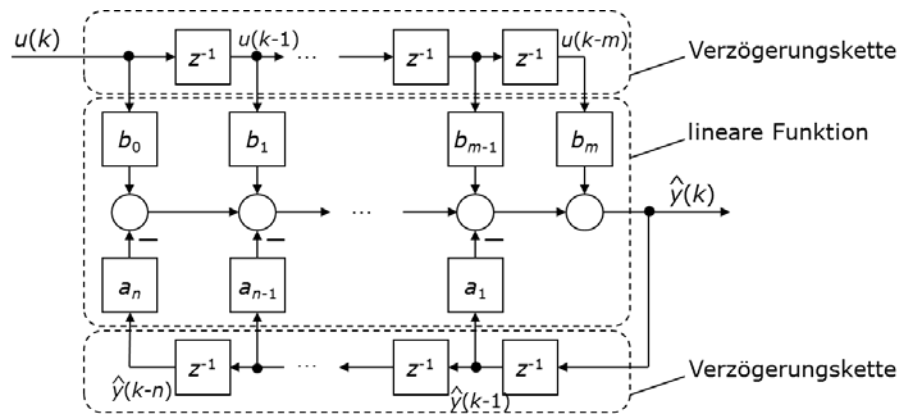
Die bisher vorgestellten statischen Modellansätze werden um eine externe Dynamik (als „Gedächtnis“) erweitert. Dies kann z.B. realisiert werden, in dem der Eingang $u(k)$ bzw. der Ausgang $y(k)$ über Verzögerungsketten gespeichert wird:



Damit können die bisher verwendeten statischen Modellansätze, die in den Abschnitten 4.1 bis 4.4 vorgestellt wurden, hier wieder genutzt werden! Der statische Modellansatz dient also dazu, auf Basis des aktuellen und der zurückliegenden Eingangswerte $u(k), \dots, u(k-m)$ sowie der zurückliegenden Ausgängen $\hat{y}(k-1), \dots, \hat{y}(k-n)$ eine Vorhersage (Prädiktion) für den nächsten Ausgangswert $\hat{y}(k)$ zu machen.

Allerdings müssen die Speicherlängen n , m sowie die Abtastzeit T geeignet gewählt werden, so dass einerseits alle dynamischen Eigenschaften erfasst werden und andererseits die Zahl der $(n+m+1)$ Eingänge nicht zu groß wird.

Beispielsweise kann entsprechend Abschnitt 4.2 ein polynomialer Ansatz mit nur linearen Gliedern verwendet werden. Eine spezielle Wahl der Parameter zeigt folgendes Blockschaltbild:



$$\hat{y}(k) = \sum_{j=1}^n -a_j \hat{y}(k-j) + \sum_{i=0}^m b_i u(k-i)$$

Dieser Modellansatz entspricht der linearen, zeitdiskreten Übertragungsfunktion

$$G(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_m z^{-m}}{a_0 + a_1 z^{-1} + \dots + a_n z^{-n}}.$$

Teil B: Identifikation

5 Identifikation statischer Modelle

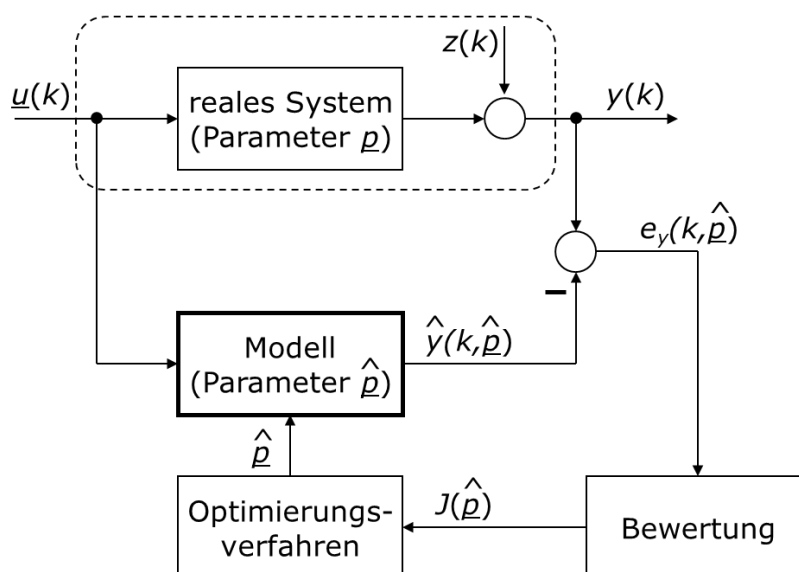
5.1 Aufgabenstellung

Ein reales System besitzt unbekannte Modellparameter, die im Parametervektor \underline{p} zusammengefasst werden. Für die Identifikation dieser Parameter wird parallel ein Modell des realen Systems betrieben, das eine Schätzung $\hat{\underline{p}}$ des Parametervektors verwendet.

Es werden insgesamt $k = 1, \dots, N$ Messungen durchgeführt. Für in jeder Messung wird bei gegebenem Eingang $u(k)$ die Messung des realen Systems $y(k)$ mit dem Modellausgang $\hat{y}(k, \hat{\underline{p}})$ verglichen und der Ausgangsfehler $e_y(k, \hat{\underline{p}})$ bestimmt.

Mit Hilfe einer Kostenfunktion $J(\hat{\underline{p}})$ wird der Ausgangsfehler über alle N Messungen hinweg bewertet. In der Regel wird die Kostenfunktion so festgelegt, dass ihr Wert zunimmt, falls der Ausgangsfehler größer wird. Es ist dann die Aufgabe eines nachgeordneten Optimierungsverfahrens $J(\hat{\underline{p}})$ in Abhängigkeit von $\hat{\underline{p}}$ zu minimieren, um so das Verhalten des Modells an das des realen Systems anzupassen. Somit ist die Aufgabe der Parameteridentifikation auf ein Optimierungsproblem zurückgeführt.

Darstellung im Blockschaltbild:



In der Wahl der Kostenfunktion ist man prinzipiell frei. Sehr häufig wird der quadratische, über alle N Messungen summierte Ausgangsfehler verwendet:

$$J(\hat{\underline{p}}) := \sum_{k=1}^N e_y^2(k, \hat{\underline{p}})$$

Nach Karl Friedrich Gauß wird dies auch als „Methode der kleinsten Quadrate“ bezeichnet (im Englischen: „Least squares“).

5.2 Lineare Optimierung: Direkte Lösung

Lineare Parameterschätzung

Es wird vorausgesetzt, dass der zu bestimmende Parametervektor \underline{p} linear in die Berechnung der Ausgangsgröße y eingeht:

$$y(k, \underline{p}) = \underline{m}^T(k) \cdot \underline{p}$$

Entsprechend wird auch das Modell angesetzt:

$$\hat{y}(k, \underline{\hat{p}}) = \underline{m}^T(k) \cdot \underline{\hat{p}}$$

Für den Ausgangsfehler zwischen dem Modell \hat{y} und der realen Messung y gilt folglich:

$$e_y(k, \underline{\hat{p}}) = y(k) - \hat{y}(k, \underline{\hat{p}}) = y(k) - \underline{m}^T(k) \cdot \underline{\hat{p}}$$

Es werden (z.B. zeitlich nacheinander) mehrerer Messungen $k = 1, \dots, N$ durchgeführt und zeilenweise untereinander vektoriell zusammengefasst:

$$\underline{e}_y(\underline{\hat{p}}) = \begin{bmatrix} e(1, \underline{\hat{p}}) \\ \vdots \\ e(N, \underline{\hat{p}}) \end{bmatrix}, \quad \underline{y} = \begin{bmatrix} y^T(1) \\ \vdots \\ y^T(N) \end{bmatrix}, \quad \underline{M} = \begin{bmatrix} \underline{m}^T(1) \\ \vdots \\ \underline{m}^T(N) \end{bmatrix}$$

Entsprechend lässt sich der Ausgangsfehler aller Messungen schreiben als:

$$\underline{e}_y(\underline{\hat{p}}) = \underline{y} - \underline{M} \cdot \underline{\hat{p}}$$

Als Kostenfunktion J wird die Summe der Fehlerquadrate

$$J(\underline{\hat{p}}) = \underline{e}_y^T(\underline{\hat{p}}) \cdot \underline{e}_y(\underline{\hat{p}})$$

minimiert (nach Karl Friedrich Gauß als „Methode der kleinsten Quadrate“ bezeichnet)

Daraus ergibt sich als Lösung für den optimalen Parametervektor $\underline{\hat{p}}$:

$$\underline{\hat{p}} = (\underline{M}^T \cdot \underline{M})^{-1} \underline{M}^T \cdot \underline{y}$$

Das System ist identifizierbar, falls $(\underline{M}^T \cdot \underline{M})$ invertierbar ist. Ist dies nicht der Fall, wurden häufig nicht genug oder zu gleichartige Messungen durchgeführt.

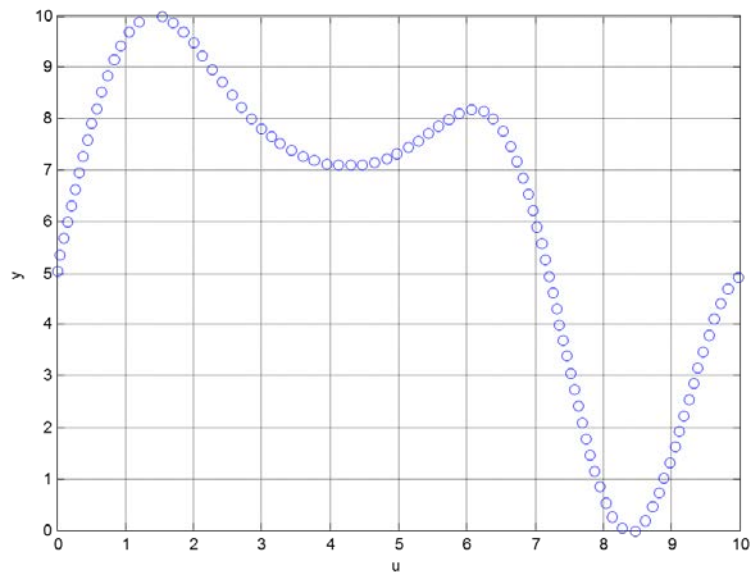
Die Kostenfunktion $J(\underline{\hat{p}})$ lässt sich zur Bewertung der Qualität der Identifikation angeben.

Daraus lässt sich auch der *root mean squared error (RMSE)* als Fehlermaß berechnen:

$$RMSE(\underline{\hat{p}}) = \sqrt{\frac{1}{N} J(\underline{\hat{p}})}$$

Beispiel 5-1: Identifikation eines polynomialen Modells

Für den folgenden dargestellten Datensatz mit $N = 94$ Messungen soll ein Blackbox-Modell identifiziert werden (aus Matlab: `simplefit_dataset`):



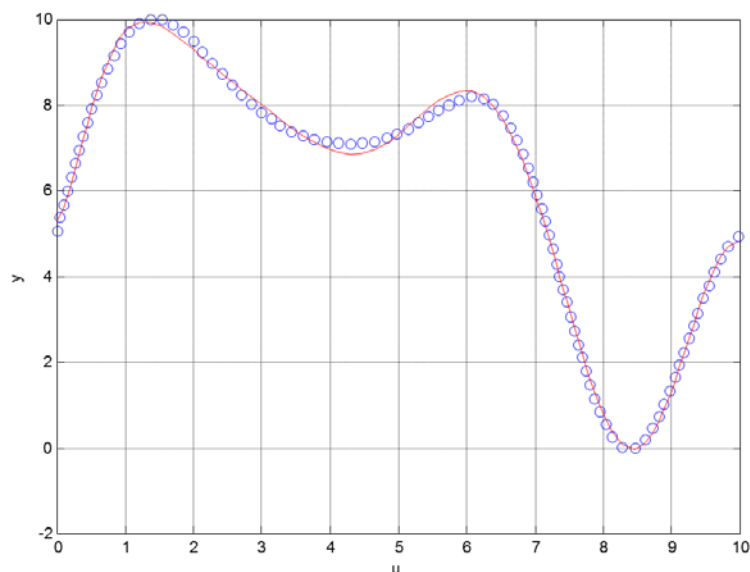
Ansatz eines polynomialen Modells:

$$\hat{y}(k, \underline{\hat{p}}) = a_0 + a_1 u(k) + \dots + a_n u^n(k) = \underbrace{\begin{bmatrix} 1 & u(k) & \dots & u^n(k) \end{bmatrix}}_{\underline{m}^T(k)} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}}_{\underline{\hat{p}}}$$

Damit ist:

$$\underline{M} = \begin{bmatrix} \underline{m}^T(1) \\ \vdots \\ \underline{m}^T(N) \end{bmatrix} = \begin{bmatrix} 1 & u(1) & \dots & u^n(1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & u(N) & \dots & u^n(N) \end{bmatrix}$$

Nachfolgend ist der Modellausgang \hat{y} für ein Polynom der Ordnung $n = 10$ gezeigt:



Beispiel 5-2: Identifikation eines Modells mit radialen Basisfunktionen

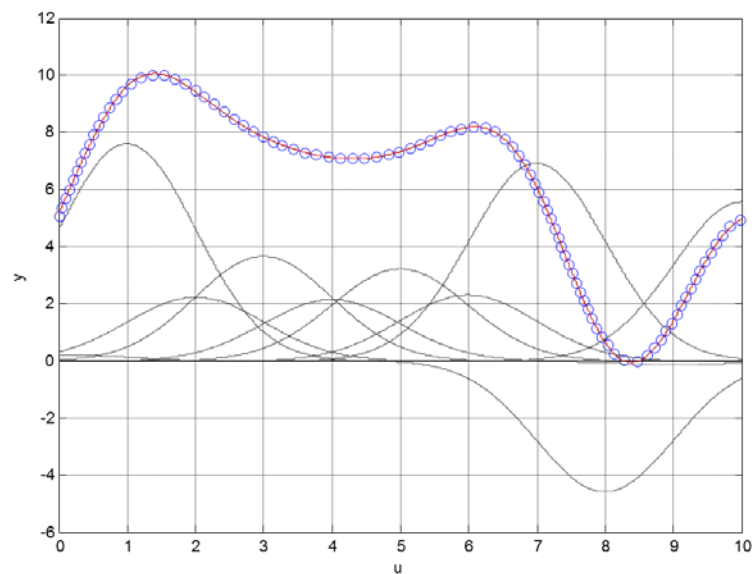
Ansatz mittels radialer Basisfunktionen:

$$\hat{y}(k, \underline{\hat{p}}) = a_1 \theta_1(u(k)) + a_2 \theta_2(u(k)) + \dots + a_n \theta_n(u(k)) = \underbrace{\begin{bmatrix} \theta_1(u(k)) & \theta_2(u(k)) & \dots & \theta_n(u(k)) \end{bmatrix}}_{\underline{m}^T(k)} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}}_{\underline{\hat{p}}}$$

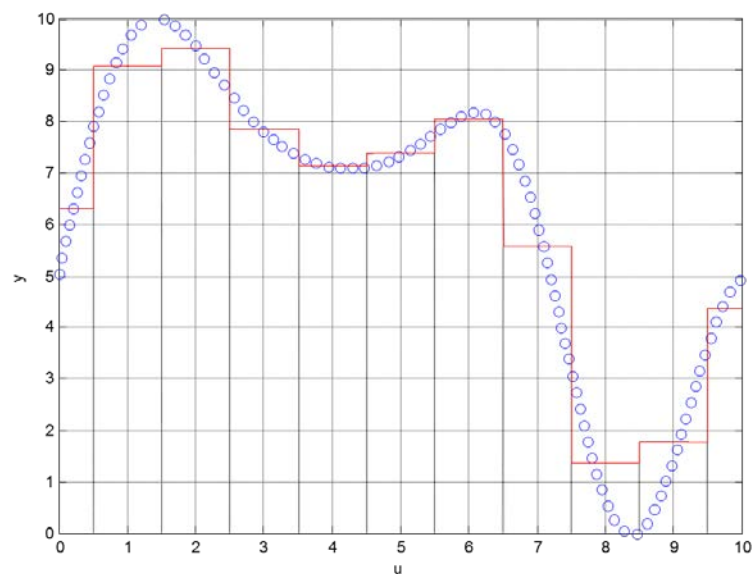
Damit gilt:

$$\underline{M} = \begin{bmatrix} \underline{m}^T(1) \\ \vdots \\ \underline{m}^T(N) \end{bmatrix} = \begin{bmatrix} \theta_1(u(1)) & \theta_2(u(1)) & \dots & \theta_n(u(1)) \\ \vdots & \vdots & \ddots & \vdots \\ \theta_1(u(N)) & \theta_2(u(N)) & \dots & \theta_n(u(N)) \end{bmatrix}$$

Der Modellausgang \hat{y} bei Verwendung von $n = 11$ Basisfunktionen: Mit Gaussfunktionen wird eine gute Approximation erreicht (einzelne Basisfunktionen sind punktiert dargestellt):



Werden lokale Rechteckfunktionen verwendet, kann ein Look-up Table identifiziert werden:



5.3 Lineare Optimierung: Rekursive Lösung

Um die geschlossene Lösung für $\hat{\underline{p}}$ nach Abschnitt 5.2 bestimmen zu können, müssen alle Messungen in \underline{y} vorliegen. Dies ist kein Problem, wenn $\hat{\underline{p}}$ erst nach Abschluss der Messungen „offline“ berechnet werden soll.

Soll die Information jeder neuen Messung $y(k)$ unmittelbar (also „online“) in eine Verbesserung der Lösung für $\hat{\underline{p}}(k)$ einfließen, so ist eine rekursive Lösung am besten geeignet.

Anfangswert der Rekursion:

$$\hat{\underline{p}}(0), \underline{P}(0)$$

Bei unsicheren Anfangswerten des Parametervektors $\hat{\underline{p}}(0)$ ist $\underline{P}(0)$ positiv definit und „groß“ zu wählen, z.B. $\underline{P}(0) = 10^{10} \cdot \underline{I}$.

Rekursion von Zeitpunkt (k) zu $(k+1)$:

1. Schritt:	$\underline{K}(k+1) = \frac{\underline{P}(k) \underline{m}(k+1)}{1 + \underline{m}^T(k+1) \underline{P}(k) \underline{m}(k+1)}$
2. Schritt:	$\hat{\underline{p}}(k+1) = \hat{\underline{p}}(k) + \underline{K}(k+1) \cdot [y(k+1) - \underline{m}^T(k+1) \hat{\underline{p}}(k)]$
3. Schritt:	$\underline{P}(k+1) = [\underline{I} - \underline{K}(k+1) \underline{m}^T(k+1)] \cdot \underline{P}(k)$

Damit wird die Kostenfunktion $J(\hat{\underline{p}}) = \underline{e}_y^T(\hat{\underline{p}}) \cdot \underline{e}_y(\hat{\underline{p}})$ (wie in Abschnitt 5.2) minimiert.

Exponentiell nachlassendem Gedächtnis

Bei zeitvarianten Prozessen ist es sinnvoll, lange zurückliegende Messungen geringer zu gewichten als aktuelle. Mit dem Vergessensfaktor $0 < \lambda < 1$ (z. B. $\lambda = 0,99$) wird die Kostenfunktion entsprechend modifiziert:

$$J(\hat{\underline{p}}) = \underline{e}_y^T(\hat{\underline{p}}) \cdot \underline{Q} \cdot \underline{e}_y(\hat{\underline{p}}) \quad \text{mit} \quad \underline{Q} = \text{diag}(\lambda^{N-1}, \dots, \lambda^2, \lambda, 1)$$

Zugehörige Rekursion mit von Zeitpunkt (k) zu $(k+1)$:

1. Schritt:	$\underline{K}(k+1) = \frac{\underline{P}(k) \underline{m}(k+1)}{\lambda + \underline{m}^T(k+1) \underline{P}(k) \underline{m}(k+1)}$
2. Schritt:	$\hat{\underline{p}}(k+1) = \hat{\underline{p}}(k) + \underline{K}(k+1) \cdot [y(k+1) - \underline{m}^T(k+1) \hat{\underline{p}}(k)]$
3. Schritt:	$\underline{P}(k+1) = \frac{1}{\lambda} [\underline{I} - \underline{K}(k+1) \underline{m}^T(k+1)] \cdot \underline{P}(k)$

5.4 Nichtlineare Optimierung

Falls kein linearer Zusammenhang zwischen dem Parametervektor \underline{p} und dem Ausgang y besteht, müssen nichtlineare Optimierungsverfahren eingesetzt werden, um den optimalen Parametervektor zu finden. Diese sind numerisch aufwendiger. Anstelle einer direkten Lösung wird ein *iteratives Suchverfahren* durchgeführt.

Darüber hinaus besteht die Gefahr, nicht das tatsächliche Minimum (=Optimum), sondern nur ein *lokales Minimum* (=suboptimale Lösung) zu finden. Dies wird maßgeblich von den *Startwerten* bestimmt, mit denen man das Suchverfahren beginnt. Daher ist eine gute Wahl der Startwerte für den Erfolg des Verfahrens wesentlich.

Nachfolgend werden zwei weit verbreitete Verfahren vorgestellt.

Simplex-Suche (nach Nelder und Mead)

Dieses Optimierungsverfahren benötigt keine Informationen über die Ableitungen der Kostenfunktion $J(\underline{p})$ und ist daher gut für *diskontinuierliche* und *stark nichtlineare* Funktionen geeignet. Es ist allerdings *weniger effizient* als Gradientenmethoden. Die Funktion `fminsearch` in Matlab implementiert diesen Algorithmus.

Die n Komponenten des Parametervektors $\underline{p} = [p_1, \dots, p_n]^T$, die durch die Minimierung der Kostenfunktion $J(\underline{p})$ bestimmt werden sollen, spannen einen n -dimensionalen Vektorraum auf. Darin wird ein Simplex S mit $n+1$ Ecken durch die Eckpunkte \underline{p}_1 bis \underline{p}_{n+1} definiert:

$$S = \{\underline{p}_1, \dots, \underline{p}_n, \underline{p}_{n+1}\}$$

Im Fall $n = 2$ ist dies ein Dreieck, im Fall $n = 3$ ein Tetraeder.

Durch heuristische Regeln wird die Lage des Simplex im Parameterraum schrittweise modifiziert, so dass sich der Simplex im Minimum „zusammenzieht“.

Algorithmus:

- (1) Wahl der Eckpunkte des Startsimplex $S^{(0)}$,
der die vermutlichen Anfangswerte der Parameter umschließt

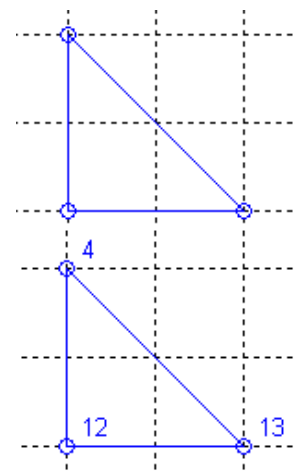
- (2) Bestimmung der Funktionswerte aller Eckpunkte
 $J(\underline{p}_1), \dots, J(\underline{p}_{n+1})$

- (3) Bestimmung des Eckpunktes \underline{p}_{\max} bzw. \underline{p}_{\min} mit
der maximalen bzw. minimalen Kostenfunktion:

$$J_{\max} = \max_i J(\underline{p}_i) \quad \text{bzw.} \quad J_{\min} = \min_i J(\underline{p}_i)$$

- (4) Bestimmung des Schwerpunktes des Simplex ohne \underline{p}_{\max} :

$$\underline{p}_c = \frac{1}{n} \cdot \sum_{\substack{i=1 \\ i \neq \max}}^{n+1} \underline{p}_i$$



(5) Teste **Reflexion**: Spiegele auf neuen Punkt

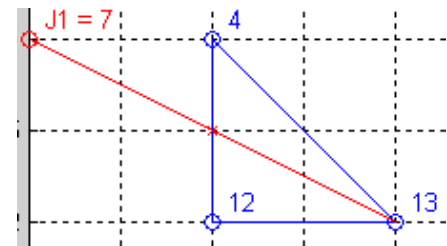
$$\hat{\underline{p}} = \underline{p}_C + \alpha \cdot (\underline{p}_C - \underline{p}_{\max}) \quad \text{mit } \alpha = 1$$

und bestimme Funktionswert $J(\hat{\underline{p}})$

Falls $J(\hat{\underline{p}}) < J_{\min}$: weiter bei (6) Expansion,

falls $J(\hat{\underline{p}}) > J_{\max}$: weiter bei (7) Kontraktion,

sonst: nimm $\hat{\underline{p}}$ statt \underline{p}_{\max} als neuen Eckpunkt an,
weiter bei (8).



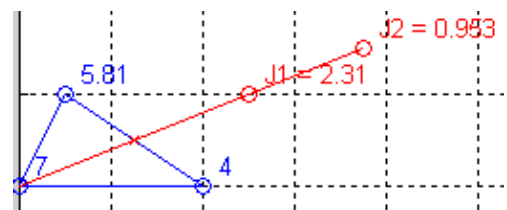
(6) Teste **Expansion**: Erweitere auf neuen Punkt

$$\tilde{\underline{p}} = \underline{p}_C + \gamma \cdot (\hat{\underline{p}} - \underline{p}_C) \quad \text{mit } \gamma = 2$$

und bestimme Funktionswert $J(\tilde{\underline{p}})$

Falls $J(\tilde{\underline{p}}) < J_{\min}$: nimm $\tilde{\underline{p}}$ statt \underline{p}_{\max} als
neuen Eckpunkt an, weiter bei (8),

sonst: nimm $\hat{\underline{p}}$ statt \underline{p}_{\max} als neuen Eckpunkt an,
weiter bei (8).



(7) Teste **Kontraktion**: Verkürze auf neuen Punkt

$$\tilde{\underline{p}}_p = \underline{p}_C + \beta \cdot (\underline{p}_{\max} - \underline{p}_C) \quad \text{mit } \beta = 0,5$$

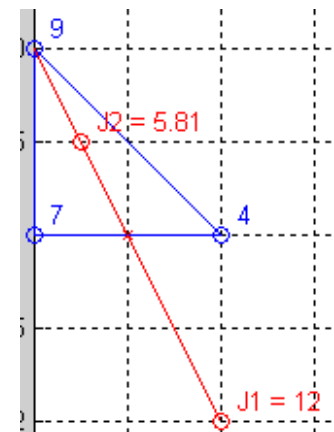
und bestimme Funktionswert $J(\tilde{\underline{p}})$

Falls $J(\tilde{\underline{p}}) < \min(J(\hat{\underline{p}}), J_{\min})$: nimm $\tilde{\underline{p}}$ statt \underline{p}_{\max} als
neuen Eckpunkt an,

sonst: verkleinere Simplex (Notlösung!)

$$\underline{p}_i = (\underline{p}_i - \underline{p}_{\min})/2 + \underline{p}_{\min} \quad \text{für alle Eckpunkte } i,$$

weiter bei (8).



(8) Falls Abbruchkriterium nicht erfüllt, weiter bei (3).

Als Abbruchkriterium kann gewählt werden:

- Das Simplexvolumen unterschreitet einen gewählten Schwellwert.
- Eine maximale Anzahl von Iterationen ist überschritten.

Liniensuche*Allgemeine Struktur des iterativen Optimierungsalgorithmus*

- (1) Bestimmung des Startwertes für den Parametervektor $\underline{p}^{(0)}$
- (2) Berechnung der Abstiegsrichtung $\underline{\eta}^{(k)}$
- (3) Berechnung der Schrittweite $\alpha^{(k)}$
- (4) Iterationsschritt: $\underline{p}^{(k+1)} = \underline{p}^{(k)} + \alpha^{(k)} \cdot \underline{\eta}^{(k)}$
- (5) Abbruch, falls Kriterium $\|\underline{\eta}^{(k)}\| \leq \varepsilon$
- (6) $k := k + 1$ und weiter bei Schritt (2)

Generell besteht die Gefahr, dass der Algorithmus bei ungünstig gewähltem Startwert $\underline{p}^{(0)}$ nicht das globale, sondern nur ein lokales Minimum der Kostenfunktion $J(\underline{p})$ findet.

Abstiegsrichtungen

Negativer Gradient:
$$\underline{\eta} = - \frac{\partial J(\underline{p})}{\partial \underline{p}}$$

mit den Elementen des Gradientenvektors: $\eta_i = - \frac{\partial J(\underline{p})}{\partial p_i}$ für $i = 1, \dots, n$ Komponenten in \underline{p} .

Der negative Gradient ist eine sichere Abstiegsrichtung.

Newton-Richtung:
$$\underline{\eta} = -[\underline{H}(\underline{p})]^{-1} \frac{\partial J(\underline{p})}{\partial \underline{p}}$$

mit den Elementen der Hessematrix: $H_{i,j}(\underline{p}) = \frac{\partial^2 J(\underline{p})}{\partial p_i \partial p_j}$ für $i = 1, \dots, n$ und $j = 1, \dots, n$

Die Newton-Richtung ist nur dann eine Abstiegsrichtung, falls die Hesse-Matrix positiv definit ist: $\underline{H} > 0$.

Modifizierte Newton-Richtung:
$$\underline{\eta} = -[\underline{H}(\underline{p}) + \delta \underline{I}]^{-1} \frac{\partial J(\underline{p})}{\partial \underline{p}}$$

Zur Hesse-Matrix wird ein Term $\delta \underline{I}$ hinzuaddiert. Die Konstante $\delta > 0$ wird so groß gewählt, so dass gilt $[\underline{H}(\underline{p}) + \delta \underline{I}] > 0$ und somit ein sicherer Abstieg erreicht wird.

Schrittweitensteuerung

Optimale Schrittweite: (führt auf ein weiteres Optimierungsproblem!):

$$J(\underline{p}^{(k)} + \alpha^{(k)} \cdot \underline{\eta}^{(k)}) \rightarrow \min$$

Armijo-Regel: Wähle $\alpha^{(k)} = \beta^m$, wobei m die kleinste ganze Zahl ist, für die

$$J(\underline{p}^{(k)}) - J(\underline{p}^{(k)} + \beta^m \cdot \underline{\eta}^{(k)}) \geq \sigma \cdot \beta^m \left[\frac{\partial J(\underline{p})}{\partial \underline{p}} \bigg|_{\underline{p}^{(k)}} \right]^T \cdot (-\underline{\eta}^{(k)})$$

erfüllt ist.

Für die Suche von m Starte mit $m = 0$ und erniedrige, falls die Ungleichung erfüllt ist bzw. erhöhe, falls die Ungleichung nicht erfüllt ist.

Wähle $\beta = 1/4$ und $\sigma = 1/10$.

Beschränkte Optimierung

Für den Parametervektor $\tilde{\underline{p}}$ muss eine untere Schranke \underline{p}_u bzw. eine obere Schranke \underline{p}_o eingehalten werden. Dies kann durch eine nichtlineare Transformation auf ein unbeschränktes Optimierungsproblem für den Parametervektor \underline{p} zurückgeführt werden, z.B.:

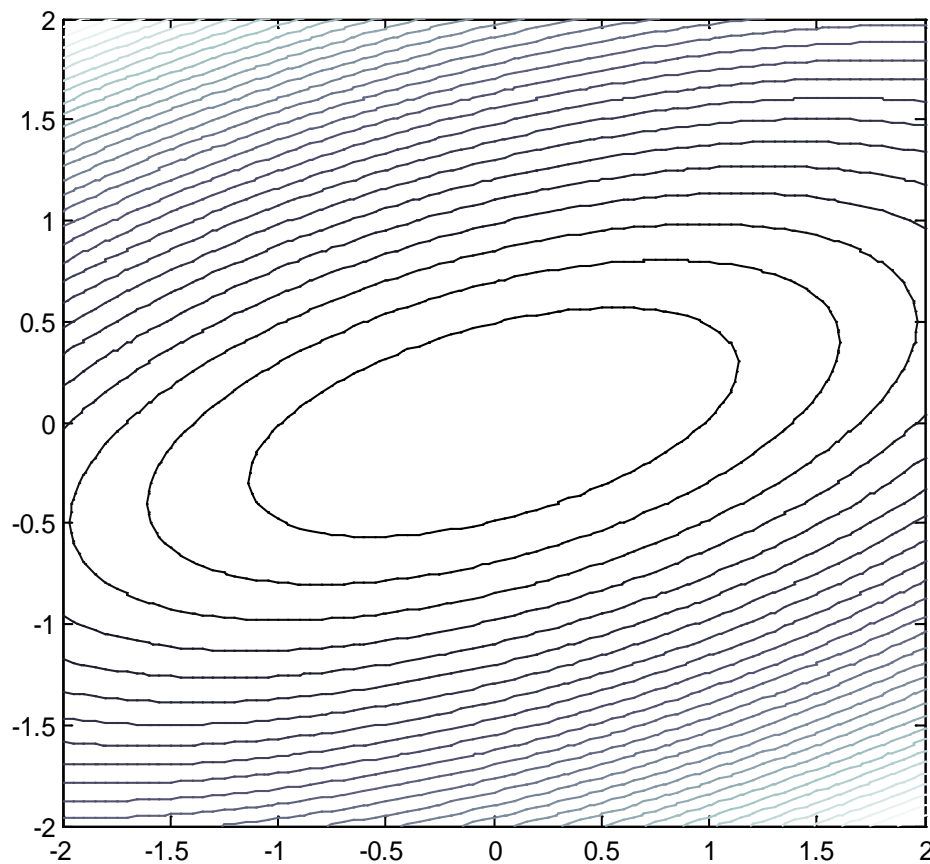
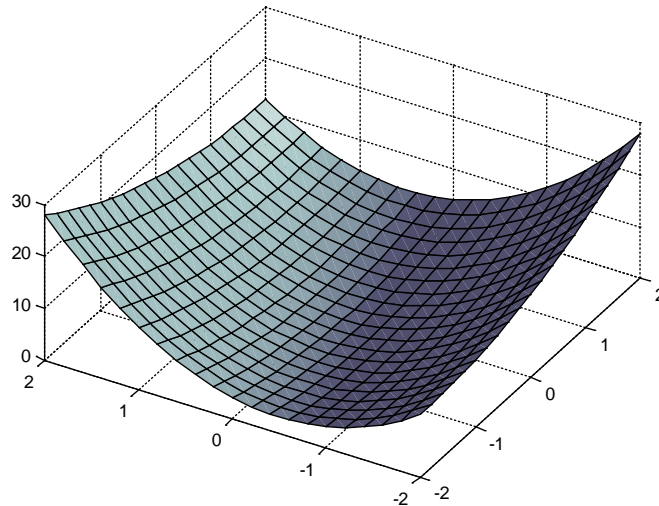
nur untere Grenze: $\underline{p}_u \leq \tilde{\underline{p}}$	$\tilde{p}_i = p_{u,i} + p_i^2$
nur obere Grenze: $\tilde{\underline{p}} \leq \underline{p}_o$	$\tilde{p}_i = p_{o,i} - p_i^2$
obere und untere Grenze: $\underline{p}_u \leq \tilde{\underline{p}} \leq \underline{p}_o$	$\tilde{p}_i = p_{u,i} + \frac{p_{o,i} - p_{u,i}}{2} (\sin p_i + 1)$

Bsp: Optimierung einer quadratischen Funktion mittels Liniensuche

Die Kostenfunktion

$$J(\underline{\hat{p}}) = \underline{\hat{p}}^T \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} \underline{\hat{p}}$$

soll mit dem Startwert $\underline{\hat{p}}^{(0)} = [-1,5 \ -1,5]^T$ beginnend optimiert werden. Nachfolgend ist die Funktion $J(\underline{\hat{p}})$ dreidimensional bzw. durch Höhenlinien dargestellt:



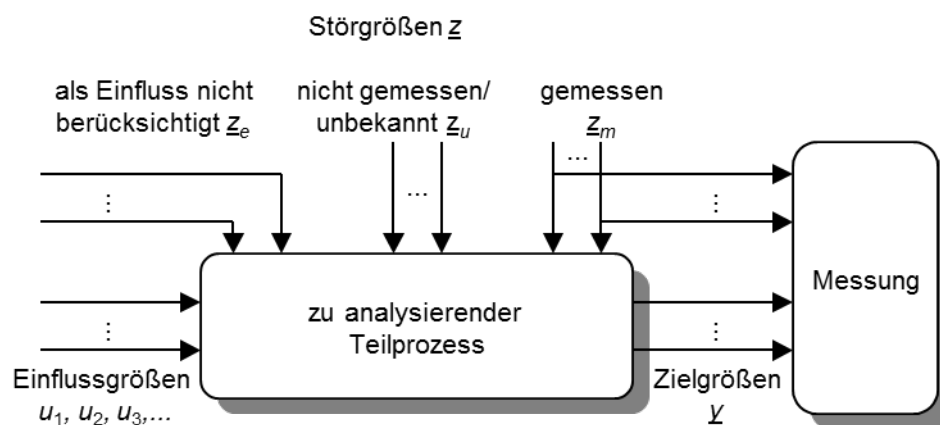
6 Versuchsplanung

6.1 Grundlagen

Möchte man Kenntnisse über einen ganz oder teilweise unbekannten Teilprozess gewinnen, wird man experimentelle Untersuchungen an diesem Teilprozess durchführen (evtl. begleitet durch simulative und theoretische Analysen). Die (*statistische*) *Versuchsplanung* – im als Englischen „*Design of Experiment (DOE)*“ bezeichnet – bietet Werkzeuge, um solche experimentellen Untersuchungen möglichst effektiv zu gestalten.

Systemtheoretische Sicht

Für die Darstellung der Zusammenhänge ist das systemtheoretischen Modell im nachstehenden Bild hilfreich: Am Ausgang des Teilprozesses charakterisieren die zu messenden *Zielgrößen* y den Prozess mit seinem Ergebnis. Im Fall eines gefertigten Bauteils kann das ein Durchmesser, die Rundheit oder die Oberflächenrauheit sein; diese Information kann auch diskret sein, z.B. „Bauteil funktioniert“, „Bauteil ist 2. Wahl“, etc. Mit Methoden der Versuchsplanung können diese Zielgrößen analysiert werden (Abschnitt 6.3).



Diese Zielgrößen werden nun durch die auf das System wirkenden Eingangsgrößen bestimmt: Alle Eingangsgrößen, deren Einfluss man untersuchen möchte und die man definiert vorgeben kann, werden als *Einflussgrößen* (oder Einflussfaktoren) u_1, u_2, u_3, \dots bezeichnet. Dies sind beispielsweise einstellbare Prozessparameter, wie Prozesstemperaturen oder -zeiten. Eine Einflussgröße muss nicht wertkontinuierlich sein, sondern kann auch eine endliche Menge diskreter Werte annehmen, z.B. das Werkstück ist aus 100Cr6, aus 16MnCr5 oder aus 20MnCr5. Die Versuchsplanung liefert Werkzeuge, um effektive Versuchspläne für die Wahl dieser Einflussgrößen aufzubauen (Abschnitt 6.2).

Alle anderen Einflüsse auf den Teilprozess werden formal den *Störgrößen* z zugezählt, die sich wiederum unterscheiden lassen:

- Es kann sich dabei um Größen z_m handeln, die man zwar nicht einstellen kann (z.B. ein chargenabhängiger Materialparameter oder die Raumtemperatur), die man aber trotzdem messtechnisch erfassen kann. Diese gemessenen Störgrößen können dadurch immerhin „kontrolliert“, d.h. in die Auswertung mit einbezogen werden. Im günstigsten Fall zeigt sich, dass die gemessene Störgröße konstant ist, sonst ist eine Blockbildung (s. nächster Abschnitt) möglich.

- Es können unbekannte oder nicht messbare Größen z_u sein, die sich somit der Analyse des Teilprozesses völlig entziehen. Entweder streuen die Messungen dadurch „zufällig“ und es sind statistische Methoden notwendig (s. nächster Abschnitt „Wiederholen von Versuchspunkten“), oder die Messung wird durch einen systematischen Fehler überlagert (s. nächster Abschnitt „Randomisierung“).
- Schließlich können Einflüsse, die man als gering erachtet, bewusst nicht zu den Einfluss-, sondern zu den Störgrößen gezählt werden, um das Modell einfach zu halten und kleine Abweichungen in Kauf zu nehmen. Abhängig von der Fragestellung der experimentellen Untersuchung unterliegt es letztlich der subjektiven Bewertung, ob eine Eingangsgröße eine Einfluss- oder eine Störgröße z_e ist!

Prinzipien der Versuchsplanung

Nachfolgend sollen einige Grundprinzipien der Versuchsplanung nach (Scheffler, 1986) dargestellt werden. Zunächst einige allgemeine Prinzipien zur experimentellen Untersuchung:

- *Wiederholen von Versuchspunkten* – Um die Wirkung zufälliger Störungen z_u zu verringern, werden Versuchspunkte wiederholt. Nur dadurch kann eine Prozessstreuung angegeben werden. Eine erwartungstreue Schätzung für die empirische Standardabweichung S der m Einzelwerte bestimmt sich aus:

$$s^2 = \frac{1}{m-1} \cdot \sum_{i=1}^m (y_i - \bar{y})^2$$

Darin berechnet sich bekanntlich der Mittelwert über die Wiederholungen als

$$\bar{y} = \frac{1}{m} \sum_{i=1}^m y_i$$

Die Standardabweichung $S\{\bar{y}\}$ dieses Mittelwertes sinkt für eine wachsende Zahl m von Wiederholungen gegenüber der Standardabweichungen des Einzelwertes σ :

$$S\{\bar{y}\} = \frac{1}{\sqrt{m}} \cdot S$$

- *Randomisierung* – Ein kurzes Fallbeispiel soll betrachtet werden: In einem Experiment wird ein Geometrieparameter eines Werkstücks von Tag zu Tag schrittweise stärker geändert und jeweils eine Verzugsgröße bestimmt. Zur Freude der Arbeitsgruppe lässt sich eine hohe Korrelation bestimmen. Tatsächlich führte aber ein Verschleiß am Messgerät zu einem täglich wachsenden Messfehler, so dass sich als Folge „vermengter Effekte“ eine Scheinkorrelation ergab, die aber keinem kausalen Zusammenhang entspricht! Durch eine zufällige Streuung (Randomisierung) der Einflussgröße über den Untersuchungszeitraum hätte sich diese Scheinkorrelation vermeiden lassen. Natürlich wäre die hohe Streuung der Messwerte aufgefallen, die dann aber einen Hinweis auf Probleme mit dem Messgerät gegeben hätte.
- *Blockbildung* – Kann eine messbare Störgröße in z_m nicht konstant gehalten werden, muss versucht werden, den Gesamtversuch in mehrere Blöcke zu unterteilen, innerhalb derer diese messbare Störgröße dann jeweils möglichst konstant ist. Sind z.B. einzelne Chargen sehr unterschiedlich, ist eine Aufteilung in je einen Block pro Charge sinnvoll. Die statistischen Auswertungen dürfen dann nur innerhalb dieser Blöcke

durchgeführt werden (um nicht „Äpfel mit Birnen“ zu vergleichen). Innerhalb der Blöcke kann und sollte nach Möglichkeit wiederum randomisiert werden.

Mit den nachfolgend genannten Prinzipien lassen sich die Zahl der erforderlichen Versuche reduzieren:

- *Nutzung von Symmetrien* – Liefern symmetrische Einflussgrößen auch symmetrische Zielgrößen? Dann können solche Symmetrien genutzt werden, um die Zahl der Versuche zu reduzieren.
- *Nutzung des Raums der unabhängigen Einflussgrößen* – Die gleichzeitige Variation mehrerer Einflussgrößen verringert die Versuchszahl (Abschnitt 6.2).
- *Vermengen* – Die systematische Überlagerung von wesentlichen Effekten und unwesentlichen Nebeneffekten verringert die Versuchsanzahl (Abschnitt 6.2, Teilfaktorielle Versuchspläne).
- *Sequentielles Experimentieren* – Um die Zahl der Experimente so gering wie möglich zu halten, kann es sinnvoll sein, die Versuchsplanung abhängig von vorhergehenden Versuchsergebnissen zu gestalten. Dadurch wird die Sequenz „Planung – Experiment – Analyse“ immer wieder durchlaufen.

Generelle Vorgehensweise

Zum Abschluss der Einführung soll die generelle Vorgehensweise der Versuchsplanung nach (Schmitt und Pfeifer, 2010) zusammengefasst werden:

1. *Systemanalyse* – Stellt sicher, dass Fachwissen in die Untersuchung einfließt, um Fehlplanungen und -interpretationen zu vermeiden: Welcher Teilprozess soll optimiert werden? Wie wird das Projektteam zusammengesetzt? Was sind die Einfluss-, Stör- und Zielgrößen? Was sind mögliche Zusammenhänge und Wechselwirkungen?
2. *Versuchsstrategie* – Festlegung des Versuchsplans (Abschnitt 6.2).
3. *Versuchsdurchführung* – streng nach Versuchsplan.
4. *Versuchsauswertung* – mit Methoden der Versuchsplanung (Abschnitt 6.3). Bei sequentiellem Experimentieren wird ggf. wieder bei Punkt 2. eingestiegen.

6.2 Versuchsplanung

Für jede der Einflussgrößen u_1, u_2, u_3, \dots müssen *Stufen* festgelegt werden, in denen die Experimente durchgeführt werden. Diese Festlegung muss aus der Erfahrung derjenigen, die das Experiment planen, getroffen werden. Da alle Werte, die außerhalb dieser Stufen liegen, nicht experimentell überprüft werden, ist diese Festlegung ganz entscheidend dafür, ob die relevanten Effekte beobachtet werden können.

Der häufigste Fall ist die Festlegung von $s = 2$ Stufen je Einflussgröße. Dies ist z.B. sinnvoll, wenn eine Variation gegen eine Standardeinstellung getestet werden soll oder ein Produkt mit neuen Eigenschaften dem bisherigen Produkt gegenüber gestellt werden soll. Für die weitere Durchführung der Experimente werden die gewählten Stufen normiert: Üblicher Weise wird die Stufe des Standards auf den Punkt „-1“, die Stufe der Variation auf „+1“ normiert. Um diesen Schritt zu dokumentieren, kann folgende Tabelle ausgefüllt werden: Am Beispiel des Drehprozesses werden für drei Einflussgrößen jeweils zwei Stufen festgelegt und auf „-1“ bzw. „+1“ normiert.

Einflussgrößen		Stufen des Originalsystems		Stufen des normierten Systems	
u_1	Schnitttiefe a_e	0,5 mm	1,0 mm	- 1	+1
u_2	Vorschub f	0,05 mm	0,1 mm	- 1	+1
u_3	Schnittgeschwindigkeit v_c	2 000 min ⁻¹	3 000 min ⁻¹	- 1	+1

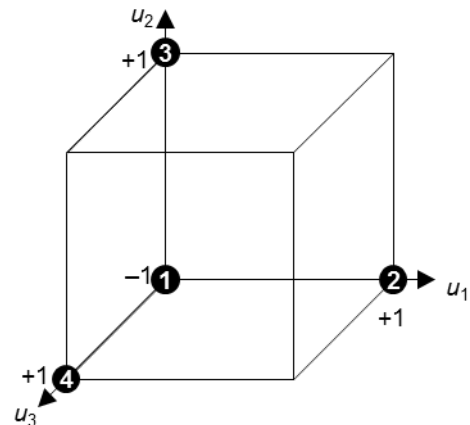
Wertet man die Wirkung der Einflussgrößen in zwei Stufen aus, können bei der Auswertung der Experimente nur lineare Beziehungen bestimmt werden; man spricht daher auch von *Versuchsplänen 1. Ordnung*. Definiert man hingegen $s = 3$ Stufen pro Einflussgröße (normiert auf „-1“, „0“, „+1“), lassen sich in einem *Versuchsplan 2. Ordnung* quadratische Beziehungen bestimmen. Natürlich steigt mit jeder neuen Stufe pro Einflussgröße die Zahl der notwendigen Experimente erheblich, so dass im Folgenden hauptsächlich Faktorpläne 1. Ordnung diskutiert werden sollen.

Einfaktormethode

Die Einfaktormethode beschreibt ein experimentelles Vorgehen, das man möglicher Weise auch intuitiv wählen würde: Ausgehend von einem ersten Versuch mit den Stufen des Standards wird dann in je einem Versuch die Stufe einer Einflussgröße variiert, während alle anderen Einflussgrößen auf der Stufe des Standards stehen. Nachfolgende Tabelle zeigt einen solchen Versuchsplan für drei Einflussgrößen.

In der Grafik rechts daneben ist die Wahl der Stufen für die Experimente entsprechend im Raum der Einflussgrößen grafisch dargestellt. Man erkennt, dass nicht alle möglichen Eckpunkte durch die Versuche belegt sind. Der große Nachteil dieses Vorgehens ist, dass mögliche Wechselwirkungen zwischen den Einflussgrößen nicht überprüft werden. Dennoch kann die Einfaktormethode für Vorversuche (Screening-Versuche) nützlich sein, um bei einer großen Zahl von Einflussgrößen die dominanten herauszufiltern. Der Aufwand für die Durchführung von Experimenten mit der Einfaktormethode steigt lediglich linear mit der Zahl der Einflussgrößen.

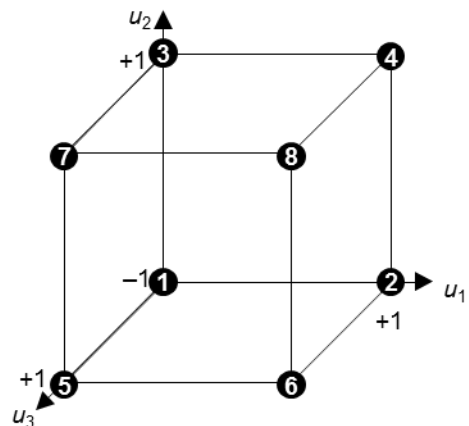
Versuchsnummer	Wahl der normierten Stufen		
	u_1	u_2	u_3
1	-1	-1	-1
2	+1	-1	-1
3	-1	+1	-1
4	-1	-1	+1



Vollfaktorielle Versuchspläne

Vollfaktorielle Versuchspläne erfassen alle möglichen Kombinationen der Stufen der Einflussgrößen. Nachfolgend sind die notwendigen Versuche für drei Einflussgrößen mit je zwei Stufen dargestellt.

Versuchsnummer	Wahl der normierten Stufen		
	u_1	u_2	u_3
1	-1	-1	-1
2	+1	-1	-1
3	-1	+1	-1
4	+1	+1	-1
5	-1	-1	+1
6	+1	-1	+1
7	-1	+1	+1
8	+1	+1	+1



Nachteilig ist, dass mit wachsender Zahl k von Einflussgrößen die Zahl der notwendigen Experimente explodiert und somit ab etwa $k = 5$ mit entsprechend $2^5 = 32$ Experimenten eine Analyse kaum noch praktisch durchführbar ist. Entsprechend werden die vollfaktoriellen Versuchspläne als 2^k -Versuchspläne bezeichnet: Das obige Beispiel ist ein 2^3 -Versuchsplan (1. Ordnung). Würde noch für jede Einflussgröße eine dritte Stufe hinzugenommen, erhielte man einen 3^3 -Versuchsplan (2. Ordnung).

Der Vorteil vollfaktorieller Analysen ist, dass alle Wechselwirkungen zwischen den Einflussgrößen bestimmt werden können. In nachstehender Tabelle wurde der 2^3 -Versuchsplan um die möglichen Wechselwirkungen $u_1 u_2$, $u_1 u_3$, $u_2 u_3$, $u_1 u_2 u_3$ ergänzt. Die Einträge in den jeweiligen Spalten entstehen formal aus der elementweisen Multiplikation der normierten Stufen der Einflussgrößen und werden später zur Auswertung herangezogen (Abschnitt 6.3). Die Wirkung der einzelnen Einflussgrößen wird als *Haupteffekt* bezeichnet, weiterhin ergeben sich *2-Faktor-Wechselwirkungen* und eine *3-Faktor-Wechselwirkung*.

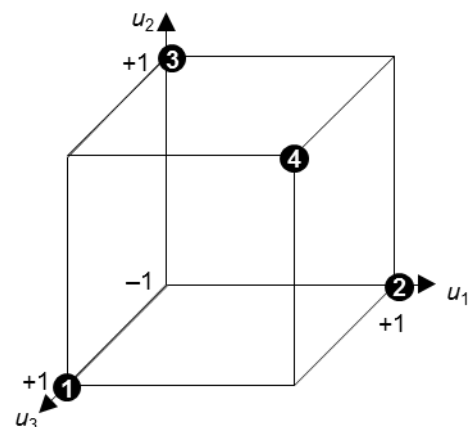
Versuchsnummer	Wahl der normierten Stufen			mögliche Wechselwirkungen			
	u_1 (HE)	u_2 (HE)	u_3 (HE)	$u_1 u_2$ (2 FW)	$u_1 u_3$ (2 FW)	$u_2 u_3$ (2 FW)	$u_1 u_2 u_3$ (3 FW)
1	-1	-1	-1	+1	+1	+1	-1
2	+1	-1	-1	-1	+1	-1	+1
3	-1	+1	-1	-1	-1	+1	+1
4	+1	+1	-1	+1	-1	-1	-1
5	-1	-1	+1	+1	-1	-1	+1
6	+1	-1	+1	-1	-1	+1	-1
7	-1	+1	+1	-1	+1	-1	-1
8	+1	+1	+1	+1	+1	+1	+1

(HE = Haupteffekt, 2 FW = 2-Faktor-Wechselwirkung, 3 FW = 3-Faktor-Wechselwirkung)

Teilfaktorielle Versuchspläne

Kann man aufgrund von Vorüberlegungen über das zu analysierende System Wechselwirkungen ausschließen, ist es nicht notwendig, diese Wechselwirkung berechnen zu können. Vielmehr kann dann anstelle dieser Wechselwirkung eine neue Einflussgröße in den Versuchsplan eingeführt werden.

Versuchsnummer	normierten Stufen		Wechselwirkung	Neue Einflussgröße
	u_1 (HE)	u_2 (HE)	$u_1 u_2$ (2 FW)	u_3 (HE)
1	-1	-1	+1	+1
2	+1	-1	-1	-1
3	-1	+1	-1	-1
4	+1	+1	+1	+1



Versuchspläne mit 4 Versuchen

Obige Tabelle zeigt dies für einen zunächst vollfaktoriellen 2^2 -Versuchsplan: Die Stufen der neuen Einflussgröße u_3 werden gleich den Stufen der Wechselwirkung $u_1 u_2$ gewählt. Dadurch entsteht ein Versuchsplan, der in nur 4 Versuchen die drei Haupteffekte untersucht, aber keine Wechselwirkung berücksichtigt.

Teilfaktorielle Versuchspläne werden als 2^{k-p} -Versuchspläne bezeichnet, wobei p die Anzahl der durch Überlagerung neu eingeführten Einflussgrößen ist. In obigem Beispiel handelt es sich also um einen 2^{3-1} -Versuchsplan.

Fasst man die Spalten der Versuchspläne als Listen auf, die man elementweise multipliziert (Matlab-Operator $\cdot *$), lässt sich formal schreiben:

$$\underline{u}_1 \cdot * \underline{u}_2 = \underline{u}_3$$

oder unter Verwendung der Eins-Liste $\underline{1}$ (deren Elemente alle +1 sind):

$$\underline{u}_1 \cdot * \underline{u}_2 \cdot * \underline{u}_3 = \underline{1}$$

Alle p Wechselwirkungen eines Versuchsplans, die $\underline{1}$ ergeben und über die eine Einflussgröße eingeführt wurde, werden als *Generatoren* G_1 bis G_p bezeichnet; hier ist also $G_1 = u_1 u_2 u_3$.

Versuchspläne mit 8 Versuchen

Mit 8 Versuchen lassen sich vollfaktoriell drei Einflussgrößen $u_1 u_2 u_3$ analysieren. Mit dem Generator $G_1 = u_1 u_2 u_3 u_4$ kann eine vierte Einflussgröße eingeführt werden. Nachstehende Tabelle zeigt den entsprechenden 2^{4-1} -Versuchsplan. Man erkennt, dass in Folge der Einführung des Faktors u_4 auch Vermengungen in anderen Spalten hervorgerufen werden (als *Aliasing* bezeichnet), die sich formal berechnen lassen (es gilt allgemein $\underline{X}^2 = \underline{1}$ und $\underline{X} \cdot * \underline{1} = \underline{X}$), z.B.:

$$\begin{aligned} \underline{u}_1 \cdot * \underline{u}_2 \cdot * \underline{u}_3 \cdot * \underline{u}_4 = \underline{1} &\Leftrightarrow \underline{u}_1 \cdot * \underline{u}_2 \cdot * (\underline{u}_3 \cdot * \underline{u}_4)^2 = \underline{u}_3 \cdot * \underline{u}_4 \Leftrightarrow \underline{u}_1 \cdot * \underline{u}_2 \cdot * \underline{1} = \underline{u}_3 \cdot * \underline{u}_4 \\ &\Leftrightarrow \underline{u}_1 \cdot * \underline{u}_2 = \underline{u}_3 \cdot * \underline{u}_4 \end{aligned}$$

Versuchsnummer	Wahl der normierten Stufen			mögliche Wechselwirkungen			
	u_1 (HE)	u_2 (HE)	u_3 (HE)	$u_1 u_2 =$ $u_3 u_4$ (2 FW)	$u_2 u_3 =$ $u_1 u_4$ (2 FW)	$u_1 u_3 =$ $u_2 u_4$ (2 FW)	$u_1 u_2 u_3 =$ u_4 (3 FW, HE)
1	-1	-1	-1	+1	+1	+1	-1
2	+1	-1	-1	-1	+1	-1	+1
3	-1	+1	-1	-1	-1	+1	+1
4	+1	+1	-1	+1	-1	-1	-1
5	-1	-1	+1	+1	-1	-1	+1
6	+1	-1	+1	-1	-1	+1	-1
7	-1	+1	+1	-1	+1	-1	-1
8	+1	+1	+1	+1	+1	+1	+1

In diesem Versuchsplan gibt es weder Vermengung von Haupteffekten (HE) noch Vermengungen von Haupteffekten (HE) und 2-Faktor-Wechselwirkungen (2 FW). Dieses wird als Lösungstyp IV bzw. als 2^{4-1}_{IV} -Versuchsplan bezeichnet.

Versuchspläne mit 16 und mehr Versuchen

Führt man eine weitere Einflussgröße u_5 ein, kann man dafür beispielsweise den Generator $G_2 = u_1 u_2 u_5$ wählen (u_5 wird $u_1 u_2$ überlagert), damit werden aber Haupteffekte und 2-Faktor-Wechselwirkungen vermengt, so dass nur noch ein Lösungstyp III bzw. ein 2^{5-2}_{III} -Versuchsplan erreicht wird. Mit Lösungstyp III lassen sich dann übrigens noch zwei weitere

Einflussgrößen einführen (z.B. mit $G_3 = u_1 u_3 u_6$ und $G_4 = u_2 u_3 u_7$), so dass man einen 2_{III}^{7-4} -Versuchsplan erhalten kann.

In analoger Weise lassen sich größere Versuchspläne aufbauen. Für 16 Versuche ist dies in (Scheffler, 1986) tabellarisch dargestellt; für größere Versuchspläne ist eine numerische Bestimmung sinnvoll (auch in „Matlab“ mit „Statistics“ Toolbox, Befehle „fullfact“ und „fracfact“). Mögliche Versuchspläne des Lösungstyps III und IV stellt nachstehende Tabelle zusammen.

Lösungstyp			Anzahl der Versuche N				
	getrennt sind	vermengt sind	4	8	16	32	64
III	HE von HE	HE mit 2 FW	$k=3$	$k=5\dots7$	$k=9\dots15$	$k=17\dots31$	$k=33\dots63$
IV	HE von 2 FW	HE mit 3 FW 2 FW mit 2 FW		$k=4$	$k=6\dots8$	$k=7\dots16$	$k=9\dots32$
Vollfaktoriell	Alle HE und alle FW	nichts	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$

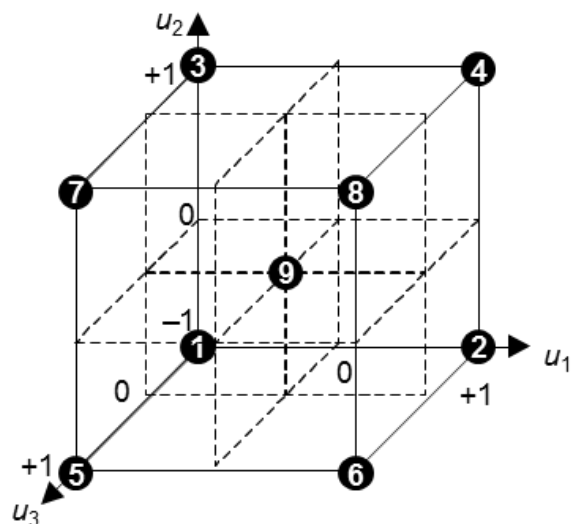
Für Prozesse mit sehr vielen Einflussfaktoren ($k = 100 \dots 1000$) wurde spezielle Versuchsstrategien entwickelt, die z.B. Screening-Versuche sequenziell vorschalten und stark vermengte Versuchspläne nutzen (z.B. nach Shainin oder Taguchi).

Überprüfung des linearen Verhaltens

Am Beginn des Abschnitts 6.2 wurde festgestellt, dass mit zwei Stufen pro Einflussgröße nur lineare Zusammenhänge erfasst werden können. Ohne gleich jeweils drei Stufen einführen zu müssen und den experimentellen Aufwand dadurch maßgeblich zu erhöhen, gibt es die Möglichkeit, durch ein einziges Experiment, das in der normierten Stufe „0“ jeder Einflussgröße liegt, die Linearität zu prüfen. Nebstehendes Bild zeigt die Lage des neuen Sternpunktes.

Für das Beispiel der Drehbearbeitung liegen die Einflussgrößen für dieses Experiment in $a_e = 0,75 \text{ mm}$, $f = 0,075 \text{ mm}$ und $v_c = 2\,500 \text{ min}^{-1}$.

Durch diese Überprüfung kann dann entschieden werden, ob drei Stufen und damit ein Versuchsplan 2. Ordnung notwendig ist.



6.3 Versuchsauswertung

Bisher wurde die effektive Wahl der Einflussgrößen in Vorbereitung auf die Versuchsdurchführung diskutiert. Um nach der Durchführung die Auswertung durchführen zu können, wird jetzt die Zielgröße y in die Betrachtung mit eingeschlossen.

Blackbox-Modell

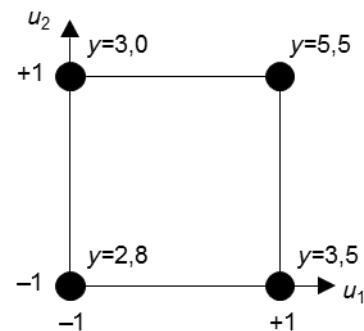
Auf Basis der durchgeführten Messungen kann ein Blackbox-Modell des Prozesses erstellt und identifiziert werden. Die Methoden dazu wurden in Kapitel 4 und 5 eingeführt. Das Modell kann dann zur Analyse des Prozesses, zu dessen Vorhersage oder zur Diagnose von Prozesses herangezogen werden.

Der Prozess kann durch ein statisches Blackbox-Modell (auch: Regressionsmodell) beschrieben werden (vgl. Abschnitt 4.1 bis 4.4). Am häufigsten werden polynomiale Ansätze verwendet. Bei einem linearen Versuchsplan werden lineare Glieder (für Haupteffekte) sowie Produkte aus Einflussfaktoren (für Wechselwirkungen) berücksichtigt. Bei höheren Versuchsplänen können entsprechend auch höhere Glieder identifiziert werden.

Beispiel: Drehbearbeitung

In Abhängigkeit der Einflüsse Schnitttiefe u_1 und Vorschub u_2 wird die für die Werkstückoberfläche erreichte Rautiefe gemessen und als Zielgröße y verwendet. Nachfolgend ist der Versuchsplan dargestellt:

Versuchsnummer	normierten Stufen		Wechselw.	Zielgröße y (Rautiefe in μm)
	u_1 (Schnitttiefe)	u_2 (Vorschub)	$u_1 u_2$	
1	-1	-1	+1	2,8
2	+1	-1	-1	3,5
3	-1	+1	-1	3,0
4	+1	+1	+1	5,5



Es wird ein polynomiales Modell angesetzt:

$$\hat{y} = a_0 + a_1 u_1 + a_2 u_2 + a_{1,2} u_1 u_2 = \underbrace{\begin{bmatrix} 1 & u_1 & u_2 & u_1 u_2 \end{bmatrix}}_{\underline{m}^T} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_{1,2} \end{bmatrix}}_{\underline{\hat{p}}}$$

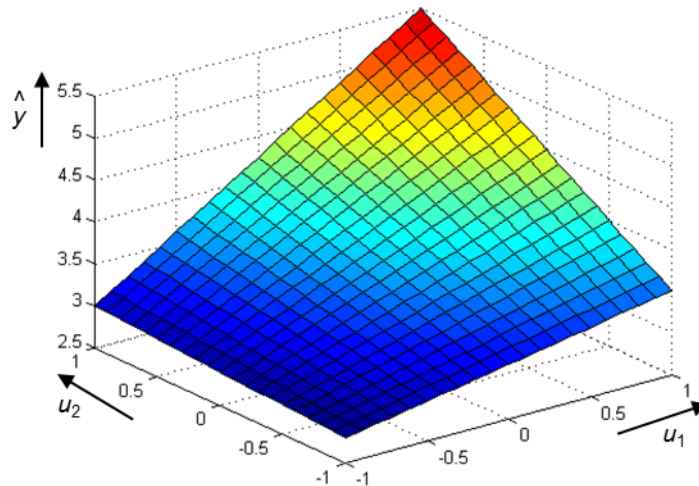
Daraus wird die Matrix \underline{M} bestimmt. Man erkennt, dass diese dem Versuchsplan entspricht, der links um eine „+1“-Spalte ergänzt wurde:

$$\underline{M} = \begin{bmatrix} \underline{m}^T(1) \\ \underline{m}^T(2) \\ \underline{m}^T(3) \\ \underline{m}^T(4) \end{bmatrix} = \begin{bmatrix} +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 \end{bmatrix}$$

Als Lösung erhält man folgende Werte für die Modellparameter:

$$\hat{\underline{p}} = \left(\underline{M}^T \cdot \underline{M} \right)^{-1} \underline{M}^T \cdot \underline{y} = \begin{bmatrix} 3,70 \\ 0,80 \\ 0,55 \\ 0,45 \end{bmatrix}$$

Nachfolgendes Bild stellt die identifizierte Modellfunktion $\hat{y} = f(u_1, u_2)$ dar. In diesem Beispiel sind es 4 Messungen und 4 Koeffizienten, so dass die Aufgabe eindeutig lösbar ist und die Eckpunkte exakt den Messungen entsprechen.



Lineare Effekte und Wechselwirkungen

Die Wirkung des Wechsels von Stufe „-1“ auf „+1“ einer Einflussgröße auf die Zielgröße \hat{y} wird auch als *Effekt* dieser Maßnahme bezeichnet. Man kann *lineare Effekte* oder *höhere Effekte* einer Einflussgröße sowie eine *Wechselwirkung* zwischen mehreren Einflussgrößen unterscheiden.

Diese Größen können auf Basis eines identifizierten Modells bestimmt werden. Für lineare Effekte gilt

$$e(u_i) = 2 a_i$$

und analog für Wechselwirkungen

$$e(u_{i,j}) = 2 a_{i,j} .$$

(Der Faktor 2 ergibt sich aus der Distanz von -1 bis +1. Damit muss die Steigung a_i multipliziert werden, um auf die Gesamtdifferenz zu kommen.)

Im vorstehenden Beispiel ist der lineare Effekt des Wechsels der Schnitttiefe $e(u_1) = 1,60$ und der lineare Effekt der Veränderung des Vorschubs $e(u_2) = 1,10$. D.h. dass bei beiden Maßnahmen die Rautiefe zunimmt. Ein Effekt von Null würde bedeuten, dass die Maßnahme keine Wirkung auf die Zielgröße hat, ein negativer Wert beschreibt eine Abnahme der Zielgröße beim Wechsel von „-1“ auf „+1“.

Die Wechselwirkung ist im Beispiel $e(u_{1,2}) = 0,90$. Diese positive Wechselwirkung bedeutet, dass das gleichzeitige Wirken der Maßnahmen u_1 und u_2 einen zusätzlichen positiven Effekt

auf die Zielgröße der Rautiefe hat, der nicht auf die linearen Effekte $e(u_1)$ bzw. $e(u_2)$ allein zurückzuführen ist.

Eine Überlagerung durch eine weitere Einflussgröße u_3 würde in diesem Fall zu Fehlinterpretationen führen. Dafür wäre es notwendig, dass die Wechselwirkung $e(u_1 u_2)$ nahe Null ist, d.h. dass keine Wechselwirkung besteht.

Signifikanz der linearen Effekte und Wechselwirkungen

Neben der absoluten Höhe der Effekte oder Wechselwirkungen ist innerhalb der Versuchsauswertung die Frage interessant, ob diese Effekte bzw. Wechselwirkungen *signifikant* im Sinne eines statistischen Tests sind.

Innerhalb einer Spalte eines Versuchsplans mit $s = 2$ Stufen bilden alle Versuche mit „-1“ die Kontrollgruppe, alle Versuche mit „+1“ die Versuchsgruppe. Es wird die Frage beantwortet, mit welcher Signifikanz ein Unterschied zwischen diesen beiden Gruppen nicht zufällig zustande gekommen ist. Dazu wird die Testhypothese „Die Mittelwerte beider Gruppen sind gleich“ aufgestellt und versucht, diese signifikant zu widerlegen. Das *Signifikanzniveau* gibt die Wahrscheinlichkeit an, mit der diese Hypothese widerlegt werden kann, siehe Tabelle:

Klasse	Aussage	p	α
–	indifferent	$< 0,95$	$> 0,05$
*	Tendenz (mehr Daten?)	$\geq 0,95$	$\leq 0,05$
**	Unterschied signifikant	$\geq 0,99$	$\leq 0,01$
***	Unterschied hoch signifikant	$\geq 0,999$	$\leq 0,001$

Um die Hypothese signifikant zu widerlegen, muss die Differenz der beiden Gruppenmittelwerte – was dem Effekt linearen $e(u_i)$ entspricht – betragsmäßig einen Grenzeffekt $e_{\text{grenz}}(p)$ überschreiten:

$$|e(u_i)| \geq e_{\text{grenz}}(p)$$

Der Grenzeffekt ist das Produkt aus der *Gesamtstandardabweichung* S der Messwerte und dem *Tabellenwert der t-Verteilung* t_{tab} :

$$e_{\text{grenz}}(p) = S \cdot t_{\text{tab}}(p, \nu)$$

Der Versuchsplan besteht aus $i = 1, \dots, N$ Zeilen, die jeweils einem Experiment entsprechen. Innerhalb jedes Experimentes soll zusätzlich eine Wiederholung von $j = 1, \dots, m$ Messungen möglich sein. Dies ist häufig sinnvoll, um zufällige Fehler zu reduzieren (siehe „Prinzipien der Versuchsplanung“ in Abschnitt 6.1). Für den Fall gleich großer Gruppen („balanciert“) gilt für die Gesamtvarianz

$$S^2 = \frac{4}{m \cdot N^2} \sum_{i=1}^N S_i^2$$

mit den Einzelvarianzen entsprechend Abschnitt 6.1

$$S_i^2 = \frac{1}{m-1} \cdot \sum_{j=1}^m (y_{i,j} - \bar{y}_i)^2 \quad \text{und} \quad \bar{y}_i = \frac{1}{m} \sum_{j=1}^m y_{i,j} .$$

Der Tabellenwert $t_{\text{tab}}(p, \nu)$ der t-Verteilung kann nachstehender Tabelle entnommen werden (alternativ in Excel: = TINV(1-p; nu)).

Da es sich um einen zweiseitigen Test handelt (der Grenzwert kann über- oder unterschritten werden), ist für p der zweiseitige Fall zu Grunde zulegen. Schließlich wird die Zahl der *Freiheitsgrade* ν benötigt. Für jeden Einzelversuch gilt: Freiheitsgrade = Messwiederholungen – 1; insgesamt gilt also:

$$\nu = N \cdot (m - 1)$$

Perzentile der t -Verteilung:

einseitig: $p =$										
	0,6	0,75	0,90	0,95	0,975	0,99	0,995	0,9975	0,999	0,9995
				*		**			***	
zweiseitig:										
ν	0,3	0,50	0,80	0,90	0,95	0,98	0,99	0,995	0,998	0,999
					*		**			***
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	5.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.812	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
200	.254	.676	1.286	1.653	1.972	2.345	2.601	2.839	3.131	3.340
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Beispiel: Drehbearbeitung (erweitert)

Versuchsplan und Ergebnisse für Zielgröße mit $m = 2$ Messwiederholungen:

	Versuchsplan			Zielgröße			
	u_1	u_2	$u_1 u_2$	$y_{i,1}$	$y_{i,2}$	Mittelwert	Varianz S^2
1	-1	-1	1	2,60	3,00	2,80	0,080
2	1	-1	-1	3,40	3,60	3,50	0,020
3	-1	1	-1	3,00	3,00	3,00	0,000
4	1	1	1	5,10	5,90	5,50	0,320
Effekte =	1,60	1,10	0,90			Summe =	0,420
	**	**	*				

Berechnung:

- Gesamtvarianz: $S^2 = \frac{4}{2 \cdot 4^2} \cdot 0,42 = 0,053$ daraus Standardabweichung: $S = 0,229$
- Freiheitsgrade: $\nu = 4 \cdot (2 - 1) = 4$
- Grenzeffekte:
 für Signifikanz * : $t_{tab}(\nu = 4, p_{zweis} = 0,95) \cdot S = 2,776 \cdot 0,229 = 0,636$
 für Signifikanz ** : $t_{tab}(\nu = 4, p_{zweis} = 0,99) \cdot S = 4,604 \cdot 0,229 = 1,055$
 für Signifikanz *** : $t_{tab}(\nu = 4, p_{zweis} = 0,999) \cdot S = 8,610 \cdot 0,162 = 1,973$
 Durch den Vergleich mit den Effekten kann die Signifikanz in der letzten Zeile obiger Tabelle bestimmt werden.

Test auf Adäquatheit des Modellansatzes

Bei der Wahl eines Modellansatzes werden häufig Vereinfachungen gemacht, um das Modell so einfach wie möglich zu halten. Beispielsweise werden Wechselwirkungen im Ansatz nicht berücksichtigt.

In Form eines statistischen Tests wird die These „Das Modell ist adäquat“ untersucht. Dazu wird ein statistischer Test auf Varianzunterschied mit Hilfe des F-Tests durchgeführt.

- Zum einen wird die durch die Modellungenauigkeit entstehende Varianz („Reststreuung“) bestimmt:

$$S_1^2 = \frac{1}{\nu_1} \sum_{i=1}^N (\bar{y}_i - \hat{y}_i)^2 \quad \text{mit } \nu_1 = N - k - 1$$

Darin nummeriert $i = 1, \dots, N$ die Versuche entsprechend der Zeilen im Versuchsplan durch. Die Anzahl der Einflussgrößen wird mit k bezeichnet. Weiterhin ist \bar{y}_i der Messwert der Zeile i des Versuchsplans (evtl. gemittelt aus mehreren Einzelmessungen) und \hat{y}_i der durch das Modell bestimmte Ausgangswert.

- Zum anderen wird die *Streuung in der experimentellen Durchführung* bestimmt. Dazu werden die Experimente (in jeder Zeile des Versuchsplans) m mal wiederholt. Die mittlere Varianz über die Einzelexperimente ist dann:

$$S_2^2 = \frac{1}{\nu_2} \sum_{i=1}^N \sum_{j=1}^m (y_{i,j} - \bar{y}_i)^2 \quad \text{mit } \nu_2 = N(m-1)$$

- Um den Test durchzuführen, werden beide Varianzen ins Verhältnis gesetzt. Die Testgröße ist:

$$F_{\text{exp}} = \frac{S_1^2}{S_2^2}$$

- Die zu widerlegende Hypothese lautet: „Die Streuungen σ_1 und σ_2 sind gleich.“ Dann wäre das Modell adäquat, da sich die Streuung σ_1 des Modells in derselben Größenordnung wie die Streuung σ_2 der Messungen bewegt. Um die Adäquatheit zu widerlegen, muss σ_1 signifikant größer sein als σ_2 .
- Dazu liest man den Tabellenwert $F_{\text{tab},p}(\nu_1, \nu_2)$ ab (siehe nachfolgend). Falls $F_{\text{exp}} > F_{\text{tab}}$ gilt, ist $\sigma_1 > \sigma_2$ mit der Signifikanz p . Dann ist die Hypothese widerlegt und das Modell ist nicht adäquat.

Beispiel: Drehbearbeitung (erweitert)

Das Beispiel der Drehbearbeitung wird um die Analyse der Adäquatheit fortgesetzt.

Fall 1: Verwendet man das auf S. 6-9 und 6-10 ermittelte Modell, werden alle 4 durchgeführten Messungen exakt durch das Modell wiedergegeben: $\bar{y}_i = \hat{y}_i$ für $i = 1, \dots, 4$. Dann gilt $S_1 = 0$ und $F_{\text{exp}} = 0$, so dass die These, das Modell sei adäquat, nicht widerlegt wird.

Fall 2: Das Modell wird reduziert, in dem die Wechselwirkung nicht berücksichtigt wird. Man erhält nun das Modell $\hat{y} = 3,7 + 0,8 u_1 + 0,55 u_2$ und den folgenden Versuchsplan:

	Versuchsplan			Zielgröße				Modell
	u_1	u_2	$u_1 u_2$	$y_{i,1}$	$y_{i,2}$	Mittelwert	Varianz S^2	
1	-1	-1	1	2,60	3,00	2,80	0,080	2,350
2	1	-1	-1	3,40	3,60	3,50	0,020	3,950
3	-1	1	-1	3,00	3,00	3,00	0,000	3,450
4	1	1	1	5,10	5,90	5,50	0,320	5,050

- Die Reststreuung ist:

$$S_1^2 = \frac{1}{\nu_1} \sum_{i=1}^N (\bar{y}_i - \hat{y}_i)^2 = 0,81 \text{ mit } \nu_1 = N - k - 1 = 4 - 2 - 1 = 1$$

- Die Einzelstreuung ist:

$$S_2^2 = \frac{1}{\nu_2} \sum_{i=1}^N \sum_{j=1}^m (y_{i,j} - \bar{y}_i)^2 = 0,105 \text{ mit } \nu_2 = N(m-1) = 4 \cdot 1 = 4$$

- Damit ist die Testgröße $F_{\text{exp}} = \frac{S_1^2}{S_2^2} = \frac{0,81}{0,105} = 07,714$.

Aus der Tabelle liest man ab: $F_{\text{tab},0,95}(\nu_1 = 1, \nu_2 = 4) = 7,71$. Damit ist die Hypothese mit einer Signifikanz von 95% (knapp) widerlegt. Das heißt das Modell ist nicht adäquat.

Perzentile der F-Verteilung:

 $F_{tab\ 0,95}(\nu_1, \nu_2)$ (für Signifikanz *, $p=0,95$, $\alpha=0,05$):

ν_2	ν_1 Freiheitsgrad im Zähler										ν_2	ν_1 Freiheitsgrad im Zähler								
	1	2	3	4	5	6	7	8	9	10		12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	1	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	2	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	3	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	4	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	5	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	6	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	7	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	8	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	9	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	10	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	11	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	12	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	13	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	14	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	15	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	16	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	17	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	18	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	19	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	20	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	21	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	22	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	23	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	24	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	25	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	26	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	27	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	28	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	29	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	30	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	40	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	60	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	120	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	∞	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Perzentile der F-Verteilung $F_{tab\ 0,99}(\nu_1, \nu_2)$ (für Signifikanz **, $p=0,99$, $\alpha=0,01$):

ν_2	ν_1 Freiheitsgrad im Zähler										ν_2	ν_1 Freiheitsgrad im Zähler								
	1	2	3	4	5	6	7	8	9	10		12	15	20	24	30	40	60	120	∞
1	4052.	5000.	5403.	5625.	5764.	5859.	5928.	5981.	6022.	6056.	1	6106.	6157.	6209.	6235.	6261.	6287.	6313.	6339.	6366.
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	2	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	3	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	4	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	5	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	6	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	7	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	8	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	9	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	10	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	11	4.40	4.25	4.10	4.03	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	12	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	13	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	14	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	15	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	16	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	17	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	18	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	19	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	20	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	21	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	22	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	23	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	24	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	25	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	26	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	27	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	28	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	29	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	30	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	40	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	60	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	120	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	∞	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

7 Identifikation von Systemantworten und Signalen

In den Kapiteln 5 und 6 wurde die Identifikation von statischen Funktionszusammenhängen gezeigt. Dies soll nun auf Zeitfunktionen

$$y(t, \underline{p})$$

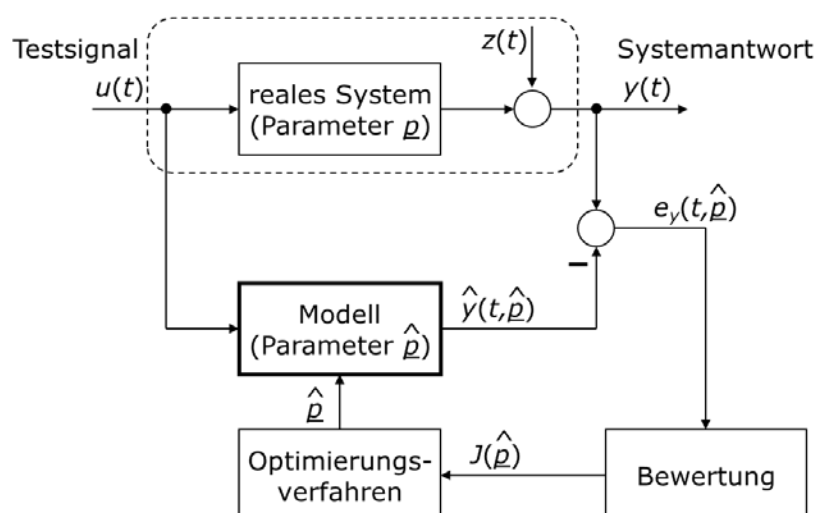
übertragen werden. Sie beschreiben also einen von der Zeit t abhängigen Verlauf, der auch von (unbekannten) Parametern in \underline{p} abhängt. Solche Zeitfunktionen können insbesondere Systemantworten (Abschnitt 7.1) oder Signale (Abschnitt 7.2) sein.

7.1 Systemantworten

Es sollen die Parameter eines realen dynamischen Systems identifiziert werden. Dazu wird das System mit einer bekannten Testfunktion $u(t)$ angeregt. Die Identifikation erfolgt nun über den Vergleich der *Systemantworten*: Dazu wird die Systemantwort $\hat{y}(t, \hat{\underline{p}})$ z.B. auf Basis der Übertragungsfunktion $G(s, \hat{\underline{p}})$ des System und der Testfunktion $U(s)$ berechnet

$$\hat{y}(t, \hat{\underline{p}}) = L^{-1}\{G(s, \hat{\underline{p}}) U(s)\}$$

und dann mit den gemessenen Werten der Systemantwort des realen System verglichen:



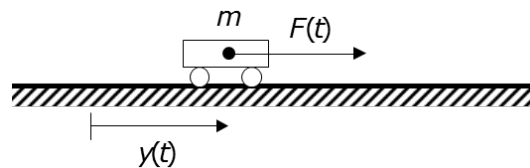
Zur Bewertung kann wieder eine geeignete Kostenfunktion frei gewählt werden. Sehr häufig wird der quadratische, über ein Zeitintervall T_0 integrierte Ausgangsfehler verwendet:

$$J(\hat{\underline{p}}) := \int_0^{T_0} e_y^2(t, \hat{\underline{p}}) dt$$

Praktisch werden diskrete Messzeitpunkte t_k mit $k = 1, \dots, N$ verwendet, so dass eine Summation anstelle des Integrals tritt:

$$J(\hat{\underline{p}}) := \sum_{k=1}^N e_y^2(t_k, \hat{\underline{p}})$$

Damit ist diese Identifikationsaufgabe auf das aus Kapitel 5 bekannte Problem zurückgeführt, so dass auch die Algorithmen zur Optimierung von dort genutzt werden können.

Beispiel 7-1: Parameteridentifikation eines gleichmäßig beschleunigten Fahrzeugs

Ein Fahrzeug mit unbekannter Masse m wird ab $t = 0$ mit einer konstanten Kraft F_0 beschleunigt:

$$F(t) = F_0 \cdot \sigma(t) \quad \text{mit} \quad F_0 = 4 \text{ N}$$

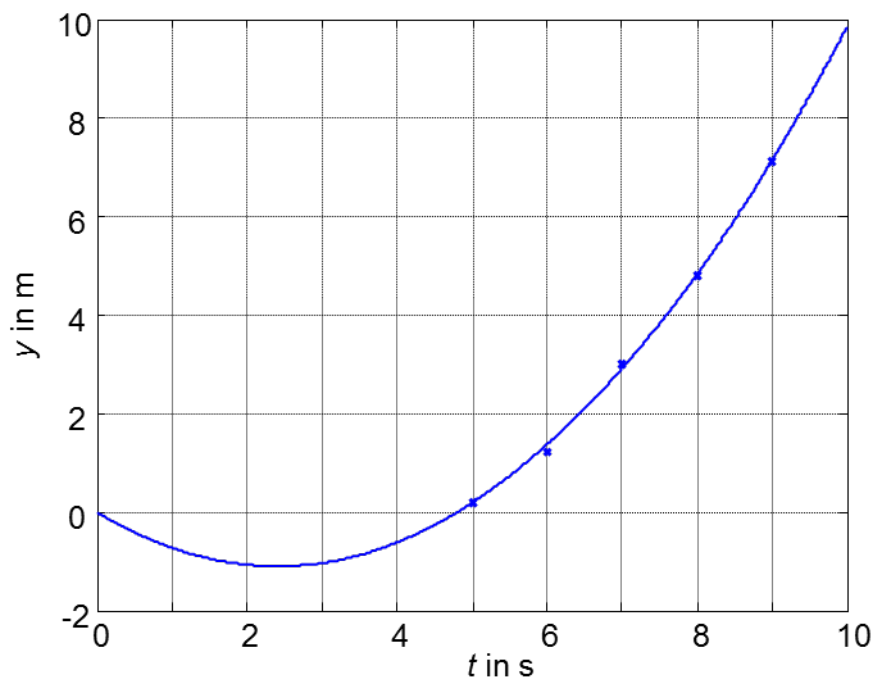
Zu den Zeitpunkten t_k wird die Position des Fahrzeugs $y(t_k)$ gemessen:

k	1	2	3	4	5
t_k in s	5,0	6,0	7,0	8,0	9,0
$y(t_k)$ in m	0,2	1,2	3,0	4,8	7,1

Für die Modellbildung soll das 2. Newtonsche Bewegungsgesetz zugrunde gelegt werden.

Das Fahrzeug befindet sich anfangs in $y(0) = 0$. Bestimmen Sie die unbekannte Masse m und die unbekannte Anfangsgeschwindigkeit $\dot{y}(0) = v_0$ des Fahrzeugs durch Parameteridentifikation.

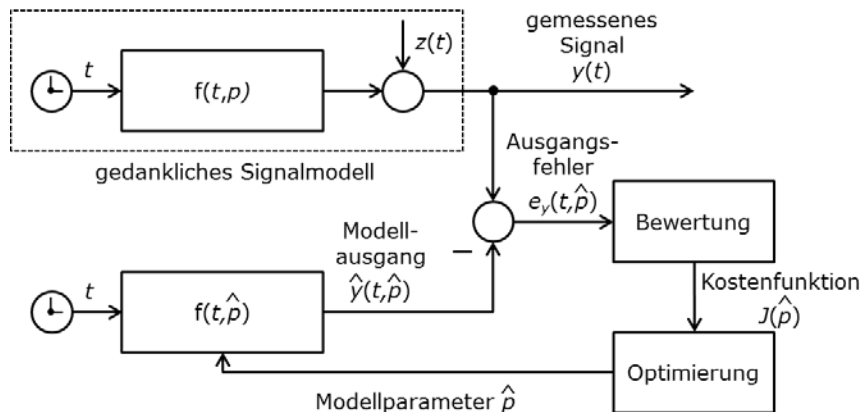
Nachstehend ist die Antwort des identifizierten Modells (durchgezogene Linie) im Vergleich zu den Messwerten (Kreuze) dargestellt:



7.2 Signale

Das Modell eines Signals charakterisiert die Signaleigenschaften und ermöglicht eine im Vergleich zu den Rohdaten kompaktere Darstellung. Es kann u.A. zur Vorhersage des zukünftigen Signalverlaufs oder zur Signalanalyse genutzt werden.

Im Sinne einer Modellvorstellung denkt sich das gemessene Signal $x(t)$ aus einer Signalquelle heraus entstanden. Darin erzeugt eine Zeitfunktion f ein Signal, das ggf. nach Störung durch $z(t)$, als Signal $y(t)$ gemessen werden kann:



Diese gedachte Signalquelle wird durch ein paralleles Signalmodell mit der Zeitfunktion \hat{f} nachgebildet, so dass das gemessene Signal $y(t)$ mit dem Modellausgang $\hat{y}(t)$ verglichen werden kann. Die Modellparameter des Signalmodells werden im Parametervektor $\underline{\hat{p}}$ zusammengefasst. Der Ausgangsfehler ist:

$$e_y(t, \underline{\hat{p}}) = y(t) - \hat{y}(t, \underline{\hat{p}})$$

Die zu lösende Aufgabe besteht nun wiederum darin, den Parametervektor $\underline{\hat{p}}$ zu finden, der den Ausgangsfehler $e_y(t, \underline{\hat{p}})$ bestmöglich minimiert.

Bewertung

In der Regel wird eine quadratische Kostenfunktion über dem Zeitintervall T_0 verwendet:

$$J(\underline{\hat{p}}) := \int_0^{T_0} e_y^2(t, \underline{\hat{p}}) dt$$

In der Praxis steht der Ausgangsfehler nur in *zeitdiskreten* Messzeitpunkten $t = kT$ mit $k=1, \dots, N$ und der Abtastzeit T zur Verfügung. Für zeitdiskrete Signale beschränken wir uns zukünftig auf eine kurze Notation nur unter Verwendung des Zählers k :

$$y(k) := y(t = kT)$$

Als Kostenfunktion kann dann

$$J(\underline{\hat{p}}) := \sum_{k=1}^N e_y^2(kT, \underline{\hat{p}}) = \underline{e}_y^T(\underline{\hat{p}}) \cdot \underline{e}_y(\underline{\hat{p}}) \quad \text{mit} \quad \underline{e}_y(\underline{\hat{p}}) = \begin{bmatrix} e_y(1T, \underline{\hat{p}}) \\ \vdots \\ e_y(NT, \underline{\hat{p}}) \end{bmatrix}$$

verwendet werden.

Damit ist auch diese Identifikationsaufgabe auf das aus Kapitel 5 bekannte Problem zurückgeführt, so dass auch die Algorithmen zur Optimierung von dort genutzt werden können. Nachfolgend sollen die beiden Abschnitt 4.5 eingeführten Ansätze für Signalmodelle identifiziert werden.

Beispiel 7-2: Polynomiales Modell

Es wird angenommen, das ungestörte Signal entspräche einer polynomialen Zeitfunktion:

$$f(kT) = \sum_{i=0}^n a_i (kT)^i$$

Eine Zerlegung in einen linearen Ansatz ist hier möglich

$$f(kT) = y(k, p) = \underline{m}^T(k) \cdot \underline{p} = \begin{bmatrix} 1 & kT & \dots & (kT)^n \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}$$

und es gilt folglich für N Messungen:

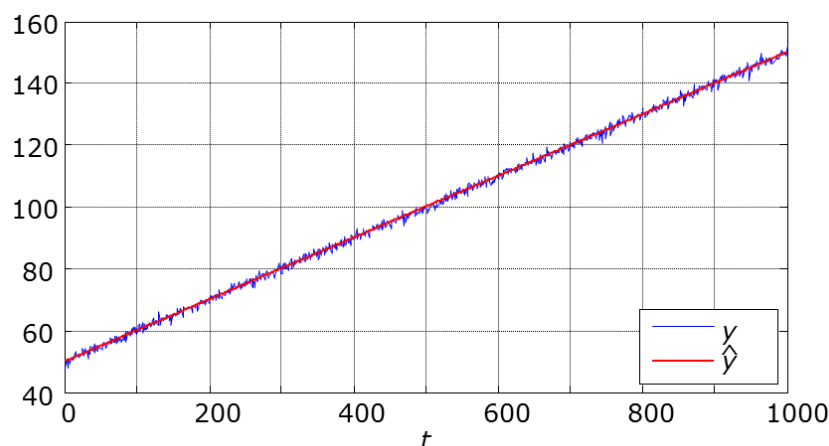
$$\underline{M} = \begin{bmatrix} 1 & 1T & \dots & (1T)^n \\ 1 & 2T & \dots & (2T)^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & NT & \dots & (NT)^n \end{bmatrix}$$

Beispiel

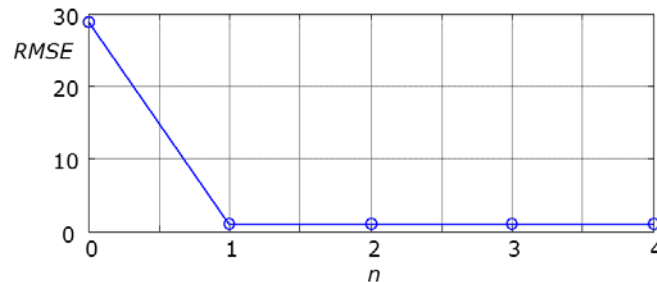
Ein gestörter, zeitlich linearer Funktionsverlauf $y(k)$ mit $k=1, \dots, 1000$ und $T=1$ wird gemessen (siehe Diagramm). Es wird eine linearer Polynomansatz ($n=1$) identifiziert, und man erhält beispielsweise

$$\underline{\hat{p}} = \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \end{bmatrix} = \begin{bmatrix} 50,034 \\ 0,099 \end{bmatrix}.$$

Nachstehend ist auch der Modellausgang $\hat{y}(k, \hat{p}) = \underline{m}^T(k) \cdot \underline{\hat{p}}$ dargestellt.



Für gemessene Daten ist häufig nicht klar, welche Ordnung n für den Modellansatz gewählt werden soll. Da es sich beim Signalmodell immer um eine näherungsweise Beschreibung handelt, wird es häufig auch keine „richtige“ oder „falsche“ Systemordnung geben. Man kann z.B. als Fehlermaß den RMSE über der Ordnung n auftragen, und entscheiden, welchen maximalen Fehler man nicht überschreiten möchte. In diesem einfachen Beispiel erkennt man, dass eine Erhöhung über $n=1$ keine weitere Verringerung des Fehlers bringt:



Beispiel 7-3: Periodisches Modell

Für harmonische Signalverläufe kann folgender Modellansatz gewählt werden:

$$f(kT) = \sum_{i=1}^n [a_i \cos(\omega_i kT) + b_i \sin(\omega_i kT)]$$

Darin sind ω_i die für das Signal wesentlichen Kreisfrequenzen. Diese müssen bekannt sein! Häufig kann man diese aus zusätzlichem Wissen über den Prozess ableiten. Bei einer Außentemperaturmessung kann z.B. vermutet werden, dass der Zeitverlauf die Periode T_1 eines Tages und die Periode T_2 eines Jahres besitzen wird. Entsprechend können $\omega_1 = 2\pi/T_1$ und $\omega_2 = 2\pi/T_2$ gewählt werden. Alternativ kann man vorab eine Signalanalyse im Frequenzbereich durchführen, um relevante Kreisfrequenzen ω_i zu finden.

Eine Zerlegung in einen linearen Ansatz ist auch hier möglich:

$$f(kT) = x(k, p) = \underline{m}^T(k) \cdot \underline{p} = [\cos(\omega_1 kT) \quad \dots \quad \cos(\omega_n kT) \quad | \quad \sin(\omega_1 kT) \quad \dots \quad \sin(\omega_n kT)] \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Und es ist:

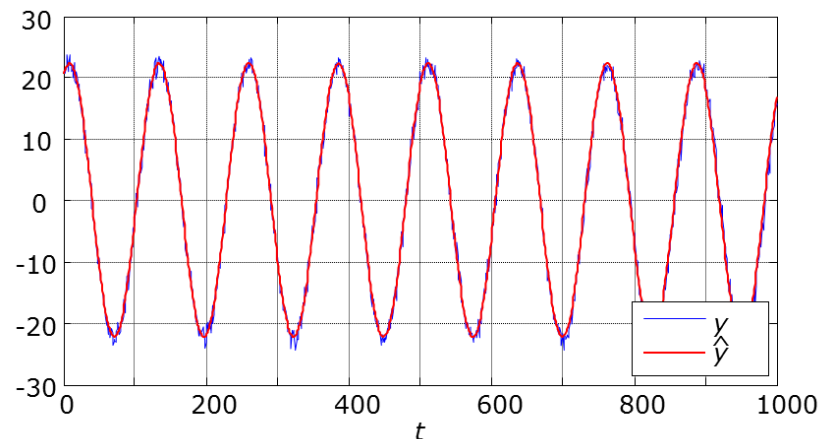
$$\underline{M} = \begin{bmatrix} \cos(\omega_1 1T) & \dots & \cos(\omega_n 1T) & | & \sin(\omega_1 1T) & \dots & \sin(\omega_n 1T) \\ \vdots & & \vdots & & \vdots & & \vdots \\ \cos(\omega_1 NT) & \dots & \cos(\omega_n NT) & | & \sin(\omega_1 NT) & \dots & \sin(\omega_n NT) \end{bmatrix}$$

Beispiel

Ein gestörtes, harmonisches Signal $x(k)$ wird mit $k=1, \dots, 1000$ und $T=1$ gemessen (siehe Diagramm). Es wird eine Schwingung mit $\omega_1 = 0,05$ identifiziert. Man erhält beispielsweise

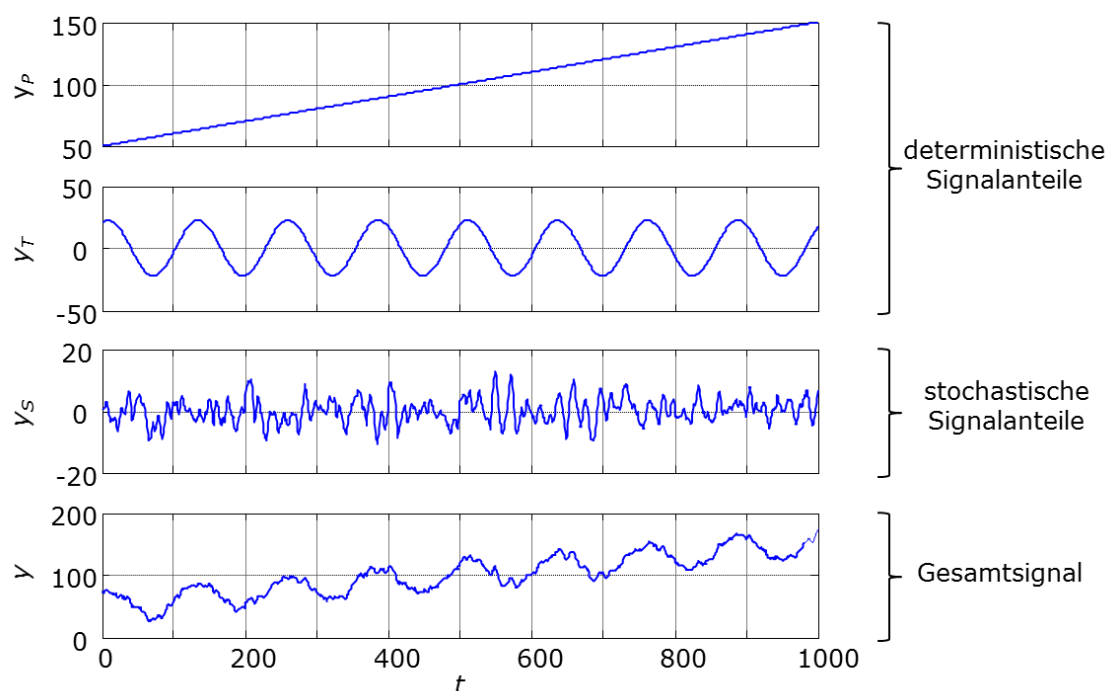
$$\underline{\hat{p}} = \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix} = \begin{bmatrix} 19,99 \\ 9,91 \end{bmatrix}$$

Nachstehend ist auch der Modellausgang $\hat{y}(k, \hat{p}) = \underline{m}^T(k) \cdot \underline{\hat{p}}$ dargestellt.



Zusammengesetztes Signalmodell

Als Modellvorstellung wird angenommen, dass sich ein gemessenes Signal als Gesamtsignal aus der Superposition mehrerer Signalanteile ergibt. Das nachfolgende Beispiel zeigt, wie das Gesamtsignal x aus der Addition von drei Signalanteilen x_P , x_T und x_S entsteht.



Um zu einem Signalmodell zu gelangen, wird in folgender Weise vorgegangen:

1. Auf Basis des vorliegenden gemessenen Signals wird entschieden, aus welchen wesentlichen Signalanteilen es aufgebaut ist. Für die Signalanteile werden jeweils Modellansätze festgelegt.
2. Die Modellparameter der Teilmodelle werden identifiziert.
3. Die Superposition der nun bekannten Teilmodelle bildet das gesuchte Modell des Gesamtsignals.

8 Identifikation dynamischer Systeme

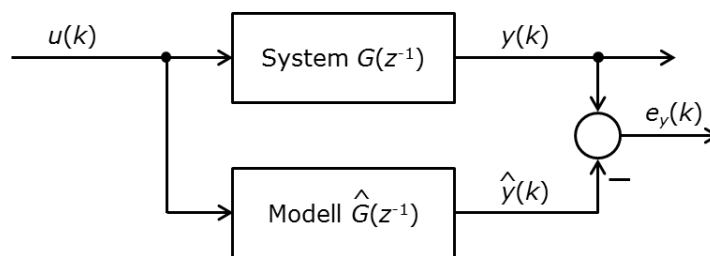
Ein unbekanntes System wird mit einem zeitdiskreten Eingangssignal $u(k)$ angeregt und antwortet mit dem Ausgangssignal $y(k)$. Die Aufgabe ist nun, das unbekannte Übertragungsverhalten dieses Systems aus den Signalen $u(k)$ und $y(k)$ zu bestimmen.

Zur Charakterisierung eines linearen und zeitinvarianten Systems wird eine rationale zeitdiskrete Übertragungsfunktion

$$\hat{G}(z^{-1}) = \frac{\hat{b}_0 + \hat{b}_1 z^{-1} + \dots + \hat{b}_m z^{-m}}{\hat{a}_0 + \hat{a}_1 z^{-1} + \dots + \hat{a}_n z^{-n}} = \frac{\hat{B}(z^{-1})}{\hat{A}(z^{-1})}$$

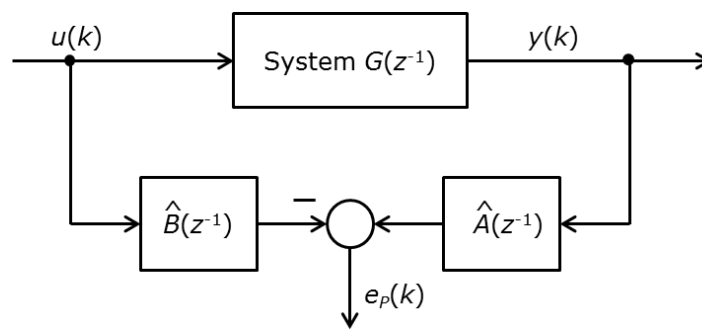
als Modell angesetzt (vergleiche Abschnitt 4.6). Es müssen also die Koeffizienten \hat{a}_i und \hat{b}_i identifiziert werden. Die Übertragungsfunktion kann durch \hat{a}_0 dividiert werden, so dass ohne Beschränkung der Allgemeinheit $\hat{a}_0 = 1$ gesetzt werden kann. Die $(m+n+1)$ unbekannten Parameter können zu einem Parametervektor $\underline{\hat{p}} = [\hat{b}_0 \dots \hat{b}_m \mid \hat{a}_1 \dots \hat{a}_n]^T$ zusammengefasst werden. Mit diesem Modellansatz ist die Aufgabe auf die Bestimmung der Modellparameter $\underline{\hat{p}}$ reduziert.

Die naheliegende Methode ist die Berechnung eines *Ausgangsfehlers* $e_y(k)$ als Differenz des gemessenen Ausgangs $y(k)$ und dem Modellausgang $\hat{y}(k)$:



Möchte man den Modellausgang $\hat{y}(k)$ auf Basis der messbaren Größen $u(k)$, $y(k)$ und des Modells $\hat{G}(z^{-1})$ berechnen, gehen die Parameter in $\underline{\hat{p}}$ nichtlinear in diesen Ansatz ein. Die Bestimmung von $\underline{\hat{p}}$ ist dann aber ein nichtlineares Optimierungsproblem, für das keine geschlossene Lösung angegeben werden kann.

In diesem Punkt erweist sich die Einführung eines modifizierten Fehlers als zielführend: Formal wird der bisherige Ausgangsfehler durch das System $\hat{A}(z^{-1})$ übertragen, um die neue Fehlerdefinition $E_G(z) = \hat{A}(z^{-1})E_y(z)$ zu erhalten. Im Blockschaltbild kann man den Block $\hat{A}(z^{-1})$ und die Summationsstelle tauschen, dabei steht $\hat{A}(z^{-1})$ dann jeweils im Zweig von $y(k)$ sowie von $\hat{y}(k)$. In letzterem kürzt sich $\hat{A}(z^{-1})$ mit dem Nenner von $\hat{G}(z^{-1})$ und man erhält folgende Struktur:



Aus dem Blockschaltbild liest man ab:

$$E_G(z) = Y(z) \hat{A}(z^{-1}) - U(z) \hat{B}(z^{-1})$$

Der Fehler $e_p(k)$ ist also eine Linearkombination des Eingangs und seiner zurückliegenden Werte (gewichtet mit $\hat{B}(z^{-1})$) sowie des Ausgangs und seiner zurückliegenden Werte (gewichtet mit $\hat{A}(z^{-1})$). Die Parameter in \hat{p} gehen daher alle linear ein.

Der Fehler $e_p(k)$ lässt sich als *Prädiktionsfehler* interpretieren: Man wandelt die Fehlergleichung in den diskreten Zeitbereich

$$e_p(k) = 1 y(k) + \hat{a}_1 y(k-1) + \dots + \hat{a}_n y(k-n) - \hat{b}_0 u(k) - \dots - \hat{b}_m u(k-m)$$

und in eine vektorielle Darstellung:

$$e_p(k, \hat{p}) = y(k) - \underbrace{\begin{bmatrix} u(k) & \dots & u(k-m) & - & y(k-1) & \dots & - & y(k-n) \end{bmatrix}}_{\underline{m}^T(k)} \cdot \underbrace{\begin{bmatrix} \hat{b}_0 \\ \vdots \\ \hat{b}_m \\ \hat{a}_1 \\ \vdots \\ \hat{a}_n \end{bmatrix}}_{\hat{p}} = y(k) - \underbrace{\underline{c}^T(k)}_{\tilde{y}(k, \hat{p})} \cdot \hat{p}$$

Darin kann man das Signal $\tilde{y}(k, \hat{p})$ (das nicht im Blockschaltbild auftaucht!) als eine Prädiktion für $y(k)$ verstanden werden. Bei bekannten Eingängen und Modellparametern wird auf Basis der zurückliegenden Ausgänge $y(k-1), \dots, y(k-n)$ eine Schätzung um einen Zeitschritt für den aktuellen Wert $y(k)$ bestimmt. Somit ist $e_p(k, \hat{p}) = y(k) - \tilde{y}(k, \hat{p})$ der Gleichungsfehler bei der Prädiktion um einen Zeitschritt.

Man kann diese Gleichung nun für mehrere diskrete Zeitpunkte $k=N_0, \dots, N$ zeilenweise untereinander schreiben. Da die Parameter \hat{p} in jedem Zeitschritt gleich sein sollen, können sie nach rechts ausmultipliziert werden:

$$\underbrace{\begin{bmatrix} e_G(N_0, \underline{p}) \\ \vdots \\ e_G(N, \underline{p}) \end{bmatrix}}_{\underline{e}_p(\underline{\hat{p}})} = \underbrace{\begin{bmatrix} y(N_0) \\ \vdots \\ y(N) \end{bmatrix}}_{\underline{y}} - \underbrace{\begin{bmatrix} u(N_0) & \dots & u(N_0 - m) & | & -y(N_0 - 1) & \dots & -y(N_0 - n) \\ \vdots & & \vdots & | & \vdots & & \vdots \\ u(N) & \dots & u(N - m) & | & -y(N - 1) & \dots & -y(N - n) \end{bmatrix}}_{\underline{M}} \cdot \underbrace{\begin{bmatrix} \hat{b}_0 \\ \vdots \\ \hat{b}_m \\ \hat{a}_1 \\ \vdots \\ \hat{a}_n \end{bmatrix}}_{\underline{\hat{p}}}$$

Anmerkung: Beginnt man in der ersten Zeile erst mit Zeitpunkt $N_0 = \max(m, n) + 1$, stehen genug Vergangenheitswerte von u und y zur Verfügung, so dass nur positive Zeitschritte verwendet werden.

Die optimale Lösung für $\underline{\hat{p}}$ im Sinne einer minimalen Kostenfunktion J , die als kleinste Fehlerquadrate („Least square“) definiert wird,

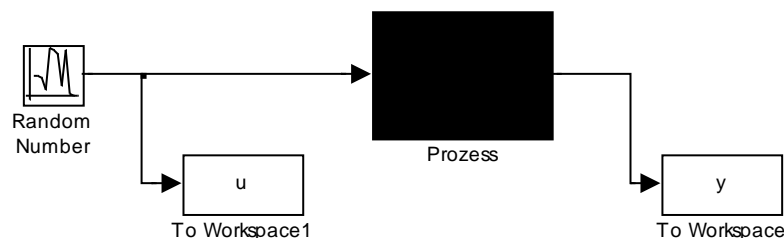
$$J(\underline{\hat{p}}) = \sum_{k=1}^N (y(k) - \tilde{y}(k, \underline{\hat{p}}))^2 = \underline{e}_p^T(\underline{\hat{p}}) \cdot \underline{e}_p(\underline{\hat{p}})$$

liefert schließlich:

$$\underline{\hat{p}} = (\underline{M}^T \underline{M})^{-1} \underline{M}^T \underline{y}$$

Beispiel 8-1: Identifikation eines unbekannten zeitdiskreten Prozesses

Ein zunächst unbekanntes System (schwarzer Block im nachstehenden Simulink-Blockschaltbild) wird am Eingang u mit einem stochastischen Prozess angeregt. Es werden $N = 10000$ Abtastwerte des Eingangs u und des (ungestörten) Ausgangs y gespeichert.



Nachfolgend ist der Matlab-Algorithmus zur Schätzung des Parametervektors $\underline{\hat{p}}$ dargestellt:

```
% Festlegung der Ordnung der ÜF:
m = 2;
n = 3;

% Anzahl N der Messungen aus Ausgangsvektor y:
N = length(y);

% Erster Messwert:
N0 = max(m, n) + 1;

% Vordefinitionen:
M = zeros(N - N0 + 1, m + 1 + n);

% Aufbau Matrix M zeilenweise:
for k = N0 : N % Zeitschritte
```

```

M(k-N0+1,:) = [ u(k:-1:k-m)' -y(k-1:-1:k-n)' ];
end

% Lösung für ges. Parameter:
p_dach = M\Y(N0:end)

```

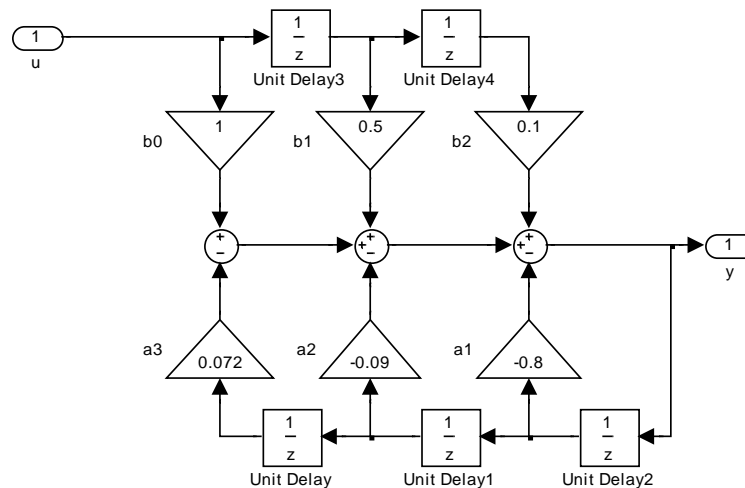
Man erhält als Ergebnis:

```

p_dach =
    1.0000
    0.5000
    0.1000
   -0.8000
   -0.0900
    0.0720

```

Nachfolgend ist das Blockschaltbild des zunächst unbekannten Prozesses dargestellt; die Parameter wurden also korrekt identifiziert:



Teil C: Simulation

9 Explizite Verfahren

9.1 Grundlagen

Aufgabestellung

Gegeben ist ein i.A. nichtlineares dynamisches System in Zustandsraum-Darstellung (siehe Abschnitt 2.6). Die Darstellung kann auch von einem identifizierten Parametervektor \underline{p} (siehe Teil B) sowie der aktuellen Zeit t abhängen:

$$\dot{\underline{x}}(t) = \underline{f}(\underline{x}(t), \underline{u}(t), \underline{p}, t)$$

(Der Übersichtlichkeit halber wird nachfolgend auf die Argumente $\underline{u}(t)$ und \underline{p} verzichtet.) Weiterhin ist der Anfangswert bekannt:

$$\underline{x}(0) = \underline{x}_0$$

Gesucht ist die Näherung $\tilde{\underline{x}}(t)$ der Lösung der Zustandsdifferentialgleichung in einem Zeitintervall $t = [0, t_{\text{end}}]$.

Euler-Verfahren

Ein einfaches Lösungserfahren erhält man, wenn man $\dot{\underline{x}}(t)$ durch den Differenzenquotienten mit einer endlichen Zeitschrittweite h annähert:

$$\dot{\underline{x}}(t) \approx \frac{\underline{x}(t+h) - \underline{x}(t)}{h}$$

Einsetzen in die Zustandsdifferentialgleichung und Auflösen nach $\underline{x}(t+h)$:

$$\underline{x}(t+h) \approx \underline{x}(t) + h \cdot \underline{f}(\underline{x}(t), t)$$

Diese Gleichung kann für die Bestimmung einer Näherungslösung $\tilde{\underline{x}}(t)$ genutzt werden. Beginnend im Zeitschritt $t_0 = 0$ mit $\tilde{\underline{x}}(t_0) = \underline{x}_0$ berechnet man schrittweise die Näherungslösung in den Zeitschritten t_k für $k = 1, 2, 3, \dots$ bis t_{end} erreicht wird:

$$\tilde{\underline{x}}(t_{k+1}) = \tilde{\underline{x}}(t_k) + h \cdot \underline{f}(\tilde{\underline{x}}(t_k), t_k)$$

Explizite Einschrittverfahren

Das Euler-Verfahren kann verallgemeinert werden, in dem statt der Funktion $\underline{f}(\dots)$ eine allgemeine *Verfahrensfunktion* $\underline{\Phi}(\dots)$ angesetzt wird:

$$\tilde{\underline{x}}(t_{k+1}) = \tilde{\underline{x}}(t_k) + h \cdot \underline{\Phi}(\tilde{\underline{x}}(t_k), t_k, h)$$

Simulationsverfahren mit dieser Rekursionsgleichung werden als *Einschrittverfahren* bezeichnet, da sie zur Berechnung nur auf den gerade zurückliegenden Zeitschritt (von t_k bis t_{k+1}) zurückgreifen. Demgegenüber verwenden *Mehrschrittverfahren* auch Werte aus vorhergehenden Zeitschritten – sollen hier aber nicht weiter betrachtet werden.

In nächsten Abschnitt 9.2 soll die Verfahrensfunktion $\underline{\Phi}(\dots)$ so entworfen werden soll, dass die Näherungslösung verbessert werden kann.

Beispiel 9-1: Numerische Simulation eines Zwei-Massen-Schwingers

Nachfolgend ist die Implementierung des Euler-Verfahrens in Matlab dargestellt:

```
% Parameter Modell:
m = 0.5;           % Masse des Schwingers
c = 5.0;           % Federkonstante

% Parameter Simulation:
h = 0.001;         % Schrittweite
t_end = 10;        % Zeithorizont der Simulation
x0 = [1 0 0 0]';   % Anfangswert

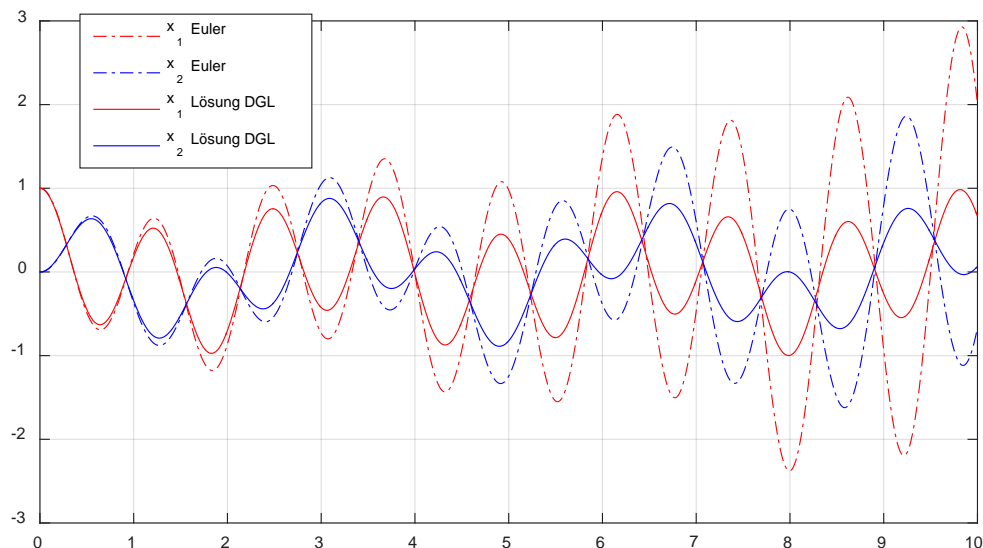
% Euler-Verfahren:
t = 0:h:t_end;     % Zeitgitter
x = zeros(4,length(t)); % Allokation
x(:,1) = x0;       % Anfangswert

% Rechte Seite der DGL als "function handle" definieren:
f = @(x) [ x(2)
           c/m*(-2*x(1)+x(3))
           x(4)
           c/m*(x(1)-x(3)) ];

% Hauptschleife:
for k=2:length(t)
    x(:,k) = x(:,k-1) + h * f(x(:,k-1));
end

% Plot:
plot(t,x(1,:),'-r',t,x(3,:),'-b')
```

Der nachstehende Plot zeigt neben der Näherungslösung des Euler-Verfahrens (strichpunkt-
tiert) auch die exakte Lösung (durchgezogen) für die Auslenkung beider Massepunkte (rot
und blau).



Um die starke Abweichung zu reduzieren, sind zwei Maßnahmen möglich:

- Reduktion der Zeitschrittweite h
- Wahl einer „besseren“ Verfahrensfunktion $\Phi(\dots)$. Dies wird Gegenstand des folgenden Abschnitts sein.

9.2 Lokaler Diskretisierungsfehler

Zunächst wird der *Differenzenquotient der exakten Lösung* der Zustandsdifferentialgleichung bestimmt: In einem beliebigen Startpunkt $\underline{x}(t_0)$ zum Zeitpunkt t_0 wird mit

$$\dot{\underline{z}}(t) = \underline{f}(\underline{z}(t), t) \text{ und } \underline{z}(t_0) = \underline{x}(t_0)$$

eine Zustandstrajektorie $\underline{z}(t)$ erzeugt. Der zugehörige Differenzenquotient ist:

$$\underline{\Delta}(\underline{x}, t_0, h) = \begin{cases} \frac{\underline{z}(t_0 + h) - \underline{x}(t_0)}{h} & , h \neq 0 \\ \underline{f}(\underline{x}, t_0) & , h = 0 \end{cases}$$

Der *lokale Diskretisierungsfehler* ist nun die Differenz zwischen diesem Differenzenquotienten der exakten Lösung und der Verfahrensfunktion:

$$\underline{\tau}(\underline{x}, t_0, h) = \underline{\Delta}(\underline{x}, t_0, h) - \underline{\Phi}(\underline{x}, t_0, h)$$

Eine sinnvolle Verfahrensfunktion erfüllt die Forderung für den lokalen Diskretisierungsfehler:

$$\lim_{h \rightarrow 0} \underline{\tau}(\underline{x}, t_0, h) = 0$$

Ein Verfahren besitzt die Ordnung p , falls sich der lokale Diskretisierungsfehler als

$$\underline{\tau}(\underline{x}, t_0, h) = h^p (\underline{c}_0 + \underline{c}_1 h + \dots)$$

schreiben lässt. D.h. im lokalen Diskretisierungsfehler finden sich erst Terme mit der Potenz h^p oder höher. Ziel ist es also Verfahren mit höherer Ordnung p zu finden, da dann bei einer gegebenen Schrittweite $h < 1$ der lokale Diskretisierungsfehler geringer wird.

Konstruktion von Verfahren höherer Ordnung

Um Verfahren höherer Ordnung zu konstruieren, kann man Differenzenquotient der exakten Lösung $\underline{\Delta}$ in eine Taylor-Reihe entwickeln. Dann wählt man die Verfahrensfunktion $\underline{\Phi}$ so, dass sie den Gliedern dieser Taylor-Reihe bis zur Potenz $p-1$ entspricht.

Man erhält z.B.:

$$p=1: \quad \underline{\Phi}(\underline{x}, t, h) = \underline{f}(\underline{x}(t), t) \quad (\text{Euler-Verfahren})$$

$$p=2: \quad \underline{\Phi}(\underline{x}, t, h) = \underline{f}(\underline{x}(t), t) + \frac{h}{2} \left[\frac{\partial \underline{f}}{\partial \underline{x}}(\underline{x}, t) \underline{f}(\underline{x}, t) + \frac{\partial \underline{f}}{\partial t}(\underline{x}, t) \right]$$

Nachteil dieser Verfahrensfunktionen für $p \geq 2$ ist, dass Gradienten der Zustandsdifferentialgleichungen bestimmt werden müssen. Um das zu vermeiden, können Ansätze für Verfahrensfunktionen gefunden werden, die stattdessen nur zusätzliche Funktionsauswertungen von \underline{f} erfordern, z.B.:

$$p=2: \quad \underline{\Phi}(\underline{x}, t, h) = a_1 \underline{f}(\underline{x}(t), t) + a_2 \underline{f}(\underline{x}(t), t + p_1 h)$$

$$p=4: \quad \underline{\Phi}(\underline{x}, t, h) = a_1 \cdot \underline{k}_1 + a_2 \cdot \underline{k}_2 + a_3 \cdot \underline{k}_3 + a_4 \cdot \underline{k}_4 \quad \text{mit}$$

$$\underline{k}_1 = \underline{f}(\underline{x}, t),$$

$$\underline{k}_2 = \underline{f}(\underline{x} + p_1 h \underline{k}_1, t + p_2 h),$$

$$\underline{k}_3 = \underline{f}(\underline{x} + p_3 h \underline{k}_2, t + p_4 h),$$

$$\underline{k}_4 = \underline{f}(\underline{x} + p_5 h \cdot \underline{k}_3, t + p_6 h)$$

Die darin enthaltenen noch unbekannten Parameter werden durch den Vergleich der Taylor-Reihen von \underline{x} bis zur Potenz $p-1$ ermittelt. Häufig verwendete Verfahren, die auf diesem Wege konstruiert werden können, werden nachfolgend zusammengefasst:

$p=1$: Verfahren von Euler

$$\Phi(\underline{x}, t, h) = \underline{f}(\underline{x}, t)$$

$p=2$: Verfahren von Heun:

$$\Phi(\underline{x}, t, h) = \frac{1}{2} [\underline{f}(\underline{x}, t) + \underline{f}(\underline{x} + h \cdot \underline{f}(\underline{x}, t), t + h)]$$

$p=2$: Verfahren von Collatz:

$$\Phi(\underline{x}, t, h) = \underline{f}\left(\underline{x} + \frac{h}{2} \cdot \underline{f}(\underline{x}, t), t + \frac{h}{2}\right)$$

$p=4$: Verfahren von Runge-Kutta:

$$\Phi(\underline{x}, t, h) = \frac{1}{6} \cdot \underline{k}_1 + \frac{2}{6} \cdot \underline{k}_2 + \frac{2}{6} \cdot \underline{k}_3 + \frac{1}{6} \cdot \underline{k}_4$$

mit

$$\underline{k}_1 = \underline{f}(\underline{x}, t),$$

$$\underline{k}_2 = \underline{f}\left(\underline{x} + \frac{1}{2} \cdot h \cdot \underline{k}_1, t + \frac{1}{2} h\right),$$

$$\underline{k}_3 = \underline{f}\left(\underline{x} + \frac{1}{2} \cdot h \cdot \underline{k}_2, t + \frac{1}{2} h\right),$$

$$\underline{k}_4 = \underline{f}(\underline{x} + h \cdot \underline{k}_3, t + h)$$

9.3 Globaler Diskretisierungsfehler und Schrittweitensteuerung

Der *globale Diskretisierungsfehler* ist die Differenz zwischen der exakten Lösung $\underline{x}(t)$ und der Näherungslösung des Verfahrens $\tilde{\underline{x}}(t, h)$ zu einem festen Zeitpunkt t_0 :

$$\underline{e}(t_0, h) = \tilde{\underline{x}}(t_0, h) - \underline{x}(t_0)$$

Ein sinnvolles Verfahren ist konvergent, d.h. es gilt:

$$\lim_{n \rightarrow 0} \underline{e}(t_0, \frac{t_0}{n}) = 0$$

Weiterhin entspricht die Fehlerordnung des globalen Diskretisierungsfehler der Ordnung p des lokalen Diskretisierungsfehlers:

$$\underline{e}(t_0, h) = \underline{c}_p(t_0) h^p + \dots$$

Fehlerabschätzung

Im Allgemeinen ist die exakte Lösung $\underline{x}(t_0)$ und damit auch $\underline{e}(t_0, h)$ nicht bekannt. Für die nachfolgende Schrittweitensteuerung wird eine Methode benötigt, diesen Fehler wenigstens abschätzen zu können.

Der globale Diskretisierungsfehler zu einem Verfahren der Ordnung p lässt sich als

$$\underline{e}(t_0, h) = \tilde{\underline{x}}(t_0, h) - \underline{x}(t_0) = \underline{c}_p(t_0) h^p + \dots$$

darstellen. Hilfsweise wird zusätzlich die Näherungslösung nach der halben Schrittweite bestimmt und daraus erhält man:

$$\underline{e}(t_0, \frac{h}{2}) = \tilde{\underline{x}}(t_0, \frac{h}{2}) - \underline{x}(t_0) = \underline{c}_p(t_0) \left(\frac{h}{2}\right)^p + \dots$$

Subtrahiert man beide Gleichungen, so eliminiert man $\underline{x}(t_0)$ und erhält

$$\tilde{\underline{x}}(t_0, h) - \tilde{\underline{x}}(t_0, \frac{h}{2}) = \underline{c}_p(t_0) \left(\frac{h}{2}\right)^p (2^p - 1) \text{ bzw. } \underline{c}_p(t_0) \left(\frac{h}{2}\right)^p = \frac{\tilde{\underline{x}}(t_0, h) - \tilde{\underline{x}}(t_0, \frac{h}{2})}{2^p - 1}.$$

Setzt man dies in die Gleichung für den Fehler mit halber Schrittweite ein, erhält man die folgende Abschätzung:

$$\underline{e}(t_0, \frac{h}{2}) = \tilde{\underline{x}}(t_0, \frac{h}{2}) - \underline{x}(t_0) = \frac{\tilde{\underline{x}}(t_0, h) - \tilde{\underline{x}}(t_0, \frac{h}{2})}{2^p - 1} h^p$$

Schrittweitensteuerung

Man betrachtet nun einen einzelnen Schritt. Der Fehler am Beginn des Schrittes sei Null, der Fehler am Ende lässt sich in eine Reihe entwickeln:

$$\underline{e}(t_0 + h, h) = \underline{c}_p(t_0 + h) h^p = \left[\underbrace{c_p(t_0)}_{=0} + \underline{c}'_p(t_0) h + \dots \right] h^p$$

Mit dem Ergebnis aus der obigen Fehlerabschätzung gilt mit $h/2$ am Ende des Schrittes:

$$\underline{e}(t_0 + h, \frac{h}{2}) = \frac{\tilde{\underline{x}}(t_0 + h, h) - \tilde{\underline{x}}(t_0 + h, \frac{h}{2})}{2^p - 1} = \underline{c}_p(t_0 + h) \left(\frac{h}{2}\right)^p = \underline{c}'_p(t_0) h \left(\frac{h}{2}\right)^p$$

Dies lässt sich für $\underline{c}'_p(t_0)$ lösen:

$$\underline{c}'_p(t_0) = \frac{1}{h^{p+1}} \frac{2^p}{2^p - 1} \left[\tilde{\underline{x}}(t_0 + h, h) - \tilde{\underline{x}}(t_0 + h, \frac{h}{2}) \right]$$

Für die Norm des Fehlers am Ende des Schrittes wird eine Vorgabe $\varepsilon > 0$ gemacht; dazu wird die passende Schrittweite h^* gesucht:

$$\left\| \underline{c}'_p(t_0) (h^*)^{p+1} \right\| = \varepsilon$$

Für $\underline{c}'_p(t_0)$ einsetzen und nach h^* lösen liefert schließlich:

$$h^* = h \left(\frac{2^p}{2^p - 1} \frac{\left\| \tilde{\underline{x}}(t_0 + h, h) - \tilde{\underline{x}}(t_0 + h, \frac{h}{2}) \right\|}{\varepsilon} \right)^{\left(\frac{1}{p+1} \right)}$$

Auf Basis der aktuellen Schrittweite h und der Vorgabe des globalen Fehlers ε für einen Schritt wird so näherungsweise eine angepasste Schrittweite h^* bestimmt.

Nachfolgend ist der entsprechende Matlab-Code auszugsweise dargestellt, in dem $\tilde{x}(t_0 + h, \frac{h}{2})$ in zwei Schritten mit je $h/2$ berechnet wird. Damit sind für die Schrittweitensteuerung also insgesamt 3 zusätzliche Funktionsauswertungen erforderlich.

```
while 1
    % Schrittweitensteuerung:
    x0 = x + h * phi(x,u);
    x1 = x + h/2 * phi(x,u);
    x2 = x1 + h/2 * phi(x1,u);
    h = h * (2^p/(2^p-1)*abs(x0-x2)/epsilon)^(-1/(p+1));

    % Schritt:
    x = x + h * phi(x,u);
    t = t + h;

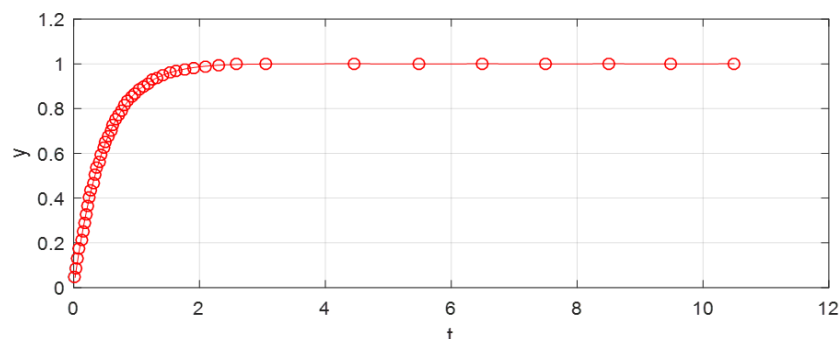
    % Abbruch?
    if t>t_end, break, end
end
```

Beispiel 9-2: Schrittweitensteuerung

Für das Verzögerungssystem

$$G(s) = \frac{1}{1 + Ts} \quad \text{mit } T = 0,5$$

wird die Sprungantwort mit einem Euler-Verfahren mit Schrittweitensteuerung mit $\varepsilon = 10^{-3}$ durchgeführt. Nachfolgend ist die Sprungantwort mit Stützstellen dargestellt:



Diskontinuierliche Systeme

Ein nichtlineares System kann ggf. *diskontinuierliche* („schaltende“) Funktionszusammenhänge aufweisen. Typische Beispiele sind:

- Mechanische Kontakte: Ein „springender Ball“ berührt eine Fläche und wird dadurch elastisch gestoßen.
- Reibung: Losbrechen aus der Haftreibung und Übergang in Gleitreibung.

Wird ein solches System mit einer festen oder zu großen Schrittweite h simuliert, entstehen Diskretisierungsfehler, die oszillieren und aufklingen können.

Bei solchen Systemen sind zusätzliche Algorithmen sinnvoll, die zunächst erkennen, ob ein solcher Schaltvorgang innerhalb des nächsten Simulationsschrittes $[t_0, t_0+h]$ auftreten wird. Ist das der Fall, muss der Schaltzeitpunkt t_1 gefunden werden („zero-crossing detection“) und der Schritt h^* mit $t_0+h^*=t_1$ so gewählt werden, dass nicht innerhalb eines Schrittes umgeschaltet werden müsste.

9.4 Stabilität

Die Stabilität von Simulationsverfahren ist eine wesentliche Grundforderung. Um das Stabilitätsverhalten zu untersuchen, wird ein autonomes, lineares System in Zustandsraum-Darstellung betrachtet:

$$\dot{\underline{x}}(t) = \underline{A} \underline{x}(t)$$

Das zugehörige Einschrittverfahren lautet dann z.B. für das Euler-Verfahren:

$$\tilde{\underline{x}}(t_{k+1}) = \tilde{\underline{x}}(t_k) + h \cdot \Phi(\tilde{\underline{x}}(t_k), t_k) = \tilde{\underline{x}}(t_k) + h \underline{A} \tilde{\underline{x}}(t_k) = (\underline{I} + h \underline{A}) \tilde{\underline{x}}(t_k)$$

Damit das Verfahren stabil ist, müssen alle $i=1, \dots, n$ Eigenwerte λ_i von $(\underline{I} + h \underline{A})$ betragsmäßig kleiner eins sein:

$$|\lambda_i(\underline{I} + h \underline{A})| < 1 \quad \text{bzw.} \quad |1 + h \lambda_i(\underline{A})| < 1 \quad \text{für } i=1, \dots, n$$

Mit der Substitution $w_i = h \lambda_i(\underline{A})$ lautet die Bedingung für die Stabilität des Verfahrens:

$$|g(w_i)| < 1 \quad \text{mit} \quad g(w) = 1 + w \quad \text{und} \quad i=1, \dots, n$$

Das Stabilitätsgebiet, in dem alle $w_i = h \lambda_i(\underline{A})$ liegen müssen, entspricht einem Kreis mit Radius 1 um den Mittelpunkt -1 und ist umseitig (in Rot) grafisch dargestellt.

In analoger Weise lassen sich die Funktionen $g(w)$ und damit die Stabilitätsgebiete für Verfahren höherer Ordnung bestimmen (grafische Darstellung umseitig):

- Verfahren 1. Ordnung (z.B. Eulerverfahren):
 $|g_1(w)| = |1 + w| < 1$
- Verfahren 2. Ordnung (z.B. Verfahren von Heun oder Collatz):
 $|g_2(w)| = \left| 1 + w + \frac{1}{2} w^2 \right| < 1$
- Verfahren 3. Ordnung:
 $|g_3(w)| = \left| 1 + w + \frac{1}{2} w^2 + \frac{1}{6} w^3 \right| < 1$
- Verfahren 4. Ordnung (z.B. Runge-Kutta-Verfahren):
 $|g_4(w)| = \left| 1 + w + \frac{1}{2} w^2 + \frac{1}{6} w^3 + \frac{1}{24} w^4 \right| < 1$

Die Stabilität hängt also von der Ordnung p und der Schrittweite h des Verfahrens, aber durch $\lambda_i(\underline{A})$ auch vom zu simulierenden System ab.

Ein Verfahren heißt *absolut stabil*, falls $|g(w_i)| < 1$ für alle w_i mit $\text{Re}(w_i) < 0$ gilt.

Beispiel 9-3: Stabilität der Simulation des 2-Massen-Schwingers (aus Bsp. 9-1)

Das System ist mit $c = 5,0$ und $m = 0,5$ durch

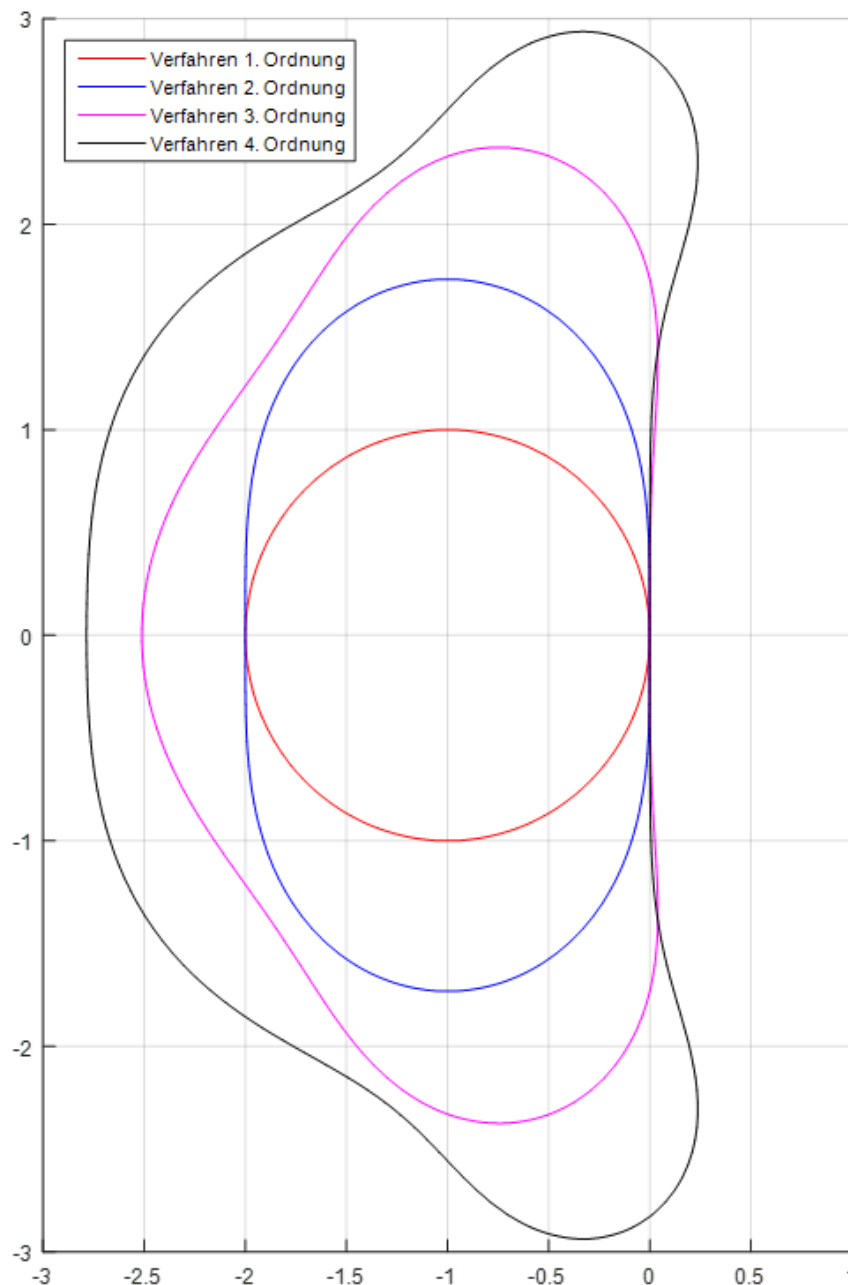
$$\underline{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -2c/m & 0 & c/m & 0 \\ 0 & 0 & 0 & 1 \\ c/m & 0 & -c/m & 0 \end{bmatrix}$$

gegeben. Seine $n = 4$ Eigenwerte liegen in $\lambda_{1,2} = \pm 5,117j$ und $\lambda_{3,4} = \pm 1,954j$.

Die Betrachtung der Stabilitätsgebiete zeigt, dass für Verfahren 1. und 2. Ordnung keine stabile Lösung für $h > 0$ gefunden werden kann. Für ein Runge-Kutta-Verfahren ($p=4$) und $0 < h < 5,117/2,83 = 1,81$ wird Stabilität hingegen (gerade) erreicht.

Stabilitätsgebiete

Die Stabilitätsgebiete für (explizite) Einschrittverfahren der Ordnung $p = 1, 2, 3, 4$ liegen im Inneren der folgenden Gebiete in der komplexen w -Ebene:



10 Implizite Verfahren

10.1 Grundlagen

Die bisher vorgestellten expliziten Verfahren haben nur ein beschränktes Stabilitätsgebiet. Lassen sich Verfahren finden, die ein größeres Stabilitätsgebiet besitzen?

Der bisherige Ansatz für Einschrittverfahren wird dahingehend erweitert, dass die Verfahrensfunktion Φ nun zusätzlich auch noch vom Zustand im Endzeitpunkt des Schrittes $\tilde{x}(t+h)$ abhängen kann:

$$\tilde{x}(t+h) = \tilde{x}(t) + h \cdot \Phi(\tilde{x}(t), \tilde{x}(t+h), t, h)$$

Mit diskreten Zeitpunkten t_k lässt sich alternativ auch schreiben:

$$\tilde{x}(t_{k+1}) = \tilde{x}(t_k) + h \cdot \Phi(\tilde{x}(t_k), \tilde{x}(t_{k+1}), t_k, h)$$

Es entsteht ein implizites Gleichungssystem, da $\tilde{x}(t_{k+1})$ sowohl als Argument als auch als Ergebnis einer Funktion vorkommt. Man spricht also von einem *impliziten Einschrittverfahren*.

Beispiele für implizite Verfahren

Um das Euler-Verfahren einzuführen, wurde in Abschnitt 9.1 für das gegebene System

$$\dot{x}(t) = f(x(t), u(t), p, t)$$

die zeitliche Ableitung des Zustands mit dem (Vorwärts-)Differenzenquotienten

$$\dot{x}(t) \approx \frac{x(t+h) - x(t)}{h}$$

angenähert. Mit gleichem Recht kann man auch den (Rückwärts-)Differenzenquotienten

$$\dot{x}(t) \approx \frac{x(t) - x(t-h)}{h}$$

verwenden. Dann erhält man (mit der Zeitverschiebung $t_{k+1} = t$ und $t_k = t-h$) das *implizite Euler-Verfahren*:

$$\tilde{x}(t_{k+1}) = \tilde{x}(t_k) + h \cdot f(\tilde{x}(t_{k+1}), t_{k+1})$$

Mittelt man die Verfahrensfunktionen des expliziten und impliziten Euler-Verfahrens, erhält man weiterhin das *Trapezverfahren*:

$$\tilde{x}(t_{k+1}) = \tilde{x}(t_k) + h \cdot \frac{1}{2} [f(\tilde{x}(t_{k+1}), t_{k+1}) + f(\tilde{x}(t_k), t_k)]$$

Lösung durch Funktionsiteration

Ein offensichtlicher Nachteil von impliziten Verfahren ist es, dass i.A. keine direkte Lösung existiert. Eine Möglichkeit ist die iterative Lösung durch folgende Funktionsiteration (auch Fixpunktiteration):

1. Startwert $\underline{z}^{(0)}$ festlegen, etwa $\tilde{\underline{x}}(t_k)$ oder ein Schritt eines expliziten Verfahrens, und $j := 0$
2. Iteration: $\underline{z}^{(j+1)} = \tilde{\underline{x}}(t_k) + h \cdot \underline{\Phi}(\tilde{\underline{x}}(t_k), \underline{z}^{(j)}, t_k, h)$
3. Falls Abbruchbedingung $k \geq k_{\max}$ oder $\|\underline{z}^{(j+1)} - \underline{z}^{(j)}\| \leq \varepsilon$ erfüllt, weiter bei 4, sonst $j := j+1$ und weiter bei 2.
4. Ausgabe von $\tilde{\underline{x}}(t_{k+1}) := \underline{z}^{(j+1)}$

Beispiel 10-1: Implizites Euler-Verfahren (Fortsetzung von Beispiel 9-1)

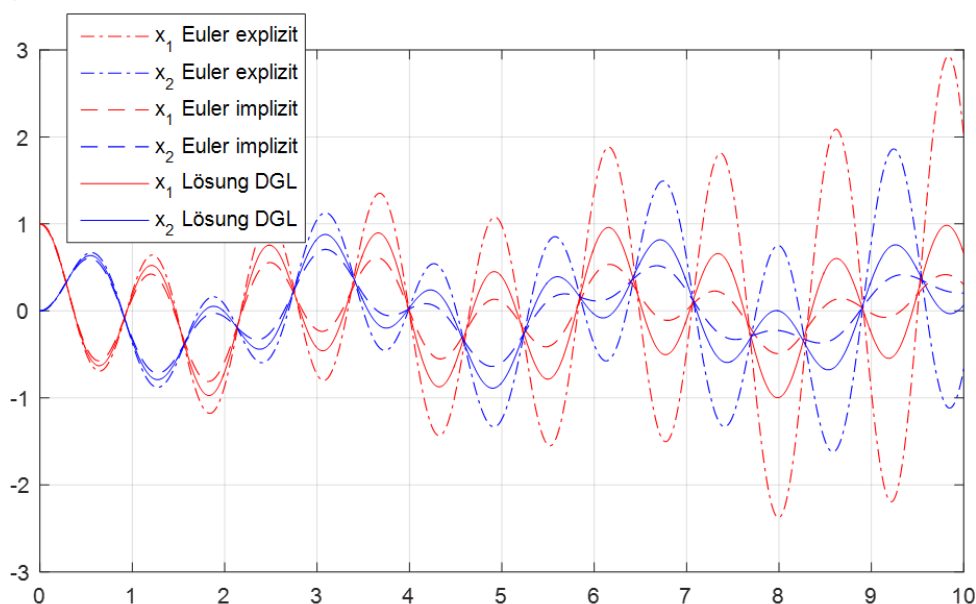
Das System eines Zwei-Masse-Schwingers aus Beispiel 9-1 soll mit einem impliziten Euler-Verfahren simuliert werden. In der Matlab-Implementierung wird die Hauptschleife in folgender Weise modifiziert:

```
% Hauptschleife:
for k=2:length(t)
    x(:,k) = funiter(x(:,k-1),phi,h);
end
```

Darin wird nun die Funktion „funiter“ aufgerufen, welche eine Funktionsiteration ausführt und wie folgt implementiert werden kann:

```
function x1 = funiter(x0,phi,h)
% FUNITER Funktionsiteration für implizite Verfahren
epsilon = 1e-6;
j_max = 100;
z = x0;
for j=1:j_max
    z_old = z;
    z = x0 + h * phi(z);
    if norm(z-z_old)<epsilon, break, end
end
x1 = z;
end
```

Der nachstehende Plot ergänzt die Lösung des impliziten Verfahrens (gestrichelt) für die Auslenkung beider Massepunkte (rot und blau). Man erkennt, dass das implizite Verfahren zwar stabil, aber auch fehlerbehaftet ist.



10.2 Stabilität

Analog zu Abschnitt 9.4 können auch für implizite Verfahren Stabilitätsgebiete bestimmt werden.

Für ein autonomes, lineares System $\dot{\underline{x}}(t) = \underline{A} \underline{x}(t)$ gilt mit $w_i = h \lambda_i(\underline{A})$

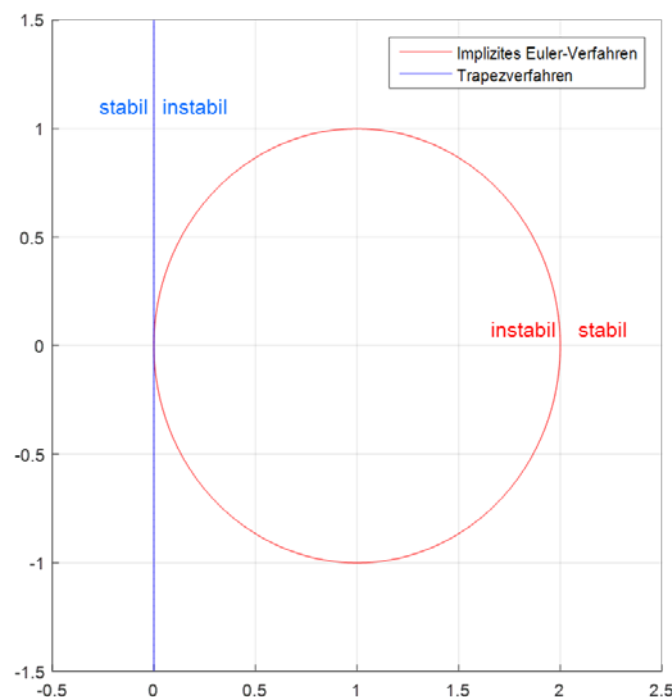
- für das implizite Euler-Verfahren:

$$|g(w)| = \left| \frac{1}{1-w} \right| < 1$$

- für das Trapezverfahren:

$$|g(w)| = \left| \frac{2+w}{2-w} \right| < 1$$

Die Stabilitätsgebiete sind nachfolgend grafisch dargestellt:



Beide Verfahren sind absolut stabil (nach Definition siehe Abschnitt 9.4)!

Man kann zeigen, dass es kein Verfahren mit der Ordnung $p > 2$ gibt, das absolut stabil ist.

10.3 Steife Systeme

Steife Systeme sind dadurch gekennzeichnet, dass sie sowohl sehr langsame, als auch sehr schnelle Systemteile umfassen. Bei linearen Systemen kann das Verhältnis des betragsmäßig größten zum betragsmäßig kleinsten Eigenwert der Systemmatrix als Maß betrachtet werden:

$$\frac{|\operatorname{Re}(\lambda_n(\underline{A}))|}{|\operatorname{Re}(\lambda_1(\underline{A}))|} \text{ mit sortierten Eigenwerten } |\lambda_1(\underline{A})| \leq |\lambda_2(\underline{A})| \leq \dots \leq |\lambda_n(\underline{A})|$$

Die Simulation solcher Systeme mit Hilfe von expliziten Verfahren ist problematisch: Die Schrittweite h muss klein gewählt werden, so dass auch die schnelle Systemdynamik ohne

zu große Fehler simuliert wird, allerdings sind dann sehr viele Schritte notwendig, um den Verlauf der langsamen Systemdynamik wiederzugeben.

Hier ist die Verwendung von impliziten Verfahren von großem Vorteil, da sie auch bei größerer Schrittweite h stabil bleiben.

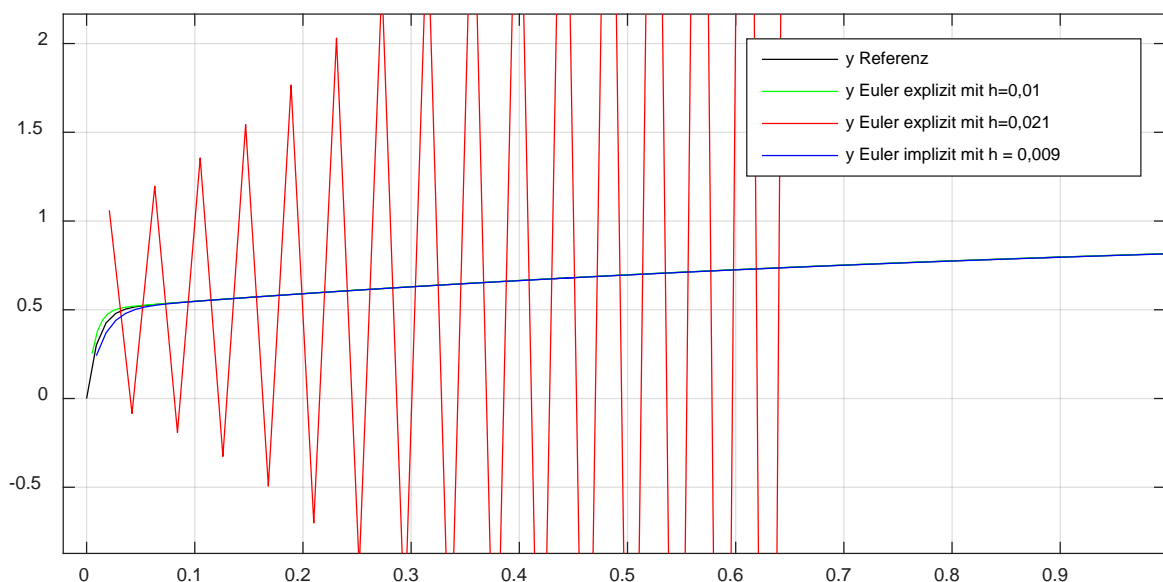
Beispiel 10-2: Steifes System 2. Ordnung

Für das folgende lineare System soll eine Sprungantwort simuliert werden.

$$\dot{\underline{x}}(t) = \begin{bmatrix} -1 & 0 \\ 0 & -100 \end{bmatrix} \underline{x}(t) + \begin{bmatrix} 1 \\ 100 \end{bmatrix} u(t), \quad y = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \underline{x}(t)$$

- Die Eigenwerte liegen also bei $\lambda_1(\underline{A}) = -1$ und $\lambda_2(\underline{A}) = -100$. Damit handelt es sich um ein *steifes System*.
- Verwendet man ein *explizites Euler-Verfahren*, muss $w_i = h \lambda_i(\underline{A})$ für beide Eigenwerte $i = 1, 2$ im Stabilitätsgebiet des gewählten Verfahrens liegen. Entsprechend dem Plot in Abschnitt 9.4 muss $w_i > -2$ gelten. Während λ_1 also eine stabile Simulation für $h < 2$ zuließe, beschränkt λ_2 den Bereich auf $h < 0,02$.
- Das *implizite Euler-Verfahren* ist zwar (absolut) stabil und h wäre frei wählbar – allerdings zwingt in diesem Beispiel die Funktionsiteration dazu, eine Schrittweite von $h < 0,01$ zu verwenden, da das Verfahren sonst nicht konvergiert.

Der nachfolgende Plot zeigt die Simulationsergebnisse im Vergleich (die Referenz entspricht der Lösung der linearen Zustandsdifferentialgleichung):



Anhang

A. Statistische Tests

Ausgangspunkt – Es wurden Experimente durchgeführt. Dabei wurden 2 Versuchsgruppen gebildet, die m_1 bzw. m_2 Einzelversuche umfassen. In jedem Einzelversuch wurde das Merkmal y gemessen. Daraus wird in beiden Versuchsgruppen der Mittelwert \bar{y}_i berechnet:

Gruppe	Umfang	Daten	Mittelwert
1	m_1	$y_{11}, y_{12}, \dots, y_{1\ m1}$	\bar{y}_1
2	m_2	$y_{21}, y_{22}, \dots, y_{2\ m2}$	\bar{y}_2

Aufgabe – Es soll die Frage beantwortet werden, ob die Unterschiede zwischen den Mittelwerten \bar{y}_i der Gruppen zufällig zu erklären sind, oder ob sie die systematische Folge von gruppenweise unterschiedlichen Einflussfaktoren sind. Nimmt man eine systematische Folge an, möchte man die Verlässlichkeit (Signifikanz) der Aussage angeben können.

Statistische Tests – Es wird eine *Hypothese* aufgestellt, die eine Behauptung über die Eigenschaften der Zufallsvariablen y macht. Ein statistischer Test überprüft nun diese Hypothese auf Basis einer experimentellen Stichprobe. Als Hypothese wird z. B. formuliert, dass die Mittelwerte beider Proben gleich sind:

$$\bar{y}_1 = \bar{y}_2$$

In der Regel wird man bestrebt sein, diese Hypothese (signifikant) zu widerlegen!

Kommt der Test zu dem Ergebnis, dass die Hypothese abgelehnt wird, so gibt es einen nicht „zufälligen“ Unterschied zwischen den getesteten Gruppen.

Die Wahrscheinlichkeit dafür, dass der Test abgelehnt wurde, obwohl die Hypothese richtig gewesen wäre (d.h. es wird unberechtigt ein Unterschied festgestellt), wird als α -Fehler bezeichnet.

Die Gegenwahrscheinlichkeit $p = 1 - \alpha$ ist die *Signifikanz* des Tests. Der Test ist umso aussagekräftiger, desto kleiner α und desto größer die Signifikanz p ist.

Üblich sind die folgenden Bewertungen der Signifikanz:

Klasse	Aussage	p	α
–	kein Hinweis auf einen Unterschied	$< 0,95$	$> 0,05$
*	indifferent („verdächtig“), mehr Daten?	$\geq 0,95$	$\leq 0,05$
**	Unterschied signifikant	$\geq 0,99$	$\leq 0,01$
***	Unterschied hoch signifikant	$\geq 0,999$	$\leq 0,001$

Je nach Voraussetzungen und Aufgabe muss ein *geeigneter Test* gewählt werden:

	Es kann angenommen werden, dass die Daten y normalverteilt sind	Es ist keine Verteilung der Daten y bekannt
Zwei Gruppen werden getestet: Ist $\bar{y}_1 = \bar{y}_2$? (Versuchsplan mit 2 Stufen)	t-Test	Wilcoxon-Rangsummen-Test
Zwei Gruppen werden getestet: Ist $\sigma_1 < \sigma_2$?	F-Test	
Mehrere Gruppen werden getestet: Ist $\bar{y}_1 = \bar{y}_2 = \bar{y}_3 = \dots$? (Versuchsplan mit >2 Stufen, aber nur ein bzw. zwei Einflussfaktoren)	Spez. Anwendung der Varianzanalyse (F-Test)	Kruskal-Wallis-Test

Test zweier Gruppenmittelwerte gegeneinander: t-Test

Die Abweichung der Gruppenmittelwerte ist

$$e = \bar{y}_1 - \bar{y}_2.$$

Auf Basis der Varianzen der beiden Gruppen S_1^2 bzw. S_2^2 wird die *Varianz S^2 der Mittelwertabweichungen* bestimmt, dabei ist zu unterscheiden, ob der Umfang der Gruppen bzw. die Varianzen der Gruppen gleich oder verschieden sind:

	Varianzen gleich: $\sigma_1^2 = \sigma_2^2$	Varianzen ungleich: $\sigma_1^2 \neq \sigma_2^2$
Umfänge gleich („balanciert“): $m_1 = m_2 = m$	$S^2 = \frac{S_1^2 + S_2^2}{m}$	
Umfänge ungleich: $m_1 \neq m_2$	$S^2 = \frac{(m_1 + m_2)[(m_1 - 1)S_1^2 + (m_2 - 1)S_2^2]}{m_1 \cdot m_2 (m_1 + m_2 - 2)}$	$S^2 = \frac{S_1^2}{m_1} + \frac{S_2^2}{m_2}$

Entsprechend der Tabelle auf Seite 6-12 wird die *Testgröße t_{exp}* als Verhältnis aus Mittelwertabweichung e zur Varianz S^2 berechnet. Je stärker diese Mittelwertsabweichung im Verhältnis zur Varianz betragsmäßig ausfällt, desto signifikanter ist ein Unterschied zwischen beiden Gruppen. Der Schwellwert für die Testgröße t_{exp} ist als $t_{tab}(p, v)$ für verschiedene Signifikanzen p tabelliert.

Üblicher Weise wird die Hypothese „beide Gruppen sind gleich“ widerlegt. Für diesen *zweiseitigen Test* muss der Tabellenwert für $\alpha_{zweis} = \frac{\alpha}{2}$ bzw. $p_{zweis} = 1 - \frac{1-p}{2}$ verwendet werden

(ist in der Tabelle auf Seite 6-12 auch zusätzlich angegeben!). Dies führt zu schärferen Testbedingungen im Vergleich zur Widerlegung einer einseitigen Hypothese wie „Gruppe 1 führt zu höheren Werten als Gruppe 2“.

zu widerlegende Hypothese	signifikant belegte Aussage	falls: (mit $t_{\text{exp}} = \frac{\bar{y}_1 - \bar{y}_2}{S}$)	oder falls: (mit $e = \bar{y}_1 - \bar{y}_2$)
Beide Gruppen haben den gleichen Mittelwert (zweiseitiger Test, häufigster Fall)	Mittelwerte beider Gruppen sind verschiedenen	$t_{\text{exp}} > t_{\text{tab}}(p_{\text{zweis}}, \nu)$ oder $t_{\text{exp}} < -t_{\text{tab}}(p_{\text{zweis}}, \nu)$ mit $p_{\text{zweis}} = 1 - \frac{1-p}{2}$	$e \geq t_{\text{tab}}(p_{\text{zweis}}, \nu) \cdot S$ oder $e \leq -t_{\text{tab}}(p_{\text{zweis}}, \nu) \cdot S$
\bar{y}_1 ist kleiner als \bar{y}_2 (einseitiger Test)	\bar{y}_1 ist größer als \bar{y}_2	$t_{\text{exp}} > t_{\text{tab}}(p, \nu)$	$e \geq t_{\text{tab}}(p, \nu) \cdot S$
\bar{y}_1 ist größer als \bar{y}_2 (einseitiger Test)	\bar{y}_1 ist kleiner als \bar{y}_2	$t_{\text{exp}} < -t_{\text{tab}}(p, \nu)$	$e \leq -t_{\text{tab}}(p, \nu) \cdot S$
Freiheitsgrade: $\nu = m_1 + m_2 - 2$			

(Die Testbedingungen in der 3. und 4. Spalte sind äquivalent; letztere sind in der Einheit der zugrunde liegenden Größe und daher evtl. anschaulicher.)

Beispiel:

Gruppe	Umfang	Daten	Mittelwert
1	$m_1 = 16$	7,2 7,7 8,0 8,1 8,3 8,4 8,4 8,5 8,6 8,7 9,1 9,1 9,1 9,8 10,1 10,3	$\bar{y}_1 = 8,71$
2	$m_2 = 10$	8,1 9,2 10,0 10,4 10,6 10,9 11,1 11,9 12,0 12,1	$\bar{y}_2 = 10,63$

Die Varianzen sind $S_1 = 0,844$ und $S_2 = 1,284$, damit erhält man $t_{\text{exp}} = -4,619$.

Nach Tabelle auf Seite 6-12 findet man für den zweiseitigen Test mit der Signifikanz $p = 0,999$ und $\nu = 24$: $t_{\text{tab}} = 3,745$.

Da $t_{\text{exp}} < -t_{\text{tab}}$ gilt, sind die Mittelwerte mit Signifikanz *** ($p = 0,999$) von einander verschieden.

Alternativ in Excel: `=1-ttest(<DatenGruppe1>;<DatenGruppe2>;2;2)` berechnet die Signifikanz p direkt.

Prüfen auf Varianzunterschied zweier Gruppen: F-Test

Es wird vermutet, dass die Varianz der Gruppe 1 größer ist als die der Gruppe 2. Es muss also die Hypothese $\sigma_1^2 = \sigma_2^2$ signifikant widerlegt werden.

Für beide Gruppen werden experimentell die Varianzen s_1^2 bzw. s_2^2 aus den Messdaten berechnet. Die Testgröße ist:

$$F_{\text{exp}} = \frac{S_1^2}{S_2^2}$$

Um nach Tabelle auf Seite 6-15 den Tabellenwert abzulesen, benötigt man die Freiheitsgrade, die sich aus den Gruppenumfängen ergeben:

$$\nu_1 = m_1 - 1, \quad \nu_2 = m_2 - 1$$

Damit liest man den Tabellenwert $F_{\text{tab},p}(\nu_1, \nu_2)$ ab. Falls $F_{\text{exp}} > F_{\text{tab}}$ gilt, kann die Hypothese widerlegt werden und es ist $\sigma_1^2 > \sigma_2^2$ mit der Signifikanz p .