

Knowledge Database
Wissensdatenbank

"THE GREATEST ENEMY OF KNOWLEDGE IS NOT
IGNORANCE, IT IS THE ILLUSION OF KNOWLEDGE."
- STEPHEN HAWKING

Author: Heufelder Heind

Erstellung datum: 17. Oktober 2016

WRITTEN IN LATEX

Inhaltsverzeichnis

I

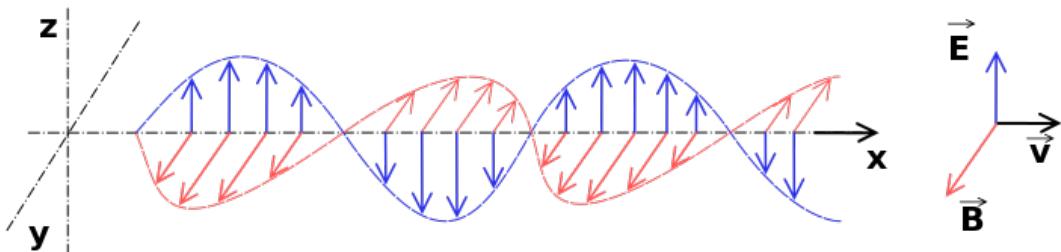
Physik

1 Astronomie

2 Elektromagnetismus

2.1 Elektromagnetische Welle

Als elektromagnetische Welle sind gekoppelte elektrischen und magnetischen Felder die sich im Raum mit Lichtgeschwindigkeit ausbreiten.

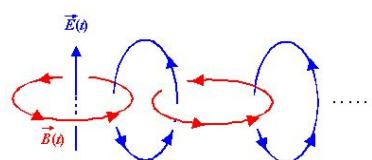


Warum breitet sich eine elektromagnetische Welle aus?

Aus den mikroskopischen Maxwellgleichungen kann nachvollzogen werden wie J. C. Maxwell auf die Existenz der elektromagnetischen Welle gestoßen ist. Aus der ersten Feldgleichung wird geschlossen, dass ein sich zeitlich veränderndes Magnetfeld, von ringförmigen elektrischen Feldlinien umgeben ist. Er folgerte aus dem zweiten Maxwell'schen Feldgesetz: Während sich ein elektrisches Feld ändert, ist es von ringförmigen geschlossenen magnetischen Feldlinien umgeben. Somit also: Veränderliche elektrische Felder erzeugen magnetische Wirbelfelder und veränderliche Magnetfelder erzeugen elektrische Wirbelfelder. Es ergibt sich also eine Kette von Veränderungen des elektrischen und magnetischen Feldes, die sich als selbstständiges Gebilde im Raum ausbreitet. Ladungen erzeugen elektrische Felder und Ströme erzeugen magnetische Felder. Das erste Feldgesetz erweitert diesen Zusammenhang mit Hilfe des zweiten Feldgesetzes nun zu einer beliebig langen Kette:

$$\begin{aligned} Q \rightarrow E \rightarrow B \rightarrow E \rightarrow \dots \\ I \rightarrow B \rightarrow E \rightarrow B \rightarrow \dots \end{aligned}$$

Elektromagnetische Wellen sind stets Transversalwellen.



2.1.1 Quellen von elektromagnetische Wellen

2.1.1.1 Hertzscher Dipol

Ein hertzscher Dipol ist eine Idealisierung eines Senders elektromagnetischer Strahlung und ist einem offenen elektrischen Schwingkreis äquivalent. Er dient somit als Grundlage für die Berechnung von Antennen. Bei der Transformation wird C und L immer kleiner, wodurch die Resonanzfrequenz steigt. Der hertzsche Dipol emittiert somit hochfrequente elektromagnetische Strahlung.

2.1.1.2 Röntgenstrahlung

Röntgenstrahlung kann durch zwei verschiedene Vorgänge entstehen:

- durch starke Beschleunigung geladener Teilchen (meistens Elektronen) – dies ist die Bremsstrahlung, deren Spektrum kontinuierlich ist –

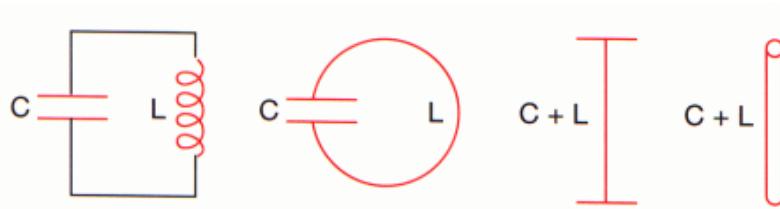


Abbildung 1: Transformation von einem Schwingkreis zum hertzschen Dipol

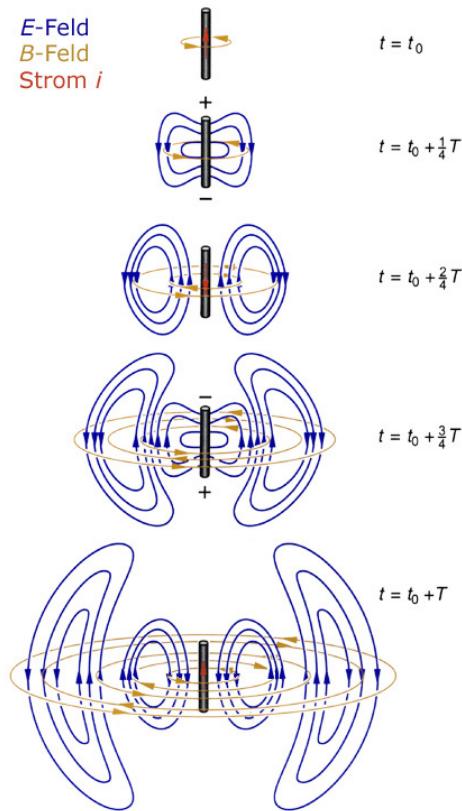


Abbildung 2: Ausbreitung der elektromagnetischen Welle eines hertzschen Dipols

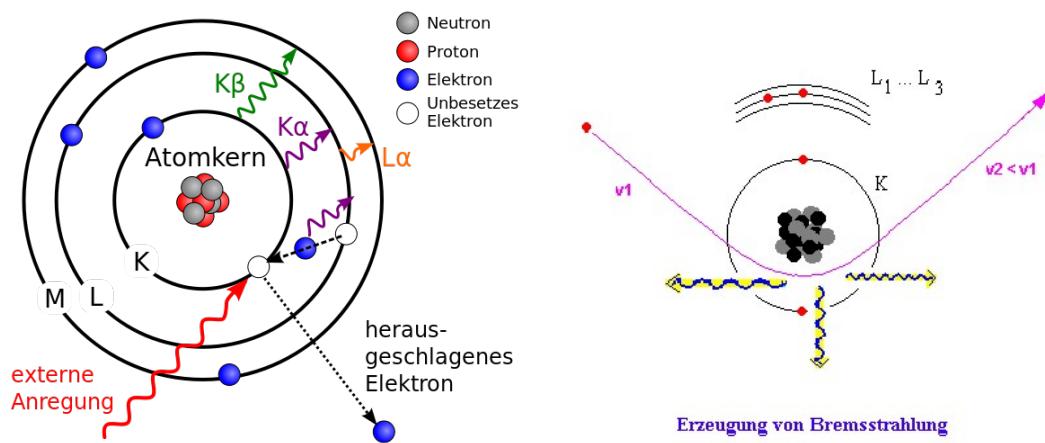
- und durch hochenergetische Übergänge in den Elektronenhüllen von Atomen oder Molekülen. Dies ist die charakteristische Röntgenstrahlung. Sie weist stets ein Linienspektrum auf.

Bremsstrahlung:

Wird eine elektrische Ladung beschleunigt, d.h. ändert sich ihr Geschwindigkeitsbetrag bzw. ihre Bewegungsrichtung, so entsteht elektromagnetische Strahlung. Die Energie der dabei auftretenden Photonen ist umso höher, je stärker die Beschleunigung ist.

Werden Elektronen auf eine Festkörperoberfläche beschleunigt und in diesem abgebremst, kommt es zu einer Beschleunigung (z.B. Richtungsänderung) durch etwa einen Atomkern. Diese Richtungsänderung führt nach obigem Satz zur Erzeugung von elektromagnetischen Wellen.

Beide Effekte werden in der Röntgenröhre ausgenutzt, in der Elektronen zunächst von einer Glühwendel (Kathode) aus beschleunigt werden (dabei setzen sie keine Röntgenstrahlung frei, weil die Beschleunigung nicht groß genug ist) und anschließend auf die Anode treffen, in der sie stark abgebremst werden. Hierbei entsteht Röntgenstrahlung (Bremsstrahlung, mit insgesamt rund 1% der eingestrahlten Energie) und Wärme (rund 99%). Außerdem werden durch Elektronenstoß Elektronen aus den Schalen der Metallatome herausgeschlagen. Ein Elektron wird z. B. durch einen Elektronenstoß aus der K-Schale entfernt und ein Elektron aus der L-Schale fällt in das Loch in der K-Schale. Die Energiedifferenz wird als charakteristische Röntgenstrahlung emittiert. Röntgenstrahlung ist ionisierend. Sie kann

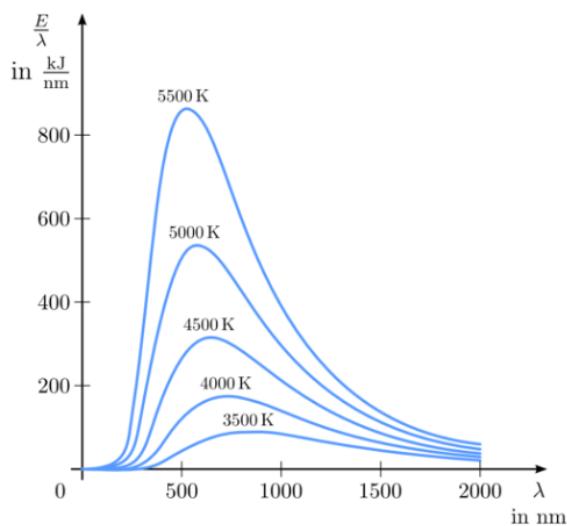


dadurch Veränderungen im lebenden Organismus hervorrufen und Schäden bis hin zu Krebs verursachen. Deshalb ist beim Umgang mit der Strahlung der Strahlenschutz zu beachten.

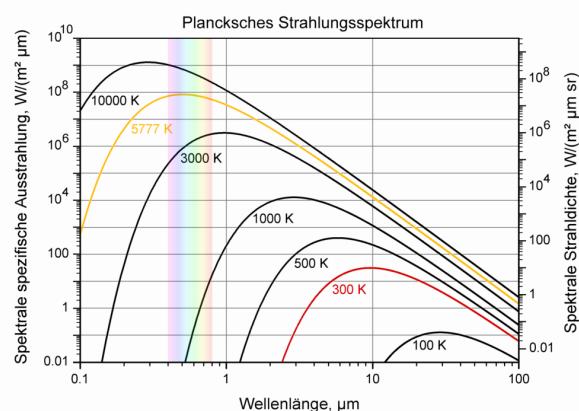
2.1.1.3 Wärmestrahlung

Genauso wie in Gasen und Flüssigkeiten eine thermische Molekularbewegung herrscht, haben auch die Atome von Festkörpern eine mikroskopische kinetische bzw. potentielle Energie. Das heißt sie schwingen um ihre Ruhelagen. Diese mikroskopischen Energien nehmen mit steigender Temperatur zu und schwingen mit höheren Frequenzen. Durch dieses Schwingen von elektrischen Ladungen ändern sich die elektrischen Felder zwischen diesen Ladungen. Es resultiert ein elektrisches Wechselfeld, welches eine elektromagnetische Welle erzeugt, die sich in Form von Wärmestrahlung im Raum ausbreitet.

Die Wärmestrahlung ist eine Art der Wärmeübertragung, bei der Energie anhand elektromagnetischer Wellen (infrarote Strahlung, infrarotes Licht) übertragen wird. Hierbei ist kein Trägermedium erforderlich, so dass also auch Wärme im Vakuum übertragen werden kann. Es ist aber eine direkte Sichtverbindung erforderlich, Hindernisse unterbrechen den Strahlungswärmeaustausch. Schwarze Körper absorbieren ein weites Spektrum an elektromagnetischen Wellen und reflektieren wenig, wodurch sie sich schneller als weiße Körper erwärmen, die mehr Strahlung reflektieren. Die physikalischen Eigenschaften des Materials bestimmen, welche Anteile einer auf das Bauteil auftreffenden Strahlung absorbiert, reflektiert oder transmittiert werden.



(a) Spektrale Energiedichte (abgestrahlte Energie je Wellenlänge) der Wärmestrahlung eines schwarzen Körpers bei verschiedenen Temperaturen



(b) Beige Kurve entspricht der Temperatur der Sonne, Rote Kurve entspricht der Umgebungstemperatur

2.2 Welle-Teilchen-Dualismus

Der Welle-Teilchen-Dualismus ist eine Erkenntnis der Quantenphysik, wonach den Objekten der Quantenphysik gleichermaßen die Eigenschaften von Wellen, sowie die von Teilchen zugeschrieben werden müssen. Je nach Art des Experiments tritt entweder der Wellen- oder der Teilchencharakter in Erscheinung, aber niemals beides gleichzeitig.

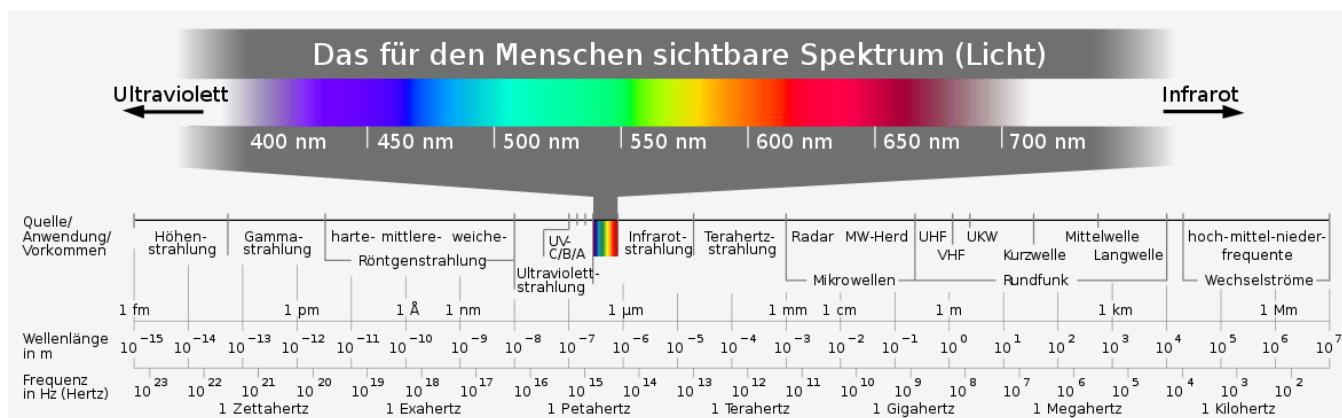
- Klassische Wellen breiten sich im Raum aus. Sie schwächen oder verstärken sich durch Überlagerung und können gleichzeitig an verschiedenen Stellen mit verschiedener Stärke einwirken.
- Ein klassisches Teilchen kann zu einem Zeitpunkt nur an einem bestimmten Ort anwesend sein. Nur dort wirkt es, aber stets mit seiner gesamten Energie, Ladung, Impuls etc.

Beide Eigenschaften scheinen sich gegenseitig auszuschließen. Trotzdem wurde in mehreren Schlüsselexperimenten für verschiedene Quantenobjekte belegt, dass beide Eigenschaften vorliegen, so dass man jedem Körper eine Materiewelle zuschreibt.

Es ist unmöglich, eine anschauliche Vorstellung zu entwickeln, die dem Welle-Teilchen-Dualismus gerecht wird. Die Frage, ob beispielsweise Elektronen oder Lichtquanten „wirklich“ Teilchen oder Wellen seien, ist nicht zu beantworten. Sie sind vielmehr eine eigene Klasse von Quantenobjekten, die je nach der Art der Messung, die man an ihnen durchführt, entweder nur ihre Wellen- oder nur ihre Teilcheneigenschaft in Erscheinung treten lassen, aber nie beide gleichzeitig.

Röntgen-, Gamma- und prinzipiell höher energetische Strahlung weist vorwiegend Teilchencharakter auf und niedriger energetische Strahlung vorwiegend Wellencharakter.

2.3 Elektromagnetisches Spektrum



II

Mathematik

3 Allgemeines

Bei Differentialgleichungen

Grad: Höchste vorkommende Potenz

Ordnung: Höchste vorkommende Ableitung

Bei üblichen Funktionen deuten diese Bezeichnungen beide auf die höchste Potenz in der Funktion hin.

4 Tensorfelder

Als Tensorfeld ist eine Abbildung zu verstehen, die jedem Punkt des Raumes eine Zahl, ein Vektor oder eben ein Tensor bestimmter Stufe zuordnet. So wird zum Beispiel der Spannungszustand eines Materials in jedem Punkt durch einen Spannungstensor zweiter Stufe beschrieben. D.h. im 3-dimensionalen Raum wird die Spannung in einem Punkt durch drei 3×3 Matrizen festgelegt.

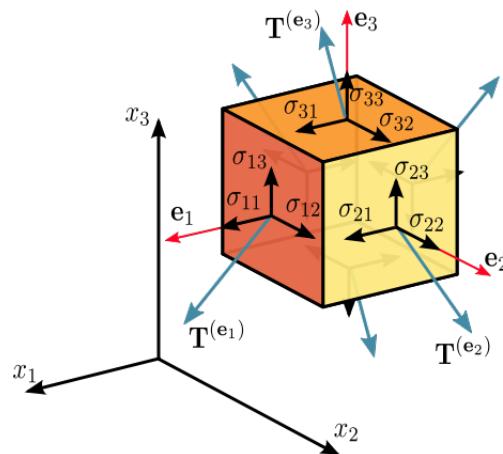


Abbildung 3: Spannungstensor in der Kontinuumsmechanik

5 Lineare Algebra

6 Fast Fourier transformation (FFT)

Die FFT ist ein Algorithmus zur effizienten Berechnung der diskreten Fourier-Transformation (DFT). Mit ihr kann ein digitales Signal in seine Frequenzanteile zerlegt und diese dann analysiert werden.

7 Mathematische Optimierung

7.1 Lineare Optimierung

Jedes Problem der linearen Optimierung kann durch eine Zielfunktion

$$z = \mathbf{c}^T \mathbf{x} \rightarrow \max!$$

und zugehörige Nebenbedingungen

$$\mathbf{A}\mathbf{x} \leq \mathbf{b},$$

$$\mathbf{x} \geq 0$$

formuliert werden.

A ... $m \times n$ -Matrix mit m Nebenbedingungen und n Variablen inkl. Schlupfvariablen

Der Lösungsbereich ist immer Polytop (Simplex) und die optimale Lösung befindet sich immer in einer Ecke. Lineare Optimierungsprobleme können deshalb durch Absuchen aller Ecken des zulässigen Bereichs gelöst werden.

Satz 1: Der zulässige Bereich M ist eine konvexe Teilmenge des \mathbb{R}^n , d.h. mit zwei Punkten \mathbf{x}, \mathbf{y} aus M ist auch jede Konvexitätskombination in M .

Satz 2: Falls der zulässige Bereich M nicht leer und beschränkt ist, so nimmt die Zielfunktion ihr Maximum in mindestens einer Ecke von M an.

Satz 3: Ein Punkt $\mathbf{x} \in M$ ist genau dann eine Ecke von M , wenn die Spaltenvektoren a_j für, welche die zugehörigen Koeffizienten x_j positiv sind, linear unabhängig sind. D.h. für positive x_j müssen die Spalten a_j von A linear unabhängig sein.

Schlupfvariable: Sie ermöglichen es die Nebenbedingungen statt als Ungleichungen, als Gleichungen anzuschreiben zu können. Dies beruht auf der Idee, dass die Ungleichung $x < b$ erfüllt ist, wenn die Gleichung $x = b - c$ für eine beliebige Zahl $c > 0$ gilt. Jedes lineare Optimierungsproblem ist in eine solche Normalform überführbar. Die Werte der Schlupfvariablen sind für die Lösung des Problems nicht von Interesse.

Erstes Eckenkriterium: Ist (x_1, \dots, x_n) eine Ecke, dann sind maximal so viele Variablen ungleich von Null, als der Rang der Matrix A (# NB). Z.B. 3 NB mit zwei Variablen x_1, x_2 und drei Schlupfvariablen. Eine Ecke kann in diesem Fall maximal drei Variablen ungleich Null besitzen. Sind weniger als drei verschieden von Null, heißt die Ecke entartet. Werden so viele Spalten aus der Matrix A gewählt, dass diese vollen Rang besitzt und damit ein Gleichungssystem angeschrieben, dann ist die Lösung eine zulässige Basislösung wenn alle Variablen größer-gleich Null sind. D.h. wählt man 3 linear unabhängige Spalten von A , dann ist dies eine zulässige Basislösung. Andernfalls ist die Basislösung nicht zulässig.

Sei m die Anzahl der Nebenbedingungen (Zeilen von A) und n die Anzahl der Spalten von A inkl. Schlupfvariablen. Verbleiben nach streichen aller l.a. Spalten noch $m=n$, dann ist die Lösung eindeutig (1 Punkt mit optimalem Zielfunktionswert). Falls \mathbf{x} weniger als m positive Komponenten hat, so spricht man von einer entarteten Ecke.

Zweites Eckenkriterium: Jede zulässige Basislösung ist eine Ecke. Umgekehrt ist jede nicht entartete Ecke, durch die entsprechenden Spaltennummern eine Basis.

Ein Lösungsalgorithmus: Man durchlufe alle Basen aus Spalten der Nebenbedingungsmatrix und berechne zugehörige Basislösungen. Ist diese zulässig, hat man eine Ecke gefunden. Auf diese Weise werden alle Ecken gewonnen. Jene mit dem maximalen Zielfunktionswert ist die optimale Lösung.

Schattenpreise: Nach einer gefundenen Lösung soll erörtert werden, ob der Zielfunktionswert weiter erhöht werden kann indem die optimale Lösung (Ecke) leicht variiert wird (Beschränkungen in \mathbf{b} ändern). Jede Nebenbedingung (i) hat einen Schattenpreis $\Delta z = \Delta z(i)$, um den sich der Gewinn erhöht, wenn die Beschränkung der entsprechenden Ressource um eine Einheit erhöht wird. Die Beschränkungen werden durch den Vektor \mathbf{b} beschrieben.

Sensitivitätsanalyse der Zielfunktion: Es wird untersucht, in welchem Rahmen die Koeffizienten der Zielfunktion variiert werden können, ohne die gewonnene optimale Lösung zu ändern.

Sensitivitätsanalyse der Einsatzmittelbeschränkung: Es wird untersucht, in welchem Rahmen die Koeffizienten des Vektors \mathbf{b} geändert werden können, sodass die gewählte Basislösung zulässig ($\mathbf{x}(i) \geq 0$) bleibt.

Das Lösen der Zielfunktion zu einem Optimum wird *primales Problem* genannt. Die Schattenpreise sind die Lösungen des *dualen Problems* oder können auch als sogenannte Lagrangemultiplikatoren erhalten werden.

$$Z = \mathbf{b}^T \mathbf{u} \rightarrow \min,$$

$$\mathbf{A}^T \mathbf{u} \geq \mathbf{c},$$

$$\mathbf{u} \geq 0$$

Zielfunktionswerte des primalen und dualen Problems stimmen stets überein. Es gilt

$$z = \mathbf{c}^T \mathbf{x} = \mathbf{b}^T \mathbf{u} = Z.$$

Daraus folgt, dass die Komponenten von \mathbf{u} gerade die Schattenpreise $\Delta z(i)$ sind. Erhöht man nämlich b_i um eine Einheit, unter Beibehaltung der zulässigen Basis, dann ergibt sich gerade

$$z_{\text{neu}} = \mathbf{b}_{\text{neu}}^T \mathbf{u} = \underbrace{\mathbf{b}_{\text{alt}}^T \mathbf{u}}_{z_{\text{alt}}} + \underbrace{1 \cdot u_i}_{\Delta z}.$$

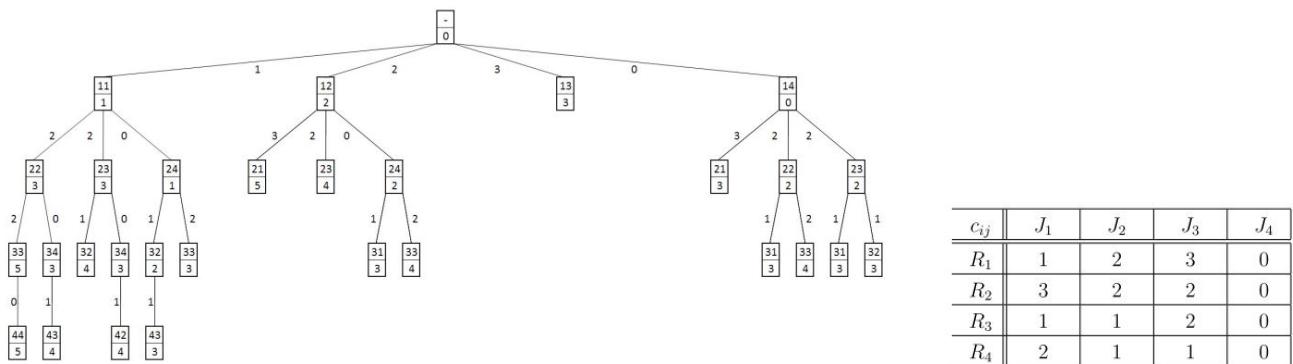
Voragangsweise bei Gleichheitsnebenbedingungen: Falls nicht m Schlupfvariablen auftreten, müssen stattdessen Hilfsvariablen eingeführt werden um eine Startecke zu finden. In der optimalen Lösungen müssen alle Spalten von Hilfsvariablen mit solchen der ursprünglichen Matrix \mathbf{A} getauscht sein, sonst hat das Problem keine Lösung. Man streicht dann die hinteren Spalten, nach Anzahl der eingeführten Hilfsvariablen und erhält eine neue Matrix \mathbf{A} mit einer Einheitsmatrix am Anfang.

Simplexalgorithmus:

7.2 Ganzzahlige Lineare Optimierung

Ein solches Problem wird über Ressourcen R_1, \dots, R_h und Jobs J_1, \dots, J_k , denen Ressourcen zugewiesen werden können beschrieben. Bekannt seien die Kosten c_{ij} einer Ressource für einen bestimmten Job. Gesucht ist der optimale, kostengünstigste Zuordnungsplan, bei dem jede Ressource höchstens einen Job und jeder Job von höchstens einer Ressource bearbeitet wird.

Da der Rechenaufwand für das Auffinden aller optimalen Lösungen sehr groß werden kann, wird über das *Branch-and-Bound-Verfahren* ein reduzierter Entscheidungsbaum erstellt.



Ein Zweig wird terminiert, falls längs dieser Zweigs kein niedriger Zielfunktionswert erzielt werden kann. Oben im Knoten ist die Indexierung und unten der Zielfunktionswert eingetragen. An den Kanten sind die Kosten aufgetragen.

Rucksackproblem

Gegeben sind n Ausstattungen mit Gewichten g_1, \dots, g_n und den Werten w_1, \dots, w_n . Gesucht ist die Kombination an Ausstattung mit dem maximalen Gesamtwert W , deren Gesamtgewicht unterhalb einer vorgeschriebenen Grenze G ist. Zur Formulierung ob der Gegenstand gewählt wurde, wird $x_j = 1$ und falls nicht $x_j = 0$ gesetzt.

$$W = \sum_{j=1}^n w_j x_j \rightarrow \max$$

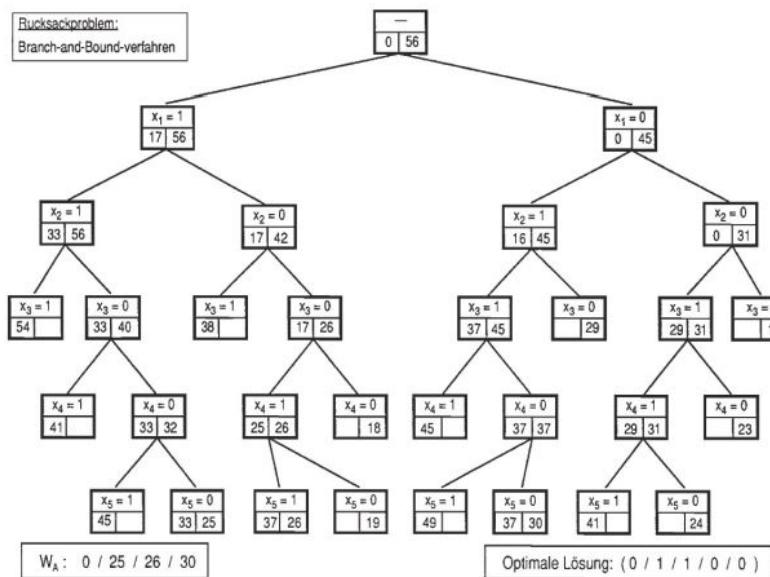
$$\sum_{j=1}^n g_j x_j \leq G$$

Gestartet wird mit der Summe der Werte aller Gegenstände $W_O = \sum_{j=1}^n w_j$ und einer unteren Gewichtsschranke $G_0 = 0$. Wird ein Gegenstand ausgewählt ($x_j = 1$), dann wird zur unteren Gewichtsschranke G_U das Gewicht g_j hinzugefügt. Falls der Gegenstand nicht verwendet wird ($x_j = 0$), dann wird der maximal erreichbare Gesamtwert W um den Wert w_j verringert. Ein Zweig wird terminiert, wenn die untere Gewichtsschranke die vorgegebene Gewichtsschranke überschreitet $G_U > G$ oder wenn eine neue obere Wertschranke W_O kleiner wird als der aktuelle Gesamtwert W_A .

Das Allgemeine Problem, Dakin-Verfahren

Das allgemeine ganzzahlige Optimierungsproblem hat die Form

$$W = \sum_{j=1}^n c_j x_j \rightarrow \max$$



j	1	2	3	4	5
g_j	17	16	21	8	12
w_j	11	14	16	8	7

Mathematische Formulierung:

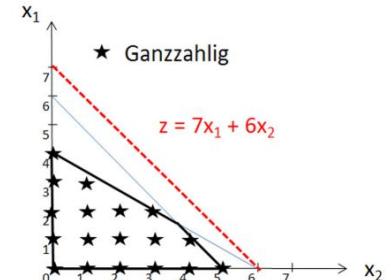
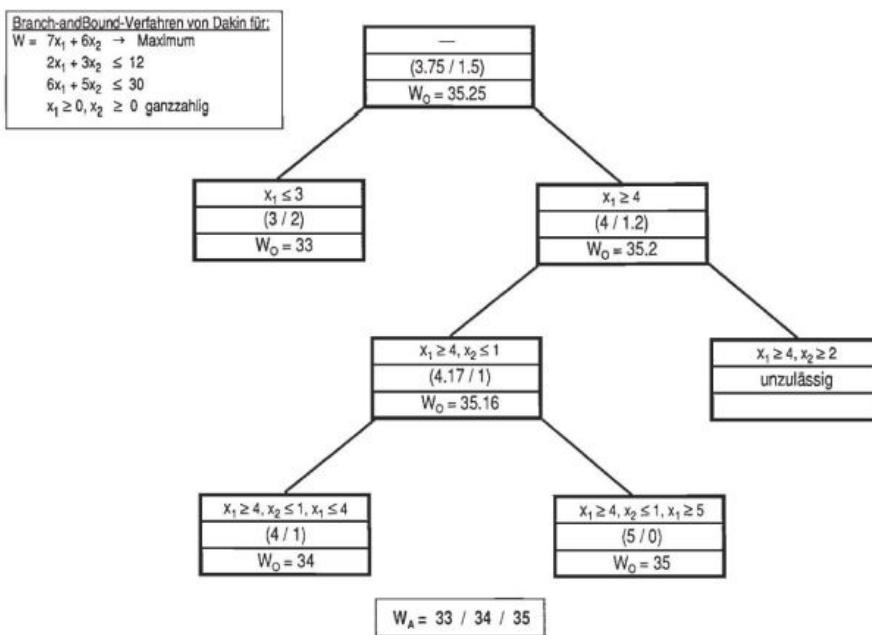
$$\begin{aligned} 11x_1 + 14x_2 + 16x_3 + 8x_4 + 7x_5 &\rightarrow \max! \\ 17x_1 + 16x_2 + 21x_3 + 8x_4 + 12x_5 &\leq 37 \end{aligned}$$

$$\sum_{j=1}^n a_{ij} x_j \leq b_j$$

$$x_j \geq 0, \text{ und ganzzahlig.}$$

Das Problem wird in mehrere Teilprobleme aufgespalten. Als erstes wird die Ganzzahligkeitsanforderung weggelassen und eine möglicherweise nichtganzzahlige Ecke aus den Nebenbedingungen berechnet. Ebenso der zugehörige Zielfunktionswert, der ebenfalls nichtganzzahlig sein kann.

Diese Lösung wird aufgespalten, indem man die erste nichtganzzahlige Variable nimmt und die größte ganzzahlige Zahl x_j nennt. Z.B. von $3.75 \rightarrow x_j = 3$. Über die jeweilige NB werden dann die restlichen Variablen berechnet. Sind alle Variablen ganzzahlig, wird terminiert. Sind jedoch noch nicht alle Variablen ganzzahlig, wird weiter aufgespalten. Man nimmt die nächste nichtganzzahlige Variable und runden diese wieder einmal *floor* und *ceil*. Z.B. aus $1.2 \rightarrow 1$ und 2 . Falls die Variablen die zugehörige NB verletzen, wird terminiert. Falls das Wählen der zweiten Variable zur Ganzzahl die erste Variable wieder nichtganzzahlig macht, wird erneut aufgespalten. Dies wird solange iteriert bis alle Zweige terminiert sind. Eine weitere Terminierungsbedingung ist, wenn eine nichtganzzahlige Lösung vorliegt und der zugehörige Zielfunktionswert kleiner als größte bereits bekannte Zielfunktionswert einer zulässigen Lösung ist.



Das Reihenfolgeproblem, Johnson-Algorithmus (*greedy-algorithm*)

Es seien n Jobs gegeben die hintereinander auf Maschine M_1 und dann auf M_2 erledigt werden müssen. Die Bearbeitungszeit des Jobs j sei auf M_1 sei a_j und auf M_2 , b_j . Gesucht ist der optimale Belegungsplan, der die optimale Reihenfolge mit der kürzesten benötigten Gesamtzeit aufweist (äquivalent: kürzeste Standzeiten der Maschinen).

Für zwei Maschinen liefert der Johnson-Algorithmus eine exakte Lösung. Bei mehreren Maschinen bzw. komplizierteren Arbeitsabläufen kann das Problem als allgemeine ganzzahlige Optimierungsaufgabe gestellt und mit dem Dakin-Verfahren gelöst werden.

Man suche $\min(a_1, \dots, a_n, b_1, \dots, b_n)$. Ist der kürzeste Arbeitsschritt a_j , wird der Job j als Erstes abgearbeitet. Ist der kürzeste Arbeitsschritt b_j , dann wird der Job j als letztes abgeschlossen. Dann wird Job j gestrichen und der Vorgang beginnt von vorne, bis man die optimale Reihenfolge der Jobs gefunden hat.

7.3 Globale Optimierung

Die allgemeine Aufgabe besteht darin, eine Funktion f über eine abgeschlossene Teilmenge D zu minimieren.

$$f(x) \rightarrow \min, \quad x \in D$$

D wird durch Nebenbedingungen in Form von Gleichungen und Ungleichungen gegeben. Maximierungsaufgabe ist wegen $\max(f) = -\min(-f)$ ebenfalls abgedeckt. Falls die Gleichungen differenzierbar sind, bietet sich die Methode der *Lagrangemultiplikatoren* an.

Lagrangemultiplikatoren, NB als Gleichungen

Für eine gegebene differenzierbare Funktion f soll die Minimierungsaufgabe

$$f(x) \rightarrow \min$$

unter den Nebenbedingungen

$$g_1(x) = 0, \dots, g_k(x) = 0$$

gelöst werden, wobei die NB differenzierbar und ihre Gradienten

$$\nabla g_1(x), \dots, \nabla g_k(x)$$

als linear unabhängig vorausgesetzt werden. Die *Lagrangemultiplikatorregel* wird wie folgt angewendet.

1. Schritt: Man bilde die Hilfsfunktion in $n+k$ Variablen $x_1, \dots, x_n, u_i, \dots, u_k$

$$L(\mathbf{x}, \mathbf{u}) = f(\mathbf{x}) + \mathbf{u}^T \mathbf{g}(\mathbf{x})$$

und berechne die Gradienten der Hilfsfunktion.

2. Schritt: Man bestimmte die stationären Punkte der Lagrangefunktion. Das sind die Lösungen des Gleichungssystems

$$\nabla L(x_1, \dots, x_n, u_i, \dots, u_k) = \mathbf{0}.$$

Das heißt,

$$\begin{aligned} \frac{\partial}{\partial x_i} L &= \frac{\partial}{\partial x_i} f(\mathbf{x}) + u_1 \frac{\partial}{\partial x_i} g_{1(x)} + \dots + u_k \frac{\partial}{\partial x_i} g_k(\mathbf{x}) \\ \frac{\partial}{\partial u_j} L &= g_j(\mathbf{x}) = \mathbf{0} \end{aligned}$$

für $i = 1, \dots, n$ und $j = 1, \dots, k$

3. Schritt: Die Maxima und Minima sind unter den Lösungen (x_1, \dots, x_n) zu finden.

Lagrangemultiplikatoren, NB als Ungleichungen

Liegen die Nebenbedingungen als Ungleichungen vor,

$$f(x) \rightarrow \min$$

$$\mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \quad \mathbf{x} \geq \mathbf{0}$$

so bildet man ebenfalls die Lagrangefunktion

$$L(\mathbf{x}, \mathbf{u}) = f(\mathbf{x}) + \mathbf{u}^T \mathbf{g}(\mathbf{x}).$$

Anstelle des Aufsuchens der stationären Punkte der Lagrangefunktion sind jetzt die Sattelpunkte zu ermitteln. Ein Paar (\mathbf{x}, \mathbf{u}) mit $\mathbf{x} \geq \mathbf{0}, \quad \mathbf{u} \geq \mathbf{0}$ heißt Sattelpunkt, falls

$$L(x, v) \leq L(x, u) \leq L(y, u)$$

für alle (\mathbf{y}, \mathbf{v}) mit $\mathbf{y} \geq \mathbf{0}, \mathbf{v} \geq \mathbf{0}$ erfüllt ist. Das heißt (\mathbf{x}, \mathbf{u}) ist ein Sattelpunkt wenn der Funktionswert beim Verlassen des Punktes in v -Richtung abnimmt ($L(x, v)$) und in y -Richtung zunimmt ($L(y, u)$).

7.3.1 Quadratische Optimierung

Das quadratische Standardproblem ist

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{b}^T \mathbf{x}$$

mit einer symmetrischen und positiv definiten Matrix \mathbf{A} . Das Lösen des linearen Gleichungssystems $\mathbf{A}\mathbf{x} = \mathbf{b}$ ist äquivalent und führt zum selben Ergebnis. Denn der Gradient ist

$$\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}.$$

Vorkonditionierung: Eine Matrix \mathbf{B} zur Vorkonditionierung dient dazu die Kondition, also die Abhängigkeit der Lösung von den Eingangsdaten, zu verringern.

Eine Iteration kann dann folgendermaßen geschrieben werden, wobei α die Distanz in die Abstiegsrichtung \mathbf{s}_k bis zur nächsten Iteration \mathbf{x}_{k+1} beschreibt.

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha \mathbf{s}_k, \quad \mathbf{B}\mathbf{s}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k \quad \Rightarrow \quad \mathbf{s}_k = -\mathbf{B}^{-1} \nabla f(\mathbf{x}_k)$$

Bei der *konjugierten Gradienten Methode* wird die Abstiegsrichtung \mathbf{p}_k nochmals mit dem Relaxationsfaktor β_k korrigiert, um ein Optimum in weniger Iterationen zu finden.

$$\mathbf{p}_k = \mathbf{s}_k + \beta_k \mathbf{p}_{k-1}$$

Sobald der Betrag des Gradienten eine gewisse Toleranz unterschreitet werden die Verfahren abgebrochen.

7.3.2 Lineare kleinste Fehlerquadrate

Das Problem überbestimmter linearer Systeme und die Minimierung der Summe der Fehlerquadrate ist

$$f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 \rightarrow \min.$$

Man kann auch

$$f(\mathbf{x}) = (\mathbf{A}\mathbf{x} - \mathbf{b})^T (\mathbf{A}\mathbf{x} - \mathbf{b}) = \mathbf{x}^T \mathbf{A}^T \mathbf{A}\mathbf{x} - 2\mathbf{x}^T \mathbf{A}^T \mathbf{b} + \mathbf{b}^T \mathbf{b}$$

schreiben. Über die partiellen Ableitungen, den Gradienten

$$\nabla f(\mathbf{x}) = 2\mathbf{A}^T \mathbf{A}\mathbf{x} - 2\mathbf{A}^T \mathbf{b}$$

und der Bedingung, dass dieser zu Null wird, können die gesuchten Parameter \mathbf{x} über das lineare Gleichungssystem

$$\mathbf{A}^T \mathbf{A}\mathbf{x} = \mathbf{A}^T \mathbf{b}$$

berechnet werden. Die Bedingung $\nabla f(\mathbf{x}) = \mathbf{0}$ folgt aus der Forderung das Fehlerquadrat zu minimieren. Da das Fehlerquadrat in jedem Punkt konstant ist, ergibt sich in den partiellen Ableitungen jedes Punktes, dieser Wert zu Null.

7.3.3 Nichtlineare kleinste Fehlerquadrate

Eine Idee ist dieses nichtlineare Problem als eine Abfolge von linearen kleinsten Fehlerquadraten iterativ zu Lösen. Die nichtlineare Funktion $\mathbf{F}(\mathbf{x})$ wird über die Taylor-Entwicklung durch quadratische Funktionen angenähert, die wiederum zum Lösen des linearen Gleichungssystems $\mathbf{Ax} - \mathbf{b}$ äquivalent sind (*Gauss-Netwon Verfahren*). Der Vorteil ist, dass die oft schwierig zu berechnenden zweiten Ableitungen nicht benötigt werden.

$$\mathbf{F}(\mathbf{x}) \approx \mathbf{F}(\mathbf{x}^k) + \mathbf{J}_F(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k)$$

Der nächste Iterationsschritt wird durch Lösung des Minimierungsproblems

$$f(\mathbf{x}) = \|\mathbf{F}(\mathbf{x}^k) + \mathbf{J}_F(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k)\|^2 \rightarrow \min$$

gewonnen. Mit $\mathbf{A} = \mathbf{J}_F(\mathbf{x}^k)$, $\mathbf{s}^k = \mathbf{x} - \mathbf{x}^k$ und $\mathbf{b} = -\mathbf{F}(\mathbf{x}^k)$ erhalten wir das Problem der linearen kleinsten Fehlerquadrate

$$\|\mathbf{As}^k - \mathbf{b}\|^2 \rightarrow \min$$

Verbesserte Varianten beheben das Problem, dass sich die Summe der Fehlerquadrate nicht mit jeder Iteration verringern. Das Problem liegt bei α der in Abstiegsrichtung ein zu großes Inkrement verweist. Dafür ist ein *line search* notwendig, das die Zielfunktion in Abstiegsrichtung minimiert. Auch bei der *trust region Gauss-Newton* Methode geht es darum, die *sufficient decrease condition* zu erfüllen. Durch das folglich reduzierte auswerten von Funktionswerten wird der Algorithmus effizienter und benötigt weniger Iterationsschritte.

7.3.4 Ableitungsfreie Methoden

Koordinatensuch-Methoden

Sie basieren auf der Suche von Funktionswerten anhand eines Rasters auf dem die Zielfunktion ausgewertet wird. Schritt für Schritt bewegt man sich mit dem Raster in Richtung absteigender Zielfunktionswerte. Wird kein kleinerer Zielfunktionswert auf dem Raster gefunden, wird das Raster verfeinert. Um die Wahrscheinlichkeit eines Verfangens in lokalen Minimas zu verringern werden zufällig gewählte Startwerte durchprobiert und der Algorithmus durchgeführt. Man wählt dann den kleinsten Zielfunktionswert aus allen *runs*.

Downhill-Simplex-Methode

Die Startkonfiguration ist eine Selektion aus $n + 1$ Punkten \mathbb{R}^n .

Schritt 1: Diese Punkte sind nach ihren Funktionswerten (beste → schlechtester) geordnet. In einer Iteration wird der schlechteste Punkt durch Reflexion, Expansion oder Kontraktion ausgetauscht.

Schritt 2: Berechne den Mittelpunkt des Dreiecks.

Schritt 3: Zuerst wird der reflektierte Punkt mit dem schlechtesten Funktionswert berechnet. Ist der reflektierte Punkt besser als der Beste, dann berechne den expandierten Punkt (Schritt 4) und ersetze den schlechtesten durch den besseren (Reflexion oder Expansion). Zurück zu Schritt 1. Ist der reflektierte Punkt nur besser als der zweitschlechteste, dann ersetze den schlechtesten durch den Reflektierten. Zurück zu Schritt 1.

Schritt 5: Falls der reflektierte Punkt nicht besser als der Zweitschlechteste (Mittlere) ist, dann berechne den kontrahierten Punkt. Falls dieser besser als der Reflektierte ist, dann ersetze den schlechtesten durch diesen. Zurück zu Schritt 1.

Schritt 6: Wird durch Reflexion, Expansion oder Kontraktion kein besserer Funktionswert mehr erreicht, kann durch Reduktion aller Punkte zum besten hin, der Zielfunktionswert noch verbessert werden.

Simmulierte Abkühlung

Grundidee ist die Nachbildung eines Abkühlungsprozesses, etwa beim Glühen in der Werkstoffkunde. Nach Erhitzen eines Metalls sorgt die langsame Abkühlung dafür, dass die Atome ausreichend Zeit haben, sich zu ordnen und stabile Kristalle zu bilden. Dadurch wird ein energieärmer Zustand nahe am Optimum erreicht.

Es wird ein Startwert \mathbf{x} und eine Temperaturfolge die gegen Null fällt gewählt. Der Wert wird mit einem Nachbarspunkt \mathbf{y} verglichen $\Delta f = f(\mathbf{x}_0) - f(\mathbf{y})$.

Falls $\Delta f > 0$ dann wird $\mathbf{x}_1 = \mathbf{y}$ gewählt.

Falls $\Delta f \leq 0$ dann wird nur mit einer Wahrscheinlichkeit der Punkt $\mathbf{x}_1 = \mathbf{y}$ übernommen und mit einer Gegenwahrscheinlichkeit wird der Schritt erneut mit einem anderen Nachbarspunkt ausgeführt.

Dies wird für jeden Temperaturschritt durchgeführt.

Verdeutlichung: Angenommen, man sucht in einer zweidimensionalen Landschaft den (global) tiefsten Punkt. Die Landschaft selbst besteht aus vielen unterschiedlich tiefen Dellen. Die einfache Suchstrategie (suche den nächsten tiefsten Punkt) entspricht dem Verhalten einer Kugel, welche in dieser Landschaft ausgesetzt wird. Sie rollt zum nächsten lokalen Minimum und bleibt dort. Bei der simulierten Abkühlung wird der Kugel immer wieder ein Stoß

versetzt, der mit zunehmender „Abkühlung“ schwächer wird. Dieser ist idealerweise stark genug, um die Kugel aus einer flachen Delle (lokales Minimum) zu entfernen, reicht aber nicht aus, um aus dem globalen Minimum zu fliehen.

Surrogate-Modelling

Es werden einfachere Modelle verwendet, die die Optimierungsfunktion annähern. Es wird dann diese Näherung optimiert.

Beispielsweise kann eine echte Optimierungsfunktion durch kubische Splines interpoliert werden. Man kann dann für jeden lokalen Spline das Optimum berechnen und erhält somit auch das genäherte Optimum der Zielfunktion.

Genetische Algorithmen

Diese Algorithmen ahmen natürliche Selektion nach.

- Es wird mit einem Set von *Parents* gestartet. $\mathbf{P} = [x_1, \dots, x_n]$
- Für jeden Parent wird eine Fitness berechnet. $F(x_1), \dots, F(x_n)$
- Bis ein Abbruckkriterium erfüllt ist
 - Selektion und Rekombination von Individuen anhand von Fitnesswerten
 - Mutation
 - Fitnessevaluation
 - Selektion einer neuen Generation

Partikelschwarm-Optimierung

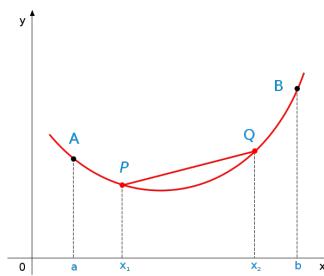
Als Partikelschwarmoptimierung (PSO) wird ein naturanaloges Optimierungsverfahren bezeichnet, das nach dem Vorbild des biologischen Schwarmverhaltens eine Lösung für ein Optimierungsproblem sucht. Analog zum natürlichen Phänomen wird eine Population von Lösungskandidaten durch den Suchraum bewegt, um eine gute Lösung für das Problem zu erhalten. In jedem Rechenschritt wird dazu die Position jedes Individuums und der zugehörige Funktionswert neu berechnet. Beim Schwarm haben die Nachbarn Einfluss auf ein Individuum. Voraussetzung ist ebenfalls das der Schwarm zusammenhält.

Neuronale Netzwerke

Ein neuronales Netz besteht aus einer Eingangsschicht, einer oder mehrerer Schichten aus künstlichen Neuronen (Übertragungsfunktionen) und einer Ausgangsschicht. Nach der Konstruktion eines Netzes folgt die Trainingsphase. Beim Lernen werden neue Verbindungen gesetzt, gelöscht, die Gewichtung verändert, Schwellwerte angepasst oder ganze Neuronen hinzugefügt oder gelöscht. Jedes Neuron hat demnach einen Threshold ab dem es aktiviert wird und ein Gewicht mit dem der Eingang, dem Ausgang übergeben wird.

7.4 Konvexe Optimierung

Der zentrale Standardfall der nichtlinearen Optimierung ist die konvexe Optimierung, bei der die Funktion f und g_i als konvex vorausgesetzt werden. f heißt *konvex*, wenn der Graph unterhalb der Verbindungsgeraden, zweier auf der Funktion liegende Punkte, verläuft.



$$f(\mu \mathbf{P} + (1 - \mu) \mathbf{Q}) \leq \mu f(\mathbf{P}) + (1 - \mu) f(\mathbf{Q}), \quad 0 \leq \mu \leq 1$$

Die NB sind Ungleichungen. Es kann jedoch aus $g_j \leq 0$ und $-g_j \leq 0$, eine Gleichheitsnebenbedingung $g_j = 0$ gebildet

werden. Falls die definierende Ungleichung strikt ($<$ statt \leq) ist, so heißt die Funktion *streu konvex*. Eine partiell differenzierbare Funktion ist genau dann konvex, wenn für alle \mathbf{x}, \mathbf{y} gilt

$$f(\mathbf{y}) \geq f(\mathbf{x}) + (\mathbf{y} - \mathbf{x})^T \nabla f(\mathbf{x}).$$

Das heißt, wenn ihr Graph für jeden Punkt oberhalb der zugehörigen Tangente verläuft. Für das Lösen eines konvexen Optimierungsproblems gehen wir davon aus, dass alle beteiligten Funktionen f und g_i konvex sind. Die NB können in einem Vektor $\mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}), \dots, g_m(\mathbf{x})]^T$ zusammengefasst werden und führen den konvexen zulässigen Bereich ein, der durch die NB definiert wird.

$$M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{x} \geq \mathbf{0}\}$$

Zur Lösung des Optimierungsproblems wird die *Lagrangefunktion* mit den *Lagrangemultiplikatoren* u_j gebildet

$$L(\mathbf{x}, \mathbf{u}) = f(\mathbf{x}) + \mathbf{u}^T \mathbf{g}(\mathbf{x}) = f(\mathbf{x}) + \sum_{j=1}^m u_j g_j.$$

Satz: Falls ein Paar (\mathbf{x}, \mathbf{u}) ein Sattelpunkt der *Lagrangefunktion* L ist, so ist \mathbf{x} die Lösung der Minimierungsaufgabe.

Slater-Bedingung: Falls der zulässige Bereich M innere Punkte \mathbf{x}' besitzt, d.h. falls es mindestens ein $\mathbf{x}' \geq \mathbf{0}$ gibt für den $\mathbf{g}(\mathbf{x}') < \mathbf{0}$ gilt.

Theorem von Kuhn-Tucker: Falls M die *Slater-Bedingung* erfüllt, dann ist $\mathbf{x} \geq \mathbf{0}$ genau dann eine Lösung des Minimierungsproblems, wenn es ein $\mathbf{u} \geq \mathbf{0}$ gibt und das Paar (\mathbf{x}, \mathbf{u}) ein Sattelpunkt von L ist.

$$\frac{\partial L}{\partial \mathbf{x}} \geq \mathbf{0}, \quad \mathbf{x}^T \partial_x L(\mathbf{x}, \mathbf{u}) = 0, \quad \mathbf{x} \geq \mathbf{0}$$

$$\frac{\partial L}{\partial \mathbf{u}} \leq \mathbf{0}, \quad \mathbf{u}^T \partial_u L(\mathbf{x}, \mathbf{u}) = 0, \quad \mathbf{u} \geq \mathbf{0}$$

Man muss also für eine Funktion f und die NB in Form von Ungleichungen prüfen ob die obigen Kuhn-Tucker Bedingungen, sowie $\mathbf{x} \geq \mathbf{0}$, $\mathbf{u} \geq \mathbf{0}$ erfüllt sind. Sind die Kuhn-Tucker Bedingungen erfüllt, dann ist (\mathbf{x}, \mathbf{u}) ein Sattelpunkt und \mathbf{x} somit eine Lösung des Minimierungsproblems.

Quadratische Optimierung mit Kuhn-Tucker

Das Standardproblem der quadratischen Optimierung mit linearen NB ist

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{C} \mathbf{x} + \mathbf{c}^T \mathbf{x} \rightarrow \min$$

$$\mathbf{A} \mathbf{x} \leq \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0}.$$

Mit positiv semidefiniter ($\mathbf{x}^T \mathbf{C} \mathbf{x} \geq 0$) Matrix \mathbf{C} . Die *Lagrangefunktion* ist dann

$$L(\mathbf{x}, \mathbf{u}) = \mathbf{x}^T \mathbf{C} \mathbf{x} + \mathbf{c}^T \mathbf{x} + \mathbf{u}^T (\mathbf{A} \mathbf{x} - \mathbf{b}).$$

Die zugehörigen Gradienten werden als \mathbf{v} und $-\mathbf{y}$ definiert

$$\partial_x L(\mathbf{x}, \mathbf{u}) = 2\mathbf{C} \mathbf{x} + \mathbf{c} + \mathbf{A}^T \mathbf{u} =: \mathbf{v},$$

$$\partial_u L(\mathbf{x}, \mathbf{u}) = \mathbf{A} \mathbf{x} - \mathbf{b} =: -\mathbf{y}$$

Die Sattelpunktbedingungen werden dann über die Kuhn-Tucker Bedingungen als

$$\mathbf{v} \geq \mathbf{0}, \quad \mathbf{x}^T \mathbf{v} = 0,$$

$$\mathbf{y} \geq \mathbf{0}, \quad \mathbf{u}^T \mathbf{y} = 0,$$

formuliert.

7.5 Variationsrechnung

Ein Funktional ist eine Funktion, die anderen Funktionen Zahlen zuordnet. Es ist also eine Abbildung $\phi : M \rightarrow \mathbb{R}$ von einer Funktionenmenge in die reellen Zahlen. Ein Funktional ordnet also jeder Funktion $u \in M$ eine Zahl $\phi(u)$ zu. Ziel der Variationsrechnung ist die Minimierung solcher Funktionale.

Fundamentallemma der Variationsrechnung: Ist f eine stetige Funktion und $\int_{x_0}^{x_1} f(x)v(x)dx = 0$ für alle stetigen Funktionen v und auch $v(x_0) = v(x_1) = 0$, dann ist auch $f(x) = 0$ für alle $x \in [x_1, x_1]$.

Die *erste Variation* ist eine verallgemeinerte Richtungsableitung eines Funktionals $\phi : M \rightarrow \mathbb{R}$. Eine Änderung $v \in V$ (V sei Funktionenraum) heißt zulässig, wenn $u + \tau v$ ebenfalls in M liegt. Für alle τ auf einem Intervall $[-\tau_0, \tau_0]$.

$$\delta\Phi(u)[v] = \frac{d}{d\tau}\Phi(u + \tau v)\Big|_{\tau=0}$$

7.6 Deterministische Optimierungsverfahren

Dia Aufzählung unter deam nur als allgemein Info ganz oba ane.

7.6.1 Nichtlineare Optimierung ohne Nebenbedingungen

7.6.1.1 Klassische Gradientenverfahren

7.6.1.2 Konjugierte Gradientenverfahren

Algorithmen zum finden eines Minimums welche den Gradienten benutzen, „wandern“ stets in Richtung des steilsten Abstiegs. Der konjugierte Gradient stellt dabei eine effizientere Methode dar, welcher eine Art verbesserte Richtung benutzt. Mit dem Verfahren des konjugierten Gradienten, soll das lineare Gleichungssystem $Ax = b$ gelöst werden. Die Systemmatrix A ist dabei symmetrisch und positiv definit. Zum lösen gilt es die die Kostenfunktion

$$F(x) = \frac{1}{2}x^T Ax - x^T b$$

zu minimieren. Das Verfahren soll im folgenden beschrieben werden. Der Startvektor x_0 ist dabei beliebig zu wählen. Mit diesem wird der erste Residuenvektor r_0 und der erste Korrekturvektor p_0 berechnet:

$$x_0; \quad r_0 = b - Ax_0; \quad p_0 = r_0$$

Die folgenden Schritte i werden iterativ berechnet. Der neue Vektor x_{i+1} wird durch den alten Vektor x_i in die Richtung von p_i bestimmt. Die Schrittlänge wird dabei durch den Faktor λ_i festgelegt:

$$\lambda_i = \frac{r_i^T p_i}{p_i^T A p_i}; \quad x_{i+1} = x_i + \lambda_i p_i$$

Für den nächsten Iterationsschritt wird nun noch ein neuer Residuenvektor r_{i+1} und ein neuer Korrekturvektor p_{i+1} berechnet:

$$r_{i+1} = r_i - \lambda_i A p_i; \quad p_{i+1} = r_{i+1} + \frac{r_{i+1}^T r_{i+1}}{r_i^T r_i} p_i$$

Das Verfahren konvergiert, abgesehen von Rundungsfehler, nach n Iterationsschritten, wobei n die Größe der quadratischen Matrix $A \in \mathbb{R}^{n \times n}$ ist. Anschaulich formuliert ist die Richtung des neuen Korrekturvektor eine verbesserte Richtung des alten Residuenvektor und zeigt somit besser in Richtung des gesuchten Ziels. In der Abbildung ist zu erkennen, dass der „rote“ Weg schneller am gesuchten Ziel X ist als der Grüne.

7.6.1.3 Newton-Verfahren

Grundlage für das Quasi-Newton-Verfahren ist das Newton-Verfahren, welches ein Iterationsverfahren ist und auf einen bestimmten Wert konvergiert. Ausgehend von einem Startwert x_0 wird anschließend folgende Iteration durchgeführt:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

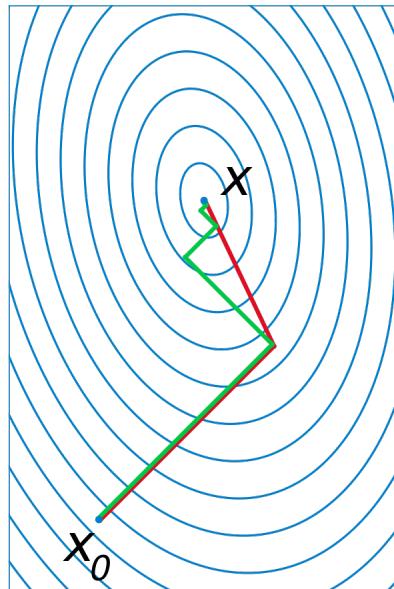


Abbildung 4: Grün: Gradient, Rot: Konjugierter Gradient

Dieses Verfahren wird zur Identifikation von Nullstellen verwendet. Um das Verfahren auf ein Optimierungsproblem anzuwenden, wird jedoch das Minimum benötigt. Die Formel ändert sich somit zu:

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$$

Im mehrdimensionalen entspricht hierbei die erste Ableitung dem Gradienten und die zweite Ableitung der Hesse Matrix. Es ergibt sich somit:

$$x_{n+1} = x_n - H^{-1}(x_n)\nabla f(x_n)$$

Dieses Verfahren hat jedoch einige Nachteile:

1. Bei jedem Iterationsschritt muss die Hesse Matrix berechnet werden, was sehr aufwendig ist.
2. Bei jedem Iterationsschritt muss die Newtongleichung

$$d = -H^{-1}(x_n)\nabla f(x_n)$$

gelöst werden. Wobei d der Suchrichtung entspricht.

7.6.1.4 Quasi-Newton-Verfahren

Das Quasi-Newton-Verfahren gehört zu den Verfahren, die zur Lösung von nichtlinearen Optimierungsproblemen verwendet werden. Um diese Nachteile zu umgehen wurden nun die Quasi-Newton-Verfahren entwickelt. Es wird hierbei zwischen dem direkten und dem inversen Quasi-Newton verfahren unterschieden. Im Folgenden wird nun näher auf das inverse Verfahren eingegangen. Anstelle der inversen Hesse Matrix selbst wird hier nur eine geeignete Approximation B_n verwendet. Es ergibt sich die Gleichung:

$$x_{n+1} = x_n - B(x_n)\nabla f(x_n)$$

An die Matrix B werden jedoch einige Anforderungen gestellt.

1. Die Matrix B_n muss symmetrisch sein, da auch die Hesse-Matrix symmetrisch ist.
2. Die Matrix B_{n+1} muss der inversen Quasi-Newton-Gleichung

$$B_{n+1}s_n = y_n$$

genügen. Wobei

$$s_n = x_{n+1} - x_n; \quad y_n = \nabla f(x_{n+1}) - \nabla f(x_n)$$

entspricht.

3. Die Matrix B_{n+1} sollte sich relativ einfach aus B_n berechnen lassen.

Für die Berechnung von B_{n+1} wird hier beispielhaft das Davidon-Fletcher-Powell Verfahren gezeigt. Bei diesem Verfahren berechnet sich B_{n+1} folgendermaßen:

$$B_{n+1} = B_n + \frac{s_n s_n^T}{s_n^T y_n} - \frac{B_n y_n y_n^T B_n}{y_n^T B_n y_n}$$

Das Verfahren konvergiert nun, wenn ein bestimmtes Abbruchkriterium erreicht ist. Ein Abbruchkriterium wäre zum Beispiel, dass der Gradient unter einem bestimmten Wert liegt.

7.6.1.5 Downhill-Simplex-Verfahren/ Nelder-Mead-Verfahren

Das Downhill Simplex Verfahren (auch Nelder Mead Algorithmus genannt) ist ein Ableitungsfreier Suchalgorithmus für lokale Minima bzw. Maxima. Er ist robust gegenüber Unstetigkeiten und besitzt lineare Konvergenz.

Simplex bezeichnet das einfachste Volumen in \mathbb{R}^n aus $n + 1$ Punkten. Beispielsweise ist ein Simplex in \mathbb{R}^1 eine Strecke oder in \mathbb{R}^2 ein Dreieck.

Der Algorithmus ist dabei Folgendermaßen aufgebaut:

1. Wähle $n + 1$ Punkte die ein Simplex bilden.
2. Sortiere sie nach Wertigkeit $f(x_0) \leq f(x_1) \leq \dots \leq f(x_n)$.
3. Berechne den Mittelpunkt $m = \frac{1}{n} \sum_{i=0}^{n-1} x_i$ von allen außer den schlechtesten Punkt.
4. Reflektiere den schlechtesten Punkt x_n am Mittelpunkt m (reflektierter Punkt $r = m + \alpha(m - x_n)$ wobei α meist 1 gewählt wird).
 - Wenn r zwischen besten und schlechtesten Wert liegt, ersetze x_n durch r und gehe zu 3. .
 - Wenn r bester Wert gehe zu 5. .
 - Wenn r schlechtester Wert gehe zu 6. .
5. Expandiere noch weiter in Richtung des Reflektierten Punktes. (expandierter Punkt $e = m + \beta(m - x_n)$ wobei $\beta > \alpha$).
 - Wenn e besser als r ersetze x_n durch e und gehe zu 3. .
 - Wenn e nicht besser oder gleich r , ersetze x_n durch r und gehe zu 3. .
6. Kontrahiere den schlechtesten Punkt x_n (kontrahierter Punkt $k = m + \gamma(m - x_n)$ wobei γ normalerweise $-\frac{1}{2}$ gewählt wird).
 - Wenn k besser als schlechtester Punkt ist, ersetze x_n durch k und gehe zu 3. .
 - sonst gehe zu 7. .
7. Komprimiere alle außer den besten Punkt $x_i = x_i + \delta(x_i - x_0)$ für $i = 1, \dots, n$ und $\delta < 1$; danach zu 3. .

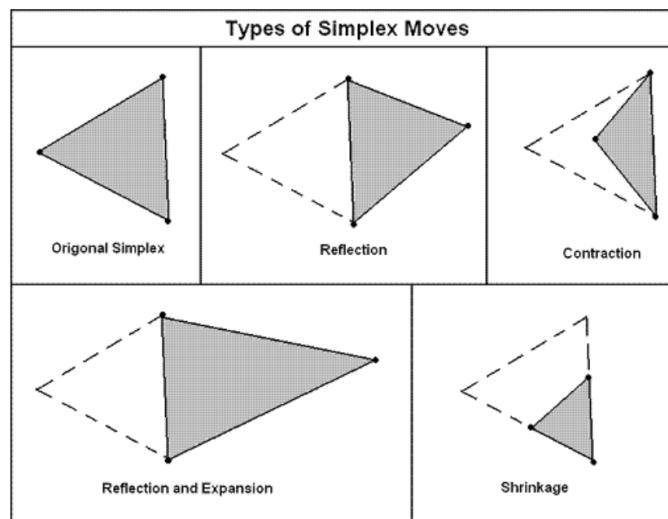
7.7 Stochastische Optimierungsverfahren

7.7.1 Evolutionäre Algorithmen

7.7.1.1 Genetische Algorithmen

7.7.1.2 Schwarmintelligenz-Verfahren

Multimodal: Funktion hat mehrere lokale Minimas.

Abbildung 5: Downhill Simplex Schritte in $n = 2$

8 Numerische Mathematik

- 8.1 Lineare Gleichungssysteme - Lösungsverfahren
- 8.2 Nichtlineare Gleichungssysteme - Lösungsverfahren
- 8.3 Numerische Integration - Lösungsverfahren
- 8.4 Numerische Approximation und Interpolation
- 8.5 Gewöhnliche Differentialgleichungen - Lösungsverfahren
- 8.6 Partielle Differentialgleichungen - Lösungsverfahren
- 8.7 Eigenwertberechnung - Lösungsverfahren
- 8.8 FFT - schnelle Fourier-Transformation

III

Mechanik

9 Begrifflichkeiten

9.1 Generalisierte Koordinaten

Die generalisierten (oder verallgemeinerten) Koordinaten bilden in der Theoretischen Mechanik und der Technischen Mechanik einen minimalen Satz von unabhängigen Koordinaten zur eindeutigen Beschreibung des räumlichen Zustands des betrachteten Systems. Sie werden so gewählt, dass die mathematische Formulierung von Bewegungen, die Zwangsbedingungen unterliegen, möglichst einfach wird. Z. B. genügt beim mathematischen Pendel statt der x- und z-Koordinate des Massenpunkts die Angabe des Auslenkwinkels, um die Lage eindeutig zu beschreiben. Die konstante Seillänge ist durch die Bindungsgleichung $l = \text{const}$ gegeben.

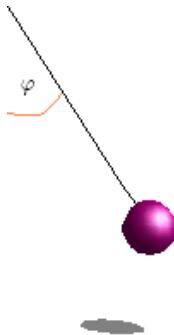


Abbildung 6: Für das Pendel in der Ebene ergibt sich ein Freiheitsgrad und somit eine generalisierte Koordinate; und zwar ϕ

Allgemein stimmt die Anzahl der generalisierten Koordinaten, die zur Beschreibung eines Systems mindestens erforderlich sind, mit der Anzahl seiner Freiheitsgrade überein. Die generalisierten Koordinaten spannen den Konfigurationsraum auf.

9.2 Arten von Kräften

Generalisierte Kräfte sind alle Kräfte die keine Zwangskräfte sind. Sie haben nicht zwangsläufig die Einheit der Kraft sondern das Produkt der generalisierte Kraft und zugehöriger generalisierter Koordinate hat die Einheit der Arbeit. So besitzt z.B. die gen. Kraft zu einem Winkel als Koordinate (dimensionslos), die Einheit Nm .

Zwangskraft ist beispielsweise die Kraft die vom Untergrund auf einen Klotz wirkt, da er durch die Erdbeschleunigung nach unten gedrückt wird.

Konservative Kräfte sind Kräfte die entlang eines geschlossenen Wegintegrals keinerlei Arbeit verrichten. D.h. der Energieinhalt ist am Ende wieder der gleiche.

Bei **Nicht-konservative Kräfte** wird Energie abgeführt (z.B. Reibung). Demnach sind Reibkräfte oder Dämpferkräfte nicht-konservativ. Aber auch Kräfte die Energie in ein System einbringen, wie etwa ein Antriebsdrehmoment sind ebenfalls nicht-konservativ.

9.3 Holonome Systeme

Ein holonomes System ist in der Mechanik ein System von Körpern, für welches die Lage der Körper über n generalisierte Koordinaten q_1, q_2, \dots, q_n beschrieben werden kann. Die Koordinaten können entweder vollständig unabhängig voneinander oder über $m < n$ Zwangsbedingungen miteinander verknüpft sein.

$$\dim(\mathbf{q}) = \dim(\text{DOF})$$

10 Kinematik

11 Kinetik

IV

Elektrotechnik

12 Grundlagen

12.1 Maxwell-Gleichungen

12.1.1 Mikroskopisch

Die mikroskopischen Maxwell-Gleichungen verknüpfen die elektrische Feldstärke \mathbf{E} und die magnetische Flussdichte \mathbf{B} mit der Ladungsdichte ρ (Ladung pro Volumen) und der elektrischen Stromdichte \mathbf{S} (Strom pro durchflossene Fläche).

1. Maxwell'sche Gleichung - Durchflutungsgesetz

$$\text{rot} \mathbf{B} = \mu_0 \mathbf{S} + \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \quad (1)$$

Während sich ein elektrisches Feld ändert, ist es von ringförmigen geschlossenen magnetischen Feldlinien umgeben.

2. Maxwell'sche Gleichung - Induktionsgesetz

$$\nabla \times \mathbf{E} = \text{rot}(\mathbf{B}) = -\frac{\partial \mathbf{B}}{\partial t} \quad (2)$$

3. Maxwell'sche Gleichung - Quellenfreiheit (Divergenzfreiheit) von magnetischen Feldern

$$\nabla \cdot \mathbf{B} = \text{div}(\mathbf{B}) = 0 \quad (3)$$

4. Maxwell'sche Gleichung - Quellenbehaftetes elektrisches Feld (Divergenz)

$$\nabla \cdot \mathbf{E} = \text{div}(\mathbf{E}) = \frac{\rho}{\varepsilon_0} \quad (4)$$

12.1.2 Makroskopisch

Bei Anwesenheit von Materie sind die mikroskopischen Maxwell-Gleichungen einerseits unhandlich, da schließlich jeder Ladungsträger in jedem Atom des Mediums berücksichtigt werden muss. Andererseits können die magnetischen Eigenschaften (beispielsweise von einem Permanentmagneten) prinzipiell nicht ohne zusätzliche physikalische Erkenntnisse der Quantenmechanik aus den mikroskopischen Maxwell-Gleichungen abgeleitet werden. Die makroskopischen Maxwell-Gleichungen berücksichtigen die Eigenschaften der Materie in Form von Materialparametern, wobei dem leeren Raum die Parameter Permittivität ε_0 und Permeabilität μ_0 zugeordnet werden.

Die Anwesenheit von Materie erfordert, dass das elektrische und das magnetische Feld jeweils durch zwei zusätzliche Vektorfelder beschrieben werden, der elektrischen Flussdichte $\mathbf{D} := \varepsilon_0 \mathbf{E} + \mathbf{P}$ und der magnetischen Feldstärke $\mathbf{H} := \frac{1}{\mu_0} \mathbf{B} - \mathbf{M}$. Dabei wird \mathbf{P} als Polarisation und \mathbf{M} als Magnetisierung bezeichnet.

Erste Maxwell'sche Gleichung - Durchflutungsgesetz

$$\nabla \times \mathbf{H} = \text{rot}(\mathbf{H}) = \mathbf{S} + \frac{\partial \mathbf{D}}{\partial t} \quad \text{bzw.} \quad \oint_S \mathbf{H} \cdot d\mathbf{s} = \int_A \left(\mathbf{S} + \frac{d\mathbf{D}}{dt} \right) dA \quad (5)$$

Elektrische Ströme, einschließlich Verschiebungsströme führen zu einem magnetischen Wirbelfeld.

Zweite Maxwell'sche Gleichung - Induktionsgesetz

$$\nabla \times \mathbf{E} = \text{rot}(\mathbf{E}) = -\frac{\partial \mathbf{B}}{\partial t} \quad \text{bzw.} \quad \oint_S \mathbf{E} \cdot d\mathbf{s} = -\frac{d}{dt} \int_A \mathbf{B} \cdot d\mathbf{A} \quad (6)$$

Die zweite Maxwellgleichung besagt, dass jedes zeitlich veränderliche magnetische Feld ein elektrisches Wirbelfeld erzeugt. Die Feldlinien des elektrischen Wirbelfeldes sind geschlossen und umgeben ringförmig die Feldlinien des sich ändernden magnetischen Feldes. Das negative Vorzeichen ergibt sich aus der Lenzschen Regel.

Dritte Maxwell'sche Gleichung - Quellenfreiheit (Divergenzfreiheit) von magnetischen Feldern

$$\nabla \cdot \mathbf{B} = \text{div}(\mathbf{B}) = 0 \quad (7)$$

Das magnetische Feld ist quellenfrei, es gibt keine magnetischen Monopole. D.h. es gibt keine magnetischen Ladungen. Deshalb sind magnetische Felder immer Wirbelfelder mit geschlossenen Feldlinien.

Vierte Maxwell'sche Gleichung - Quellenbehaftetes elektrisches Feld (Divergenz)

$$\nabla \cdot \mathbf{D} = \text{div}(\mathbf{D}) = \rho \quad (8)$$

Die Ladung ist Quelle und Senke des elektrischen Feldes. Ein solches elektrisches Feld beginnt an positiven Ladungen und endet an negativen Ladungen.

12.2 Polarisation

12.3 Permittivität - Dielektrizität

Die Permittivität ϵ gibt in der Elektrodynamik und Elektrostatisik die Durchlässigkeit eines Materials für elektrische Felder an. Die Permittivität setzt sich zusammen aus

$$\epsilon = \epsilon_0 \epsilon_r.$$

$$\begin{aligned} \epsilon_0 & \dots \text{ elektrische Feldkonstante im Vakuum } [\frac{As}{Vm}]; \epsilon_0 = \frac{1}{\mu_0 c_0^2} = 8.8514878\dots \cdot 10^{-12} \frac{As}{Vm} \\ & (c_0 \dots \text{ Vakuumlichtgeschwindigkeit}, \mu_0 \dots \text{ magnetische Feldkonstante}) \\ \epsilon_r & \dots \text{ materialabhängige relative Permittivität } [\frac{As}{Vm}]; \text{ im Vakuum } \epsilon_r = 1 \end{aligned}$$

Aus der elektrischen Flussdichte D und der Permittivität ϵ erhält man dann das elektrische Feld abhängig vom gegebenen Material

$$\mathbf{E} = \frac{\mathbf{D}}{\epsilon} = \frac{\mathbf{D}}{\epsilon_0 \epsilon_r}$$

Wichtige Werte der relativen Permittivität bei 18°C und einer Feldfrequenz von 50 Hz:

Medium	relative Permittivität ϵ_r
Vakuum	1.0
Wasser	1.77
Luft	1.00059
CaTiO ₃ (Ferroelektrikum)	150 - 180
Polyethylen (90°C)	2.4
Aluminiumoxid (Keramik)	9

Typische Dielektrika für Kondensatoren sind Polyethylen, PTFE, Keramik (z.B. Steatit, Aluminiumoxid), Glimmer (Mineralien) oder Luft.

12.4 Elektrische Feldstärke und Verschiebungsflussdichte

12.5 Magnetisierung

12.6 Magnetische Permeabilität

Auch magnetische Leitfähigkeit genannt. Sie dient als Faktor für die Abschwächung/ Verstärkung der magnetischen Flussdichte durch ein Material und ergibt sich als

$$\mu = \frac{B}{H}$$

$$\begin{aligned} \mathbf{B} & \dots \text{ magnetische Flussdichte } [\frac{Vs}{m^2}] \\ \mathbf{H} & \dots \text{ magnetische Feldstärke } [\frac{A}{m}] \end{aligned}$$

Die magnetische Permeabilität wird dabei durch eine Naturkonstante skaliert, die sogenannte magnetische Feldkonstante $\mu_0 = 4\pi \cdot 10^{-7} Vs/Am$. Sie gibt die magnetische Leitfähigkeit des Vakuums an. Für jedes Material kann die magnetische Permeabilität als Produkt aus der Permeabilitätszahl und der magnetischen Feldkonstante angegeben werden.

$$\mu = \mu_r \cdot \mu_0$$

$$\mu_r \dots \text{ relative magnetische Permeabilität (auch Permeabilitätszahl) [dimensionslos]}$$

Die Materialkonstante ist für Paramagnetika $\mu_r > 1$, Diamagnetika $\mu_r = 1$ und für Ferro-, Ferri und Antiferromagnetika eine komplizierte Funktion von Feldstärke und Vorgeschichte (μ_r bis 10^5).

Diamagnetika: Gold, CO₂, Kupfer, Wasser, Zink, ...

Paramagnetika: Mangan, Chrom, Natrium, Aluminium, ... Auch alle Ferromagnetika gehören dazu. Bei Stoffen in denen Dipole durch ein äußeres Magnetfeld ausgerichtet werden können und dieses somit verstärken.

Ferromagnetika: Eisen, Cobalt, Nickel, ...

Ferrimagnetismus: Kristallstruktur innerhalb welcher sich die magnetischen Momente der Atome abwechselnd antiparallel ausrichten. Diese heben sich gegenseitig aber nicht wie beim Antiferromagnetismus vollständig auf, sondern eine der beiden Richtungen ist stärker.

12.7 Magnetische Flussdichte und Feldstärke

12.8 Verschiebungsstrom

Verschiebungsstrom (engl. displacement current) ist eine Bezeichnung aus der Elektrodynamik und stellt die Tatsache dar, dass die zeitliche Änderung eines elektrischen Feldes bzw. der elektrischen Flussdichte ein Teil des totalen elektrischen Stromes ist. Der Begriff wurde von James Clerk Maxwell entwickelt und erweitert das Ampèresche Gesetz um einen zusätzlichen Term.

Der gesamte elektrische Strom setzt sich grundsätzlich aus zwei additiven Komponenten zusammen:

- Dem **Leistungsstrom** I_L , welcher durch den Fluss von elektrischen Ladungsträgern wie Elektronen oder Ionen getragen wird. Der Leistungsstrom wird durch das elektrische Feld und den damit auf Ladungsträger ausgeübte mechanische Kräfte verursacht.
- Der **Verschiebungsstrom** I_V wird durch die zeitliche Änderungsrate des elektrischen Flusses bestimmt und ist nicht an die Existenz eines elektrischen Leiters gebunden. Der Verschiebungsstrom ist als ein Teil der Wirkung des elektrischen Feldes zu verstehen und drückt im Prinzip die zeitliche Änderungsrate des elektrischen Flusses aus.

$$I = I_L + I_V$$

Dadurch folgt eine begriffliche Erweiterung des Ampèreschen Gesetzes (Durchflutungsgesetz), welche den gesamten elektrischen Strom in der Form

$$I = \int_A \left(\sigma \vec{E} + \varepsilon \frac{\partial \vec{E}}{\partial t} \right) \cdot d\vec{A}$$

ausdrückt. Dabei stellt der erste Summand den Leistungsstrom dar, welcher von der elektrischen Feldstärke E ausgelöst wird. Die dabei auftretende Konstante σ stellt einen Ausdruck für die elektrische Leitfähigkeit in jenem Medium dar, in welchem sich der Leistungsstrom ausbreitet. Man nennt solche Medien im Regelfall elektrische Leiter. Der zweite Summand stellt den Verschiebungsstrom mit der zeitlichen Änderungsrate der elektrischen Feldstärke und der dielektrischen Leitfähigkeit ε dar. Die dielektrische Leitfähigkeit drückt die Eigenschaft eines Mediums zur Leitung des elektrischen Flusses aus. Daher fließt der Verschiebungsstrom vor allem in Materialien mit guter dielektrischer und schlechter elektrischer Leitfähigkeit. Man nennt Materialien mit jenen Stoffkonstanten auch Isolatoren.

12.9 Wirbelströme

Wie im Induktionsgesetz beschrieben, wird in einem elektrischen Leiter ein Strom induziert wenn sich dieser in einem konstanten Magnetfeld bewegt oder sich in Ruhe in einem sich wechselnden Magnetfeld befindet (= magnetische Flussänderung). Bei einem Draht ist die Stromrichtung eindeutig vorgegeben. Handelt es sich jedoch um einen Leiter mit großer Oberfläche, dann bilden die induzierten Ströme sogenannte Ringströme/Wirbelströme.

Dies ist beispielsweise bei Transformatoren oder Generatoren der Fall. Da Wirbelströme zu Energieverlusten führen, müssen diese durch etwa die Verwendung von Blechpaketen statt einem dicken Eisenkern, reduziert werden. Dabei sind einzelne Bleche elektrisch voneinander isoliert.

Wirbelstrombremse

Die Wirbelstrombremse ist eine verschleißfreie Bremse. Befindet sich eine Metallplatte in einem inhomogenen Magnetfeld oder wird eine solchige durch ein homogenes Magnetfeld bewegt, werden Wirbelströme induziert. Da die Metallplatte einen ohmschen Widerstand hat, wird die Platte durch die Elektronenbewegung erwärmt. Dies hat zur Folge, dass die Bewegungsenergie der Relativbewegung zwischen Magnetfeld und Metallplatte in Wärmeenergie dissipiert wird. Voraussetzung für die Induktion von Wirbelströmen, ist das vorhanden sein von freien Ladungsträgern in der Metallplatte (elektrischer Leiter).

12.10 Reluktanzkraft / Maxwellkraft-Prinzip

Bei einer stromdurchflossenen Spule, die wie in der Abbildung um einen Eisenkern gewickelt ist, erzeugt die Spule einen magnetischen Fluss im Eisenkern. Es entstehen in beiden Luftspalten Anziehungskräfte auf den losen Eisenkörper (Anker). Der Effekt der Anziehung beruht darauf, dass die *Reluktanz*, also der magnetische Widerstand, reduziert und die Induktivität erhöht wird. Das ist der Fall wenn die Feldlinien durch den Anker verlaufen. Wird der Abstand zwischen dem Anker und den Magnetpolen verkürzt, verringert sich der magnetische Widerstand der Feldlinien, denn der Widerstand ist im Eisen wesentlich kleiner als in der Luft.

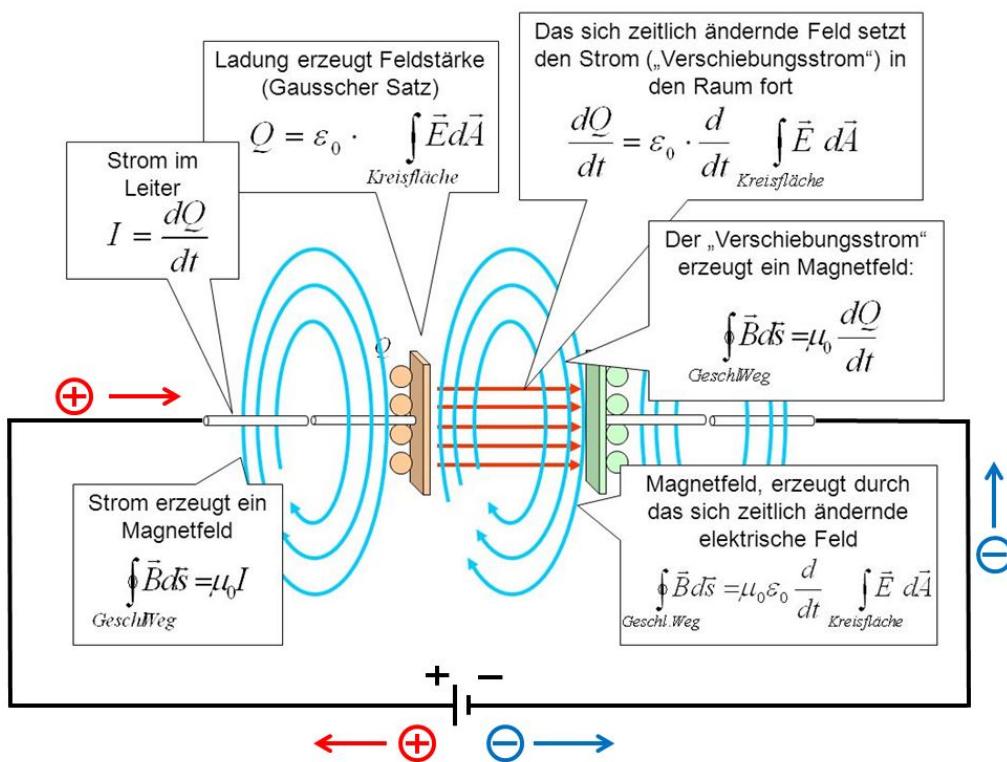


Abbildung 7: Aufladung eines Kondensators als Paradebeispiel für Verschiebungsstrom

12.11 Halbleiter

Halbleiter sind Festkörper, deren elektrische Leitfähigkeit zwischen der von elektrischen Leitern und der von Nichtleitern liegt. Da sich die Grenzbereiche der drei Gruppen überschneiden, ist der negative Temperaturkoeffizient des spezifischen Widerstandes ein weiteres wichtiges Merkmal von Halbleitern, das heißt, ihre Leitfähigkeit nimmt mit steigender Temperatur zu, sie sind sogenannte Heißleiter. Ursache hierfür ist die sogenannte Bandlücke zwischen dem Valenz- und dem Leitungsbau. Nah am absoluten Temperaturnullpunkt sind diese voll- bzw. unbesetzt und Halbleiter daher Nichtleiter. Es existieren im Gegensatz zu Metallen primär keine freien Ladungsträger, diese müssen erst z. B. durch thermische Anregung entstehen. Die elektrische Leitfähigkeit von Halbleitern steigt aber steil mit der Temperatur an, so dass sie bei Raumtemperatur, je nach materialspezifischem Abstand von Leitungs- und Valenzband, mehr oder weniger leitend sind. Des Weiteren lässt sich durch das Einbringen von Fremdatomen (Dotieren) aus einer anderen chemischen Hauptgruppe die Leitfähigkeit und der Leitungscharakter (Elektronen- und Löcherleitung) in weiten Grenzen gezielt beeinflussen.

Am bekanntesten sind die Elementhalbleiter Silicium und Germanium, die aus einem einzigen Element aufgebaut sind, und Verbindungshalbleitern wie zum Beispiel der III-V-Verbindungshalbleiter Galliumarsenid. Des Weiteren haben in den letzten Jahrzehnten organische Halbleiter an Bedeutung und Bekanntheit gewonnen, sie werden beispielsweise in organischen Leuchtdioden (OLEDs) eingesetzt.

- **Elementhalbleiter:** Si, Ge, B, Se, Te, C
- **Verbindungshalbleiter:** GaAs, ...
- **Organische Halbleiter**

12.12 Raumladungszone und p/n-Übergang

Für die Herstellung eines Halbleiterbauelements wird der Siliziumkristall verunreinigt. Bei der p-Dotierung werden Fremdatome (Bor) eingefügt die weniger Elektronen als das Silizium besitzen, wodurch sich positive Löcher im Material ausbilden. Bei der n-Dotierung wird Phosphor eingebracht, das mehr Elektronen als das Silizium besitzt und es entsteht ein Material das freie Elektronen besitzt. Dotierte Halbleiter sind in ihrem Grundzustand ungeladen. Es existieren immer gleich viele freie Ladungsträger als ortsfeste Raumladungen der ionisierten Dotierungsatome. Bringt

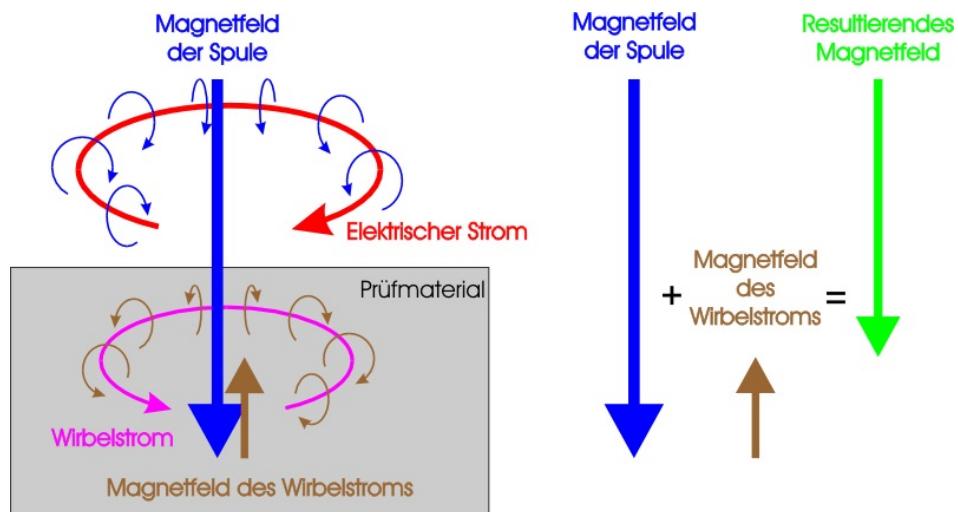


Abbildung 8: Induktion eines Wirbelstromes

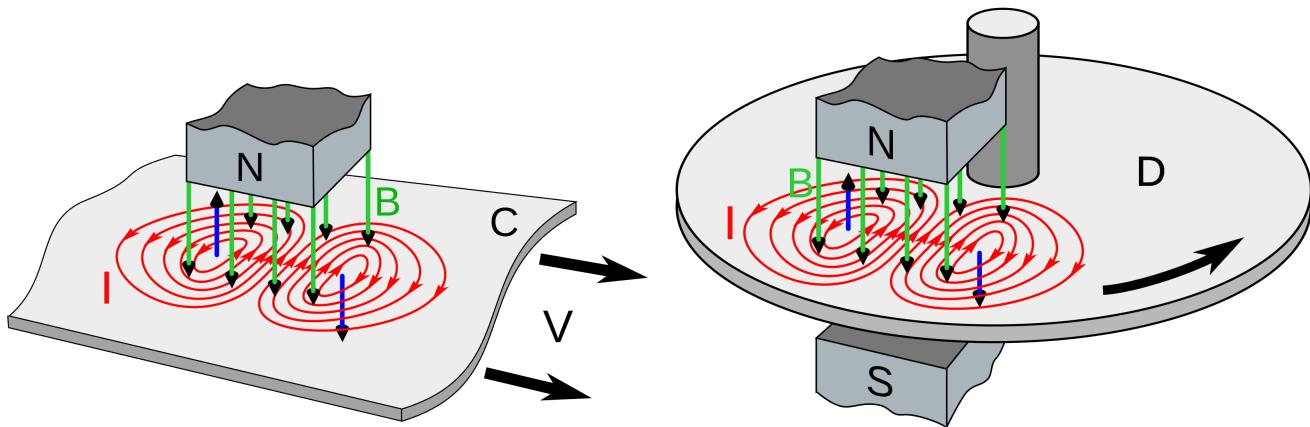
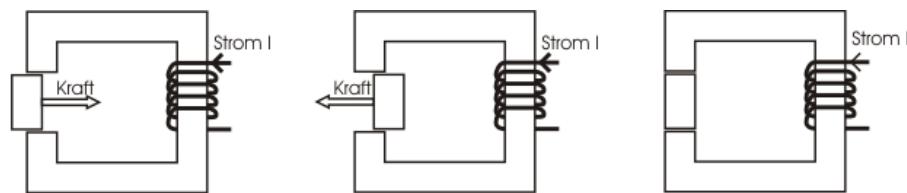


Abbildung 9: Beispiele für Wirbelstrombremsen: Zug an Schienen oder LKW an Bremsscheiben

man diese beiden Materialen aneinander, ergibt sich ein Konzentrationsunterschied von positiven und negativen Ladungsträgern zwischen den beiden Gebieten. Aufgrund von Diffusion streben positive Löcher vom p-dotierten Bereich ins n-Gebiet und negative Elektronen in den p-dotierten Bereich. Positive Löcher rekombinieren im n-Gebiet mit den Elektronen und die Elektronen rekombinieren im p-Gebiet mit den dort vorhandenen Löchern. Es entsteht eine Zone ohne freie Ladungsträger, aber mit positiven Phosphor-Ionen und negativen Bor-Ionen. In weiterer Folge bildet sich ein elektrisches Feld in der Raumladungszone aus. Ist das elektrische Feld groß genug ($0.7V$), können die Elektronen des n-Gebiets die Schicht an negativen Bor-Ionen nicht mehr durchdringen und werden von diesen abgestoßen. Das elektrische Feld wirkt der Diffusion entgegen und es hat sich ein Gleichgewicht ausgebildet.

Extern angelegte Spannung

In Durchlassrichtung wirkt das elektrische Feld der externen Spannung, dem elektrischen Feld in der Raumladungszone entgegen. Die positiven Löcher werden vom positiven Pol abgestoßen und die Elektronen vom negativen Pol. Dadurch bewegen sich Elektronen und Löcher zur Mitte Richtung Grenzschicht hin und rekombinieren dort fortlaufend. Ein elektrischer Stromfluss ist möglich. Bei Sperrrichtung liegt genau der umgekehrte Fall vor und die Raumladungszone wird dabei vergrößert.



13 Elektrische Bauelemente

13.1 Bipolar-Transistor

Funktionsweise

Im Folgenden wird die physikalische Stromrichtung (Elektronenstrom) für die Erläuterung verwendet. Durch das Anlegen einer Spannung U_{BE} von etwa 0,7 V, ist die untere Diode (Prinzip) in Durchlassrichtung geschaltet. Die Elektronen gelangen in die p-Schicht und werden von dem Plus-Pol der Spannung U_{BE} angezogen. Da die p-Schicht sehr klein ist, wird nur ein geringer Teil der Elektronen angezogen. Der größte Teil der Elektronen bewegt sich weiter in die obere Grenzschicht. Dadurch wird diese leitend und der Plus-Pol der Spannung U_{CE} zieht die Elektronen an. Es fließt ein Kollektorstrom I_C . Bei üblichen Transistoren gelangen etwa 99% der Elektronen von Emitter zum Kollektor durch. In der Basisschicht rekombinieren etwa 1% der Elektronen.

14 Elektronische Grundschaltungen

14.1 Wechselrichter

15 Filter

15.1 Filtertypen

- Tiefpassfilter
- Hochpassfilter
- Bandpassfilter
- Bandsperre, Kerbfilter

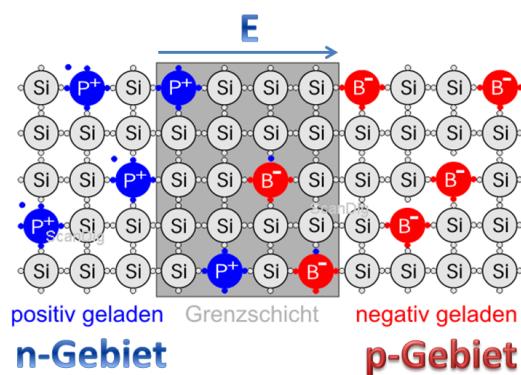
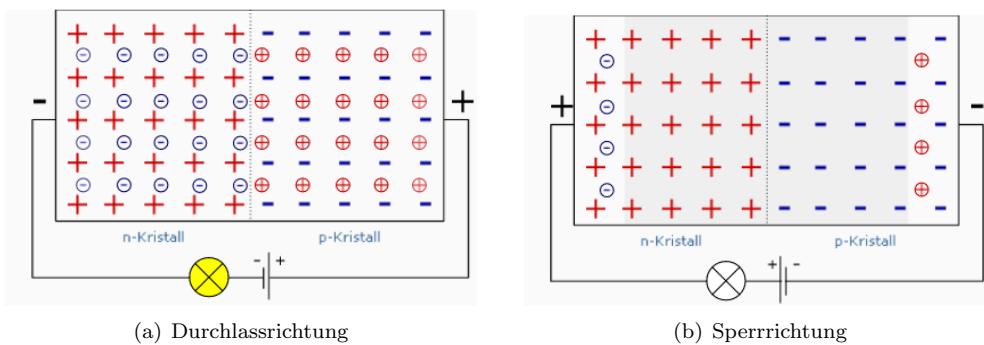


Abbildung 10: Phosphor mit 5 Protonen im Kern und Bor mit 3 Protonen im Kern, Silizium mit 4 Protonen und Elektronen



- Allpassfilter – lässt alle Frequenzen mit gleicher Verstärkung durch, kann aber eine Phasenverschiebung verursachen.

Die Filterordnung beschreibt die Dämpfung von Frequenzen unterhalb od. oberhalb der Grenzfrequenz. Z.b. bei einem Filter n. Ordnung würde die TF mit $n \cdot 6dB/Oktave$ fallen.

Weitere Einteilungen: Aktiv, Passiv, Linear, Nicht-Linear, Ordnung

15.2 Analoges aktives Filter

15.3 Analoges passives Filter

15.3.1 Topologien

<https://de.wikipedia.org/wiki/Analogfilter>
<http://elektroniktutor.de/analogtechnik/filter.html>

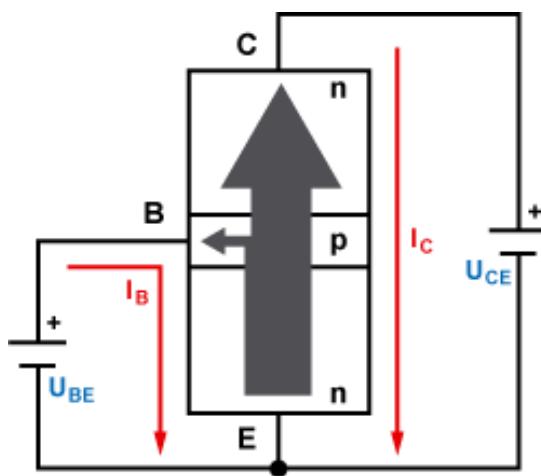
15.3.2 Tiefpassfilter

Praktische Anwendung, wie wird ein Tiefpassfilter dimensioniert – ω_c gewünschte Grenzfrequenz einstellen

15.3.3 Hochpassfilter

15.3.4 Notchfilter

Ein Kerbfilter ist ein elektronisches Filter, mit dem Frequenzen innerhalb eines engen Frequenzbereiches ausgefiltert werden können. Anschaulich wird eine Kerbe in das Frequenzdiagramm eingefügt. Kerbfilter stellen einen besonders schmalbandigen Typ von Bandsperrfilter dar, welche in der Übertragungsfunktion nur eine Nullstelle aufweisen und damit nicht ein breites Frequenzband, sondern idealerweise genau eine Frequenz möglichst stark dämpfen.



15.4 Digitales Filter

15.4.1 FIR - Finite Impulse Response

15.4.2 IIR - Infinite Impulse Response

15.5 Kalmanfilter

Ein Kalman-Filter schätzt den Wert einer ungenauen oder unsicheren Messung anhand der vorhergegangen Messwerte ab. Das Prinzip beruht auf Wahrscheinlichkeitsfunktionen. Systemzustände müssen dafür bekannt sein und fließen in die Berechnung mit ein. Anders formuliert wird eine Vorhersage des nächsten Wertes anhand des momentanen Wertes und der bekannten System-Eingangsvariablen geschätzt. Zusätzlich vergrößern die nicht abschätzbarer Parameter den Wahrscheinlichkeitsbereich, wo der nächste Wert landen könnte. Hinzu kommt nun noch eine Wahrscheinlichkeit für den Messwert des Sensors. Aufgrund von Rauschen erhalten wir einen anderen Wert als real vorhanden ist. Diese Abweichung wird wiederum als Wahrscheinlichkeit beschrieben. Nun kann man die beiden Wahrscheinlichkeiten (die des nächsten Wertes und der Korrektheit des Messwertes) multiplizieren und erhält einen neuen verbesserten Schätzwert.

Dieser verbesserte Schätzwert kann nun wieder durch den ganzen Ablauf verwendet werden und wir erhalten einen neuen Schätzwert. Dieser Vorgang kann beliebig oft vorgenommen werden. Die Filter sind sehr schnell – Echtzeitfähig – und liefern in der Praxis sehr gute Ergebnisse. Zeigt sehr gute Filtereigenschaften für Sensorrauschen (Unsicherheit der Messung ausgleichen).

<http://www.bzarg.com/p/how-a-kalman-filter-works-in-pictures/>

16 Sensoren

16.1 Übersicht von Sensoren nach deren Messgrößen

Abstands- bzw. Wegsensoren Geschwindigkeits und Drehzahlensensoren Beschleunigungssensoren

16.2 Encoder

16.2.1 TTL-Encoder

2 Signale A,B mit insgesamt 4 verschiedenen Zuständen –; Drehencoder mit 1000 ppr = 4000 Zustände pro Umdrehung.

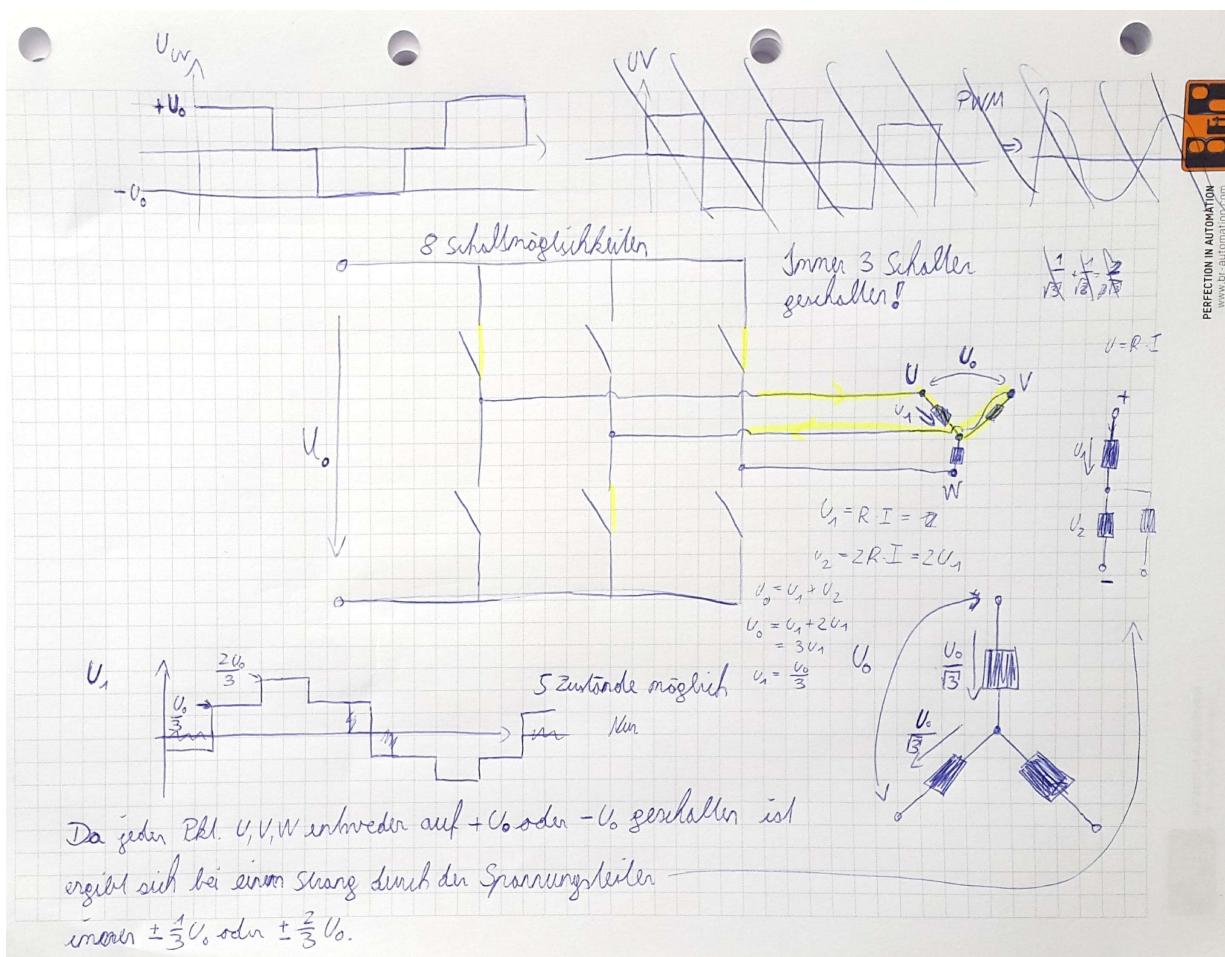


Abbildung 11:

16.2.2 SinCos-Encoder

16.3 Piezoelektrische Sensoren

16.4 Induktive Sensoren

16.5 Kapazitive Sensoren

16.6 Mechanische Sensoren

16.7 Thermoelektrische Sensoren

16.8 Resistive Sensoren

16.9 Magnetische Sensoren

16.10 Optische Sensoren

16.10.1 Interferometer

Das Messprinzip basiert auf der Überlagerung von zwei Wellen die unterschiedliche Wege zurücklegen. Grundsätzlich lässt sich mit jeder Art von Welle, seien es elektromagnetische Wellen, Schall- oder Materiewellen, eine Interferenz erzeugen.

Laserlicht eignet sich ausgezeichnet für kürzere (Meterbereich) und äußerst präzise (Pikometerbereich) Wegmessungen. Voraussetzung für eine erfolgreiche Interferenz ist, dass Wellen kohärent überlagert werden. Deshalb wird Laserlicht verwendet, da ausgesendete Lichtwellen nahezu phasengleich zueinander sind und eine sehr große Kohärenzlänge (max. Weglänge zweier Lichtstrahlen, bei der diese noch ein stabiles Interferenzmuster ergeben) - d.h. auch über große Distanzen phasengleich bleiben - besitzen.

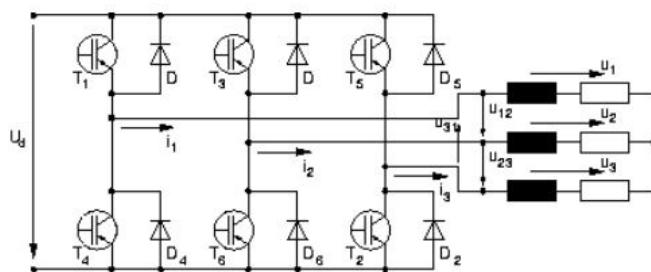


Abbildung 12:

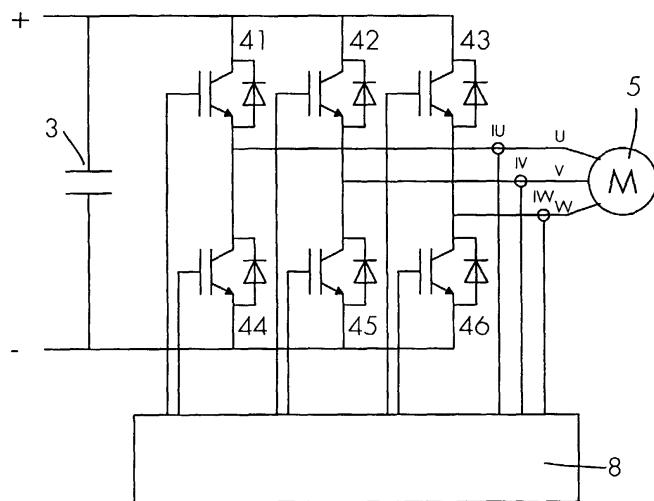


Fig 2

Abbildung 13:

Die unterschiedlichen Typen von Interferometern haben meist das selbe Funktionsprinzip. Durch Überlagerung zweier Wellen, die in getrennten optischen Bahnen geführt und von Spiegeln bzw. einem Objekt reflektiert werden, bilden ein Interferenzmuster (Interferenzstreifen oder -ringe). Dieses ist abhängig von der Distanz des Objektes. Somit können Distanzänderungen detektiert werden.

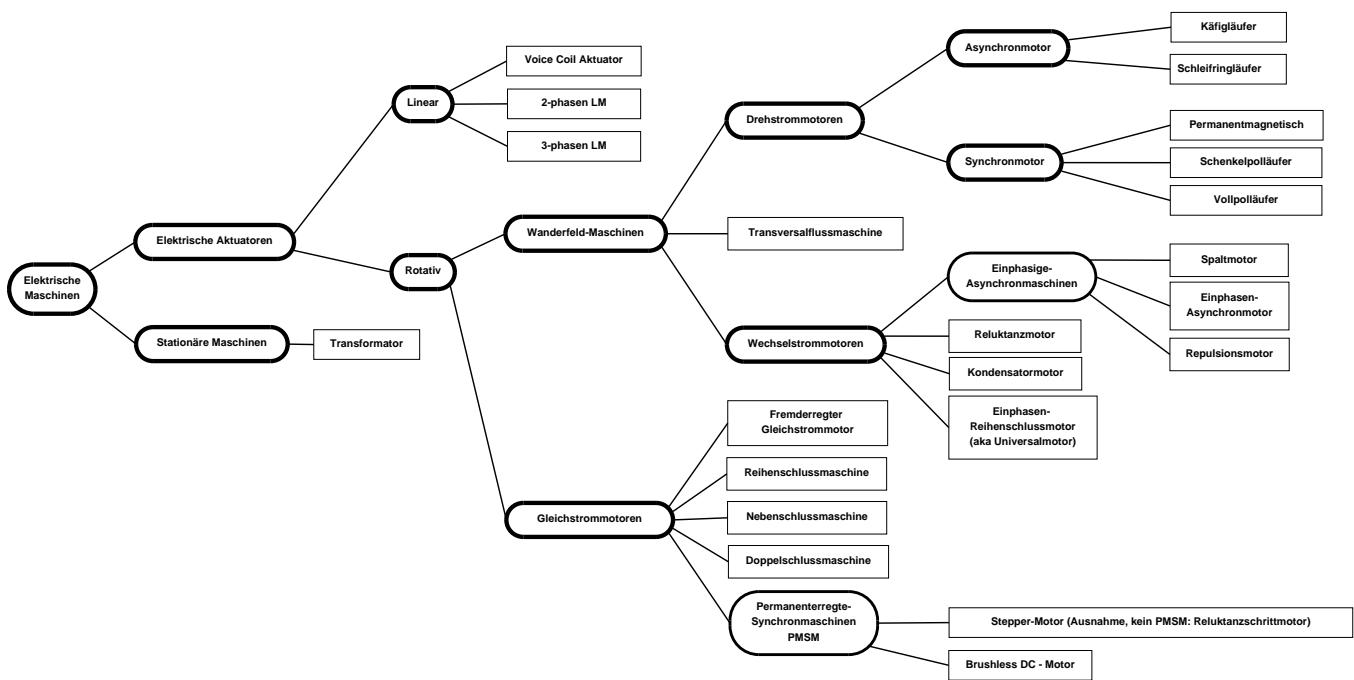
So werden beim **Michelson-Interferometer** durch einen Photosensor abwechselnd Minima und Maxima des Interferenzbildes registriert und mitgezählt. Wird die Anzahl der gemessenen Maxima mit der Wellenlänge des Laserlichts multipliziert, erhält man die Wegdifferenz. Da das Medium durch das die elektromagnetische Welle wandert, deren Wellenlänge beeinflusst, ist ein wichtiger Faktor, die Eigenschaften dieses Mediums zu berücksichtigen (Luftdruck, Luftfeuchtigkeit, Temperatur).

Bei Vielstrahlinterferometern wie dem **Fabry-Perót-Interferometer** wird die Längenänderung gemessen indem die Länge des optischen Resonators verändert wird. D.h. ein optischer Ausgang/Messkopf des Interferometers und das Objekt bilden gemeinsam einen optischen Resonator. Martin Zech von Attocube hat ein solches Interferometer entwickelt, das Genauigkeiten im sub-Nanometerbereich bei Raumtemperatur erreicht. Wird der Sensor mit flüssigem Helium auf 4 K gekühlt kann durch Rauschunterdrückung eine Genauigkeit von 1 pm erreicht werden.

Vorteil:

- Höhere Empfindlichkeit
- Längenmessungen bis 10^{-12} m Genauigkeit können gemessen werden

17 Elektrische Maschinen



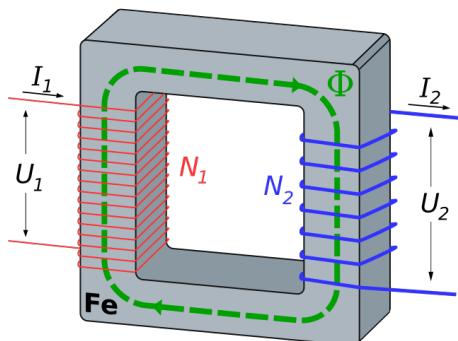
17.1 Stationäre Maschinen

17.1.1 Transformator

Ein Transformator besteht meist aus zwei oder mehr Spulen (Wicklungen), die in der Regel aus Kupferdraht gewickelt sind und sich auf einem gemeinsamen Ferrit- bzw. Eisenkern befinden. Ein Transformator wandelt eine Eingangswechselspannung, die an einer der Spulen angelegt ist, in eine Ausgangswechselspannung um, die an der anderen Spule abgegriffen werden kann. Dabei entspricht das Verhältnis von Eingangs- und Ausgangsspannung dem Verhältnis der Windungszahlen der beiden Spulen. So wird zum Beispiel bei einem Windungsverhältnis von 20 zu 1 eine Eingangsspannung von 240 Volt in eine Ausgangsspannung von 12 Volt transformiert. Je nach Auslegung des Transformators kann die Ausgangsspannung somit kleiner, größer oder gleich der Eingangsspannung sein.

Wirkprinzip

Eine Wechselspannung auf der Primärseite des Transformators bewirkt entsprechend dem Induktionsgesetz einen wechselnden magnetischen Fluss im Kern. Der wechselnde magnetische Fluss wiederum induziert auf der Sekundärseite des Transformators eine Spannung (Spannungstransformation)



Spannungstransformation beim idealen Transformator:

$$\frac{U_2}{U_1} = \frac{N_2}{N_1} \quad \text{bzw.} \quad U_2 = \frac{N_2}{N_1} U_1$$

Wird an die sekundäre Wicklung ein **Verbraucher angeschlossen**, so entnimmt dieser der Sekundärspule elektrische Energie. Dabei kommt ein Strom auf der Sekundärseite zustande und der Primärstrom vergrößert sich. Im Gegensatz zu den Spannungen an den Wicklungen sind die Ströme in den Wicklungen jedoch entgegengesetzt gerichtet: Wenn der Primärstrom bezogen auf den Kern rechtsherum durch die Spule fließt, fließt der Sekundärstrom linksherum und umgekehrt (Lenzsche Regel).

$$\frac{I_2}{I_1} = \frac{N_1}{N_2} \quad \text{bzw.} \quad I_2 = \frac{N_1}{N_2} I_1$$

Realer Transformator

Ein realer Transformator hat demgegenüber Übertragungsverluste durch den ohmschen Widerstand der Wicklung, durch Wirbelstrombildung im Kern, Ummagnetisierungsverluste und durch andere Effekte. Bei großen Transformatoren muss die Verlustleistung gegebenenfalls durch geeignete Kühlung abgeführt werden. Bei starker Überlastung kann sich ein Transformator überhitzen und durchbrennen.

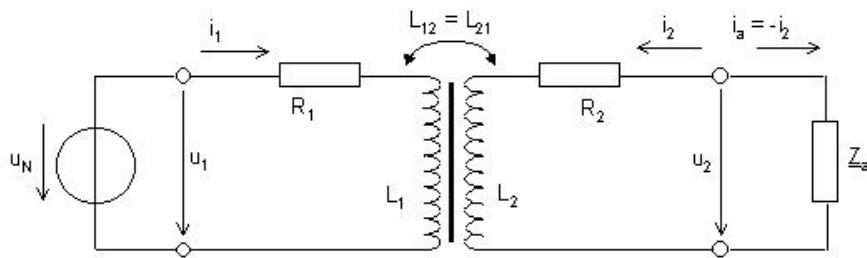


Abbildung 14: Ersatzschaltbild eines realen Transformators

Die Größen der Sekundärseite können auf die Primärseite mithilfe des Übertragungsfaktors γ umgerechnet werden.

$$\gamma = \frac{N_1}{N_2} \quad \rightarrow \quad U'_2 = U_2 \gamma, \quad I'_2 = \frac{I_2}{\gamma}, \quad L'_{\sigma 2} = L_{\sigma 2} \gamma^2, \quad Z' = Z \gamma^2$$

Das heißt, ein sekundärseitiger Widerstand R , wirkt primärseitig als $R \gamma^2$.

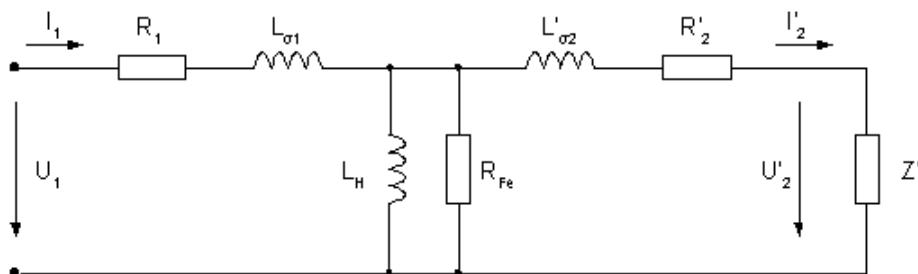


Abbildung 15: Ersatzschaltbild eines realen Transformators mit transformierten sekundärseitigen Größen

- R_1 ... ohmscher Widerstand der Primärwicklung
- R'_2 ... transformierter ohmscher Widerstand der Sekundärwicklung
- U_1 ... primärseitige Spannungsquelle
- U'_2 ... transformierte Ausgangsspannung
- Z' ... transformierte Lastimpedanz
- $L_{\sigma 1}$... Streuinduktivität der Primärseite

- $L'_{\sigma 2}$... transformierte Streuinduktivität der Sekundärseite
 L_H ... Hauptinduktivität, die den Magnetisierungsstrom führt
 R_{fe} ... Eisenverluste des Kerns

17.2 Elektrische Aktuatoren

Unterschiede zwischen Eisenlose und eisenbehaftete Motoren

Bei eisenlosen Motoren werden die Wicklungen mit Epoxid vergossen (Luftspaltwicklung). Diese Motoren sind für sehr gleichförmige Bewegungen geeignet. Bei ihnen wird keine magnetische Anziehungskraft erzeugt. Bei eisenbehafteten Motoren hingegen werden die Wicklungen in einem Eisenkäfig fixiert. Eisenbehaftete Motoren nutzen das Eisen, um den magnetischen Fluss zu bündeln und können so eine sehr hohe Kraftdichte erzeugen.

Vorteile eisenlose Motoren:

- Keine Eisenverluste (Eisenmagnet muss nicht umgepolt werden, die Hysteresekurve muss nicht durchlaufen werden, was Energie kostet)
- Kompakter und leichterer Forcer/Rotor
- Kleine Rotor Induktivität (weniger Elektromagnetische Störungen, schnelle Reaktionszeit des Stroms)
- Bei Sinuskommutierung erhält man ein gleichförmiges Moment bzw. eine konstant wirkende Kraft

Vorteile eisenbehaftete Motoren:

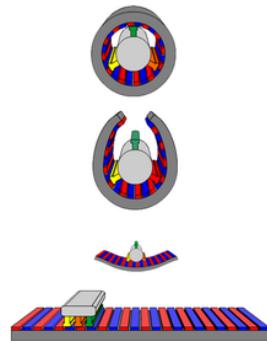
- sehr hohe Drehmomentdichte
- Maximierung der magnetischen Kraft
- höhere Spitzenkraft als eisenlose Motoren der selben Baugröße (\rightarrow kompaktere Ausführungen möglich)

17.2.1 Linearaktuatoren

Allgemein

Grundsätzlich arbeiten Linearmotoren wie rotatorische Motoren. Man kann sie sich so vorstellen, dass man einen beliebigen Motor aufschneidet und in die Ebene abwickelt.

Wobei die ursprünglich kreisförmig angeordneten elektrischen Erregerwicklungen (Stator) auf einer



ebenen Strecke angeordnet sind. Der Läufer, der im Drehstrommotor rotiert, wird beim Linearmotor von dem längs bewegten Magnetfeld über die Fahrstrecke bewegt. Analog dazu funktionieren tubulare Linearmotoren ebenfalls wie ihre rotativen Verwandten, jedoch ist bei Linearmotoren der Unterschied einer begrenzten Stator-Strecke. Die Idee bei Linearmotoren ist, dass die durch den Motor erzeugte Kraft direkt auf eine zu bewegende Masse wirkt ohne Verluste in Getrieben, Spindeln, Riementrieben, usw. in Kauf nehmen zu müssen.

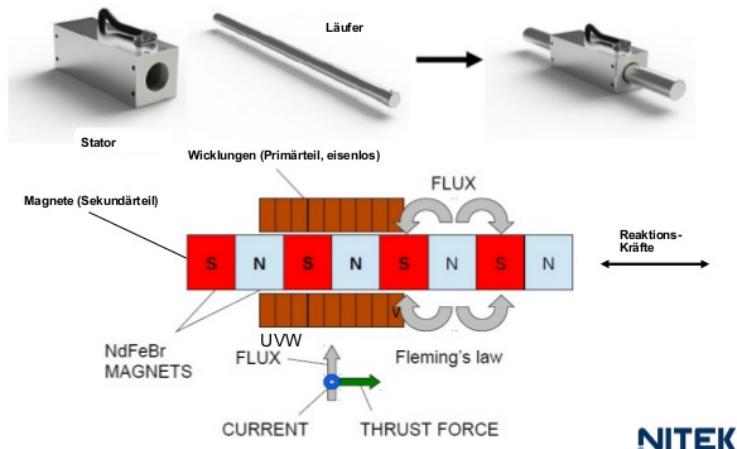


Abbildung 16: Symbol für Linearmotoren

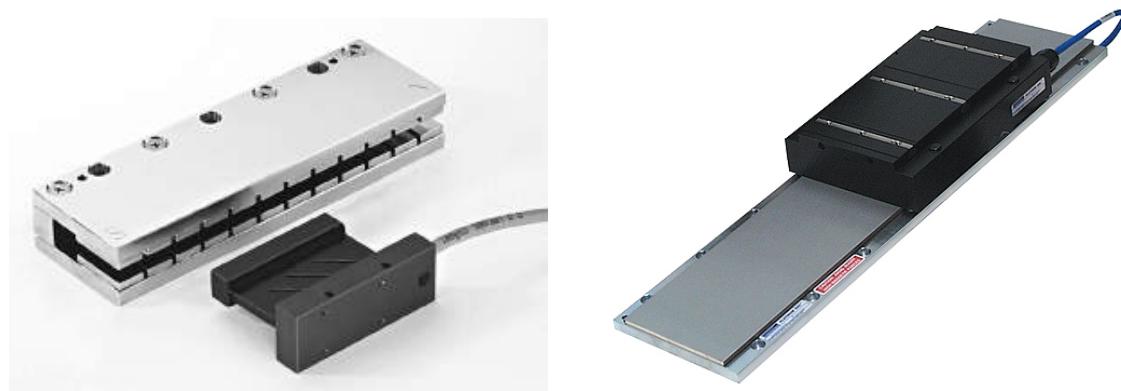
Einteilung



Tubulare Linearmotoren Unter einem tubularen Linearmotor versteht man einen elektrischen Direktantrieb, bei dem die lineare Bewegung direkt aufgrund der elektromagnetischen Kraftwirkung erzeugt wird. Das heißt, die translatorische Bewegung wird nicht durch eine mechanische Umwandlung über eine Spindel, Riemen oder Kurvenscheibe erzeugt, sondern basiert direkt auf den Wechselwirkungen zwischen elektromagnetischen Kräften. Grundsätzlich können derartige Motoren sowohl nach dem Lorenzkraftprinzip wie auch nach dem Maxwellkraftprinzip (siehe Kap. ??) aufgebaut werden und gehören zu der Gattung der Linearmotoren. Der Unterschied zu den flachen oder U-förmigen Linearmotoren besteht darin, dass die Erregerwicklung des Stators rohrförmig (tubular) und die Magnete des Läufers stabförmig sind. Konstruktiv gesehen entsteht so ein Antriebselement, das Ähnlichkeiten zu einem pneumatischen oder hydraulischen Zylinder aufweist. Im Bild ist ein permanenterregter Läufer abgebildet. Im asynchronen Fall besteht der Läufer statt aus Permanentmagneten, aus kurzgeschlossenen Kupferwindungen wie beim bekannten rotativen Asynchronomotor.



Flache, U-förmige Linearmotoren Bei flachen bzw. U-förmigen Linearmotoren befinden sich bei PMSM die Permanentmagneten am Stator des Motors. Beim asynchronen Linearmotor hält der Stator hingegen die Erregerwicklungen, die um weichmagnetisches Eisen gewickelt sind und somit Magnetfelder erzeugt. Ein Linearmotor besteht aus nur zwei Komponenten: einem Wicklungspaket (Forcer bzw. Primärteil) und einem Träger (Magnetschiene bzw. Sekundärteil), auf dem die Permanentmagneten fixiert sind. Die Kupferwindungen des Wicklungspakets sind entweder in Epoxidharz oder Eisen eingebettet. Die Kupferwindungen führen den gesamten Strom eines Linearmotors. Das Magnetassembly besteht aus



(a) U-förmiger Linearmotor

(b) flacher Linearmotor

Seltenen-Erde-Magneten, die in abwechselnder Polarität auf einem Stahlträger montiert sind. Sie erzeugen ein magnetisches Feld senkrecht zum Träger. Wenn in den Kupferwindungen Strom fließt, ergibt sich nach dem Lorentz'schen Gesetz eine Kraft $F = I \times B$, die zur Beschleunigung der Masse benutzt wird. Eine Linearmotorachse besteht wiederum aus einem Linearmotor mit Lagerung und einem integrierten Positions-Gebersystem (Encoder). Der Forcer wird üblicherweise an den bewegten Teilen der Maschine befestigt. Das Magnetassembly wird am statischen Teil der Maschine fixiert.

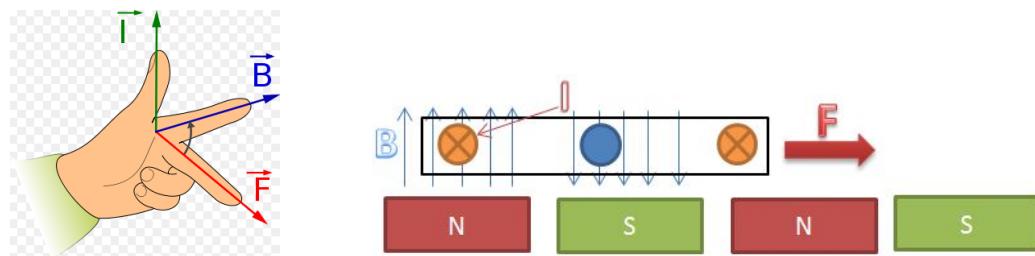


Abbildung 17: Linke-Hand-Regel für die Wirkrichtung der Lorentzkraft

Asynchroner Linearmotor

Beim linearen Asynchronomotor wird die bewegende Kraft durch ein wanderndes Magnetfeld das mit einem Leiter interagiert, erzeugt. In jedem Leiter, sei es eine Windung, Spule oder einfach nur eine leitende Metallplatte, die im Feld platziert wird, werden Ströme induziert. Durch Wechselwirkung des magnetischen Wanderfeldes und dem induzierten Strom resultiert jene Kraft, die für eine Linearbewegung verantwortlich ist. Asynchrone Linearmotoren werden für Anwendungen die große Kräfte benötigen verwendet. Zum Beispiel elektrisch Schiebetüren oder Magnetschwebbahnen.

PSM als Linearmotor

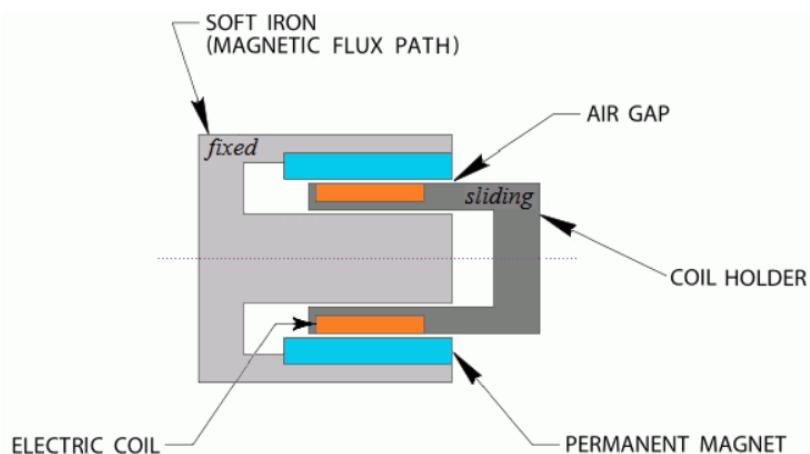
Die permanenterregte Synchronmaschine (PSM) als Linearmotor benötigt im Forcer einen sinusförmigen Stromverlauf um ohne rucken bewegt werden zu können. Dies kann entweder durch Einspeisung von Wechselstrom erreicht werden oder durch Sinuskommutierung einer Gleichspannung.

Sinuskommutierung

Ein hoher Gleichlauf kann erreicht werden, indem die Phasenströme graduell angeglichen werden. Der Linearmotor ist geometrisch so aufgebaut, dass sich die Back-EMF in Form einer Sinusspannung ergibt. Daher ist die beste Lösung einen sinusförmigen Stromverlauf vor zu geben, der dann zu einer gleichmäßigen Bewegung führt. Das erzeugte Drehmoment/ die erzeugte Kraft ist dann konstant. Einen sinusförmigen Strom nachzubilden verlangt aber eine höhere Positionsauflösung als von den Hallsensoren erhalten werden kann. Der Strom in den 3 Phasen muss viel häufiger angepasst werden. Darum verwendet die Sinuskommutierung meist Encoder zur präzisen Lagebestimmung des Rotors. Sinuskommutierung ermöglicht einen gleichmäßigen Motorbetrieb und ergibt sogar eine bessere Performance.

17.2.1.1 1-phasen Linearmotor/ Linear Voice Coil Motor/ Tauchspulenaktuator

Der Voice Coil Aktor ist ein zweipoliger nichtkommuterter Antriebsmechanismus mit limitiertem Weg. Er besteht aus zwei Komponenten: einem Dauermagneten am fixierten Teil und einer Spule, die auf einen im Luftspalt beweglichen nichtmagnetischen Spulenkörper gewickelt ist. Im eingebauten Zustand befindet sich die Spule im Luftspalt des Magnetfeldes. Richtung und Amplitude der Lorentzkraft werden dabei von der Stromstärke und -richtung bestimmt. Sie arbeiten nach dem Lorentz-Kraft-Prinzip und sind dadurch bidirektional aktiv ansteuerbar. Die konstante Kraftverteilung über dem Hub ermöglicht eine hohe Startkraft, die verbunden mit der geringen Masse der bewegten Spule höchste Dynamik erlaubt. Zudem lässt sich eine extrem geringe Hysterese erreichen. Durch diese Eigenschaften empfohlen sich Tauchspulenaktuatoren vor allem für dynamische und präzise Regelungsanwendungen.



17.2.1.2 2- und 3-phasen Linearmotor

Bsp. 2-phasen Das Funktionsprinzip ist das selbe wie im Kapitel Linearmotoren erklärt wird. In der Abbildung *Linearmotor* funktioniert der eisenbehaftete LM nach dem Reluktanzkraft-Prinzip (Kap. ??).

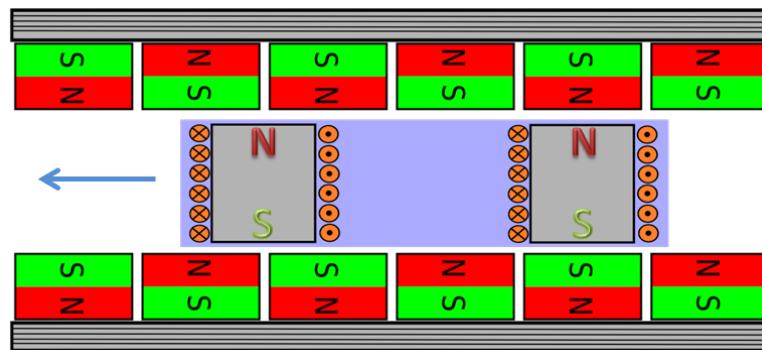


Abbildung 18: 2-phasiger, eisenbehafteter, U-förmiger Linearmotor

Die Magnetfelder des Läufers und die des Stators werden immer kombiniert geschaltet (mit Strom versorgt), dass der Läufer vom nächsten Permanentmagnetpol in Bewegungsrichtung ein Wegstück „nach vorne“ gezogen wird (und vom Magnetfeld hinter sich abgestoßen wird). Hat er die Position erreicht, zu der er gezogen wurde, so wird umgepolzt, und der Läufer wird von dieser Position nun weggedrückt und zur nächsten Magnetspule/Permanentmagnet hingezogen. Dadurch, dass der Läufer zwei etwas versetzte Magnetfeld-Erzeuger besitzt, befindet sich immer mindestens einer davon gerade „auf halbem Weg“, was eine Festlegung der Laufrichtung (vorwärts oder rückwärts) ermöglicht. Dieser Motor ist die lineare Variante einer permanenterregten Synchronmaschine.

Bsp. 3-phasen Die eisenlosen 3-phasigen elektronisch kommutierten AC synchron Linearmotoren sind für die anspruchsvollen Anforderungen mit hohen Beschleunigungen sehr gut geeignet. Vorteil der 3-phasigen Linearmotoren ist, dass ein gleichmäßigerer Lauf möglich ist bzw. eine konstante Kraft geliefert werden kann.

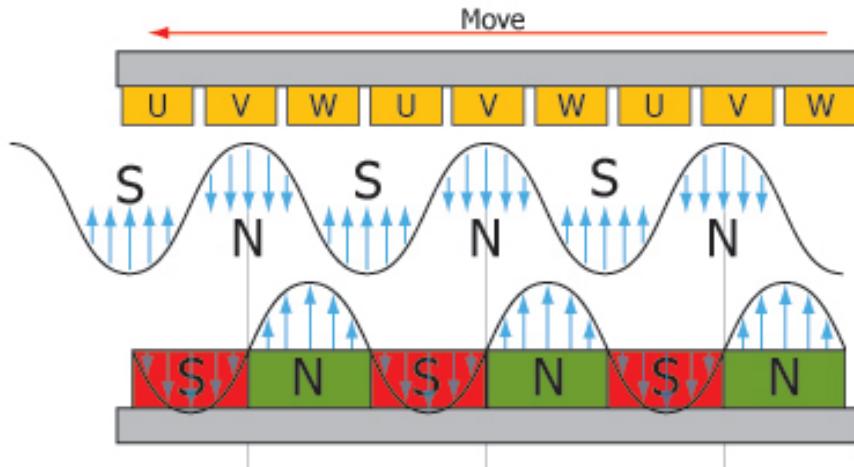


Abbildung 19: eisenloser, permanenterregter, 3-phasen synchron Linearmotor

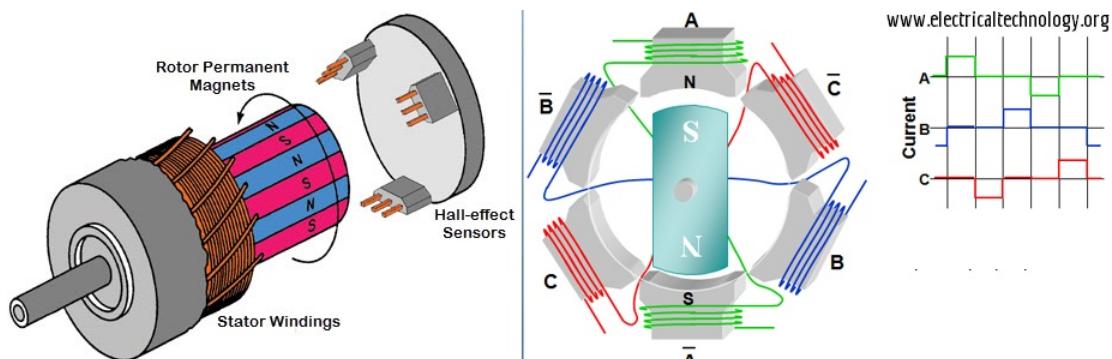
17.2.2 Rotationsaktuatoren

17.2.2.1 Gleichstrommotoren

Brushless DC-Motor

Aufbau

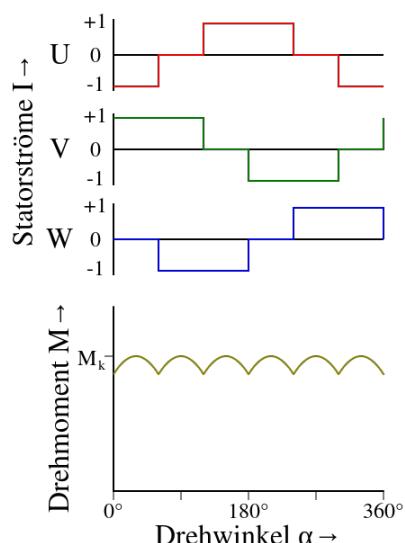
Der *brushless DC-Motor* basiert entgegen der Namensgebung nicht auf dem Funktionsprinzip der Gleichstrommaschine, sondern ist aufgebaut wie eine Drehstrom-Synchronmaschine mit Erregung durch Permanentmagnete. Die (oft dreisträngige) Drehstromwicklung wird elektronisch kommutiert (z.B. Block- oder Sinuskommutierung). Dadurch entsteht ein sich drehendes magnetisches Feld, welches den permanenterregten Rotor mitzieht.



Construction, Working Principle and Operation of BLDC Motor (Brushless DC Motor)

Kommutierungsarten

Bei einer Blockkommutierung wird der Strom in den jeweiligen Wicklungen zwischen 3 Zuständen geschaltet. Das sich dabei ergebende Drehmoment weist eine Welligkeit auf. Bei einer Sinuskommutierung ergibt sich hingegen ein konstantes Drehmoment. Die Kommutierung kann entweder mit oder ohne Sensoren erfolgen.



Sensorgesteuerte Kommutierung

Bei ersterem werden entweder Hallsensoren oder optische Sensoren verwendet. Entsprechend dieser Stellungsinformation werden über geeignete Leistungstreiber von der Steuerelektronik die Wicklungen angesteuert, die im Rotor ein Drehmoment erzeugen. Der Vorteil ist, dass die sensorgesteuerte Kommutierung auch bei sehr geringen Drehzahlen bzw. im Stand funktioniert. Gewöhnlich werden bei dieser Kommutierung nicht alle Phasen zugleich bestromt. Bei den Drehphasenmotoren ist üblicherweise zu jedem Zeitpunkt jeweils eine Phase stromlos.

Sensorlose Kommutierung

Bei der sensorlosen Kommutierung erfolgt die Erfassung der Rotorposition über die in den Spulen des Stators induzierte Spannung, welche von der elektronischen Steuerschaltung ausgewertet

wird. Allerdings ist zur Auswertung der Induktionsspannung eine gewisse Mindestdrehzahl erforderlich. Sensorlose EC-Motoren müssen daher wie Synchronmotoren bzw. Schrittmotoren bis zum Erreichen der Mindestdrehzahl blind geschaltet werden. Mittlerweile gibt es allerdings Verfahren, mit denen ein EC-Motor auch unterhalb dieser Mindestdrehzahl nicht blind gesteuert wird. Dazu werden bei Stillstand kurze Stromimpulse gesendet, die den Motor zwar nicht bewegen, aber durch das magnetische Feld des Rotors beeinflusst werden. Das Magnetfeld mindert oder verstärkt den Stromfluss und verändert so die Zeit, die ein Stromimpuls benötigt, um eine Schwelle zu überschreiten. Diese Zeiten werden gemessen, und man kann damit die Rotorposition schon bei Stillstand bestimmen.

Geeignetste Kommutierung

Man erhält für EC-Motoren (electronically commutated) das beste Ergebnis (konst. Kraft), wenn man sie mit ihrer BEMF-Spannung (back electromotive force) ansteuert. Diese Spannung kann an bzw. zwischen den Phasen gemessen werden, wenn man den Motor von Hand bewegt.

Unterschied zur PMSM

Die BEMF ergibt bei PMSM einen sinusoidalen Spannungsverlauf. Bei BLDC hingegen einen trapezförmigen Spannungsverlauf der Gegegeninduktionsspannung. Die Verläufe folgern aus den unterschiedlichen Statorwicklungen. Bei BLDC sind die Wicklungen konzentriert und bei PMSM hingegen viel gleichmäßiger auf den Umfang verteilt, was zu einer sinusoidalen Spannung führt. Dadurch unterscheidet sich auch die Ansteuerung der beiden Motoren, da man die höchste Performance erreicht indem man den Motor mit der selbigen Spannung als seine BEMF ansteuert.

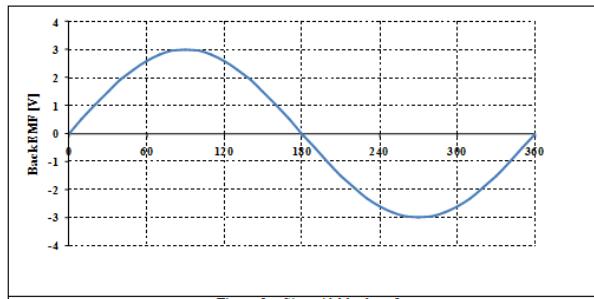


Figure 1a: Sinusoidal back emf

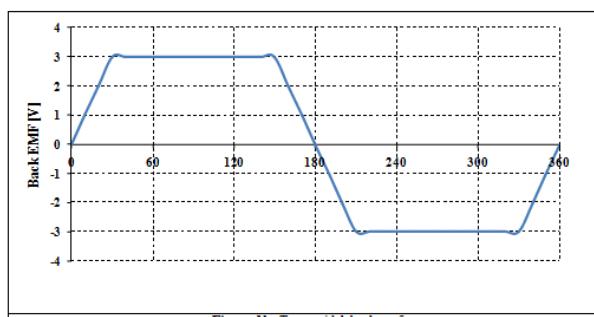


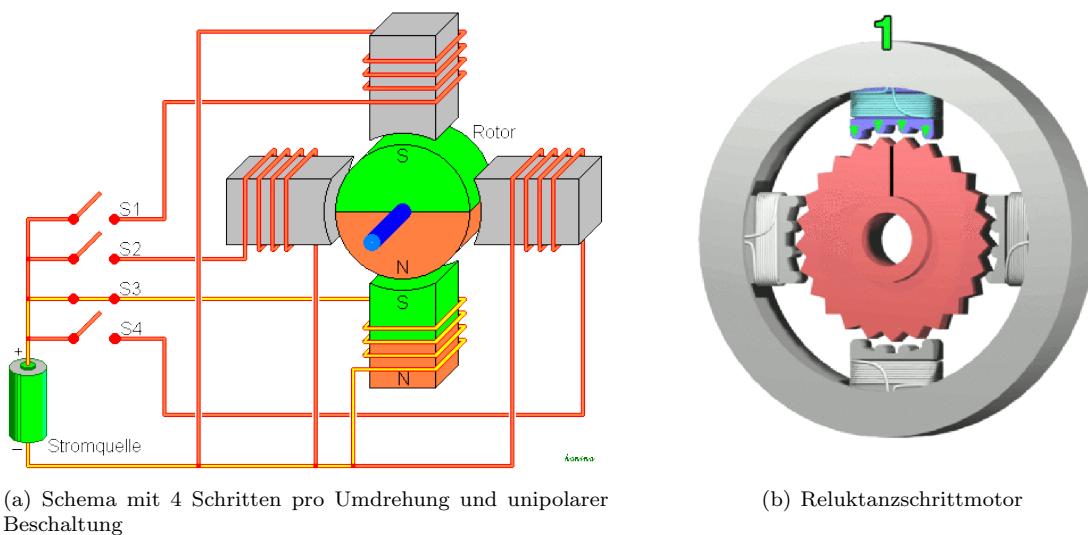
Figure 1b: Trapezoidal back emf

Abbildung 20: PMSM- sinusoidale BEMF

Steppermotor

Allgemein

Ein Schrittmotor ist ein Synchronmotor, bei dem der Rotor durch ein gesteuertes, schrittweise rotierendes, elektromagnetisches Feld der Statorspulen um einen minimalen Winkel (Schritt) oder sein Vielfaches gedreht werden kann. Schrittmotoren existieren auch in Form von Linearmotoren. Schrittmotoren folgen im Prinzip exakt dem außen angelegten Feld und können ohne Sensoren zur Positionsrückmeldung (Encoder, Drehgeber oder ähnliches) genau betrieben werden. Sie sind haben ein ähnliches Verhalten wie Synchronmotoren, weisen aber in der Regel eine deutlich höhere Polpaarzahl auf. Daher können sie einfacher betrieben werden als beispielsweise Servomotoren, da diese auf die gewünschte Position eingeregelt werden müssen. Für einen besonders homogenen Verlauf werden Schrittmotoren mit einem gleichförmigen Drehfeld angesteuert.



Schrittverlust

Wird ein Schrittmotor durch ein externes Lastmoment oder durch die anzutreibende Masse beim starken Beschleunigen beziehungsweise Verzögern überlastet (d. h. Lastmoment \geq Motormoment), kann der Rotor dem Drehfeld nicht mehr folgen. Es werden Schritte übersprungen, und die Information über die aktuelle Position des Rotors geht verloren. Bei diesem sogenannten Schrittverlust springt der Motor in die vorherige oder nächste Position gleicher Phase zurück. Durch die mechanische Bewegungsenergie (Trägheit) kommt es bei rasch bewegten Magnetfeldern meist zu einer Serie von verlorenen Schritten. Auftretende Schrittverluste summieren sich und führen dann zu einer ungenauen Positionierung.

Bauarten

- Reluktanzschrittmotor
- Permanentmagnetschrittmotor
- Hybridschrittmotor (Reluktanz- und Permanentmagnetschrittmotor)
- Lavet-Schrittmotor

Reluktanzschrittmotor

Beim Reluktanzschrittmotor besteht der Rotor aus einem gezahnten Weicheisenkern. Bei diesem Material verschwindet nach dem Ausschalten des Statorstromes das Magnetfeld. Bei eingeschaltetem Strom fließt der magnetische Fluss durch den Weicheisenkern des Rotors. Die Drehbewegung des Rotors kommt zustande, weil vom gezahnten Stator der nächstliegende Zahn des Rotors angezogen wird, da sich so der magnetische Widerstand verringert. Da der Reluktanzschrittmotor keine Permanentmagnete enthält, hat er daher im Gegensatz zum Permanentmagnetschrittmotor bei ausgeschaltetem Strom auch kein Rastmoment.

Permanentmagnetschrittmotor

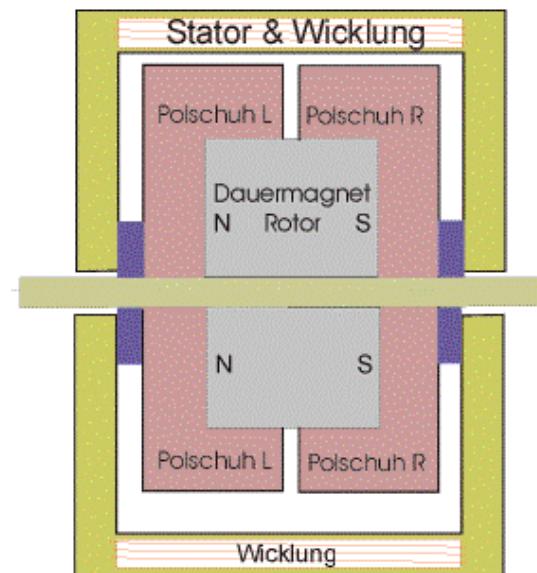
Beim Permanentmagnetschrittmotor besteht der Stator aus Weicheisen und der Rotor aus Dauermagneten, die abwechselnd einen Nord- und einen Südpol aufweisen. Mit dem Stator-Magnetfeld richtet man den dauer magnetischen Rotor so aus, dass eine Drehbewegung entsteht. Beim Permanentmagnetschrittmotor ist die Anzahl der Pole (und damit die Auflösung) begrenzt.

Hybridschrittmotor

Der Hybridschrittmotor vereint die positiven Eigenschaften beider Bauformen durch feine Schrittein teilung und gutes Drehmoment. Sein Rotor besteht aus einem axialen Permanentmagneten, an dessen Enden (Polen) gezahnte Weicheisenkränze befestigt sind. Beide sind um eine halbe Zahnbreite gegeneinander versetzt, so dass sich Nord- und Südpole abwechseln. Vorteil ist das erheblich kleinere äußere Magnetfeld. Nahezu alle heute erhältlichen Schrittmotoren sind Hybridgetoren, da er hohe mechanische Leistungen bei kleinen Schrittwinkeln und kleiner Bauform vereint. Die gebräuchlichsten Schrittauflösungen liegen zwischen 50 und 2000 Schritte pro Motorumdrehung ohne elektronische Zusatzmaßnahmen. Bezuglich der Phasenzahl sind 2-Phasenmotore am gängigsten.

Ansteuerung & Anschluss

Weiterhin kann man zwei Gruppen von Motoren bzw. Ansteuertechniken unterscheiden: Unipolare und bipolare Motoren. Unipolare Motoren verfügen über zwei Spulen mit Mittelabgriff. Sie haben fünf oder sechs Anschlüsse. Mit einem Multimeter lässt sich schnell feststellen, welche Anschlüsse



die Mittelabgriffe und welche die Spulenenden sind. Die Ansteuerung erfolgt durch wechselweises Einschalten von jeweils einem Spulenende, so daß immer nur die halbe Spule bestromt ist. Ein 2-phägiger, bipolarer Motor hat zwei Spulen, manchmal auch zwei Spulenpaare, die durch Umpolen angesteuert werden. Sind zwei Spulenpaare vorhanden (also acht Anschlüsse am Motor), können die Spulenpaare wahlweise parallel oder in Reihe geschaltet werden, woraus sich unterschiedliche (dynamische) Eigenschaften ergeben. Eine Parallelschaltung führt im allgemeinen zu mehr Drehmoment im oberen Drehzahlbereich, stellt aber auch höhere Anforderungen an den Stromregler. Bipolare Motoren mit 8 Anschlüssen können prinzipiell auch unipolar angesteuert werden, wobei man allerdings einen Teil der Motorperformance verschenkt. Bei der Beschaltung der Spulen

Wire Connection Diagrams

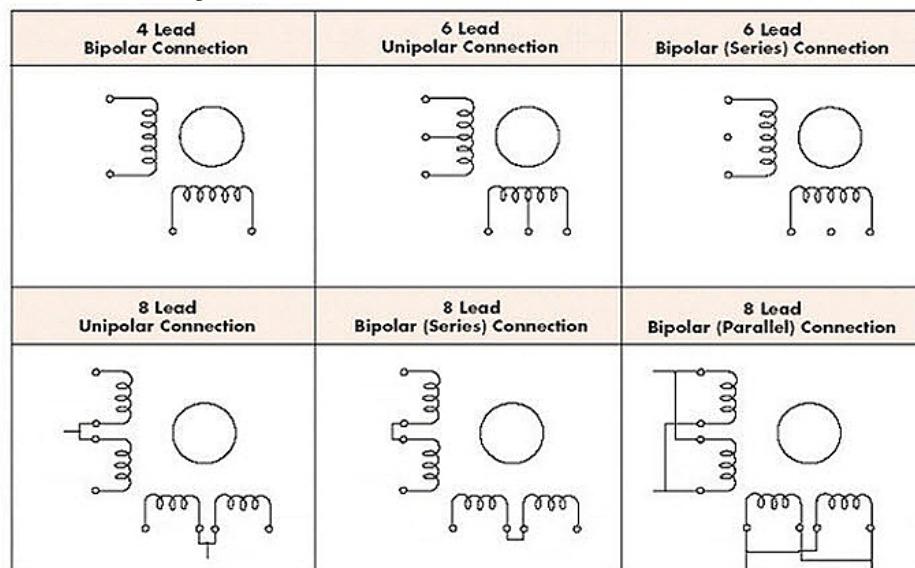


Abbildung 21: Anschlussvarianten eines 2-phägigen Steppermotors

ergeben sich drei weitere Betriebsarten: Vollschrift- Wavedrive- und Halbschriftbetrieb.

1. Normal-Betrieb: Es werden immer beide Spulen gleichzeitig bestromt. Es ergeben sich vier unterschiedliche Schrittpositionen (siehe Schaubild) pro Umlauf.
2. Wavedrive-Betrieb: Hier wird immer nur eine Spule bestromt. Die Leistungsaufnahme und damit auch das Drehmoment sind im Vergleich zu 1) geringer. Die resultierenden vier Schritte-

positionen liegen zwischen denen aus 1).

3. Halbschritt-Betrieb: Kombination aus 1) und 2). Es wird wechselweise eine, bzw. zwei Spulen bestromt. Es ergeben sich 8 Schrittpositionen. Daher kommt die Bezeichnung Halbschritt, da der physikalische Schrittewinkel des Motors halbiert wird.

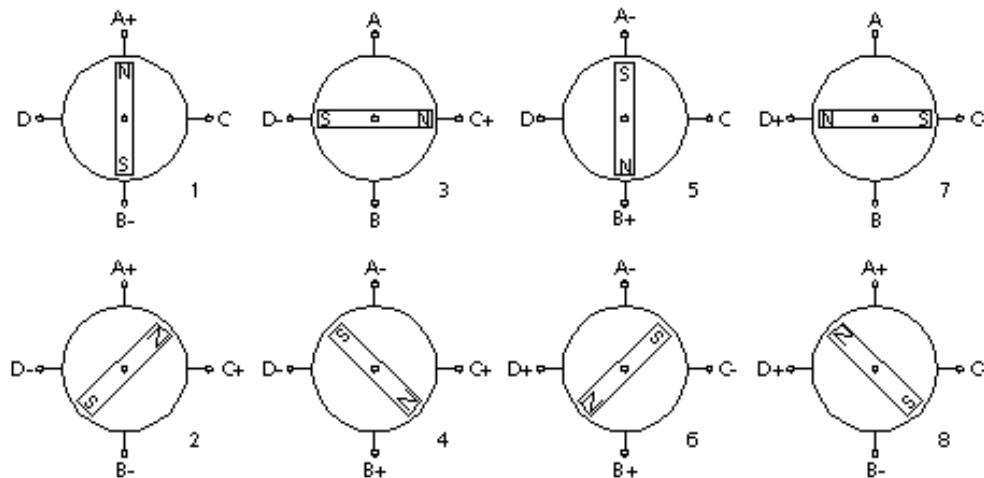


Abbildung 22: Oberen vier Zustände: Wavedrive-Betrieb, unteren vier Zustände: Vollschrittbetrieb, Halbschrittbetrieb: alle 8 Zustände

Vor- & Nachteile gegenüber BLDC

Steppermotor

- + Günstiger als BLDC (Preis-Leistung)
- + Einfache Winkel-(Weg-)messung durch Schrittzählung ohne Encoder möglich
- Moment nimmt ab einer bestimmten Drehzahl ab
- nur niedrige Drehzahlen möglich

Anwendungszwecke

Typische Anwendungsgebiete sind Drucker, vor allem Matrixdrucker, oder der Antrieb des Schreib-/Lesekopfes in einem Diskettenlaufwerk. Aufgrund ihrer hohen Genauigkeit werden sie auch in computergesteuerten Werkzeugmaschinen zur Positionierung der Werkzeuge verwendet. Durch die ständig sinkenden Kosten für die Ansteuerelektronik werden sie auch zunehmend im Konsumgüterbereich verwendet. So sind in Kraftfahrzeugen der mittleren und gehobenen Kategorie heute bis über 50 Schrittmotoren im Einsatz, die Betätigung der vielen Klappen einer automatischen Heizungs- und Klimaanlage ist dafür ein Beispiel. Schrittmotoren können bis ungefähr 1 kW wirtschaftlich eingesetzt werden.

Servomotor

Als Servomotor werden Elektromotoren beliebiger Bauart (AC, DC) bezeichnet, die die Kontrolle der Winkelposition, Drehzahl und Winkelbeschleunigung ihrer Motorwelle erlauben. Bürstenlose Servomotoren sind entweder Brushless DC-Motoren, Drehstrom-Asynchronmotoren oder Drehstrom-Synchronmotoren. Sie bestehen aus einem Elektromotor, der zusätzlich mit einem Sensor (Encoder) zur Positionsbestimmung ausgestattet ist. Die vom Sensor ermittelte Winkelposition/-geschwindigkeit der Motorwelle wird kontinuierlich an eine meist außerhalb des eigentlichen Motors angebrachte Regelelektronik übermittelt, den so genannten Servoregler, der die Bewegung des Motors entsprechend einem oder mehreren einstellbaren Sollwerten, wie etwa Soll-Winkelposition oder Solldrehzahl, regelt.

Brushless DC-Motor

- + Hohe Drehzahlen
- + Sehr gute Laufruhe
- + günstiger Momentenverlauf über den gesamten zulässigen Drehzahlbereich
- Kostet mehr als Stepper
- Encoder nötig für Weg-, Drehzahlmessung

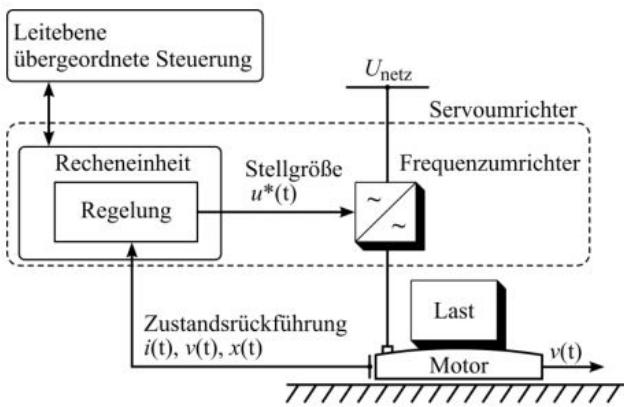
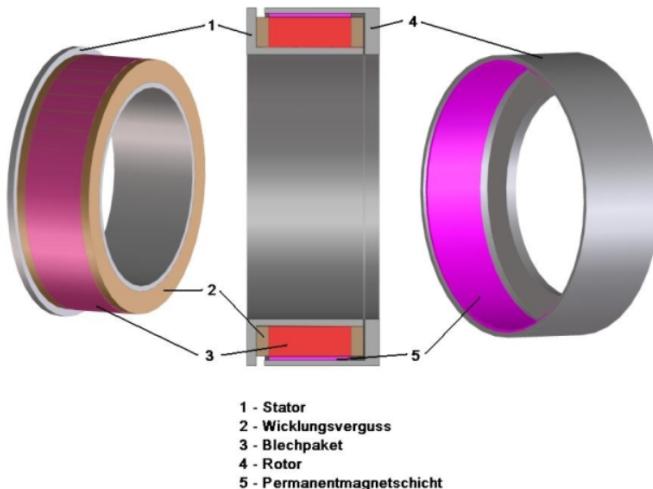


Abbildung 23: Die Kombination Servomotor und Servoregler wird als Servoantrieb bezeichnet

Torque-Motor

Torque-Motoren können als Außenläufer (Stator innen, Rotor außen) oder normale Innenläufer ausgeführt sein. Durch den verhältnismäßig großen Durchmesser, die hochpoligen Wicklungen können diese Motoren große Drehmomente bei geringen Drehzahlen erzeugen. Meist werden BLDC-Motoren als Torquemotoren gebaut, teilweise aber auch Reluktanzmotoren.

- + Große Beschleunigungen möglich → hohe Dynamik des Systems
- + Antriebssteif → kein Verdrehspiel → geringere Störgrößen → gute Regeleigenschaften
- + Kaum Verschleiß
- Starke Wärmeentwicklung (muss oft Wassergekühlt werden)
- Teuer



17.2.2.2 Drehstrommotoren

Synchronmotor

Mn und Pn Kurven, vergleich mit asynchron!!!

Asynchronmotor**17.2.2.3 Wechselstrommotoren****Reluktanzmotor**

...

V

Regelungstechnik

18 Begrifflichkeiten

Begriff	Beschreibung
eingangsaffine Systeme	sind Systeme die in u linear sind. Man sagt auch eingangslineare Systeme. Genauer bedeutet es, dass sich die wirkende Stellgröße u nur linear über die Zeit ändert.
relativer Grad δ eines Systems	bei SISO-Sytemen gibt der relative Grad gleich dem Grad der Ableitungen der Ausgangsgröße y an, bei dem die δ -mal abgeleitete Ausgangsgröße erstmals direkt von der Stellgröße u abhängt: $y_i^{(\delta)} = fcn(u)$. Ist der rel. Grad eines Systems <i>wohldefiniert</i> , dann ist es Eingangs-/Ausgangs-linearisierbar. Für lineare Systeme entspricht der relative Grad dem Differenzgrad des Zähler- und Nennerpolynoms. Der relative Grad r auch als die Anzahl der Integratoren gedeutet werden, die der Eingang u durchlaufen muss, bevor er den Ausgang y beeinflusst. In anderen Worten wird der Ausgang $y^{(\delta)}$ direkt vom Eingang u beeinflusst und der Ausgang y erst über δ Integratoren. Der relative Grad r ist somit ein Maß für die Mindestverzögerung im E/A-Verhalten. [NICHTLINEARE SYSTEME UND REGELUNGEN, JÜRGEN ADAMY S290]
Transitionsmatrix $\Phi(t)$	Die Transitionsmatrix $\Phi(t)$ beschreibt den Übergang des Zustandsvektors $\mathbf{x}(t)$ von seinem Anfangswert \mathbf{x}_0 zu seinem Wert zum Zeitpunkt t . $\mathbf{x}(t) = \Phi(t)\mathbf{x}_0 + \int_0^t \Phi(\tau - t)\mathbf{B}\mathbf{u}(\tau)d\tau$
Diffeomorphismus	Ist eine nichtlineare Zustandstransformation $\mathbf{z} = \mathbf{q}(\mathbf{x})$, $\mathbf{x} = \mathbf{q}^{-1}(\mathbf{z})$ die stetig differenzierbar und eindeutig ist. Letzteres bedeutet, dass jedem \mathbf{x} kann genau ein \mathbf{z} zugeordnet werden und umgekehrt.

19 Eigenschaften von Systemen

19.1 Flachheit

Ist ein System *flach*, so können die dynamischen Gleichungen des Systems so umgeformt werden, dass sich Eingangs- und Zustandsgrößen anhand von Funktionen darstellen die nur von den Ausgangsgrößen und deren Ableitungen abhängig sind. Dadurch ist es möglich für einen gewünschten Ausgangsgrößenverlaufs die nötigen Eingangs- und Zustandsgrößenverläufe direkt anzugeben.

Beispiel:

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= -x_1 - x_2 + u \\ y &= x_1\end{aligned}$$

Die Zustandsgrößen des Systems können als Funktionen der der Ausgangsgrößen dargestellt werden.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y \\ \dot{y} \end{bmatrix}$$

Ebenso der Eingangsvektor aus der mittleren Gleichung.

$$u = \ddot{y} + \dot{y} + y$$

Bei einem solchen System wird y als **flacher Ausgang** bezeichnet. Dabei ist es nicht relevant ob dieser Ausgang tatsächlich existiert oder nur ausgedacht - *fiktiv* ist. Existieren nur *fiktive* flache Ausgänge müssen diese für den

Entwurf einer Steuerung in den realen Ausgang umgerechnet werden.

Definition: Flachheit

Für Systeme

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}), \quad \mathbf{x} \in \mathbb{R}^n, \mathbf{u} \in \mathbb{R}^m$$

mit $m \leq n$ (weniger Eingangs- als Zustandsgrößen) für die sich keine Stellgröße u_i als Funktion anderer Stellgrößen $u_{j \neq i}$ darstellen lässt

$$\text{rang}\left(\frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}}\right) = m,$$

ist die *Flachheit* folgendermaßen definiert:

Es heißt flach wenn ein realer oder fiktiver Ausgangsvektor

$$\mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^\alpha)$$

existiert, so dass

(1) der Zustandsvektor \mathbf{x} als Funktion von \mathbf{y} und einer endlichen Anzahl β von Ableitungen von \mathbf{y}

$$\mathbf{x} = \Psi_1(\mathbf{y}, \dot{\mathbf{y}}, \dots, \mathbf{y}^{(\beta)})$$

dargestellt werden kann,

(2) der Eingangsvektor \mathbf{u} als Funktion

$$\mathbf{u} = \Psi_2(\mathbf{y}, \dot{\mathbf{y}}, \dots, \mathbf{y}^{(\beta+1)})$$

darstellbar ist,

(3) für Ein- und Ausgangsvektor

$$\dim(\mathbf{y}) = \dim(\mathbf{u})$$

gilt.

20 Systemtransformationen

20.1 Zustandstransformation

Die Wahl der Zustandsgrößen eines Systems sind keinesfalls eindeutig. Mit Hilfe einer *Zustandstransformation* kann ein System in eine neue Systemdarstellung mit neuen Zustandsgrößen transformiert werden.

20.1.1 Lineare Zustandstransformation

Durch eine *reguläre Zustandstransformation* der Form

$$\mathbf{x}(t) = \mathbf{V}\mathbf{z}(t)$$

mit der regulären Matrix \mathbf{V} (reguläre Matrix: invertierbare, quadratische Matrix) lässt sich ein *Axbu-Modell* mit dem Zustand $\mathbf{x}(t)$ in die Form

$$\dot{\mathbf{z}} = \mathbf{V}^{-1}\mathbf{A}\mathbf{V}\mathbf{z} + \mathbf{V}^{-1}\mathbf{B}\mathbf{u}$$

$$\mathbf{y} = \mathbf{C}\mathbf{V}\mathbf{z} + \mathbf{D}\mathbf{u}$$

mit dem neuen Zustand \mathbf{z} überführen. Eine solche Transformation kann beispielsweise verwendet werden um die Dynamikmatrix \mathbf{A} in eine Diagonalmatrix umzuwandeln. Damit kann auf einfache Weise die Transitionsmatrix $\Phi(t)$ berechnet und mit \mathbf{V} wieder in die ursprüngliche Zustandsdarstellung überführt werden.

$$\Phi(t) = \mathbf{V}^{-1}\tilde{\Phi}(t)\mathbf{V}$$

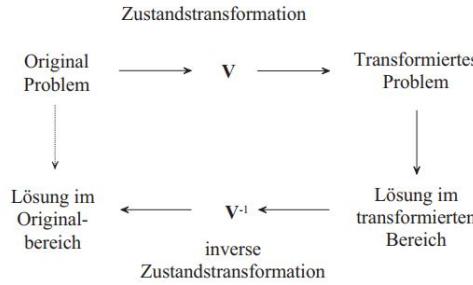


Abbildung 24: Reguläre Zustandstransformation

Auch wird die Matrix $\tilde{\mathbf{A}} = \mathbf{V}^{-1} \mathbf{A} \mathbf{V}$ gerne in eine Begleitmatrix der Form

$$\tilde{\mathbf{A}} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix}$$

umgewandelt, da diese Darstellungen nützlich für den Reglerentwurf sind (als Beispiel sei die Polvorgabe gegeben) oder Systemeigenschaften direkt abgelesen werden können.

20.1.2 Nichtlineare Zustandstransformation

Auch bei nichtlinearen Systemen

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

können Transformationen

$$\mathbf{z} = \mathbf{q}(\mathbf{x}), \quad \mathbf{x} = \mathbf{q}^{-1}(\mathbf{z})$$

sinnvoll sein, wenn es gelingt, das System in eine für den Reglerentwurf geeignete Form zu bringen. Für die Transformation wird gefordert, dass sie stetig differenzierbar und eindeutig ist. Eindeutig bedeutet, dass jedem \mathbf{x} genau ein \mathbf{z} zugeordnet ist und umgekehrt. Dadurch ist auch gefordert, dass die Anzahl der Zustände durch eine nichtlineare Zustandstransformation gleich bleibt. Eine solche Transformation nennt man Diffeomorphismus. Das transformierte System ergibt sich durch Substitution zu

$$\frac{\partial \mathbf{q}^{-1}(\mathbf{z})}{\partial t} = \mathbf{f}(\mathbf{q}^{-1}(\mathbf{z}), \mathbf{u}) \quad \Rightarrow \quad \frac{\partial \mathbf{q}^{-1}(\mathbf{z})}{\partial \mathbf{z}} \dot{\mathbf{z}} = \mathbf{f}(\mathbf{q}^{-1}(\mathbf{z}), \mathbf{u}),$$

woraus sich die transformierte Systemdarstellung

$$\Rightarrow \quad \dot{\mathbf{z}} = \left(\frac{\partial \mathbf{q}^{-1}(\mathbf{z})}{\partial \mathbf{z}} \right)^{-1} \cdot \mathbf{f}(\mathbf{q}^{-1}(\mathbf{z}), \mathbf{u}) = \tilde{\mathbf{f}}(\mathbf{z}, \mathbf{u})$$

ergibt. Da die Transformation eindeutig ist, gibt es auch eine Rücktransformation

$$\dot{\mathbf{x}} = \left(\frac{\partial \mathbf{q}(\mathbf{x})}{\partial \mathbf{x}} \right)^{-1} \cdot \tilde{\mathbf{f}}(\mathbf{q}(\mathbf{x}), \mathbf{u}) = \mathbf{f}(\mathbf{x}, \mathbf{u}).$$

Im allgemeinen kann es sein, dass die Transformationsgleichung auch eine Abhängigkeit vom Eingangsvektor \mathbf{u} und eventuell seinen zeitlichen Ableitungen aufweist. D.h. die Form

$$\mathbf{z} = \mathbf{q}(\mathbf{x}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(i)})$$

$$\mathbf{x} = \mathbf{q}^{-1}(\mathbf{z}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(i)})$$

hat. Transformiert man nun das System $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u})$ in \mathbf{z} -Koordinaten, erhält man

$$\dot{\mathbf{x}} = \frac{\partial \mathbf{q}^{-1}(\mathbf{z}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(i)})}{\partial t} = \frac{\partial \mathbf{q}^{-1}(\mathbf{z}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(i)})}{\partial \mathbf{z}} \dot{\mathbf{z}} + \sum_{j=0}^i \frac{\partial \mathbf{q}^{-1}(\mathbf{z}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(i)})}{\partial \mathbf{u}^{(j)}} \mathbf{u}^{(j+1)} = \mathbf{f}(\mathbf{q}^{-1}(\mathbf{z}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(i)}), \mathbf{u})$$

und schließlich die transformierte Systembeschreibung

$$\Rightarrow \dot{\mathbf{z}} = \left(\frac{\partial \mathbf{q}^{-1}(\mathbf{z}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(i)})}{\partial \mathbf{z}} \right)^{-1} \cdot \left(\mathbf{f}(\mathbf{q}^{-1}(\mathbf{z}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(i)}), \mathbf{u}) - \sum_{j=0}^i \frac{\partial \mathbf{q}^{-1}(\mathbf{z}, \mathbf{u}, \dot{\mathbf{u}}, \dots, \mathbf{u}^{(i)})}{\partial \mathbf{u}^{(j)}} \mathbf{u}^{(j+1)} \right).$$

Mit dem Hintergedanken dass die Zustände \mathbf{x} bzw. \mathbf{z} auch vom Eingang und seinen Ableitungen abhängt, kann die transformierte folgendermaßen dargestellt werden.

$$\Rightarrow \dot{\mathbf{z}} = \left(\frac{\partial \mathbf{q}^{-1}}{\partial \mathbf{z}} \right)^{-1} \cdot \left(\mathbf{f}(\mathbf{q}^{-1}, \mathbf{u}) - \sum_{j=0}^i \frac{\partial \mathbf{q}^{-1}}{\partial \mathbf{u}^{(j)}} \mathbf{u}^{(j+1)} \right) = \tilde{\mathbf{f}}(\mathbf{z}, \mathbf{u})$$

NICHTLINEARE SYSTEME UND REGELUNGEN, JÜRGEN ADAMY S207

20.1.2.1 Herleitung des Diffeomorphismus (Transformationsvorschrift \mathbf{q})

TODO

20.2 Eingangstransformation

21 Stabilität

21.1 Stabilitätsanalyse des charakteristischen Polynoms

Bei linearen Systemen ist das Nennerpolynom der Übertragungsfunktion das charakteristische Polynom, wobei dessen Koeffizienten die Nullstellen des Polynoms und somit die Pole des Systems darstellen. Gegeben sei die Differentialgleichung

$$\dot{y}(t) + a_0 y(t) = 0.$$

Mit dem eulerschen Ansatz $y = e^{\lambda t}$ ergibt sich

$$e^{\lambda t}(\lambda + a_0) = 0$$

mit dem charakteristischen Polynom $\lambda + a_0 = 0$. Dieses Polynom hat die Nullstelle

$$\lambda = -a_0.$$

In den Ansatz eingesetzt ergibt sich

$$y(t) = e^{-a_0 t}.$$

Dadurch ergibt sich für unendliche Zeit und einem negativen Realteil von a_0

$$\lim_{t \rightarrow \infty} (e^{-a_0 t}) = 0.$$

Die Differentialgleichung heißt dann asymptotisch stabil. Für Differentialgleichungen höherer Ordnungen werden etwa das *Hurwitz*- oder *Routh*-Kriterium verwendet um die Stabilität von DGL's zu prüfen.

22 Eigenschaften eines Regelkreises

22.1 Bandbreite

Die Bandbreite eines Regelkreises ist jene Frequenz, bei der der Amplitudenverlauf des Bode Diagramms, die -3dB Grenze unterschreitet. Höhere Bandbreite bedeutet, dass der Regelkreis (das geregelte System) seinem Sollwert bis zu höheren Frequenzen folgen kann.

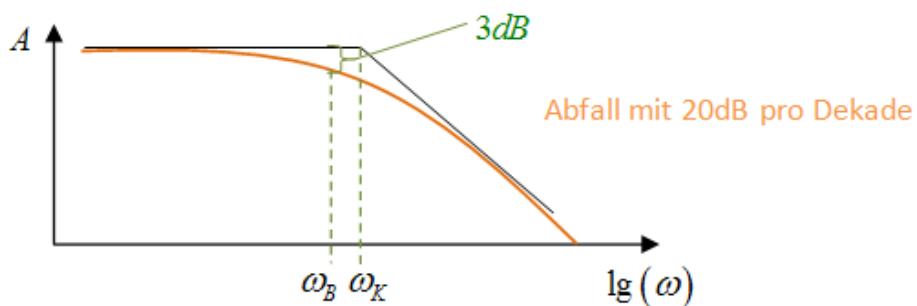


Abbildung 25: Amplitudenverlauf: ω_B - Bandbreite des Systems, ω_K - Knickfrequenz

22.2 Nyquist-Diagramm

22.3 Wurzelortskurve

22.4 Empirische Grundformeln

22.5 Führungs- und Störungsverhalten eines Regelkreises

23 Identifikation von Systemen

24 Systemdarstellung: Normal- und Regelungsformen

24.1 Normalformen

24.1.1 Regelungnormalform (auch Steuerungnormalform)

24.1.2 Beobachtungnormalform

24.1.3 Kanonische Normalform

24.1.4 Brunovsky–Normalform

24.1.5 Jordansche-Normalform

24.1.6 Nichtlineare Regelungnormalform

Nichtlineare eingangsaffine Systeme der Form

$$\dot{x} = a(x) + B(x) \cdot u = a(x) + \sum_{i=1}^m b_i(u) \cdot u_i \quad (9)$$

können in der nichtlinearen Regelungnormalform

$$\begin{bmatrix} \dot{x}_1 \\ \vdots \\ \dot{x}_{n-1} \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} x_2 \\ \vdots \\ x_n \\ \alpha(x) \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \beta(x) \end{bmatrix} u$$

$$y = x_1$$

dargestellt werden.

24.1.6.1 Zustandskoordinatentransformation mittels Lie-Ableitungen

Im folgenden wird beschrieben, wie Systeme der Form (??) in die nichtlineare Regelungnormalform transformiert werden können.

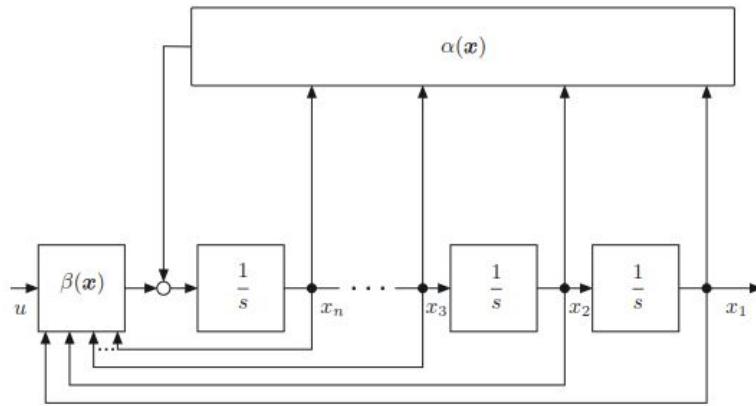


Abbildung 26: System in nichtlinearer Regelungsnormalform

Definition: Lie-Derivierter/Lie-Ableitung

Sie sind als Gradient einer skalaren Funktion $h(\mathbf{x})$ multipliziert mit einem Vektorfeld $\mathbf{f}(\mathbf{x})$ definiert

$$L_{\mathbf{f}} h(\mathbf{x}) = \frac{\partial h(\mathbf{x})}{\partial \mathbf{x}} \mathbf{f}(\mathbf{x}) = \text{grad}^T h(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})$$

Die skalare Funktion kann beispielsweise die Ausgangsfunktion $c(\mathbf{x})$ sein und das Vektorfeld z.B. der Dynamikvektor $\mathbf{a}(\mathbf{x})$.

Man bildet nun die zeitliche Ableitung der Ausgangsgröße y und erhält

$$\dot{y} = \frac{dc(\mathbf{x}(t))}{dt} = \underbrace{\frac{\partial c(\mathbf{x})}{\partial x_1} \frac{\partial x_1(t)}{\partial t} + \dots + \frac{\partial c(\mathbf{x})}{\partial x_n} \dot{x}_n}_{\dot{x}_1} = \frac{\partial c(\mathbf{x})}{\partial \mathbf{x}} \dot{\mathbf{x}}.$$

Durch einsetzen von $\dot{\mathbf{x}} = \mathbf{a}(\mathbf{x}) + \mathbf{b}(\mathbf{x}) \cdot u$ in obige Gleichung, ergibt sich

$$\dot{y} = \frac{\partial c(\mathbf{x})}{\partial \mathbf{x}} \mathbf{a}(\mathbf{x}) + \frac{\partial c(\mathbf{x})}{\partial \mathbf{x}} \mathbf{b}(\mathbf{x}) \cdot u.$$

Geschrieben mit den Lie-Derivierten:

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

Für die meisten technischen Systeme ist der zweite Term gleich Null, woraus

$$\dot{y} = L_{\mathbf{a}} c(\mathbf{x})$$

folgt. Die zweite Ableitung ergibt sich dann zu

$$\ddot{y} = \frac{dL_{\mathbf{a}} c(\mathbf{x})}{dt} = \underbrace{\frac{\partial L_{\mathbf{a}} c(\mathbf{x})}{\partial \mathbf{x}} \mathbf{a}(\mathbf{x})}_{L_{\mathbf{a}} L_{\mathbf{a}} c(\mathbf{x})} + \underbrace{\frac{\partial L_{\mathbf{a}} c(\mathbf{x})}{\partial \mathbf{x}} \mathbf{b}(\mathbf{x}) \cdot u}_{L_{\mathbf{b}} L_{\mathbf{a}} c(\mathbf{x})} = L_{\mathbf{a}} L_{\mathbf{a}} c(\mathbf{x}) + L_{\mathbf{b}} L_{\mathbf{a}} c(\mathbf{x}) \cdot u.$$

Mit dem zweiten Term wieder zu Null folgt

$$\ddot{y} = L_{\mathbf{a}} L_{\mathbf{a}} c(\mathbf{x}) = L_{\mathbf{a}}^2 c(\mathbf{x})$$

Bildet man auch die höheren Ableitungen, so ergibt sich schließlich folgendes Ergebnis:

$$\begin{aligned}
 y &= c(\mathbf{x}) \\
 \dot{y} &= L_{\mathbf{a}} c(\mathbf{x}) \\
 \ddot{y} &= L_{\mathbf{a}}^2 c(\mathbf{x}) \\
 &\vdots \\
 y^{(\delta-1)} &= L_{\mathbf{a}}^{(\delta-1)} c(\mathbf{x}) \\
 y^{(\delta)} &= L_{\mathbf{a}}^{(\delta)} c(\mathbf{x}) + L_{\mathbf{b}} L_{\mathbf{a}}^{(\delta-1)} c(\mathbf{x}) \cdot u
 \end{aligned} \tag{10}$$

Leitet man die Ausgangsfunktion δ -mal, so dass die Stellgröße u darin auftaucht, nennt man δ den relativen Grad des Systems. Es wird nun lediglich der Fall

$$\delta = n$$

betrachtet, also der Fall bei dem der relative Grad gleich der Systemordnung entspricht. Weiter führen wir die neuen Zustandskoordinaten z ein.

$$z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} y \\ \dot{y} \\ \ddot{y} \\ \vdots \\ y^{(n-1)} \end{bmatrix} = \begin{bmatrix} c(\mathbf{x}) \\ L_a c(\mathbf{x}) \\ L_a^2 c(\mathbf{x}) \\ \vdots \\ L_a^{(n-1)} c(\mathbf{x}) \end{bmatrix} = t(\mathbf{x})$$

Ist $t(\mathbf{x})$ stetig differenzierbar und es existiert die Umkehrfunktion

$$t^{-1}(t(\mathbf{x})) = \mathbf{x},$$

dann nennt man t einen *Diffeomorphismus*. Durch Differenziation der Transformationsgleichung $bmt(\mathbf{x})$ erhalten wir die nichtlineare Regelungsnormalform.

$$\begin{aligned} \dot{z} &= \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \\ \vdots \\ \dot{z}_n \end{bmatrix} = \begin{bmatrix} \dot{y} \\ \ddot{y} \\ \dddot{y} \\ \vdots \\ y^{(n)} \end{bmatrix} = \begin{bmatrix} z_2 \\ z_3 \\ z_4 \\ \vdots \\ z_{n+1} \end{bmatrix} = \dot{t}(\mathbf{x}) = \begin{bmatrix} L_a c(\mathbf{x}) \\ L_a^2 c(\mathbf{x}) \\ \vdots \\ L_a^{(n-1)} c(\mathbf{x}) \\ L_a^{(n)} c(\mathbf{x}) + L_b L_a^{(n-1)} c(\mathbf{x}) \cdot u \end{bmatrix} \\ &\Rightarrow \begin{bmatrix} \dot{z}_1 \\ \vdots \\ \dot{z}_{n-1} \\ \dot{z}_n \end{bmatrix} = \begin{bmatrix} z_2 \\ \vdots \\ z_n \\ \alpha(\mathbf{x}) \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \beta(\mathbf{x}) \end{bmatrix} u, \end{aligned}$$

mit

$$\begin{aligned} \alpha(\mathbf{x}) &= L_a^{(n)} c(\mathbf{x}) \\ \beta(\mathbf{x}) &= L_b L_a^{(n-1)} c(\mathbf{x}) \end{aligned}$$

24.1.7 Byrnes-Isidori-Normalform

25 Arten von Regelstrecken

- linear, nichtlinear
- statisch, dynamische
- zeitvariant, zeitinvariant
- SISO, MIMO

26 Arten von Reglern

- Realisierung als Analog- und digitaler Regler
- Stetige lineare Regler
 - PID-Regler
 - Zustandsregler
- Unstetige Regler
 - Zweipunktregler
 - Dreipunktregler

- Mehrpunktregler
- Nichtlineare Regler
 - Adaptiver Regler
 - Fuzzy-Regler

26.1 Standardregelkreis

Der Standardregelkreis ist ein Regelkreis mit Ausgangsrückführung. Dabei wird lediglich das Eingangs-/Ausgangsverhalten des Systems betrachtet. Die Regelstrecke kann auch als Zustandsraummodell mit Zustandsvariablen beschreiben werden, jedoch wird Regler können anhand von Frequenzlinienverfahren oder durch vorgeben der Zähler- und Nenner-

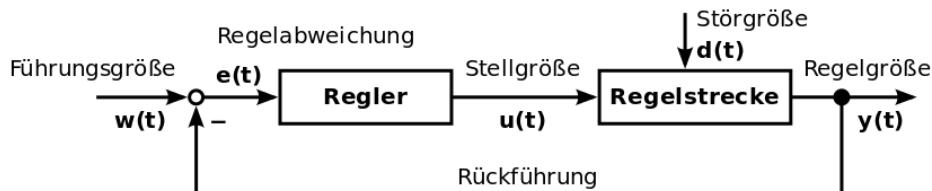


Abbildung 27: Standardregelkreis mit Ausgangsrückführung

polynome der Übertragungsfunktion (Polvorgabe) entworfen werden. Bei vorhandener Übertragungsfunktion kann mittels inverser Laplace-Transformation der zeitliche Verlauf der Ausgangsgröße errechnet und durch die Parameterwahl vorgegeben werden.

26.2 Zustandsregler

Der *Standardregelkreis* und dazugehörige Frequenzkennlinienverfahren beruhen auf einer Regelkreisstruktur, bei der eine Größe, die so genannte Ausgangsgröße, gemessen wird, und auf deren Kenntnis gemeinsam mit der vorgegebenen Führungsgröße der Regler als dynamisches System die Stellgröße errechnet. Daher werden Regelkreise dieser Art auch als Ausgangsregelungen bezeichnet. Setzt man nun voraus, dass der gesamte Zustand eines Systems messtechnisch erfassbar ist, dann ist es möglich, einen so genannten *Zustandsregler* zu entwerfen. Die Zustandsvariablen eines Systems sind frühzeitiger als der Ausgang vorhanden, was zu einer schnellen Regelung führt.

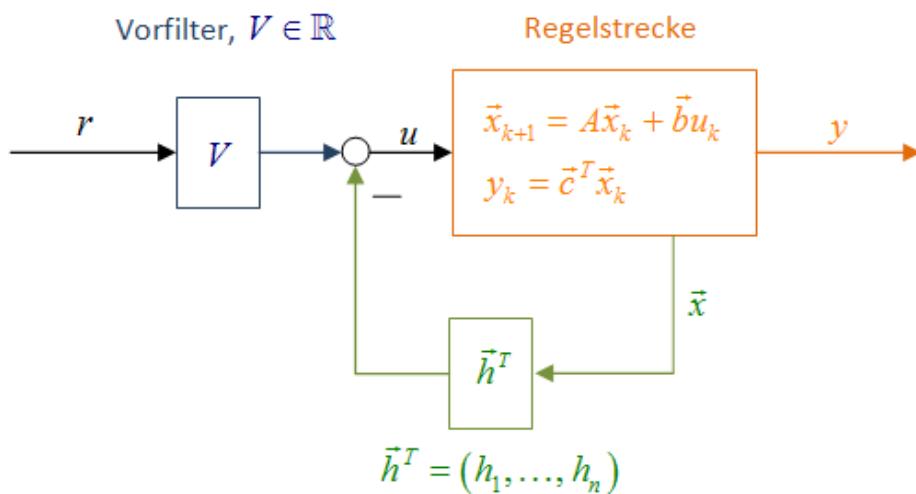


Abbildung 28: Zustandsregler mit h^T zur Vorgabe der Systemdynamik und V zur Elimination des statischen Fehlers

Durch die Hintereinanderschaltung der Integratoren ist nur die Zustandsvariable $x_1(t) = y(t)$ eine stationäre Größe, wenn die Eingangsgröße $u(t)$ konstant ist. Alle anderen Zustandsvariablen – eine stabile Regelstrecke vorausgesetzt – streben gegen den Wert null. Nach Einstellung und Optimierung des Faktors k_1 ergibt sich ein stabiler Regelkreis bestimmter Dämpfung mit einem Proportionalfehler der Regelgröße $y(t)$ gegenüber $w(t)$. Die anderen Faktoren der

Zustandsvariablen werden hintereinander beispielsweise zur Optimierung des Übergangsverhaltens eingestellt. Ein Vorfilter vor dem Soll-Ist-Vergleich korrigiert den statischen Fehler zwischen $w(t)$ und $y(t)$. Durch Einfügen eines überlagerten PI-Reglers verschwinden die Nachteile des einfachen Zustandsreglers. Das Vorfilter wird dann nicht mehr benötigt.

Durch die Zustandsrückführung kann keine stationäre Genauigkeit der Regelgröße zum Sollwert erreicht werden. Selbst bei einer Regelstrecke ohne Nullstellen, also ohne differenzielle Anteile, ist die Zustandsvariable x_1 nach der Regelungsnormalform nicht identisch mit der Ausgangsgröße $y(t)$. Deshalb wird die Zustandsrückführung häufig mit einem Vorfilter V erweitert. Die Führungsgröße $w(t)$ wirkt direkt auf das Vorfilter. Für einfache Anforderungen kann mittels eines Faktors in dem Vorfilter eine Korrektur durchgeführt werden, damit im stationären Zustand $w(t) = y(t)$ erreicht wird.

27 Auslegung von Reglern

27.1 Reglerauslegung durch Polvorgabe

27.2 Frequenzkennlinienverfahren

28 Arten von Rückführungen

28.1 Ausgangsrückführung

Die klassische Ausgangsrückführung $u = \mathbf{K}(w - y)$ stand am Anfang der Entwicklung der Regelungstechnik. Diese Art der Rückführung wird meist im LAPLACE-Bildbereich untersucht und es spielt dabei nur das Eingangs-/Ausgangsverhalten eine Rolle.

28.2 Zustandsrückführung

Für komplexere Regelungsaufgaben kommen Zustandsrückführungen zum Einsatz.

$$u = R_{u1}x + R_{u2}v$$

Sie ist die allgemeinere Rückführungsart der Beiden und umfasst somit auch die Ausgangsrückführung. Betrachtungen von Zustandsmodellen finden meist im Zeitbereich statt. Hauptanwendungen sind *exakte Linearisierungen* oder *Entkopplungen* die das System in mehrere Teilsysteme aufteilt, die dann unabhängig voneinander gesteuert oder geregelt werden können.

28.2.1 Statische Zustandsrückführung

Bei der statischen Zustandsrückführung ist charakteristisch, dass die Eingangsgröße u von den Zustandsgrößen und der neuen Eingangsgröße v abhängig ist.

$$u = r(x) + R(x)v$$

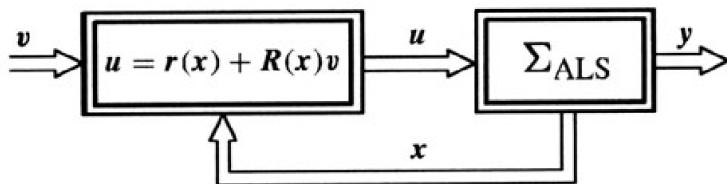


Abbildung 29: Statische Zustandsrückführung; Σ_{ALS} bedeutet *analytisches Modell mit linear eingehender Steuerung*

28.2.2 Quasi-statische Zustandsrückführung

Bei der quasi-statischen Rückführung ist u zusätzlich von den Ableitungen des neuen Eingangs v abhängig. Jedoch wird der Zustandsvektor nicht wie bei der dynamischen Rückführung erweitert, sondern behält seine Dimension.

28.2.3 Dynamische Zustandsrückführung

Bei der dynamischen Rückführung treten zusätzliche Differentialgleichungen in einer neuen internen Variable z auf. Ist die neue Eingangsgröße v linear, ergibt sich die Stellgröße zu

$$\dot{z} = r_z(x, z) + R_z(x, z)v$$

$$u = r_u(x, z) + R_u(x, z)v$$

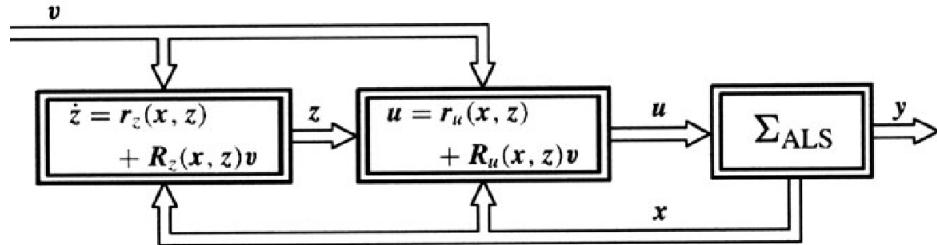


Abbildung 30: Dynamische Zustandsrückführung

29 Stellgrößenbeschränkung

29.1 Anti-Windup Maßnahmen bei Reglern mit I-Anteil

30 Regelungen für nichtlineare Strecken

30.1 Linearisierungsmethoden

30.1.1 Taylor-Linearisierung

30.1.2 Carleman-Linearisierung

30.1.3 Feedback-Linearisierung

30.1.4 Harmonische-Linearisierung

30.2 Reglerentwurf mittels exakter Eingangs-/Ausgangslinearisierung

30.2.1 Bei maximalem relativen Grad $\delta = n$ von SISO-Systemen

Grundidee ist es direkt und nicht über den Umweg einer linearen Näherung (wie vorherige Kapitel) einen Regler für das nichtlineare System zu entwerfen. Grundgedanke hierbei ist einen nichtlinearen Regler auszulegen der das nichtlineare Verhalten der Strecke kompensiert und so insgesamt ein linearer Regelkreis entsteht. Die exakte Linearisierung wird auch Eingangs-/Ausgangslinearisierung genannt und linearisiert demnach das Verhalten zwischen Eingangs- und Ausgangsgröße. Die interne Dynamik bleibt dabei nichtlinear. Voraussetzung ist ein eingangsaffines nichtlineares System das in nichtlinearer Regelungsnormalform vorliegt oder in eine solche überführt werden kann.

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

$$y = c(x)$$

In der *nichtlinearen Regelungsnormalform*

$$\Rightarrow \begin{bmatrix} \dot{z}_1 \\ \vdots \\ \dot{z}_{n-1} \\ \dot{z}_n \end{bmatrix} = \begin{bmatrix} z_2 \\ \vdots \\ z_n \\ \alpha(x) \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \beta(x) \end{bmatrix} u,$$

$$y = z_1$$

Mit dem System in nichtlinearer Regelungsnormalform, erhält man mit dem Regelgesetz

$$u = -\frac{\alpha(\mathbf{x}) + \sum_{i=1}^n a_{i-1} \cdot z_i}{\beta(\mathbf{x})} + \frac{V}{\beta(\mathbf{x})} w$$

den linearen geschlossenen Regelkreis in Regelungsnormalform

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \vdots \\ \dot{z}_{n-1} \\ \dot{z}_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{n-1} \\ z_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ V \end{bmatrix} w,$$

für den das Verhalten wie gewohnt durch Polvorgabe gewählt werden kann.

Erweiterung auf allgemeinere Systeme:

Für den gezeigten Fall muss das System in nichtlinearer Regelungsnormalform vorliegen. Da dies nicht oft so ist, muss eine Zustandskoordinatentransformation durchgeführt werden, die das System in die nichtlineare Regelungsnormalform transformiert. Für diesen Zweck werden die *Lie-Dervierte* oder *Lie-Ableitungen* verwendet. (Siehe Kapitel ??)

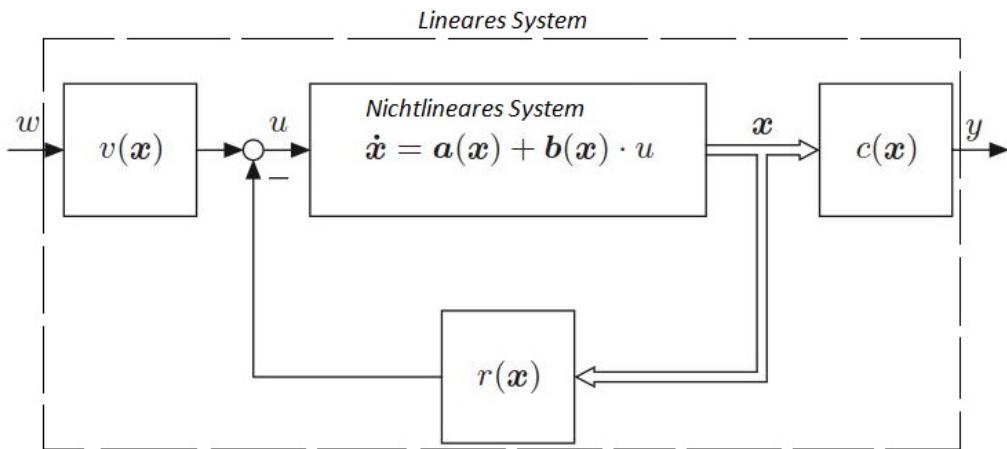


Abbildung 31: Strukturbild des Regelkreises mit insgesamt linearer Dynamik

Man nennt die beschriebene Vorgehensweise **exakte Linearisierung bei maximalem relativen Grad**. Mit exakter Linearisierung ist hierbei genau genommen die *exakte Eingangs-/Ausgangslinearisierung* gemeint.

Für die Differentialgleichung der Ausgangsgröße y ergibt sich folgendes lineares Verhalten aus obigem Zustandsraummodell:

$$y^{(n)} + a_{n-1}y^{(n-1)} + \dots + a_1\dot{y} + a_0y = V \cdot w.$$

Setzt man in obiges Regelgesetz α und β ein und ersetzt z mit

$$z = t(\mathbf{x}) = \begin{bmatrix} c(\mathbf{x}) \\ L_a c(\mathbf{x}) \\ L_a^2 c(\mathbf{x}) \\ \vdots \\ L_a^{(n-1)} c(\mathbf{x}) \end{bmatrix},$$

erhält man die *nichtlineare Ackermann-Formel*

Definition: nichtlineare Ackermann-Formel

$$u = -\frac{L_a^{(n)} c(\mathbf{x}) + a_{n-1} L_a^{(n-1)} c(\mathbf{x}) + \dots + a_1 L_a c(\mathbf{x}) + a_0 c(\mathbf{x})}{L_b L_a^{(n-1)} c(\mathbf{x})} + \frac{V}{L_b L_a^{(n-1)} c(\mathbf{x})} w$$

[NICHTLINEARE SYSTEME UND REGELUNGEN, JÜRGEN ADAMY S292]

30.2.2 Bei reduziertem relativen Grad $\delta < n$ von SISO-Systemen

Ist der relative Grad $\delta < n$, dann bleibt eine nichtlineare interne Dynamik übrig. Die Zustandsvariablen der internen Dynamik verhalten sich nichtlinear, wirken sich jedoch nicht auf die Ausgangsgröße aus. Voraussetzung für die Stabilität des Gesamtsystems ist jedoch auch die Stabilität der internen Dynamik.

30.3 Exakte Zustandslinearisierung

Auch **exakte Eingangs-Zustandslinearisierung** genannt.

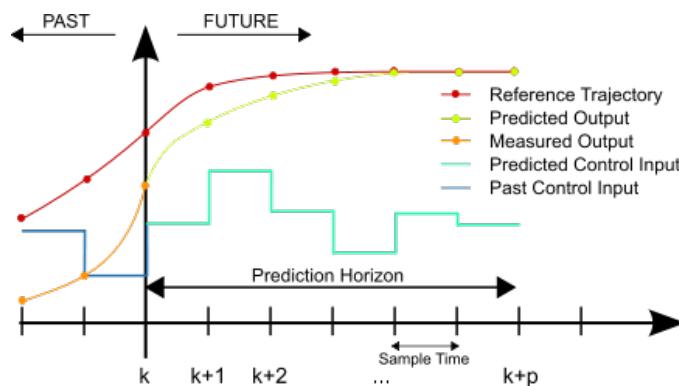
Bei der exakten Eingangs-/Ausgangslinearisierung ist das System exakt linearisierbar, wenn $\delta = n$ gilt. Ist der relative Grad jedoch kleiner als die Systemordnung, dann kann die Linearisierung nicht für alle Zustände durchgeführt werden. Es bleibt somit eine nichtlineare interne Dynamik. Bei der exakten Zustandslinearisierung wird ein *fiktiver* Ausgang y gesucht für den wieder $\delta = n$ gilt und die exakte Linearisierung somit durchführbar ist.

Systeme ohne Ausgang??

Brunovsky-Normalform [NICHTLINEARE SYSTEME UND REGELUNGEN, JÜRGEN ADAMY S337]

30.4 Nonlinear Model predictive control (NMPC)

Auf Basis eines zeitdiskreten dynamischen Systems wird der zukünftige Verlauf vorhergesagt. Dieser ist abhängig von den Eingangssignalen, die so gewählt werden, dass sich der vorhergesagte Verlauf einer gewünschten Kurve annähert. Die Berechnung der optimalen Eingangssignale erfolgt über eine Kostenfunktional. Durch dessen Minimierung erhält man den optimalen Ausgangsverlauf. Die Optimierung wird in jedem Zeitschritt erneut, mit neuen Messdaten durchgeführt → ressourcenintensiv.



Die nichtlineare Modellprädiktive Regelung bietet eine Möglichkeit für die Prozessmodellierung, wenn diese Prozesse nicht linearisiert werden sollen.

Vorteile & Nachteile:

- + keine Linearisierung nötig
- Zeitintensive Berechnung der Optimierungsfunktion im nichtlinearen Fall. Daher auch nur für langsame Regelstrecken (z.B. Prozessindustrie) einsetzbar.
- Optimierung des Kostenfunktional muss in jedem Schritt durchgeführt werden

30.5 Flachheitsbasierter Ansatz

Bei der Regelung von Robotern ist man mit der Aufgabe der Folgeregelung bzw. Bahnverfolgung konfrontiert. Der Endeffektor soll eine vorgegebene Bahn abfahren, während dieses Bewegungsvorgangs ändert sich das Systemverhalten kontinuierlich. Infolge des großen Arbeitsbereichs ist die Annahme eines linearisierten Modells nicht ausreichend, deshalb ist eine nichtlineare Modellbildung erforderlich. Für nichtlineare Modelle, die als flach charakterisiert werden können, vereinfacht sich die Trajektorienregelung wesentlich. Denn bei diesem Verfahren wird durch eine Rückführung die Dynamik des Folgefehlers linearisiert. Der Dynamik des geregelten Systems wird dagegen kein lineares Verhalten aufgeprägt. Daraus resultiert ein geringer Aufwand bei der Auslegung leistungsfähiger Regler, außerdem können eine Vielzahl technischer Anlagen als flache Systeme modelliert werden.

Nichtlineare flache Systeme können in geeigneten Koordinaten wie lineare Systeme in linearen Räumen dargestellt werden. Die neuen Koordinaten stehen mit den Originalkoordinaten in einem (formal eindeutig umkehrbaren) nichtlinearen Zusammenhang. Die so eingeführten nichtlinearen Koordinatentransformationen können als dynamische Rückführungen interpretiert werden.

Der flachheitsbasierte Entwurf einer nichtlinearen Folgeregelung geschieht in zwei Stufen: Zunächst erfolgt eine Planung der Trajektorie zur Beschreibung des Sollverhaltens. Dann wird die Folgebewegung entlang der Solltrajektorie durch eine Regelung stabilisiert.

In erster Linie kann durch den flachen Ausgang die Steuerung in einfacher Weise entworfen werden, indem die Eingangsgrößen anhand der Ausgangsgrößen berechnet werden. Die Ausgangsgröße ist der gewünschte Trajektorienverlauf, den man beliebig wählen kann.

$$u = \Psi_2(y, \dot{y}, \dots, y^{(\beta+1)})$$

Bei Einwirkung von Störgrößen weicht y vom gewünschten Verlauf y_{soll} ab. Daher ist eine Regelung Strecke nötig

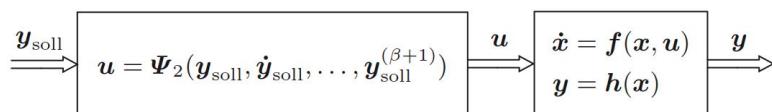


Abbildung 32: Steuerung eines flachen Systems

um den vorgegebenen Verlauf zu ermöglichen.

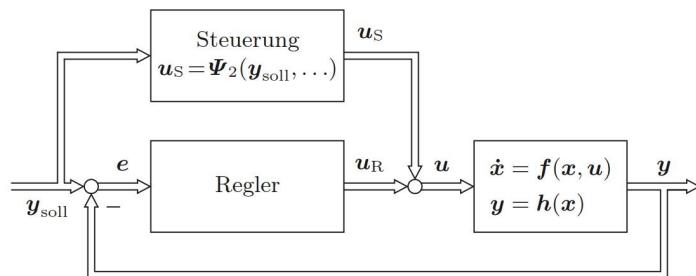


Abbildung 33: Flachheitsbasierte Steuerung und Regelung eines flachen Systems

30.6 Gain-scheduling-Regler

Ein nichtlineares System der Form ($d \dots$ einwirkende Störgröße)

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

kann in einem Arbeitspunkt linearisiert werden und für das linearisierte System kann dann ein Regler entworfen werden der das System lokal stabilisiert. Je weiter man sich jedoch von diesem Arbeitspunkt weg bewegt, desto schlechter funktioniert der Regler.

Wird ein nichtlineares System nicht nur um einen Arbeitspunkt betrieben, ist es sinnvoll Regler zu entwerfen die nicht nur lokale Gültigkeit haben.

Beim *Gain-scheduling-Regler* werden mehrere Arbeitspunkte des Systems linearisiert (parametrisierte Linearisierungsfamilie) und für jeden Arbeitspunkt wird mittels linearer Entwurfsmethoden ein entsprechender Regler entworfen. Dafür kommen verschiedene Reglertypen in Frage, z.B. lineare Zustandsregler, PID-Regler, usw.

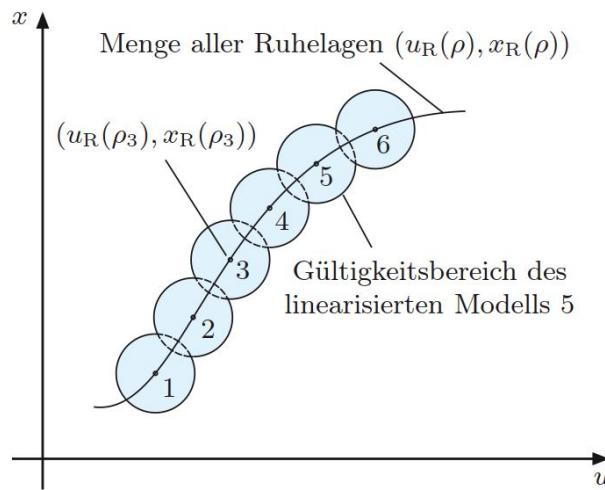


Abbildung 34: Geltungsbereiche der parametrisierten Linearisierungsfamilie

Zwischen den Reglern kann entweder hin und her geschaltet werden, was zu einer abrupten und unerwünschten Änderung der Stellgröße führt oder die Regelparameter werden zwischen den AP'en interpoliert. Die Interpolation kann beispielsweise auf Basis des gewichteten Mittelwertes

$$\mathbf{u} = \frac{\sum_{i=1}^p \mu_i \cdot (\overbrace{\mathbf{u}_R(\boldsymbol{\rho}_i) + \Delta \mathbf{u}_i}^{\mathbf{u}})}{\sum_{i=1}^p \mu_i}$$

erfolgen. Dabei ist μ_i die Gewichtungsfunktion (z.B. Gauß-Funktion), $\mathbf{u}_R(\boldsymbol{\rho}_i)$ die Stellgröße für den i ten Arbeitspunkt und $\Delta \mathbf{u}_i = \mathbf{u} - \mathbf{u}_R(\boldsymbol{\rho}_i)$.

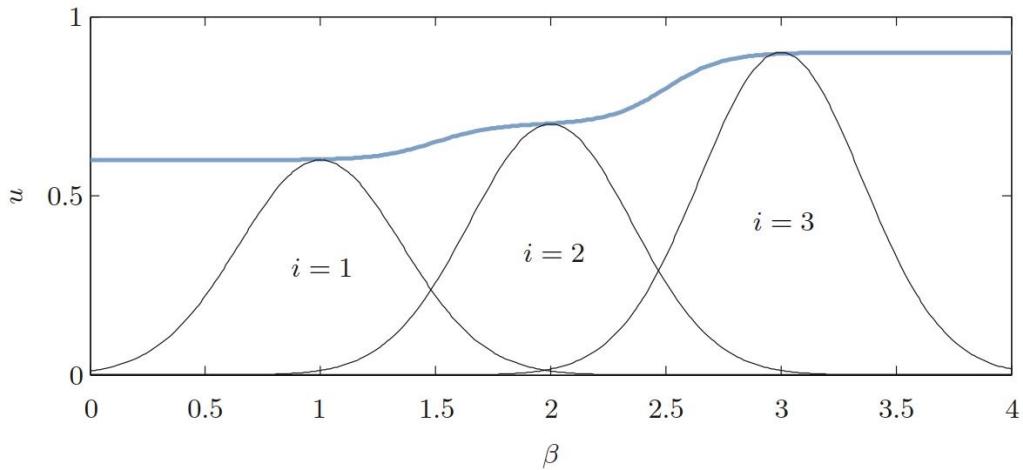


Abbildung 35: Die Interpolation \mathbf{u} der Reglerausgangswerte $\mathbf{u}_R(\boldsymbol{\rho}_i) + \Delta \mathbf{u}_i$ auf Basis des gewichteten Mittelwertes ist als blaue Kurve exemplarisch dargestellt. Die schwarzen Kurven entsprechen den Funktionen $\mu_i \cdot (\mathbf{u}_R(\boldsymbol{\rho}_i) + \Delta \mathbf{u}_i)$ für $i = 1, \dots, 3$.

Zusammenfassung:

Schritt 1: Berechne parametrierte Linearisierungsfamilie

Schritt 2: Entwerfe die linearen Teilregler mittels linearer Entwurfsmethoden.

Schritt 3: Lege das Auswahl- bzw. Interpolationsgesetz für die Teilregler in Abhängigkeit der Schedulingparameter fest.

Schritt 4: Simuliere den gesamten Regelkreis und prüfe Stabilität und Regelgüte.

Man beachte: Die Stabilität der einzelnen linearen Teilregelkreise reicht nicht aus, da das gesamte System nichtlinear ist. Das heißt, dass die Stabilität im Allgemeinen nicht analytisch gesichert werden kann, sondern simuliert werden muss.

Gain-scheduling Regler: [NICHTLINEARE SYSTEME UND REGELUNGEN, JÜRGEN ADAMY S273]

30.7 Control-Ljapunov-Funktionen

30.8 Backstepping-Verfahren

Das Backstepping-Verfahren ermöglicht es, Regler und Ljapunov-Funktionen für nichtlineare Regelstrecken in *strenger Rückkopplungsform* zu bestimmen.

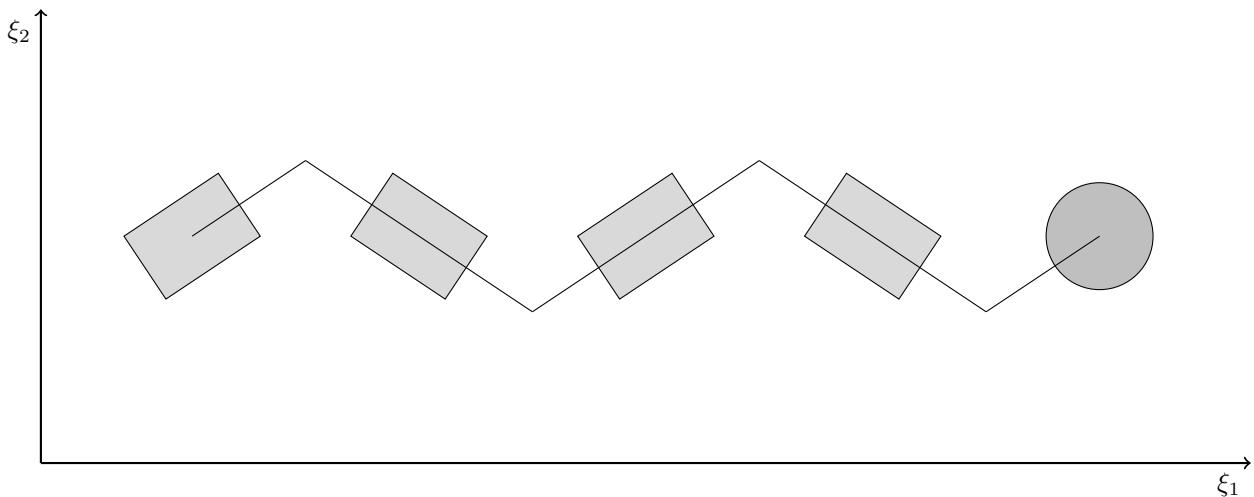
Definition: *System vom Typ strenger Rückkopplungsform (engl. strict feedback)*

$$\begin{aligned}\dot{x}_1 &= f_1(\mathbf{x}_1) + h_1(\mathbf{x}_1) \cdot x_2 \\ \dot{x}_2 &= f_2(\mathbf{x}_1, x_2) + h_2(\mathbf{x}_1, x_2) \cdot x_3 \\ &\vdots \\ \dot{x}_k &= f_k(\mathbf{x}_1, x_2, \dots, x_k) + h_k(\mathbf{x}_1, x_2, \dots, x_k) \cdot u\end{aligned}$$

mit $\mathbf{x}_1 \in \mathbb{R}^n$ und $x_2, \dots, x_k \in \mathbb{R}$.

30.9 Sliding-Mode-Regelung

- + Robustheit gegenüber Regelstreckenparameter-Änderungen, also Änderungen der Dynamikmatrix \mathbf{A} . Ebenfalls sehr Robust gegenüber von außen einwirkenden Störungen \mathbf{d} . Genauer wird das Regelgesetz so gewählt, dass die Dynamik des geschlossenen Regelkreises unabhängig von \mathbf{A} und \mathbf{d} ist.
- Durch Schalten des Stellgliedes mit sehr hoher Frequenz während des Gleitens zur Ruhelage folgt ein hoher Verschleiß.



31 Regelstrukturen

31.1 Kaskadenregelung

32 Vektorregelung

32.1 Adaptive Regelung

VI

Informatik

33 Matlab

VII

Berühmte Wissenschaftler

15'th century

33.1 Leonardo da Vinci 1452-1519

33.2 Nikolaus Kopernikus 1473-1543

16'th century

33.3 Galileo Galilei 1564-1642

33.4 Johannes Keppler 1571-1630

17'th century

33.5 Isaac Newton 1643-1727

18'th century

33.6 Carl Friedrich Gauß 1777-1855

33.7 Michael Faraday 1791-1867

19'th century

33.8 Charles Darwin 1809-1882

33.9 Gustav Robert Kirchhoff 1824-1887

33.10 James Clerk Maxwell 1831-1879

33.11 Thomas Alva Edison 1847-1931

Wer der Erfinder der Glühbirne ist, ist nach wie vor ein Rätsel. Anscheinend soll der deutsche Uhrenmacher Heinrich Göbel im Jahre 1854 die erste Glühlampe gebaut und erfunden haben. Jedoch verfügte dieser weder über das Know-How noch die technischen Geräte um ein derartiges Projekt fertigzustellen. 1879 hat Edison eine Kohlefaden-Lampe gebaut und darauf ein Patent erhalten. Wer aber die Idee dazu ist unbekannt.

- Thomas Edison meldete für über tausend Erfindungen Patente an
- Edison galt als skrupelloser Geschäftsmann
- Er versuchte den Gleichstrom bei der Elektrifizierung für seinen eigenen Profit durchzusetzen, gewonnen hat der Wechselstrom von Westinghouse. Er richtete sogar öffentliche Tiere durch Wechselstrom hin um zu zeigen, dass dieser gefährlich sei und sein Gleichstrom nicht. Als Höhepunkt brachte er einen Elefanten um.

33.12 Nikola Tesla 1856-1943

Ausbildung: Technische Universität Graz

Herkunft: Serbien

In Österreich bis 1891 gelebt, danach bis zum Tod in Amerika.

Erfindungen:

- Drehstrom-Asynchronmaschine
- rotierendes magnetisches Feld
- Tesla-Spule (el. Energie durch Luft übertragen: hochspannung, niedriger Strom, hohe Frequenz)
- erstes ferngesteuertes Boot

33.13 Heinrich Hertz 1857-1894

33.14 Max Planck 1858-1947

33.15 Albert Einstein 1879-1955

33.16 Niels Bohr 1885-1962

20'th century

33.17 Werner Heisenberg 1901-1976

33.18 Stephen Hawking 1942-heute