

Mathematik C

Vorlesung an der
Universität des Saarlandes

SS 2018

Klaus Schindler

Version 18.1
©K. Schindler 2018

Inhaltsverzeichnis

Inhaltsverzeichnis	2
1. Differenzengleichungen	5
1.1. Differenzen	5
Differenzenoperator	5
1.2. Differenzengleichungen	10
Beispiele	10
Implizite und explizite Differenzengleichungen	14
Lösung von Differenzengleichungen	15
Lineare Differenzengleichungen	17
2. Differentialgleichungen	37
Autonome Differentialgleichungen	38
Differentialgleichungen und Differentialgleichungssysteme	40
Existenz- und Eindeutigkeitssätze	41
2.1. Spezielle Typen von DGLen erster Ordnung	44
2.1.1. DGLen mit getrennten Variablen	44
2.1.2. DGLen mit homogenen Funktionen	45
2.1.3. Gebrochen rationale DGLen	45
2.1.4. Exakte DGLen	46
2.1.5. Bernoullische Differentialgleichungen	49
2.1.6. Riccatische Differentialgleichungen	50
2.2. Lineare DGLen	50
2.3. Laplace-Transformation	66
3. Stabilität dynamischer Systeme	75
Eigenwerte und Stabilität	79
3.1. Ljapunov-Funktionen	87
3.2. Richtungsfeld, Phasendiagramm	92
4. Dynamische Optimierung	111
4.1. Variationsrechnung	122
4.1.1. Nebenbedingungen	136
4.1.2. Bedingungen zweiter Ordnung	142
4.2. Dynamische Programmierung	146
Die Gleichung von Bellman	148

4.3. Optimale Kontrolltheorie	154
Die Hamilton-Funktion	155
Das Maximumprinzip	157
Literaturverzeichnis	167
Anhang	168
A. Differentialrechnung	169
B. Banachscher Fixpunktsatz	179
C. Komplexe Zahlen	181
C.1. Eigenschaften komplexer Zahlen	184
C.2. Darstellung komplexer Zahlen	185
D. Lineare Gleichungssysteme, Eigenwerte, Eigenvektoren	189
E. Quadratische Formen, Definitheit	195
F. Statische Optimierung	199
Index	205

1.1. Differenzen

Die Dynamik ökonomischer und naturwissenschaftlicher Größen wird durch ihr Änderungsverhalten beschrieben. Hierzu wurde der Begriff der Ableitung bzw. des Differentials einer Funktion y als Grenzwert des Differenzenquotienten eingeführt, d.h.¹

$$y'(t) = \lim_{\Delta t \rightarrow 0} \frac{y(t+\Delta t) - y(t)}{\Delta t} .$$

Für infinitesimale (marginale) Änderungen dt kann damit $dy = y(t+dt) - y(t)$, d.h. die resultierende Änderung des Funktionswertes bestimmt werden. Näherungsweise gilt

$$\Delta y = y'(t) \cdot \Delta t .$$

Nachteilig am Differentialbegriff ist, dass er nicht verwendet werden kann, wenn die zu untersuchende Funktion $y(t)$ einen diskreten Definitionsbereich (z.B. $\mathbb{N}$) besitzt, d.h. wenn eine Folge vorliegt und dadurch der Grenzübergang $\Delta t \rightarrow 0$ nicht möglich ist. Ökonomische Beispiele für solche Funktionen, die Werte nur zu bestimmten Zeitpunkten liefern, sind etwa Konsum, Investition, Einkommen oder Aktienkurse. Da diese Größen in der Regel nur in zeitlich konstanten Abständen gemessen werden, liegt es nahe, diese Zeitperiode auf 1 zu normieren, d.h. $\Delta t=1$ zu setzen, eine Vereinbarung, die man auch in der Finanzmathematik für eine Zinsperiode trifft. Das Änderungsverhalten von $y(t)$ wird dann nicht mehr durch den Grenzwert des Differenzenquotienten, sondern direkt durch den Differenzenquotienten gemessen. Dieser reduziert sich wegen $\Delta t=1$ auf den Wert

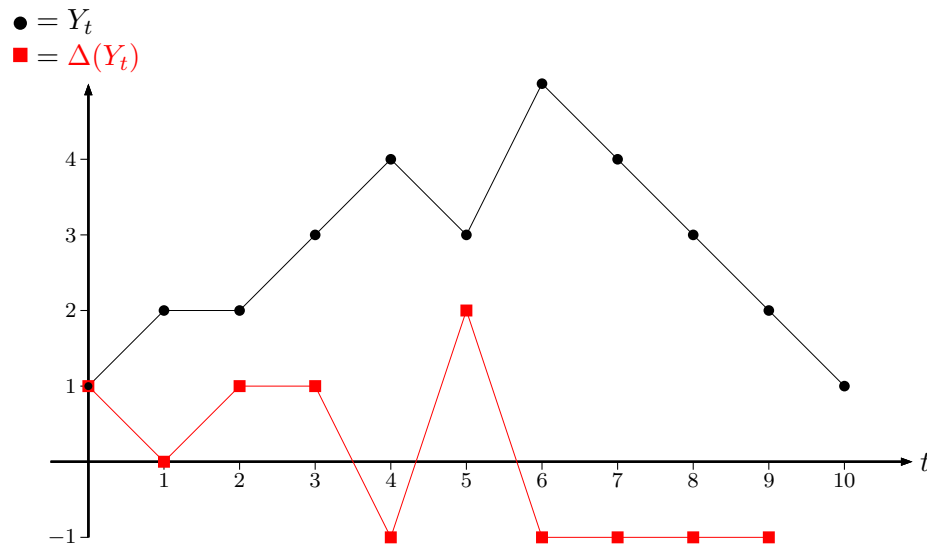
$$\frac{y(t+\Delta t) - y(t)}{\Delta t} = y(t+1) - y(t) ,$$

die wir zukünftig als Differenz von y zum Zeitpunkt t bezeichnen und mit $(\Delta y)(t)$ oder $(\Delta y)_t$ notieren. Δy tritt bei Folgen damit an die Stelle der 1. Ableitung.

Abbildung 1.1 zeigt die Größe Y_t zu den Zeitpunkten $t = 0, 1, \dots, 10$ und die zugehörigen Veränderungen (*Differenzen*).

Wir wollen zunächst den Begriff der „Differenz“ genau definieren.

¹Die Verwendung der Variablen t soll andeuten, dass die Dynamik i.d.R. zeitgetrieben ist.

Abbildung 1.1.: Werte Y_t und Differenzen $(\Delta Y)_t$ einer Folge $(Y_t)_{t=0}^{\infty}$ **1.1 Definition**

Sei $\mathbf{Y} : \mathbb{D} \rightarrow \mathbb{R}^m$ eine Funktion mit $\mathbb{D} \subset \mathbb{R}$.

- a) $\Delta \mathbf{Y}$, die 1. Differenz von $\mathbf{Y}$ in t mit Schrittweite h , ist definiert durch

$$(\Delta \mathbf{Y})_t := \mathbf{Y}_{t+h} - \mathbf{Y}_t.$$

Δ heißt Differenzenoperator.

- b) $F\mathbf{Y}$, die 1. Verschiebung von $\mathbf{Y}$ in t mit Schrittweite h , ist definiert durch

$$(F\mathbf{Y})_t := \mathbf{Y}_{t+h}.$$

F heißt Forwardoperator.

- c) Die n -te Differenz von $\mathbf{Y}$ in t mit Schrittweite h bezeichnet man mit $\Delta^n \mathbf{Y}$. Induktiv definiert man sie als erste Differenz der $(n-1)$ -ten Differenz, d.h. durch

$$(\Delta^n \mathbf{Y})_t := \Delta(\Delta^{n-1} \mathbf{Y})_t = (\Delta^{n-1} \mathbf{Y})_{t+h} - (\Delta^{n-1} \mathbf{Y})_t.$$

Analog bezeichnet $F^n \mathbf{Y}$ die n -te Verschiebung von $\mathbf{Y}$ in t mit Schrittweite h , d.h.²

$$(F^n \mathbf{Y})_t := F(F^{n-1} \mathbf{Y})_t = \mathbf{Y}(t+n \cdot h).$$

□

²Wie üblich trifft man die Konventionen $F^0 \mathbf{Y} := \mathbf{Y}$ bzw. $\Delta^0 \mathbf{Y} := \mathbf{Y}$.

1.2 Beispiel

i) $\mathbb{D} = \mathbb{Z}$ und $Y_t := 2t^2 - 3t + 2$, $h \in \mathbb{Z}$ konstant. Dann gilt

$$\begin{aligned}(\Delta Y)_t &= Y_{t+h} - Y_t = 4th - 3h + 2h^2 \\(\Delta^2 Y)_t &= (\Delta Y)_{t+h} - (\Delta Y)_t = 4h^2 \\(\Delta^3 Y)_t &= 0\end{aligned}$$

ii) $\mathbb{D} = \mathbb{N}_0$, mit $Y_t := t^2$, $h = 1$. Dann gilt

$$\begin{aligned}(\Delta Y)_t &= Y_{t+1} - Y_t = (t+1)^2 - t^2 = 2t + 1 \\(\Delta^2 Y)_t &= (\Delta Y)_{t+1} - (\Delta Y)_t = 2 \\(\Delta^3 Y)_t &= 0\end{aligned}$$

□

1.3 Bemerkung

i) h wird als *Differenzenintervall* bezeichnet. Ist h aus dem Kontext nicht ersichtlich, sollte die Bezeichnung Δ_h verwendet werden. Δ^n kann genau dann gebildet werden, wenn die Punkte $t, t+h, \dots, t+n \cdot h$ im Definitionsbereich $\mathbb{D}$ von Y liegen.

ii) Die Operatoren Δ und F sind linear, d.h. es gilt für $\mathbf{X}, \mathbf{Y} : \mathbb{D} \rightarrow \mathbb{R}^m$ und $\alpha, \beta \in \mathbb{R}$:

$$\begin{aligned}\Delta(\alpha \cdot \mathbf{X} + \beta \cdot \mathbf{Y}) &= \alpha \cdot \Delta \mathbf{X} + \beta \cdot \Delta \mathbf{Y} \\F(\alpha \cdot \mathbf{X} + \beta \cdot \mathbf{Y}) &= \alpha \cdot F \mathbf{X} + \beta \cdot F \mathbf{Y}\end{aligned}$$

iii) Der Identitätsoperator $\mathbb{1}$ mit $\mathbb{1}(\mathbf{Y}) := \mathbf{Y}$ liefert $\Delta = F - \mathbb{1}$ bzw. $F = \Delta + \mathbb{1}$.

iv) Ist Y differenzierbar, gilt

$$Y'(t) = \lim_{h \rightarrow 0} \frac{Y(t+h) - Y(t)}{h} = \lim_{h \rightarrow 0} \frac{(\Delta_h Y)_t}{h},$$

woraus sich für kleine h

$$Y'(t) \approx \frac{(\Delta_h Y)_t}{h}$$

ergibt. In diesem Sinne kann im Fall $h=1$ die Differenz ΔY als diskrete Ableitung interpretiert werden. Dies erklärt auch, warum der Differentialoperator D und der Differenzenoperator Δ ähnliche Eigenschaften haben.

v) Der inverse Forwardoperator F^{-1} wird auch als *Backshift-* oder *Lagoperator* L bezeichnet. Dieser erzeugt die vorangehenden Elemente, d.h.

$$L^n(\mathbf{Y})_t = \mathbf{Y}_{t-n \cdot h}$$

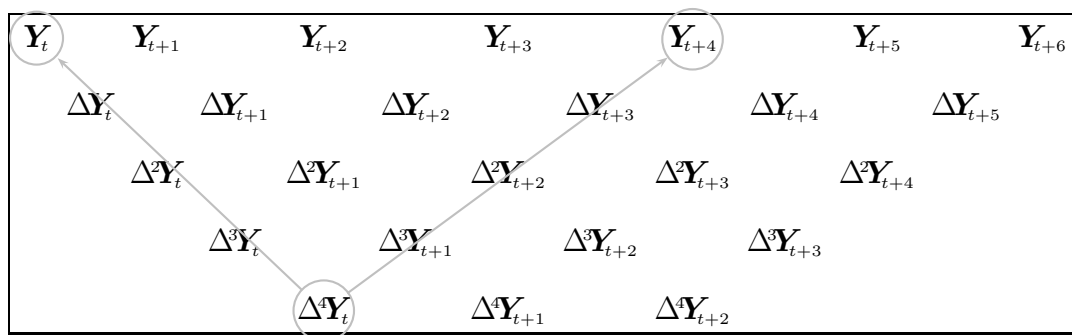
- vi) **Wir beschränken uns im Folgenden auf den Fall $\mathbb{D}=\mathbb{N}_0$ und $h=1$** , da durch $t+n \cdot h \mapsto n$ eine Bijektion zwischen den natürlichen Zahlen $\mathbb{N}_0=\{0, 1, 2, \dots\}$ und der Menge $\{t, t+h, t+2 \cdot h, \dots\}$ definiert wird. Aus diesem Grund verwendet man auch die Indexschreibweise $(\Delta^n Y)_t$ statt $(\Delta^n Y)(t)$. $\square$

Der folgende Satz liefert in Teil a) die Möglichkeit, Differenzen höherer Ordnung direkt mit Hilfe der Ausgangsfolge zu berechnen, d.h. ohne den durch die rekursive Definition bedingten Umweg über Differenzen niedrigerer Ordnung. Teil b) ist das diskrete Analogon zum Satz von Taylor. Er gestattet die Berechnung der Ausgangsfolge aus den Differenzen.

1.4 Satz

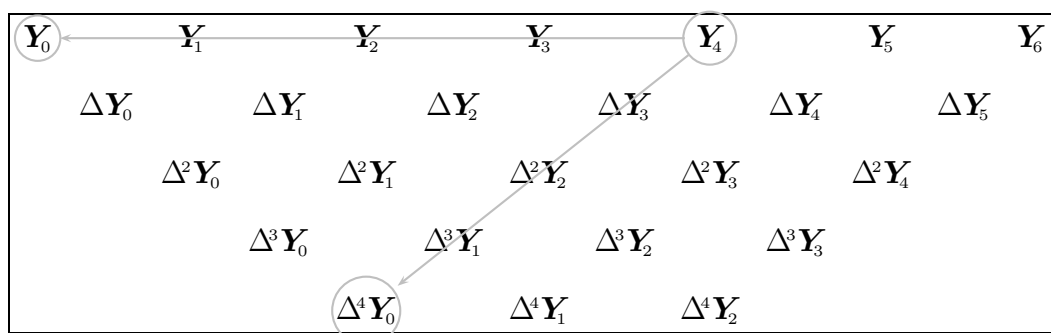
Sei $Y : \mathbb{N}_0 \rightarrow \mathbb{R}^m$ eine Abbildung. Dann gilt für alle k und s aus $\mathbb{N}_0$

$$a) \quad \Delta^{k+s} Y_t = \sum_{i=0}^k \binom{k}{i} (-1)^{k-i} \Delta^s Y_{t+i}$$



Für $s=0$ erhält man die explizite Darstellung der Differenzen $\Delta^k Y_t$ mit Hilfe der Funktionswerte $Y_t, \dots, Y_{t+k}$ (siehe Graphik für $k=4$).

$$b) \quad Y_{t+k+s} = \sum_{i=0}^k \binom{k}{i} \Delta^i Y_{t+s}$$



Für $s=t=0$ erhält man eine explizite Darstellung aller Funktionswerte Y_k mit Hilfe der Differenzen $\Delta^0 Y_0, \dots, \Delta^k Y_0$ (siehe Graphik für $k=4$). $\square$

Beweis:

- a) Wegen der Linearität von Δ kann o.B.d.A. $s=0$ angenommen werden. Mit dem binomischen Lehrsatz³

$$(F - \mathbb{1})^k = \sum_{i=0}^k \binom{k}{i} F^i (-\mathbb{1})^{k-i}$$

ergibt sich dann

$$\Delta^k \mathbf{Y}_t \stackrel{\Delta=F-\mathbb{1}}{=} (F - \mathbb{1})^k \mathbf{Y}_t = \sum_{i=0}^k \binom{k}{i} (-1)^{k-i} F^i \mathbf{Y}_t = \sum_{i=0}^k \binom{k}{i} (-1)^{k-i} \mathbf{Y}_{t+i}$$

- b) Der Beweis verläuft analog zu Teil a), denn es gilt

$$\mathbf{Y}_{t+k} = F^k(\mathbf{Y})_t = (\Delta + \mathbb{1})^k \mathbf{Y}_t$$

■

1.5 Beispiel

- i) Betrachtet man $\Delta^3 Y_t + 2\Delta Y_t + Y_t$, so erhält man mit Satz 1.4 a)

$$\begin{aligned} \Delta^3 Y_t &= Y_{t+3} - 3Y_{t+2} + 3Y_{t+1} - Y_t \\ \Delta Y_t &= Y_{t+1} - Y_t \end{aligned}$$

und es folgt durch Einsetzen

$$\Delta^3 Y_t + 2\Delta Y_t + Y_t = Y_{t+3} - 3Y_{t+2} + 5Y_{t+1} - 2Y_t$$

Ist umgekehrt $Y_{t+3} - 3Y_{t+2} + 5Y_{t+1} - 2Y_t$ vorgegeben, liefert Teil b) von Satz 1.4

$$\begin{aligned} Y_{t+3} &= \Delta^3 Y_t + 3\Delta^2 Y_t + 3\Delta Y_t + Y_t \\ Y_{t+2} &= \Delta^2 Y_t + 2\Delta Y_t + Y_t \\ Y_{t+1} &= \Delta Y_t + Y_t \end{aligned}$$

und wie zu erwarten folgt durch Einsetzen

$$Y_{t+3} - 3Y_{t+2} + 5Y_{t+1} - 2Y_t = \Delta^3 Y_t + 2\Delta Y_t + Y_t .$$

- ii) Die rekursiv definierte Fibonacci-Folge

$$Y_{t+2} = Y_{t+1} + Y_t$$

kann äquivalent definiert werden durch

$$\Delta^2 Y_t + \Delta Y_t - Y_t = 0 .$$

□

³Wesentlich ist dabei, dass die Operatoren F und $\mathbb{1}$ kommutieren!

1.2. Differenzengleichungen

Zwar sind die Differential- und Differenzenrechnung von großem Nutzen bei der genauen Untersuchung von Funktionen bzw. Folgen, jedoch ist bei den meisten dynamischen Systemen die explizite Funktionsvorschrift der interessierenden Größen unbekannt. Bekannt ist i.d.R. nur das direkt zu beobachtende momentane Änderungsverhalten der Größen und deren Zusammenhang. Die Frage ist, ob aus diesem lokalen Änderungsverhalten auf die expliziten globalen Funktionsvorschriften geschlossen werden kann. Eine Gleichung, die den Zusammenhang zwischen den interessierenden Größen und deren Änderungsverhalten beschreibt, wird als „*Differenzengleichung*“ (ΔGL) oder „*Differentialgleichung*“ (DGL) bezeichnet, je nachdem, ob die Änderungen mittels Differenzen oder Differentialen (=Ableitungen) beschrieben werden. Diese Gleichungen liefern implizite Beschreibungen, da sie nur das momentane Verhalten des betrachteten Systems angeben.

Eine der bekanntesten Gleichungen dieses Typs ist die Newtonsche DGL „Kraft = Masse mal Beschleunigung“

$$F = m \cdot y'' .$$

Hier wird bei gegebener Kraft F , die auf ein Objekt der Masse m wirkt, der zeitabhängige Ort $y(t)$ des betrachteten Körpers implizit über die zweite Ableitung (=Beschleunigung) y'' des Körpers beschrieben. Die Differenzengleichung

$$(\Delta K)_t = i \cdot K \cdot \Delta t \tag{1.1}$$

ist ein entsprechendes Beispiel aus dem Bereich der Finanzmathematik. Gleichung (1.1) beschreibt die momentane Wertänderung eines Kapitals in Abhängigkeit der Zeit t .

In der ökonomischen Theorie kommen ΔGL vor allem zum Einsatz, um die zeitliche Entwicklung ökonomischer Größen zu analysieren. Konjunktur und Wachstum werden häufig in Form von ΔGL en modelliert. Hierbei geht man davon aus, dass z.B. das Bruttoinlandsprodukt sich auf einem bestimmten Pfad hin zu einem langfristigen Gleichgewicht entwickelt, in dem alle Kapazitäten ausgelastet sind.

In der Zeitreihenanalyse lassen sich mit ΔGL en die Werte einer Zeitreihe, die rekursiv zusammenhängen, berechnen.

Vor einer genauen Definition sollen zwei ausführlichere ökonomische Beispiele den Begriff der ΔGL erläutern. Sie zeigen, dass man in natürlicher Weise relativ schnell und einfach im Bereich der Ökonomie zur Beschreibung mittels ΔGL en gelangt.

1.6 Beispiel

- i) Das grundlegende dynamische Verhalten in der Ökonomie ist das Verhalten von Angebot S (*Supply*) und Nachfrage D (*Demand*) unter Konkurrenz. Es bezeichne $S(P_t)$ bzw. $D(P_t)$ das Marktangebot bzw. die Marktnachfrage bei einem Preis P_t zum Zeitpunkt t . Wie üblich nehmen wir an, dass Nachfrage und Angebot auf Preisveränderungen

reagieren und dass S bzw. D in Abhängigkeit von P monoton wachsend bzw. fallend sind. Wegen der Zeitverzögerung in der Produktion gehen wir außerdem davon aus, dass die Angebotsreaktion Q auf die veränderte Nachfrage erst eine Periode später einsetzt, d.h.

$$Q_{t+1} = S(P_t) \quad \text{und} \quad Q_t = D(P_t) .$$

Die Bedingung „Angebot = Nachfrage“ führt dann zu folgender Dynamik:

$$D(P_{t+1}) = Q_{t+1} = S(P_t)$$

Nimmt man vereinfachend an, dass D und S affin linear (Geraden) sind, d.h.

$$D(P) = a + b \cdot P \quad \text{und} \quad S(P) = c + d \cdot P \quad (b < 0, d > 0),$$

so erhält man die Dynamik

$$a + b \cdot P_{t+1} = c + d \cdot P_t \iff P_{t+1} = \frac{c-a}{b} + \frac{d}{b} \cdot P_t \iff \Delta P_t = \frac{c-a}{b} + \left(\frac{d}{b} - 1\right) \cdot P_t .$$

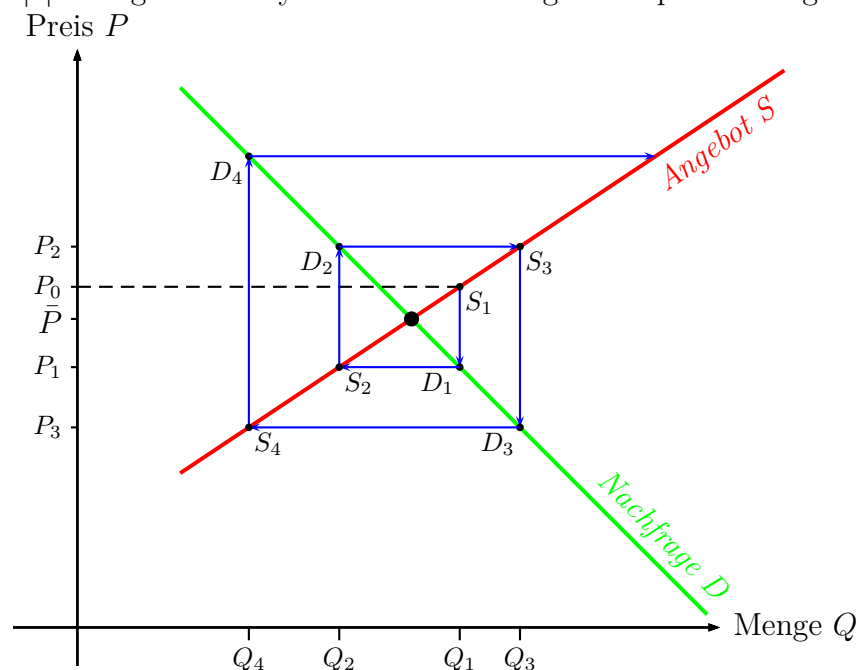
Wie wir später noch sehen werden, lautet die allgemeine Lösung mit der abkürzenden Bezeichnung $\bar{P} := \frac{c-a}{b-d}$:

$$P_t = (P_0 - \bar{P}) \cdot \left(\frac{d}{b}\right)^t + \bar{P} .$$

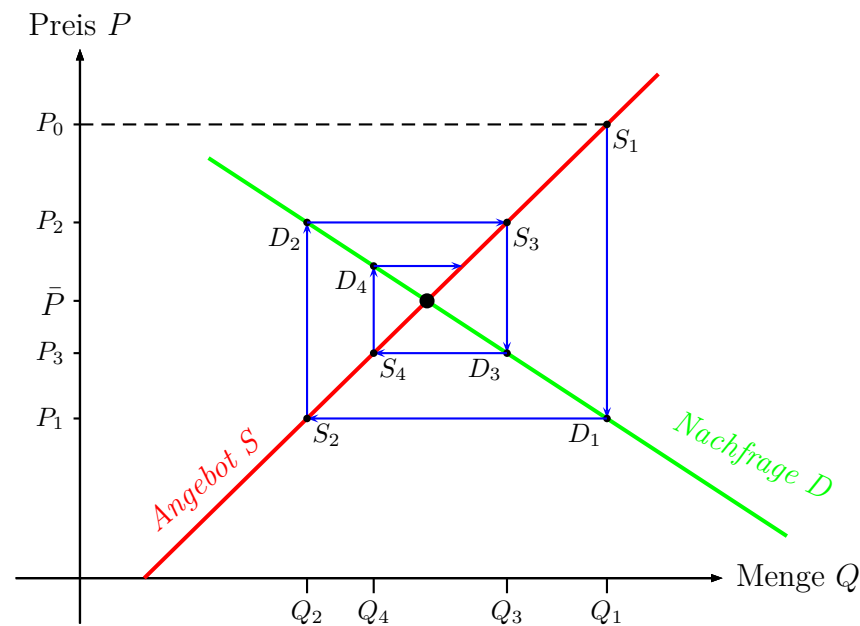
$\bar{P}$ ist der statische Gleichgewichtspreis, denn er entspricht der Lösung des statischen d.h. zeitunabhängigen Angebots- und Nachfragemodells

$$D(P) = a + b \cdot P, \quad S(P) = c + d \cdot P, \quad D = S .$$

Im Fall $d > |b|$ divergiert die Dynamik vom Gleichgewichtspreis $\bar{P}$ weg.



Im Fall $d < |b|$ wird die Preisdynamik graphisch durch eine „Rechteckspirale“ beschrieben, bei der der Preis gegen $\bar{P}$ konvergiert.



Die resultierenden Graphiken erklären auch die in elementarökonomischen Texten häufig anzutreffenden Begriffe *Spinnweb-Modell* bzw. *Cobweb-Diagramm*⁴.

ii) Bezeichnet man in Periode t ($t \in \mathbb{N}_0$) mit

Y_t das Volkseinkommen

S_t die (Netto-)Sparsumme

I_t die geplante (Netto-)Investition,

so geht das *Wachstumsmodell von Harrod* für das Volkseinkommen Y von folgenden Annahmen aus:

1.) Die Sparrate s ist in jeder Periode konstant, d.h.⁵

$$S_t = s \cdot Y_t. \quad (1.2)$$

2.) Es liegt ein konstanter Akzelerator $g > 0$ vor, d.h.

$$I_t = g \cdot (Y_t - Y_{t-1}). \quad (1.3)$$

Im Gleichgewicht muss die gewünschte Investition I_t gleich der tatsächlichen Investition S_t sein, d.h. es gilt

$$I_t = S_t \quad (1.4)$$

⁴Die bekannteste empirische Bestätigung für das Spinnweb-Modell ist der Schweinezyklus.

⁵Eine übliche Modifikation arbeitet mit einem time-lag: $S_t = s \cdot Y_{t-1}$

Einsetzen von Gleichungen (1.2) und (1.3) in Gleichung (1.4) liefert

$$g \cdot (Y_t - Y_{t-1}) = s \cdot Y_t \quad (t \geq 1)$$

oder äquivalent

$$g \cdot (Y_{t+1} - Y_t) = s \cdot Y_{t+1} \quad (t \geq 0)$$

Wegen $Y_{t+1} = \Delta Y_t + Y_t$ folgt die äquivalente ΔGL

$$(g-s) \cdot \Delta Y_t - s \cdot Y_t = 0 \quad (t \geq 0) .$$

□

1.7 Definition

Sei $\mathcal{F} : \mathbb{N}_0 \times \mathbb{R}^{m+1} \rightarrow \mathbb{R}$ eine Funktion. Eine Gleichung der Form

$$\mathcal{F}(t, Y_t, \Delta Y_t, \dots, \Delta^m Y_t) = 0 , \tag{1.5}$$

die eine Beziehung zwischen einer unbekannten Funktion $y : \mathbb{N}_0 \rightarrow \mathbb{R}$ und ihren Differenzen $\Delta Y_t, \dots, \Delta^m Y_t$ herstellt, heißt *Differenzengleichung der Ordnung m*.

Eine Funktion $\varphi : \mathbb{N}_0 \rightarrow \mathbb{R}$ heißt *Lösung der ΔGL* , wenn $\varphi_t, \Delta \varphi_t, \dots, \Delta^m \varphi_t$ die ΔGL für alle $t \in \mathbb{N}_0$ erfüllen, d.h. wenn gilt

$$\forall t \in \mathbb{N}_0 : \mathcal{F}(t, \varphi_t, \Delta \varphi_t, \dots, \Delta^m \varphi_t) = 0 .$$

□

1.8 Bemerkung

Kann die ΔGL (1.5) nach der höchsten auftretenden Differenz aufgelöst werden⁶, d.h. existiert eine geeignete Funktion f , so dass Gleichung (1.5) in der Form

$$\Delta^m Y_t = f(t, Y_t, \Delta Y_t, \dots, \Delta^{m-1} Y_t) ,$$

darstellbar ist, spricht man von einer *expliziten*, andernfalls von einer *impliziten* ΔGL . □

1.9 Beispiel

Die Gleichung

$$[\Delta^2 Y(t)]^3 + [\Delta Y(t)]^4 + [Y(t)]^5 - t = 0$$

ist eine ΔGL der Ordnung 2. Die größte Potenz, in der $\Delta^2 Y$ auftritt, wird als Grad der ΔGL bezeichnet. Hier liegt also eine ΔGL der Ordnung 2, vom Grad 3 vor. Diese ΔGL ist äquivalent zur expliziten ΔGL

$$\Delta^2 Y(t) = \sqrt[3]{t - [\Delta Y(t)]^4 - [Y(t)]^5}$$

□

⁶Diese Eigenschaft besagt, dass $\mathcal{F}$ nach der letzten Variablen auflösbar ist. Mit Hilfe des impliziten Funktionensatzes kann diese Auflösbarkeit zumindest lokal in kleinen Umgebungen erreicht werden.

1.10 Bemerkung

Satz 1.4 gestattet, jede ΔGL $\mathcal{F}(t, Y_t, \Delta Y_t, \dots, \Delta^m Y_t) = 0$ in die rekursive Form

$$\widetilde{\mathcal{F}}(t, Y_{t+m}, Y_{t+m-1}, \dots, Y_t) = 0$$

umzuwandeln (und umgekehrt). Daher gibt die Ordnung m einer ΔGL an, dass der Wert der Größe Y zum Zeitpunkt t durch die m vorangehenden Werte $Y_{t-1}, Y_{t-2}, \dots, Y_{t-m}$ von Y beeinflusst wird. Durch eine ΔGL wird daher die Folge $(Y_t)_t$ rekursiv definiert, d.h., dass jedes Folgenglied eine Funktion der vorangehenden Folgenglieder ist. Aus diesem Grund wird eine ΔGL auch als *Rekursionsgleichung* bezeichnet. Für die ΔGL

$$\Delta^3 Y_t + 2\Delta Y_t + Y_t = 0$$

aus Beispiel 1.5 i) lautet die rekursive Form

$$Y_{t+3} - 3Y_{t+2} + 5Y_{t+1} - 2Y_t = 0.$$

Mit dem Forwardoperator F ergibt sich für die ΔGL eine übersichtlichere Darstellung:

$$(F^3 - 3F^2 + 5F - 2)Y_t = 0$$

□

1.11 Definition

Sei $\mathcal{F} : \mathbb{D} \times (\mathbb{R}^n)^{m+1} \rightarrow \mathbb{R}^n$ eine $\mathbb{R}^n$ -wertige Abbildung ($\mathbb{D} \subset \mathbb{R}$). Eine Gleichung

$$\mathcal{F}(t, \mathbf{Y}, \Delta \mathbf{Y}, \dots, \Delta^m \mathbf{Y}) = \mathbf{0}, \quad (1.6)$$

die eine Beziehung zwischen den Werten einer vektorwertigen Abbildung $\mathbf{Y} : \mathbb{D} \rightarrow \mathbb{R}^n$, den Argumenten $t \in \mathbb{D}$ und einer oder mehreren Differenzen $\Delta \mathbf{Y}(t), \dots, \Delta^m \mathbf{Y}(t)$ herstellt, heißt *implizites Differenzengleichungssystem der Ordnung m* über der Menge $\mathbb{D}$. Jede Abbildung $\Phi : \mathbb{D} \rightarrow \mathbb{R}^n$, für die die Differenzen $\Delta \Phi(t), \dots, \Delta^m \Phi(t)$ existieren und die das ΔGLs -System (1.6) für alle $t \in \mathbb{D}$ erfüllt, d.h. mit

$$\mathcal{F}(t, \Phi(t), \Delta \Phi(t), \dots, \Delta^m \Phi(t)) = \mathbf{0},$$

heißt eine *Lösung des Differenzengleichungssystems*.

Existiert eine Funktion $\mathbf{f} : \mathbb{D} \times (\mathbb{R}^n)^m \rightarrow \mathbb{R}^n$, so dass (1.6) äquivalent ist zu

$$\Delta^m \mathbf{Y}(t) = \mathbf{f}(t, \mathbf{Y}(t), \Delta \mathbf{Y}(t), \dots, \Delta^{m-1} \mathbf{Y}(t)),$$

nennt man (1.6) ein *explizites Differenzengleichungssystem der Ordnung m* . □

1.12 Beispiel

Das folgende Gleichungssystem ist ein Beispiel für ein explizites Differenzengleichungssystem der Ordnung 1 bestehend aus $n=3$ Gleichungen.

$$\begin{aligned} \Delta \phi_1(t) &= t\phi_2(t) + \phi_3(t) - 2 \\ \Delta \phi_2(t) &= \sin(t)\phi_1(t) + \phi_3(t) + 5 \\ \Delta \phi_3(t) &= \ln(1+t)\phi_1(t) + \phi_2(t) + \phi_3(t) + t^2 \end{aligned}$$

In Vektorschreibweise lautet es mit $\Phi(t) := (\phi_1(t), \phi_2(t), \phi_3(t))^T$

$$\Delta \Phi(t) = \begin{pmatrix} 0 & t & 1 \\ \sin(t) & 0 & 1 \\ \ln(1+t) & 1 & 1 \end{pmatrix} \cdot \Phi(t) + \begin{pmatrix} -2 \\ 5 \\ t^2 \end{pmatrix} \quad \square$$

1.13 Lemma

Jede explizite Δ GL m -ter Ordnung lässt sich in ein äquivalentes Δ GLs-System erster Ordnung überführen. Genauer gilt: Die Lösungen der Δ GL

$$\Delta^m Y = f(t, Y, \Delta Y, \dots, \Delta^{m-1} Y) \quad (1.7)$$

entsprechen bijektiv die Lösungen des folgenden Δ GLs-Systems erster Ordnung

$$\begin{aligned} \Delta \phi_1 &= \phi_2 \\ \Delta \phi_2 &= \phi_3 \\ &\vdots \\ \Delta \phi_{m-1} &= \phi_m \\ \Delta \phi_m &= f(t, \phi_1, \phi_2, \dots, \phi_m) \end{aligned} \quad (1.8) \quad \square$$

Beweis:

Ist $\varphi : \mathbb{D} \rightarrow \mathbb{R}$ eine Lösung der Δ GL (1.7), substituiere $\Phi_j := \Delta^{j-1} \varphi$ für $j = 1, \dots, m$, d.h.

$$\Phi(t) := (\varphi(t), \Delta \varphi(t), \dots, \Delta^{m-1} \varphi(t))^T, .$$

Dann ist Φ eine Lösung von (1.8), denn für $j = 1, \dots, m-1$ gilt nach Definition von Φ

$$\Delta \Phi_j(t) = \Delta [\Delta^{j-1} \varphi(t)] = \Delta^j \varphi(t) = \Phi_{j+1}(t)$$

und

$$\Delta \Phi_m(t) = \Delta [\Delta^{m-1} \varphi(t)] = \Delta^m \varphi(t) \stackrel{(1.7)}{=} f(t, \varphi(t), \Delta \varphi(t), \dots, \Delta^{m-1} \varphi(t)) .$$

Ist umgekehrt $\Phi(t) = (\phi_1(t), \dots, \phi_m(t))^T$ eine Lösung des Δ GLs-Systems (1.8), so setze man $\varphi := \phi_1$. Aus der speziellen Gestalt von (1.8) ergibt sich sofort, dass $\Delta \varphi, \dots, \Delta^m \varphi$ existieren und dass $\Delta^{j-1} \varphi = \phi_j$ für $j = 1, \dots, m$ gilt. Die letzte Zeile von (1.8) liefert:

$$\begin{aligned} \Delta^m \varphi(t) &= \Delta [\Delta^{m-1} \varphi(t)] = \Delta \phi_m(t) \stackrel{(1.8)}{=} f(t, \phi_1(t), \phi_2(t), \dots, \phi_m(t)) \\ &= f(t, \varphi(t), \Delta \varphi(t), \dots, \Delta^{m-1} \varphi(t)) \quad \blacksquare \end{aligned}$$

1.14 Bemerkung

Der letzte Satz gilt analog für Δ GLen vektorwertiger Abbildungen $\mathbf{Y} : \mathbb{D} \rightarrow \mathbb{R}^n$. Das äquivalente Δ GLs-System besteht dann allerdings aus $m \cdot n$ Gleichungen. $\square$

1.15 Beispiel

Gegeben sei die homogene lineare Δ GL der Ordnung 3

$$Y_{t+3} - 6Y_{t+2} + 11Y_{t+1} - 6Y_t = 0.$$

Diese ist nach Satz 1.4 äquivalent zu:

$$(*) \quad \Delta^3 Y_t - 3\Delta^2 Y_t + 2\Delta Y_t = 0$$

Die Substitution

$$\phi_1 := Y$$

$$\phi_2 := \Delta Y = \Delta \phi_1$$

$$\phi_3 := \Delta^2 Y = \Delta \phi_2$$

liefert das äquivalente lineare Δ GL-System der Ordnung 1

$$\Delta \phi_1 = \phi_2$$

$$\Delta \phi_2 = \phi_3$$

$$\Delta \phi_3 = -2\phi_2 + 3\phi_3$$

Man beachte, dass die dritte Gleichung der ursprünglichen Δ GL (*) entspricht!

Mit der vektorwertigen Funktion $\Phi := \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix} = \begin{pmatrix} Y \\ \Delta Y \\ \Delta^2 Y \end{pmatrix}$ lautet das Gleichungssystem

$$\Delta \Phi_t = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -2 & 3 \end{pmatrix} \cdot \Phi_t$$

oder in rekursiver Form

$$\Phi_{t+1} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -2 & 3 \end{pmatrix} \cdot \Phi_t + \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_A \cdot \Phi_t = \underbrace{\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & -2 & 4 \end{pmatrix}}_A \cdot \Phi_t$$

$\square$

Die Theorie der Δ GLen und der DGLen befasst sich mit folgenden Fragen:

- (1) Man gebe Kriterien für die Existenz von Lösungen.
- (2) Man verschaffe sich einen Überblick über die Lösungsgesamtheit einer Δ GL.
- (3) Man gebe Bedingungen an, die die Eindeutigkeit der Lösung garantieren.

Wie schon bei „normalen“ Gleichungen nehmen die numerischen Probleme schnell zu, wenn die Δ GL „kompliziert“ ist, d.h. wenn die Funktion, die die Δ GL beschreibt, „kompliziert“ ist. Daher wollen wir uns hier auf lineare Δ GLen einschränken und uns überlegen, wie dies zur Lösung nichtlinearer Δ GLen verwendet werden kann⁷.

1.16 Definition

Ein Δ GLs-System $\mathcal{F}(t, \mathbf{Y}, \Delta \mathbf{Y}, \dots, \Delta^m \mathbf{Y}) = \mathbf{0}$ der Ordnung m heißt *linear*, wenn $\mathcal{F}$ in den Variablen $\mathbf{Y}, \Delta \mathbf{Y}, \dots, \Delta^m \mathbf{Y}$ linear ist. Das Δ GLs-System lässt sich dann in die folgende Form bringen:

$$\mathbf{A}_m(t) \cdot \Delta^m \mathbf{Y} + \mathbf{A}_{m-1}(t) \cdot \Delta^{m-1} \mathbf{Y} + \dots + \mathbf{A}_0(t) \cdot \mathbf{Y} = \mathbf{b}(t) \quad (1.9)$$

$\mathbf{Y}$ und $\mathbf{b}$ seien dabei $\mathbb{R}^n$ -wertige Abbildungen auf $\mathbb{D}$ und $\mathbf{A}_j(t)$ quadratische $(n \times n)$ -Matrizen, deren Koeffizienten Funktionen von t sind. Sind die $\mathbf{A}_j$ konstant, d.h. unabhängig von t , so heißt das Δ GLs-System (1.9) ein lineares Δ GLs-System mit konstanten Koeffizienten. Das System (1.9) heißt *homogen*, falls $\mathbf{b} \equiv \mathbf{0}$ gilt, andernfalls *inhomogen*. $\square$

1.17 Beispiel

- i) Das Δ GLs-System aus Beispiel 1.12 stellt ein lineares Δ GLs-System der Ordnung 1 mit nichtkonstanten Koeffizienten dar.
- ii) Die Δ GL aus Beispiel 1.15 ist eine lineare Δ GL der Ordnung 3 mit konstanten Koeffizienten.
- iii) Bezeichnet man im Jahr t ($t \in \mathbb{N}_0$) mit

Y_t das Volkseinkommen
 G_t die Regierungsausgaben
 C_t den Konsum
 I_t die Investition,

so hat *P. Samuelson* in einem Wachstumsmodell vorgeschlagen⁸, den jährlichen Konsum proportional zum Volkseinkommen des vergangenen Jahres und die jährliche Investition proportional zum Wachstum des Konsums im vergangenen Jahr zu betrachten. Diese Annahmen führen zu folgenden Gleichungen (b, ρ konstant):

$$\begin{aligned} Y_t &= C_t + I_t + G_t \\ C_{t+1} &= b \cdot Y_t \\ I_{t+1} &= \rho \cdot [C_{t+1} - C_t] \end{aligned}$$

⁷Erinnert sei an das Newtonverfahren, das die Linearisierung eines nichtlinearen Nullstellenproblems ist.

⁸“Interactions between the multiplier analysis and the principle of acceleration“, Review of Economic Statistics 21 (1939), 75-78

Fasst man die erste und zweite Gleichung zusammen, erhält man

$$\begin{aligned} C_{t+1} &= b \cdot [C_t + I_t + G_t] \\ I_{t+1} &= \rho \cdot [C_{t+1} - C_t] \end{aligned}$$

In vektorieller Schreibweise lautet das Δ GLs-System

$$\begin{pmatrix} 1 & 0 \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} C_{t+1} \\ I_{t+1} \end{pmatrix} = \begin{pmatrix} b & b \\ -\rho & 0 \end{pmatrix} \begin{pmatrix} C_t \\ I_t \end{pmatrix} + \begin{pmatrix} bG_t \\ 0 \end{pmatrix}$$

Stellt man das Δ GLs-System mittels Differenzen dar, erhält man

$$\begin{pmatrix} 1 & 0 \\ -\rho & 1 \end{pmatrix} \cdot \Delta \begin{pmatrix} C_t \\ I_t \end{pmatrix} + \begin{pmatrix} 1-b & -b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} C_t \\ I_t \end{pmatrix} = \begin{pmatrix} bG_t \\ 0 \end{pmatrix} \quad \square$$

1.18 Bemerkung

i) Nach Lemma 1.13 ist jede explizite lineare Δ GL der Form

$$\Delta^m Y_t + a_{m-1}(t) \Delta^{m-1} Y_t + \dots + a_1(t) \Delta Y_t + a_0(t) Y_t = b(t)$$

mit den reellwertigen Funktionen $a_0, \dots, a_{m-1}, b$ äquivalent zu dem folgenden linearen Δ GLs-System erster Ordnung

$$\begin{pmatrix} \Delta \phi_1 \\ \vdots \\ \vdots \\ \Delta \phi_{m-1} \\ \Delta \phi_m \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & \vdots \\ 0 & 0 & 0 & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 \\ -a_0 & -a_1 & \dots & \dots & -a_{m-2} & -a_{m-1} \end{pmatrix} \cdot \begin{pmatrix} \phi_1 \\ \vdots \\ \vdots \\ \vdots \\ \phi_{m-1} \\ \phi_m \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \\ b \end{pmatrix}$$

ii) Mittels Satz 1.4 kann das lineare Δ GLs-System

$$\mathbf{A}_m(t) \cdot \Delta^m \mathbf{Y}_t + \mathbf{A}_{m-1}(t) \cdot \Delta^{m-1} \mathbf{Y}_t + \dots + \mathbf{A}_0(t) \cdot \mathbf{Y}_t = \mathbf{b}(t)$$

mit geeigneten $(n \times n)$ -Matrizen B_j äquivalent in eine Rekursionsgleichung umgeformt werden zu

$$\mathbf{A}_m(t) \cdot \mathbf{Y}_{t+m} + B_{m-1}(t) \cdot \mathbf{Y}_{t+m-1} + \dots + B_0(t) \cdot \mathbf{Y}_t = \mathbf{b}(t) \quad \square$$

1.19 Satz

Sind m aufeinanderfolgende Funktionswerte $Y_{s+1}, Y_{s+2}, \dots, Y_{s+m}$ als Anfangsbedingungen vorgegeben, so besitzt die lineare Δ GL der Ordnung m

$$a_m(t) \cdot Y_{t+m} + a_{m-1}(t) \cdot Y_{t+m-1} + \dots + a_0(t) \cdot Y_t = b(t) \quad (1.10)$$

eine eindeutige Lösung, wenn $a_0(t) \neq 0$ und $a_m(t) \neq 0$ für alle t gilt. $\square$

Beweis:

Löst man Gleichung (1.10) nach Y_{t+m} auf, erhält man für alle t

$$Y_{t+m} = -\frac{1}{a_m(t)} \cdot \left[\sum_{j=0}^{m-1} a_j(t) Y_{t+j} - b(t) \right]$$

Wählt man in dieser Gleichung $t = s+1$, lässt sich der Wert von Y_{s+1+m} mit Hilfe der vorgegebenen Werte $Y_{s+1}, Y_{s+2}, \dots, Y_{s+m}$ berechnen. Induktiv ergeben sich entsprechend die Werte $Y_{s+m+1}, Y_{s+m+2}, \dots$, indem man $t = s+2, t = s+3, \dots$ wählt.

Löst man Gleichung (1.10) nach Y_t auf, lassen sich analog die Werte $Y_s, \dots, Y_0$ mit Hilfe der vorgegebenen Werte $Y_{s+1}, \dots, Y_{s+m}$ berechnen. Man erhält

$$Y_t = -\frac{1}{a_0(t)} \cdot \left[\sum_{j=1}^m a_j(t) Y_{t+j} - b(t) \right]$$

woraus sich die Werte $Y_s, \dots, Y_0$ ergeben, wenn man $t = s, t = s-1, t = s-2, \dots$ setzt. $\blacksquare$

1.20 Bemerkung

i) Die Differenzengleichung

$$Y_{t+2} = Y_t \quad \text{mit} \quad Y_0 = Y_2 = 1$$

mit den beiden sogar linear unabhängigen Lösungen

$$Y = (1, 0, 1, 0, 1, 0, \dots) \quad \text{und} \quad \tilde{Y} = (1, 1, 1, 1, 1, 1, \dots)$$

zeigt, dass in Satz 1.19 nicht auf die Forderung verzichtet werden kann, dass aufeinanderfolgende Anfangswerte vorgegeben sind.

ii) Fordert man analog die Invertierbarkeit der $(n \times n)$ -Matrizen $\mathbf{A}_m(t)$ und $\mathbf{A}_0(t)$, gilt Satz 1.19 in gleicher Form und gleichem Beweis für lineare Δ GLs-Systeme der Form

$$\mathbf{A}_m(t) \cdot \mathbf{Y}_{t+m} + \mathbf{A}_{m-1}(t) \cdot \mathbf{Y}_{t+m-1} + \dots + \mathbf{A}_0(t) \cdot \mathbf{Y}_t = \mathbf{b}(t), \quad \square$$

Der folgende Satz zeigt, dass sich der Lösungsraum linearer Δ GLs-Systeme wie der Lösungsraum linearer Gleichungssysteme verhält und berechnen lässt.

1.21 Satz

Gegeben sei die lineare Δ GL der Ordnung m

$$a_m(t) \cdot Y_{t+m} + a_{m-1}(t) \cdot Y_{t+m-1} + \dots + a_0(t) \cdot Y_t = b(t), \quad (1.11)$$

mit $a_0(t) \neq 0$ und $a_m(t) \neq 0$ für alle t . Bezeichnet $\mathbb{L}_h$ bzw. $\mathbb{L}$ die Lösungsmenge der zugehörigen homogenen bzw. inhomogenen Gleichung, so gilt:

- a) $\mathbb{L}_h$ ist ein Vektorraum der Dimension m . Eine Basis (*Fundamentalsystem*) von $\mathbb{L}_h$ ist gegeben durch die Funktionen $\varphi^{(1)}, \dots, \varphi^{(m)}$. Hierbei bezeichne $\varphi^{(j)}$ die nach Satz 1.19 eindeutige Lösung der homogenen Δ GL (1.11) mit den Anfangsbedingungen

$$(\varphi^{(j)}(1), \dots, \varphi^{(j)}(m)) = \mathbf{e}_j = \underset{\substack{j\text{-te Komp.} \\ \downarrow}}{(0, \dots, 0, 1, 0, \dots, 0)}$$

- b) Ist ψ eine spezielle (*partikuläre*) Lösung der inhomogenen Δ GL, so gilt

$$\mathbb{L} = \{\psi\} + \mathbb{L}_h := \{\psi + \varphi \mid \varphi \in \mathbb{L}_h\}$$

□

Beweis:

- a) Zur Reduzierung des Schreibaufwandes definieren wir

$$\Lambda(Y_t) := a_m(t) \cdot Y_{t+m} + a_{m-1}(t) \cdot Y_{t+m-1} + \dots + a_0(t) \cdot Y_t$$

$\mathbb{L}_h$ ist nichtleer, da der Nullvektor (Nullfunktion) in $\mathbb{L}_h$ liegt. Sind nun $\varphi^{(1)}, \varphi^{(2)} \in \mathbb{L}_h$ zwei homogene Lösungen und α, β reelle Zahlen, so gilt

$$\begin{aligned} \Lambda(\alpha \cdot \varphi_t^{(1)} + \beta \cdot \varphi_t^{(2)}) &= \sum_{j=0}^m a_j(t) \cdot [\alpha \cdot \varphi_{t+j}^{(1)} + \beta \cdot \varphi_{t+j}^{(2)}] \\ &= \alpha \cdot \sum_{j=0}^m a_j(t) \cdot \varphi_{t+j}^{(1)} + \beta \cdot \sum_{j=0}^m a_j(t) \cdot \varphi_{t+j}^{(2)} \\ &= \alpha \cdot \Lambda(\varphi_t^{(1)}) + \beta \cdot \Lambda(\varphi_t^{(2)}) \\ &= \alpha \cdot \mathbf{0} + \beta \cdot \mathbf{0} = \mathbf{0} \end{aligned}$$

Damit liegt auch $\alpha \cdot \varphi^{(1)} + \beta \cdot \varphi^{(2)}$ in $\mathbb{L}_h$, d.h. $\mathbb{L}_h$ ist ein Vektorraum.

Wir definieren nun die Abbildung $\mathfrak{I} : \mathbb{L}_h \rightarrow \mathbb{R}^m$ durch⁹

$$\mathfrak{I}(\varphi) := (\varphi_1, \dots, \varphi_m) = (\varphi(1), \dots, \varphi(m))$$

⁹ $\mathfrak{I}$ ordnet jeder Lösung der Δ GL die ersten m Folgenwerte zu.

und zeigen, dass $\mathfrak{S}$ eine bijektive, lineare Abbildung, d.h. ein Vektorraumisomorphismus ist. Da die Linearität von $\mathfrak{S}$ klar ist, bleibt nur die Bijektivität zu zeigen.

$\mathfrak{S}$ ist surjektiv, da nach Satz 1.19 zu jedem m -Tupel $(x_1, \dots, x_m)$ reeller Zahlen eine Lösung φ der homogenen Δ GL (1.11) existiert, mit $\varphi_j = x_j$, d.h. mit der Eigenschaft $\mathfrak{S}(\varphi) = (x_1, \dots, x_m)$. Da diese Lösung φ nach Satz 1.19 eindeutig bestimmt ist, muss $\mathfrak{S}$ außerdem auch injektiv sein.

Aus der linearen Unabhängigkeit der Einheitsvektoren $\mathbf{e}_j$ ($j=1, \dots, m$) und da bijektive lineare Abbildungen die lineare Unabhängigkeit erhalten, folgt damit insbesondere die lineare Unabhängigkeit der homogenen Lösungen $\varphi^{(j)} := \mathfrak{S}^{-1}(\mathbf{e}_j)$.

- b) Sei $\phi \in \mathbb{L}$ im folgenden eine fest vorgegebene spezielle (*partikuläre*) Lösung der inhomogenen Δ GL.

Ist $\varphi \in \mathbb{L}_h$ eine beliebige homogene Lösung, so gilt $\phi + \varphi \in \mathbb{L}$, wegen

$$\Lambda(\phi + \varphi) \stackrel{\Lambda \text{ linear}}{=} \Lambda(\phi) + \Lambda(\varphi) = b + \mathbf{0} = b$$

Daher ist $\{\phi\} + \mathbb{L}_h \subset \mathbb{L}$.

Ist umgekehrt $\tilde{\phi} \in \mathbb{L}$ eine beliebige inhomogene Lösung, so ist $\tilde{\phi} - \phi$ eine homogene Lösung, wegen

$$\Lambda(\tilde{\phi} - \phi) \stackrel{\Lambda \text{ linear}}{=} \Lambda(\tilde{\phi}) - \Lambda(\phi) = b - b = \mathbf{0}$$

Damit gilt $\tilde{\phi} - \phi \in \mathbb{L}_h$ bzw. $\tilde{\phi} \in \{\phi\} + \mathbb{L}_h$. Dies bedeutet $\mathbb{L} \subset \{\phi\} + \mathbb{L}_h$. ■

1.22 Bemerkung

Satz 1.21 gilt in gleicher Form und mit dem gleichen Beweis für Δ GLs-Systeme der Form

$$\mathbf{A}_m(t) \cdot \mathbf{Y}_{t+m} + \mathbf{A}_{m-1}(t) \cdot \mathbf{Y}_{t+m-1} + \dots + \mathbf{A}_0(t) \cdot \mathbf{Y}_t = \mathbf{b}(t).$$

Der einzige Unterschied ist, dass die Dimension des homogenen Lösungsraumes $n \cdot m$ ist. □

Satz 1.21 führt zu folgender Vorgehensweise bei der Lösung linearer Δ GLs-Systeme:

1. Schritt: Bestimme eine Fundamentalebasis $\varphi^{(1)}, \dots, \varphi^{(m)}$ des homogenen Lösungsraumes
2. Schritt: Bestimme eine partikuläre Lösung ϕ des inhomogenen Δ GLs-Systems.
3. Schritt: $\mathbb{L}_h = \{\phi\} + \mathbb{L}_h$, d.h. jede Lösung des inhomogenen Δ GLs-Systems hat die Form

$$\phi + \lambda_1 \cdot \varphi^{(1)} + \dots + \lambda_m \cdot \varphi^{(m)}$$

4. Schritt: Bestimme $\lambda_1, \dots, \lambda_m$ durch die gegebenen Anfangsbedingungen.

1.23 Satz (Lösung linearer Differenzengleichungen 1. Ordnung)

Gegeben sei die lineare Δ GL erster Ordnung

$$Y_{t+1} - a_t \cdot Y_t = b_t \iff [F - a_t] Y_t = b_t, \quad (1.12)$$

wobei a_t und b_t gegeben sind, mit $a_t \neq 0$ für alle t . Dann gilt:

- a) Der homogene Lösungsraum ist $\mathbb{L}_h = \{\varphi \cdot c \mid c \in \mathbb{R}\}$, mit $\varphi : \mathbb{N}_0 \rightarrow \mathbb{R}$ definiert durch¹⁰

$$\varphi_t := \prod_{\ell=0}^{t-1} a_\ell = e^{\sum_{\ell=0}^{t-1} \ln(a_\ell)}$$

- b) Eine partikuläre Lösung ϕ der inhomogenen Δ GL ist gegeben durch (φ wie in a))

$$\phi_t = \varphi_t \cdot \sum_{j=0}^{t-1} \frac{b_j}{\varphi_{j+1}} = \sum_{j=0}^{t-1} b_j \cdot \prod_{\ell=j+1}^{t-1} a_\ell$$

- c) Sind $a_t = a$ und $b_t = b$ konstant, so gilt speziell (mit dem Rentenendwertfaktor $s_t(a)$)

$$\varphi_t = a^t \quad \text{und} \quad \phi_t = b \cdot \sum_{j=0}^{t-1} a^j = b \cdot s_t(a)$$

Jede Lösung $Y \in \mathbb{L}$ hat in diesem Fall folgendes Aussehen (c konstant):

$$Y_t = b \cdot s_t(a) + c \cdot a^t$$

□

Beweis:

- a) Wir bestimmen die nach Satz 1.19 eindeutig bestimmte Lösung φ der homogenen Δ GL mit $\varphi_0 = 1$. Da die homogene Δ GL äquivalent ist zu $Y_{t+1} = a_t \cdot Y_t$ ergibt sich $\varphi_{t+1} = a_t \cdot \varphi_t$ und mit $\varphi_0 = 1$ folgt daher induktiv

$$\varphi_t = \prod_{\ell=0}^{t-1} a_\ell.$$

Da nach Satz 1.21 außerdem $\dim(\mathbb{L}_h) = 1$ gilt, ist Teil a) bewiesen.

- b) Zur Bestimmung einer partikulären Lösung ϕ der inhomogenen Δ GL verwenden wir die Methode der „*Variation der Konstanten*“. Hierbei versucht man ϕ als (nichtkonstantes)

¹⁰Die Exponentialdarstellung gilt nur im Fall $a_\ell > 0$. Man beachte hierbei die Konventionen $\sum_{i \in \emptyset} = 0$ und $\prod_{i \in \emptyset} = 1$.

Vielfaches der homogenen Lösung φ zu bestimmen, d.h. man versucht eine von t abhängige Größe $c = c(t)$ zu finden mit der Eigenschaft

$$\phi_t := c(t) \cdot \varphi_t ,$$

Setzt man dieses ϕ in die Δ GL (1.12) ein, erhält man für alle s die Forderung

$$c(s+1) \cdot \varphi_{s+1} - a_s \cdot c(s) \cdot \varphi_s \stackrel{!}{=} b_s .$$

Da φ eine Lösung der homogenen Δ GL ist, gilt $a_s \cdot \varphi_s = \varphi_{s+1}$. Einsetzen liefert eine Δ GL für c , nämlich:

$$c(s+1) - c(s) = \frac{b_s}{\varphi_{s+1}}$$

Bildet man auf beiden Seiten dieser Gleichung die Summe von 0 bis $t-1$, heben sich auf der linken Seite fast alle Summanden weg und man erhält

$$c(t) - c(0) = \sum_{s=0}^{t-1} \frac{b_s}{\varphi_{s+1}}$$

Da es für unsere Zwecke genügt, eine spezielle Lösung zu finden, können wir $c(0)$ beliebig vorgeben. Am einfachsten ist die Wahl $c(0) := 0$

c) Einfaches Einsetzen in Teil a) und b). ■

1.24 Beispiel

i) Nach Satz 1.23 ist $Y_t = C \cdot 2^t$ die allgemeine Lösung der Δ GL

$$\Delta Y_t = Y_t \iff Y_{t+1} = 2 \cdot Y_t$$

ii) Bei der *linearen Abschreibung* ist der Buchwert Y_t zu Beginn des Jahres t so zu bestimmen, dass folgende Eigenschaften erfüllt sind ($t=0, 1, \dots, n$):

$$\begin{aligned} Y_t - Y_{t+1} &= d = \text{konstant} \\ Y_0 &= K_0 \text{ (Anfangswert)} \\ Y_n &= K_n \text{ (vorgegebener Restwert, Randbedingung)} \end{aligned}$$

Die Abschreibungsfolge Y_t ist also charakterisiert durch die Δ GL

$$Y_{t+1} - Y_t = -d \iff \Delta Y_t = -d.$$

Anwendung des letzten Satzes 1.23 mit $a_t=1$ und $b_t=-d$ liefert als homogene Lösung $\varphi_t = a^t = 1$ und als inhomogene Lösung $\phi_t = b \cdot t = -d \cdot t$. Die allgemeine Lösung Y hat damit die Form

$$Y_t = \phi_t + \lambda \varphi_t = -d \cdot t + \lambda$$

Wegen $Y_0=K_0$ muss $\lambda=K_0$ gelten. Die Randbedingung $Y_n=K_n$ führt schließlich zu

$$d = \frac{K_0 - K_n}{n}.$$

Die gesuchte Abschreibungsformel lautet also

$$Y_t = K_0 - \frac{K_0 - K_n}{n} \cdot t$$

- iii) Leiht man sich einen Betrag B zu einem Jahreszinssatz i (jährlicher Zinszuschlag), so genügt der Wert S_t der Schuld in Periode t der homogenen¹¹ Δ GL

$$S_{t+1} = (1+i) \cdot S_t, \quad S_0 = B$$

Zahlt der Schuldner jährlich den gleichen Betrag A zurück, genügt S_t der Δ GL

$$S_{t+1} = (1+i) \cdot S_t - A, \quad S_0 = B.$$

Satz 1.23 c) liefert mit $a_t=q=1+i$, $b_t=-A$ die homogene Lösung $\varphi_t=q^t$ und die spezielle Lösung $\phi_t=-A \cdot s_t(q)$. Die allgemeine Lösung S_t lautet daher

$$S_t = \lambda \cdot \varphi_t + \phi_t = \lambda \cdot q^t - A \cdot s_t(q).$$

Einsetzen in die Anfangsbedingung $S_0=B$ liefert $\lambda=B$ wegen $s_0(q)=0$ und daher

$$S_t = B \cdot q^t - A \cdot s_t(q).$$

Soll die Schuld in N Jahren getilgt werden, muss A so gewählt werden, dass zusätzlich $S_N=0$ gilt. Dies liefert

$$0 = B \cdot q^N - A \cdot s_N(q).$$

Man erhält die bekannte Tilgungsformel

$$A = \frac{S_0}{s_N(q) \cdot q^{-N}} = \frac{S_0}{\text{RBF}}$$

¹¹Interessanterweise gilt $\Delta \ln(S_t) = \ln(q)$.

- iv) Das Wachstumsmodell von Harrod (ohne time-lag) liefert nach Beispiel 1.6 i) die homogene Δ GL erster Ordnung:

$$Y_{t+1} - \frac{g}{g-s} Y_t = 0$$

Anwendung von Satz 1.23 mit $a(t) = \frac{g}{g-s}$ und $b(t) = 0$ liefert als homogene Lösung $\varphi_t = \left(\frac{g}{g-s}\right)^t$. Bei Vorgabe eines Startwertes Y_0 lautet die gesuchte Lösung daher

$$Y_t = Y_0 \cdot \left(\frac{g}{g-s}\right)^t = Y_0 \cdot \left(1 + \frac{s}{g-s}\right)^t \quad \square$$

Satz 1.23 gilt mit den entsprechenden Anpassungen auch für lineare Δ GLs-Systeme. Da der im Prinzip analoge Beweis wegen der Matrizendarstellung etwas anspruchsvoller ist, soll er an dieser Stelle gesondert formuliert werden.

1.25 Satz (Lösung linearer Differenzgleichungssysteme 1. Ordnung)

Gegeben sei das lineare Δ GLs-System erster Ordnung

$$\mathbf{Y}_{t+1} - \mathbf{A}_t \cdot \mathbf{Y}_t = \mathbf{b}_t \iff (\mathbf{F} - \mathbf{A}_t) \mathbf{Y}_t = \mathbf{b}_t, \quad (1.13)$$

wobei $\mathbf{Y}, \mathbf{b} : \mathbb{D} \rightarrow \mathbb{R}^n$ vektorwertige Abbildungen und $\mathbf{A}(t)$ eine für jedes t invertierbare $(n \times n)$ -Matrix sei. Dann gilt:

- a) Für den homogenen Lösungsraum gilt $\mathbb{L}_h = \{\Phi \cdot \mathbf{c} \mid \mathbf{c} \in \mathbb{R}^n\}$, wobei $\Phi : \mathbb{N}_0 \rightarrow \mathbb{R}^{n \times n}$ definiert ist durch¹²

$$\Phi_t := \mathbf{A}_{t-1} \cdot \mathbf{A}_{t-2} \cdots \mathbf{A}_1 \cdot \mathbf{A}_0.$$

Die Spaltenvektoren der Matrix Φ_t , d.h. die Funktionen $\phi_t^{(j)} = \Phi_t \cdot \mathbf{e}_j$, sind daher ein *Fundamentalsystem* des n -dimensionalen homogenen Lösungsraumes $\mathbb{L}_h$.

- b) Eine *partikuläre* Lösung ϕ des inhomogenen Δ GLs-System ist gegeben durch

$$\phi_t = \Phi_t \cdot \sum_{j=0}^{t-1} \Phi_{j+1}^{-1} \cdot \mathbf{b}_j.$$

- c) Sind $\mathbf{A}_t = \mathbf{A}$ und $\mathbf{b}_t = \mathbf{b}$ konstant, so gilt speziell¹³

$$\Phi_t = \mathbf{A}^t \quad \text{und} \quad \phi_t = \left(\sum_{j=0}^{t-1} \mathbf{A}^j \right) \cdot \mathbf{b} = (\mathbf{A}^t - \mathbf{1}) \cdot (\mathbf{A} - \mathbf{1})^{-1} \cdot \mathbf{b} \quad \square$$

¹²Man beachte, dass i.A. $\mathbf{A}_{t-1} \cdots \mathbf{A}_0 \neq \mathbf{A}_0 \cdots \mathbf{A}_{t-1}$ gilt, da die Matrizenmultiplikation nicht kommutativ ist.

¹³Falls $\mathbf{A} - \mathbf{1}$ invertierbar ist.

Beweis:

a) Offensichtlich gilt $\Phi_{t+1} - \mathbf{A}_t \cdot \Phi_t = \Phi_{t+1} - \Phi_{t+1} = \mathbf{0}$ und daher

$$\phi_{t+1}^{(j)} - \mathbf{A}_t \cdot \phi_t^{(j)} = [\Phi_{t+1} - \mathbf{A}_t \cdot \Phi_t] \cdot \mathbf{e}_j = \mathbf{0} \cdot \mathbf{e}_j = 0$$

b) Wir verwenden wieder die Methode der Variation der Konstanten, d.h. wir nehmen an, dass eine partikuläre Lösung ϕ ein Vielfaches der homogenen Lösung ist¹⁴, d.h.

$$\phi_t = \Phi_t \cdot \mathbf{c}(t) .$$

Setzt man ϕ_s in Gleichung (1.13) ein, erhält man wegen $\mathbf{A}_s \cdot \Phi_s = \Phi_{s+1}$

$$\Phi_{s+1} \cdot (\mathbf{c}(s+1) - \mathbf{c}(s)) \stackrel{!}{=} \mathbf{b}_s .$$

Als Produkt invertierbarer Matrizen ist Φ_{s+1} invertierbar und es folgt

$$\mathbf{c}(s+1) - \mathbf{c}(s) \stackrel{!}{=} \Phi_{s+1}^{-1} \cdot \mathbf{b}_s .$$

Summiert man auf beiden Seiten von $s=0$ bis $t-1$, folgt die Behauptung, wenn man beachtet, dass $\mathbf{c}(0) = \mathbf{0}$ gewählt werden kann.

c) Ergibt sich direkt aus Teil a) und b) durch Einsetzen. ■

1.26 Bemerkung

Da jede lineare ΔGL m -ter Ordnung gemäß Lemma 1.13 äquivalent zu einem ΔGLs -System erster Ordnung ist, bietet Satz 1.25 eine scheinbar einfache Lösungsmöglichkeit für ΔGL en der Ordnung m . Die Schwierigkeit liegt jedoch in der Berechnung der Matrizenprodukte Φ_t . Selbst wenn die Matrizen $\mathbf{A}(t)$ für alle t konstant $\mathbf{A}$ sind, können die Potenzen $\Phi_t = \mathbf{A}^t$ nur in Ausnahmefällen (z.B. wenn $\mathbf{A}$ eine Diagonalmatrix ist) explizit berechnet werden (siehe hierzu auch Bemerkung 1.29). In diesem Fall bietet sich aber eine alternative Lösungsmöglichkeit für das homogene ΔGLs -System $\mathbf{Y}_{t+1} - \mathbf{A} \cdot \mathbf{Y}_t = \mathbf{0}$ an. Wählt man nämlich in Satz 1.25 speziell $\mathbf{c} := \mathbf{v}$, wobei $\mathbf{v}$ ein Eigenvektor der Matrix $\mathbf{A}$ zum Eigenwert λ ist¹⁵, so hat die homogene Lösung $\Phi_t \cdot \mathbf{v}$ ein besonders einfaches Aussehen:

$$\Phi_t \cdot \mathbf{v} = \mathbf{A}^t \cdot \mathbf{v} = \lambda \mathbf{A}^{t-1} \cdot \mathbf{v} = \dots = \lambda^t \cdot \mathbf{v}$$

Die Hauptschwierigkeit ist in diesem Fall die Berechnung der Eigenwerte (siehe hierzu auch Bemerkung 1.29) □

¹⁴Direktes Einsetzen führt zwar schneller zum Ziel, erklärt jedoch nicht, wie man die Lösung findet.

¹⁵Dies bedeutet $\mathbf{A} \cdot \mathbf{v} = \lambda \cdot \mathbf{v}$.

Als nächstes untersuchen wir lineare Δ GLen der Ordnung $m \geq 2$. Der Einfachheit halber beschränken wir uns auf Δ GLen mit konstanten Koeffizienten, d.h. auf Gleichungen

$$Y_{t+m} + a_{m-1} \cdot Y_{t+m-1} + \dots + a_1 \cdot Y_{t+1} + a_0 \cdot Y_t = b, \quad (1.14)$$

mit reellen Zahlen $a_0, \dots, a_{m-1}, b$. Stellt man (1.14) mit Hilfe des Forward-Operators F dar,

$$(F^m + a_{m-1}F^{m-1} + \dots + a_1F + a_0)Y_t = b, \quad (1.15)$$

wird die Idee zur Bestimmung der homogenen Lösungen der Δ GL schnell klar. Ist nämlich $\tilde{x}$ eine Nullstelle der zu (1.14) bzw. (1.15) gehörenden *charakteristischen Gleichung*¹⁶

$$x^m + a_{m-1} \cdot x^{m-1} + \dots + a_1 \cdot x + a_0 = 0,$$

so kann in (1.15) der Term $F - \tilde{x}$ als Faktor abgespalten werden, so dass jede Lösung der Δ GL $(F - \tilde{x})Y_t = 0$ auch eine homogene Lösung der vorgegebenen Δ GL (1.15) ist. Hierdurch wird das ursprüngliche Problem m -ter Ordnung auf die Lösung von Δ GLen erster Ordnung zurückgeführt. Neben der Berechnung der Nullstellen bleiben dann noch zwei Probleme zu klären. Erstens, was ist zu tun bei mehrfachen Nullstellen? Zweitens, liefern komplexe Nullstellen reelle Lösungen?

1.27 Satz (Lösung homogener linearer Δ GLen mit konstanten Koeffizienten)

Sei z eine Nullstelle der Vielfachheit k der charakteristischen Gleichung der homogenen linearen Δ GL mit konstanten Koeffizienten

$$Y_{t+m} + a_{m-1} \cdot Y_{t+m-1} + \dots + a_1 \cdot Y_{t+1} + a_0 \cdot Y_t = 0 \quad (1.16)$$

Dann ist für jedes Polynom P mit $\text{grad}(P) \leq k-1$ die Funktion $\varphi_t := P(t) \cdot z^t$ ebenfalls eine Lösung der Δ GL (1.16). $\square$

Beweis:

Aufgrund der Linearität genügt es zu zeigen, dass alle Funktionen der Form $t^j \cdot z^t$ für $j \in \mathbb{N}_0$ mit $j \leq k-1$ eine Lösung von (1.16) sind.

Stellt man die Δ GL (1.16) mit Hilfe des Forward-Operators dar, erhält man

$$(F^m + a_{m-1}F^{m-1} + \dots + a_1F + a_0)Y_t = 0. \quad (1.17)$$

Da z eine Nullstelle der Ordnung k ist, kann (1.17) dargestellt werden in der Form

$$P(F) \cdot (F - z)^k Y_t = 0,$$

mit einem Polynom P der Ordnung $m-k$. Offensichtlich genügt es daher $(F - z)^k (t^j z^t) = 0$ für $j \leq k-1$ zu zeigen. Der Beweis erfolgt mittels vollständiger Induktion nach k .

¹⁶Diese entsteht, wenn man Y_t entfernt und den Forward-Operator F formal durch die Variable x ersetzt.

Für $k=1$ ist die Behauptung nach Satz 1.23 klar, da dann nur $j=0$ möglich ist. Gelte die Behauptung nun für eine natürliche Zahl k und sei z eine Nullstelle der Ordnung $k+1$ (und damit insbesondere der Ordnung k). Dann folgt mit Hilfe des binomischen Satzes für alle $j \in \mathbb{N}_0$ mit $j \leq k$, d.h. $j-1 \leq k-1$

$$\begin{aligned}
 (F-z)^{k+1}(t^j z^t) &= (F-z)^k \left[(F-z)(t^j z^t) \right] \\
 &= (F-z)^k \left[(t+1)^j z^{t+1} - t^j z^{t+1} \right] \\
 &\stackrel{\text{bin. Satz}}{=} (F-z)^k \left[\sum_{\ell=0}^{j-1} \binom{j}{\ell} t^\ell z^{t+1} \right] \\
 &\stackrel{z \text{ konst.}}{=} z(F-z)^k \left[\sum_{\ell=0}^{j-1} \binom{j}{\ell} t^\ell z^t \right] \\
 &= z \left[\sum_{\ell=0}^{j-1} \binom{j}{\ell} \underbrace{(F-z)^k (t^\ell z^t)}_{=0, \text{ da } \ell \leq j-1 \leq k-1} \right] = 0 \quad \blacksquare
 \end{aligned}$$

1.28 Bemerkung

Eine Schwierigkeit ergibt sich beim Auftreten komplexer Nullstellen $z=q+ip$, da dann die Lösungsfunktion $\varphi_t=z^t$ komplexe Werte annimmt. In diesem Fall sind jedoch $\operatorname{Re}(z^t)$ und $\operatorname{Im}(z^t)$ zwei linear unabhängige reelle Lösungen. Dies ergibt sich aus folgender Überlegung. Da die Koeffizienten der charakteristischen Gleichung

$$x^m + a_{m-1}x^{m-1} + \dots + a_1x + a_0 = 0$$

reell sind, ist auch die zu z konjugiert komplexe Zahl $\bar{z}:=q-ip$ eine Nullstelle der charakteristischen Gleichung¹⁷. Daher ist auch $\phi_t=\bar{z}^t=\bar{\varphi}_t$ eine homogene Lösung. Da der homogene Lösungsraum aufgrund der Linearität der Δ GL ein Untervektorraum ist, sind auch

$$\frac{1}{2} \cdot (\varphi_t + \phi_t) = \frac{1}{2} \cdot (\varphi_t + \bar{\varphi}_t) = \operatorname{Re}(\varphi_t)$$

und

$$\frac{1}{2} \cdot i \cdot (\varphi_t - \phi_t) = \frac{1}{2} \cdot i \cdot (\varphi_t - \bar{\varphi}_t) = \operatorname{Im}(\varphi_t)$$

homogene Lösungen, da sie Linearkombinationen sind.

Mit $|z|:=\sqrt{q^2+p^2}$ und $\tan(\alpha)=\frac{p}{q}$ liefert die Formel von „Moivre“ (Anhang Satz C.5)

$$z^t = [|z| \cdot (\cos \alpha + i \sin \alpha)]^t = |z|^t \cdot [\cos(\alpha t) + i \sin(\alpha t)] ,$$

d.h. die reellen, unabhängigen homogenen Lösungen $|z|^t \cdot \cos(\alpha t)$ und $|z|^t \cdot \sin(\alpha t)$. $\square$

¹⁷Es gilt nämlich für alle komplexen Zahlen z_1, z_2 und reelle Zahlen r (siehe Anhang, Satz C.2)

$$\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2, \quad \overline{z_1 \cdot z_2} = \bar{z}_1 \cdot \bar{z}_2 \quad \text{und} \quad \bar{\bar{r}} = r .$$

1.29 Bemerkung

Das Problem bei der Lösung von homogenen Δ GLn höherer Ordnung, liegt in der Berechnung der Nullstellen der charakteristischen Gleichung, da für $m \geq 3$ dies i.a. große numerische Probleme bereitet. Wie schon in Bemerkung 1.26 i) festgestellt, ist die Transformation in ein Δ GLs-System erster Ordnung ein nur scheinbar einfacherer Lösungsweg, da die Berechnung des Matrizenproduktes A^t große Schwierigkeiten bereitet¹⁸.

Dass die numerischen Probleme immer gleich bleiben, zeigt der Vergleich der beiden Verfahren zur Bestimmung der homogenen Lösungen von linearen Δ GLn der Ordnung m , die wir in Bemerkung 1.26 ii) und Satz 1.27 kennen gelernt haben. Geben wir uns hierzu die folgende homogene lineare Δ GL der Ordnung m vor:

$$\Delta^m Y_t + \tilde{a}_{m-1} \cdot \Delta^{m-1} Y_t + \dots + \tilde{a}_1 \cdot \Delta Y_t + \tilde{a}_0 \cdot Y_t = 0 \quad (1.18)$$

1. Lösungsweg: Eigenwertmethode (siehe Bemerkung 1.26 ii))

Wir transformieren Gleichung (1.18) auf ein äquivalentes lineares Δ GLs-System erster Ordnung. Wie in Bemerkung 1.18 gezeigt, ist (1.18) äquivalent zu $\Delta \mathbf{Y}_t = \tilde{A} \cdot \mathbf{Y}_t$ mit

$$\tilde{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & \vdots \\ 0 & 0 & 0 & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 \\ -\tilde{a}_0 & -\tilde{a}_1 & \dots & \dots & -\tilde{a}_{m-2} & -\tilde{a}_{m-1} \end{pmatrix}$$

Unter Berücksichtigung von $\Delta \mathbf{Y}_t = \mathbf{Y}_{t+1} - \mathbf{Y}_t$ bringen wir das Δ GLs-System $\Delta \mathbf{Y}_t = \tilde{A} \cdot \mathbf{Y}_t$ auf die zur Eigenwertmethode erforderliche Form

$$\mathbf{Y}_{t+1} = A \cdot \mathbf{Y}_t \quad \text{mit} \quad A = \tilde{A} + \mathbf{1}$$

Die Eigenwerte von A sind die Nullstellen des charakteristischen Polynoms $P(\lambda) = \det(\lambda \mathbf{1} - A)$. Mittels vollständiger Induktion folgt

$$P(\lambda) = (\lambda-1)^m + \tilde{a}_{m-1} \cdot (\lambda-1)^{m-1} + \dots + \tilde{a}_1 \cdot (\lambda-1) + \tilde{a}_0$$

2. Lösungsweg: Nullstellenmethode (siehe Satz 1.27)

Hierzu müssen wir die Δ GL (1.18) zunächst mit Hilfe des Forward-Operators F ausdrücken. Nach Bemerkung 1.3 iii) gilt

$$\Delta^k Y_t = (F - \mathbf{1})^k Y_t.$$

Setzt man dies in die Ausgangsgleichung (1.18) ein, erhält man die äquivalente Form

$$\left[(F - \mathbf{1})^m + \tilde{a}_{m-1} \cdot (F - \mathbf{1})^{m-1} + \dots + \tilde{a}_1 \cdot (F - \mathbf{1}) + \tilde{a}_0 \right] Y_t = 0$$

¹⁸Im Prinzip ist dies nur möglich, wenn A mittels einer Basis von Eigenvektoren auf Diagonalgestalt gebracht wird. Hierzu ist jedoch wiederum die Kenntnis der Eigenwerte erforderlich.

Die zugehörige charakteristische Gleichung entsteht, wenn man Y_t streicht und den Forward-Operator F durch die Variable x ersetzt. Offensichtlich erhält man genau das gleiche Nullstellenproblem wie beim ersten Lösungsweg. $\square$

Beim Lösen linearer Δ GLen bereitet die Inhomogenität die größten Probleme, da sich bei der Bestimmung der partikulären Lösung keine Regel angeben lässt. Interpretiert man die Inhomogenität als exogene Störung, ist zu vermuten, dass das durch die Δ GL beschriebene System mit einer partikulären Lösung mit ähnlichen Eigenschaften reagiert. Dies soll beispielhaft für Polynome als Inhomogenität untersucht werden (Beachte auch Beispiel 1.34).

1.30 Lemma

Sei P ein Polynom vom Grad $m \in \mathbb{N}_0$. Dann gilt

- a) ΔP ist ein Polynom vom Grad $m-1$.
- b) $\Delta^m P \equiv c = \text{konstant}$ und $\Delta^k P \equiv 0$ für $k \geq m+1$. $\square$

Beweis:

- a) Wegen der Linearität des Differenzenoperators Δ kann o.B.d.A. $P_t = t^m$ angenommen werden. Dann folgt mit Hilfe des binomischen Lehrsatzes

$$\begin{aligned} \Delta(t^m) &= (t+1)^m - t^m \\ &= \sum_{\ell}^m \binom{m}{\ell} (-1)^{m-\ell} t^{\ell} - t^m \\ &= \sum_{\ell}^{m-1} \binom{m}{\ell} (-1)^{m-\ell} t^{\ell}. \end{aligned}$$

Also ist $\Delta(t^m)$ ein Polynom vom Grad $m-1$.

- b) Diese Aussage ist eine direkte Konsequenz von Teil a). $\blacksquare$

1.31 Satz

Die lineare inhomogene Δ GL erster Ordnung

$$(\Delta - a)Y = P \quad (a \text{ konstant}) \tag{1.19}$$

besitzt ein Polynom Q als partikuläre Lösung, wenn die Inhomogenität P ein Polynom ist. Außerdem gilt

$$\text{grad}(Q) = \begin{cases} \text{grad}(P) & \text{falls } a \neq 0 \\ \text{grad}(P) + 1 & \text{falls } a = 0 \end{cases}$$

Da Q der Gleichung $(\Delta - a)Q = P$ genügt, bezeichnen wir Q aus naheliegenden Gründen auch kurz mit $(\Delta - a)^{-1}P$ bzw. $\Delta^{-1}P$ im Fall $a = 0$. $\square$

Beweis:

Gelte zunächst $a \neq 0$. Dann definieren wir¹⁹

$$Q_t := (\Delta - a)^{-1}P_t := -\sum_{k=0}^{\infty} \frac{\Delta^k P_t}{a^{k+1}} = -\sum_{k=0}^{\text{grad}(P)} \frac{\Delta^k P_t}{a^{k+1}}$$

Nach Lemma 1.30 ist Q ein Polynom mit $\text{grad}(Q) = \text{grad}(P)$. Zu zeigen bleibt, dass Q der ΔGL (1.19) genügt. Mit $g = \text{grad}(P)$ gilt $\Delta^{g+1}P = 0$ nach Lemma 1.30 b) und damit

$$\begin{aligned} (\Delta - a)Q_t &= -\Delta \sum_{k=0}^g \frac{\Delta^k P_t}{a^{k+1}} + a \sum_{k=0}^g \frac{\Delta^k P_t}{a^{k+1}} \\ &= -\sum_{k=0}^g \frac{\Delta^{k+1} P_t}{a^{k+1}} + \sum_{k=0}^g \frac{\Delta^k P_t}{a^k} \\ &\stackrel{\text{Ind. Trafo}}{=} -\sum_{k=1}^{g+1} \frac{\Delta^k P_t}{a^k} + \sum_{k=0}^g \frac{\Delta^k P_t}{a^k} \\ &= -\frac{\Delta^{g+1} P_t}{a^k} + P_t \stackrel{1.30 \text{ b)}}{=} P_t \end{aligned}$$

Im Fall $a = 0$ definieren wir

$$Q_t := \Delta^{-1}P_t := \sum_{k=1}^{\infty} \binom{t}{k} \Delta^{k-1} P_0 = \sum_{k=1}^{\text{grad}(P)+1} \binom{t}{k} \Delta^{k-1} P_0$$

Da $\binom{t}{k}$ in der Variablen t ein Polynom vom Grade k ist (Beweis z.B. mittels vollständiger Induktion nach k) ist Q ein Polynom mit $\text{grad}(Q) = \text{grad}(P) + 1$.

Bezeichnet g wieder den Grad von P , so gilt wegen $\binom{t+1}{k} - \binom{t}{k} = \binom{t}{k-1}$ und nach Satz 1.4 b) der Vorlesung (man beachte, dass $\Delta^{k-1}P_0$ unabhängig von t ist)

¹⁹Die Darstellung beruht auf der Taylorentwicklung von $(x - a)^{-1}$ (siehe hierzu Bemerkung 1.33).

$$\begin{aligned}
\Delta Q_t &= \Delta \sum_{k=1}^{g+1} \binom{t}{k} \Delta^{k-1} P_0 \\
&= \sum_{k=1}^{g+1} [\Delta \binom{t}{k}] \Delta^{k-1} P_0 \\
&= \sum_{k=1}^{g+1} \left[\binom{t+1}{k} - \binom{t}{k} \right] \Delta^{k-1} P_0 \\
&= \sum_{k=1}^{g+1} \binom{t}{k-1} \Delta^{k-1} P_0 \\
&\stackrel{\text{Ind. Trafo}}{=} \sum_{k=0}^g \binom{t}{k} \Delta^k P_0 \stackrel{1.4 \text{ b)}}{=} P_t \quad \blacksquare
\end{aligned}$$

1.32 Folgerung

Jede inhomogene lineare Δ GL der Ordnung m mit konstanten Koeffizienten, bei der die Inhomogenität ein Polynom P ist, besitzt ein Polynom Q als partikuläre Lösung mit

$$\text{grad}(P) \leq \text{grad}(Q) \leq \text{grad}(P) + m$$

Sind die Koeffizienten der Δ GL und des Polynoms P reell, existiert auch ein reelles Polynom Q als partikuläre Lösung. $\square$

Beweis:

Gegeben sei die inhomogene lineare Δ -GL n -ter Ordnung

$$\sum_{i=0}^m a_i \Delta^i Y = a_m \Delta^m Y + a_{m-1} \Delta^{m-1} Y + \dots + a_1 \Delta Y + a_0 Y = P \quad (1.20)$$

mit konstanten reellen Koeffizienten a_i , $i = 0, \dots, m$ und dem Polynom P als Inhomogenität. O.B.d.A. gelte $a_m = 1$.

Nach dem Hauptsatz der Algebra besitzt die durch die Δ GL (1.20) erzeugte charakteristische Gleichung

$$\sum_{i=0}^m a_i x^i = 0$$

m Nullstellen, die wir (inklusive Vielfachheit) mit z_i ($i=1, \dots, m$) bezeichnen. Diese (evtl. komplexen!) Nullstellen gestatten eine Zerlegung in Linearfaktoren, d.h. es gilt

$$\sum_{i=0}^m a_i x^i = \prod_{i=1}^m (x - z_i)$$

Ersetzt man x durch den Differenzenoperator Δ , lautet die Δ GL (1.20)

$$\sum_{i=0}^m a_i \Delta^i Y = (\Delta - z_1) \cdots (\Delta - z_m) Y = P.$$

Wendet man Satz 1.31 iterativ an, erhält man als partikuläre Lösung Q der Δ GL (1.20) das Polynom

$$Q = (\Delta - z_m)^{-1} \cdots (\Delta - z_1)^{-1} P.$$

Ebenfalls nach Satz 1.31 erfüllt Q auch die Bedingung

$$\text{grad}(P) \leq \text{grad}(Q) \leq \text{grad}(P) + m$$

Es bleibt nachzuweisen, dass Q reell gewählt werden kann, wenn die Δ GL (1.20) reell ist. Hierzu zeigen wir, dass mit Q auch der Realteil von Q , also das reellwertige Polynom $R := \frac{1}{2}(Q + \overline{Q})$ eine Lösung der Δ GL ist.

Beachtet man, dass wegen $\overline{a - b} = \overline{a} - \overline{b}$ auch $\overline{\Delta Q} = \Delta \overline{Q}$ gilt, folgt nämlich

$$\begin{aligned} \sum_{i=0}^m a_i \Delta^i R &= \sum_{i=0}^m a_i \cdot \frac{1}{2} \cdot (\Delta^i Q + \Delta^i \overline{Q}) \\ &= \frac{1}{2} \sum_{i=0}^m a_i \cdot \Delta^i Q + \frac{1}{2} \sum_{i=0}^m a_i \cdot \overline{\Delta^i Q} \\ &\stackrel{a_i \text{ reell}}{=} \frac{1}{2} \sum_{i=0}^m a_i \cdot \Delta^i Q + \frac{1}{2} \sum_{i=0}^m \overline{a_i \cdot \Delta^i Q} \\ &= \frac{1}{2} P + \frac{1}{2} \overline{P} \\ &\stackrel{P \text{ reell}}{=} \frac{1}{2} P + \frac{1}{2} P = P \end{aligned}$$

■

1.33 Bemerkung

Die Konstruktion des Operators $(\Delta - a)^{-1}$ beruht auf der Idee, mit Hilfe der Taylorentwicklung (siehe Anhang A, Satz A.9) den Definitionsbereich einer Funktion $f(x)$ auch auf Operatoren oder Matrizen auszuweiten. Hierdurch lassen sich auch Terme der Form $f(\Delta)$, $f(D)$ oder $f(A)$ definieren, wobei Δ der Differenzenoperator, D der Differentialoperator und A eine beliebige quadratische Matrix ist.

Für $a \neq 0$ lautet die Taylorentwicklung von $f(x) = (x - a)^{-1}$ (Entwicklungspunkt $x_0 = 0$)

$$f(x) = (x - a)^{-1} = - \sum_{k=0}^{\infty} \frac{x^k}{a^{k+1}}.$$

Ersetzt man x durch Δ erhält man den in Satz 1.31 verwendeten Operator $(\Delta - a)^{-1}$.

Δ^{-1} stellt als Inverse des Differenzenoperators das diskrete Analogon zum Integral dar, also die Summe $\Delta^{-1} Y_t := \sum_{i=0}^{t-1} Y_i$. Die Darstellung im Beweis von Satz 1.31 ergibt sich, indem man in Satz 1.4 b) $t=s=0$ setzt. Man erhält $Y_k = \sum_{i=0}^k \binom{k}{i} \Delta^i Y_0$. Ersetzt man Y durch $\Delta^{-1} P$ folgt die gewünschte Darstellung. □

1.34 Beispiel

- i) In einer Kaninchenpopulation, vermehrt sich jedes Kaninchenpaar zweimal in seinem Leben, und zwar im Alter von einem und zwei Monaten. Dabei werde jedes Mal ein neues Kaninchenpaar geboren. Bezeichnet Y_t die Zahl von Kaninchenpaaren, die in Periode t (gerechnet in Monaten) *geboren* werden²⁰, so gilt die Δ GL

$$Y_{t+2} = Y_{t+1} + Y_t \iff \Delta^2 Y_t + \Delta Y_t - Y_t = 0 \quad (1.21)$$

Gesucht sei die Lösung mit den Anfangswerten $Y_0 := 0$ und $Y_1 := 1$.

Die Nullstellen der charakteristischen Gleichung $x^2 - x - 1 = 0$ lauten:

$$x_1 = \frac{1 + \sqrt{5}}{2}, \quad x_2 = \frac{1 - \sqrt{5}}{2}.$$

Ein Fundamentalsystem $\varphi^{(1)}, \varphi^{(2)}$ des homogenen Lösungsraumes der Δ GL (1.21) ist daher gegeben durch

$$\varphi_t^{(1)} := \left(\frac{1 + \sqrt{5}}{2}\right)^t, \quad \varphi_t^{(2)} := \left(\frac{1 - \sqrt{5}}{2}\right)^t.$$

Daher gilt $Y_t = \lambda \varphi_t^{(1)} + \mu \varphi_t^{(2)}$. Die Anfangsbedingungen $Y_0 := 0$ und $Y_1 := 1$ liefern

$$\begin{aligned} 0 &= \lambda \varphi_0^{(1)} + \mu \varphi_0^{(2)} = \lambda + \mu \\ 1 &= \lambda \varphi_1^{(1)} + \mu \varphi_1^{(2)} = \lambda \cdot \frac{1 + \sqrt{5}}{2} + \mu \cdot \frac{1 - \sqrt{5}}{2} \end{aligned}$$

mit der Lösung $\lambda = \frac{1}{\sqrt{5}}$ und $\mu = -\frac{1}{\sqrt{5}}$. Die gesuchte Lösung lautet daher²¹

$$Y_t = \frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2}\right)^t - \frac{1}{\sqrt{5}} \left(\frac{1 - \sqrt{5}}{2}\right)^t.$$

- ii) Sei $b = \rho = \frac{1}{2}$ und $G_t := \bar{G} = \text{konstant}$ in der Wachstumsgleichung von Beispiel 1.17
 iii). Löst man diese nach C auf, ergibt sich die Δ GL

$$C_{t+2} - \frac{3}{4}C_{t+1} + \frac{1}{4}C_t = \frac{1}{2}\bar{G}.$$

Die charakteristische Gleichung $x^2 - \frac{3}{4}x + \frac{1}{4} = 0$ hat die beiden komplexen Nullstellen

$$z_1 = \frac{3 + i\sqrt{7}}{8}, \quad z_2 = \frac{3 - i\sqrt{7}}{8}.$$

²⁰Ohne Berücksichtigung von Todesfällen ergibt sich dann die gesamte Population durch $\sum Y_t$.

²¹Verblüffenderweise liefert die Formel für Y_t tatsächlich nur natürliche Zahlen. Für das Verhältnis $\frac{Y_{t+1}}{Y_t}$ zweier aufeinanderfolgender Werte der Fibonaccifolge konvergiert außerdem gegen den sog. *Goldenen Schnitt* $\frac{1+\sqrt{5}}{2}$.

Nach Satz 1.27 sind mit $|z_1| = |z_2| = \frac{1}{2}$ und $\alpha = \arctan\left(\frac{\sqrt{7}}{3}\right)$ die Funktionen

$$\varphi_t^{(1)} := 2^{-t} \cos(\alpha t), \quad \varphi_t^{(2)} := 2^{-t} \sin(\alpha t)$$

eine Basis des homogenen Lösungsraumes. Folgerung 1.32 liefert die partikuläre Lösung $\psi_t = \bar{G} = \text{konstant}$, so dass die allgemeine Lösung der Δ GL lautet²²:

$$C_t = \lambda \cdot 2^{-t} \cos(\alpha t) + \mu \cdot 2^{-t} \sin(\alpha t) + \bar{G}$$

iii) Betrachte die inhomogene lineare Δ GL

$$Y_{t+2} - 2Y_{t+1} - 3Y_t = 2^t.$$

Die charakteristische Gleichung dieser Δ GL 2. Ordnung liefert die Nullstellen -1 und 3 und damit die homogenen Lösungsfunktionen $\varphi_t^{(1)} := (-1)^t$ und $\varphi_t^{(2)} := 3^t$.

Da die Inhomogenität kein Polynom ist, können wir Folgerung 1.32 nicht anwenden. Wir versuchen daher eine inhomogene Lösung ψ zu finden, die ein ähnliches Aussehen wie die Inhomogenität besitzt. Hierzu machen wir den Ansatz $\psi_t := c \cdot 2^t$.

Setzt man dieses ψ in die vorgegebene Δ GL ein, erhält man die Gleichung

$$c \cdot 2^{t+2} - 2c \cdot 2^{t+1} - 3c \cdot 2^t \stackrel{!}{=} 2^t.$$

Division durch 2^t ergibt $c = -\frac{1}{3}$. Damit lautet die allgemeine Lösung der Δ GL:

$$Y_t = \lambda \cdot (-1)^t + \mu \cdot 3^t - \frac{1}{3} \cdot 2^t.$$

□

1.35 Bemerkung

Liegt eine lineare Δ GL mit einer allgemeinen Inhomogenität $b(t)$ vor, so findet man eine spezielle Lösung ψ der Δ GL häufig dadurch, dass man ψ als Linearkombination von Funktionen ansetzt, die vom gleichen Typ wie die Inhomogenität $b(t)$ sind.

Ist z.B. $b(t) = \cos(k \cdot t)$, versucht man es mit dem Ansatz

$$\psi(t) = A \sin(k \cdot t) + B \cos(k \cdot t)$$

und versucht die Konstanten A und B so zu bestimmen, dass ψ der Δ GL genügt.

Die Idee hinter diesem Ansatz ist, dass die exogene Störung $b(t)$ in dem System zu einer Reaktion mit ähnlichen Eigenschaften führen sollte. □

²²Die durch die cos- bzw. sin-Funktion hervorgerufenen Oszillationen können als endogene Wirtschaftszyklen interpretiert werden.

2

Differentialgleichungen

Differentialgleichungen sind das kontinuierliche Gegenstück der Δ GLen aus Abschnitt 1. Ein(e) Differentialgleichung(-ssystem) ist eine Gleichung (Gleichungssystem) zwischen gesuchten Funktionen, deren zugrundeliegenden Veränderlichen und den Ableitungen der gesuchten Funktionen nach den unabhängigen Veränderlichen. Hängen die gesuchten Funktionen nur von einer unabhängigen Veränderlichen ab, spricht man von einer *gewöhnlichen Differentialgleichung* (künftig mit DGL abgekürzt). Hängen die gesuchten Funktionen von mehr als einer unabhängigen Veränderlichen ab, spricht man von einer *partiellen Differentialgleichung*. Da die Behandlung partieller Differentialgleichungen sehr kompliziert ist, sollen im Rahmen dieses Kapitels nur gewöhnliche DGLen besprochen werden.

Die Aussagen für DGLen sind häufig analog zu den Aussagen aus dem Bereich der Δ GLen, so dass wir uns größtenteils auf die Angabe der entsprechenden Sätze beschränken können. Um das Themengebiet nicht zu langweilig werden zu lassen, geben wir jedoch einige im Vergleich zu Differenzgleichungen allgemeinere Ergebnisse an.

Wie schon im Kapitel über Δ GLen wird die Hauptvariable mit t bezeichnet, da sie in Anwendungen oft die Zeit repräsentiert. Demgemäß werden wir (wie in der Physik üblich) die Ableitungen mit einem Punkt bezeichnen, also $\dot{f}, \ddot{f}$ statt f', f'' schreiben.

Im Unterschied zu Δ GLen werden wir die jetzt auftretenden stetigen Größen mit kleinen Buchstaben bezeichnen, also $\dot{y}, \ddot{y}$ statt $\dot{Y}, \ddot{Y}$ schreiben.

Analog zum Kapitel über Differenzgleichungen starten wir mit mehreren Beispielen.

2.1 Beispiel

- i) Wir wollen wie in Beispiel 1.6 die Preisdynamik eines Gutes beschreiben, die durch Angebot S und Nachfrage D entsteht. Wie üblich sei S monoton wachsend und D monoton fallend (als Funktion von P). Das *Preisfestsetzungsmodell von EVANS* nimmt an, dass Preisänderungen proportional zum Nachfrageüberschuss sind, d.h.

$$\dot{P}(t) = \gamma \cdot (D(t) - S(t)), \quad \text{mit } \gamma > 0.$$

Betrachtet man den einfachen Fall, dass D und S affin lineare Funktionen in P sind

$$D(t) = \alpha + aP(t) \text{ und } S(t) = \beta + bP(t),$$

mit $a < 0$ und $b > 0$, so lässt sich die Preisdynamik beschreiben durch:

$$\dot{P}(t) - \gamma(a - b)P(t) = \gamma(\alpha - \beta) .$$

- ii) Bei gegebenem Periodenzinssatz r gilt für ein Kapital K_t im Zeitintervall Δt die Wachstumsdynamik (siehe auch Beispiel 1.24 ii))

$$K_{t+\Delta t} = (1 + \Delta t \cdot r) \cdot K_t \quad \text{oder} \quad \frac{K(t+\Delta t) - K(t)}{\Delta t} = r \cdot K(t) .$$

Führt man den Grenzübergang $\Delta t \rightarrow 0$ durch, erhält man

$$\dot{K}(t) = r \cdot K(t) ,$$

so dass $\frac{\dot{K}(t)}{K(t)}$, d.h. die *momentane Wachstumsrate* des Kapitals, gleich r ist. Wie aus der Finanzmathematik bekannt ist, gilt $K(t) = K_0 \cdot e^{rt}$.

- iii) Dieselbe Gleichung wie vorher $\dot{y}(t) = r \cdot y(t)$ beschreibt das Wachstumsverhalten einer Population y mit konstanter Wachstumsrate r (*Malthus-Gesetz*). Die Lösung $y(t) = y_0 \cdot e^{rt}$ zeigt, dass diese Population exponentiell ohne Schranke wächst. Offensichtlich ist dieses Malthus-Modell relativ simpel. Ein realistischeres Modell sollte auch wachstumshemmende Faktoren berücksichtigen. Weil mit zunehmender Population z.B. ein zunehmender Ressourcenverbrauch einher geht, ist es naheliegend anzunehmen, dass die Wachstumsrate der Population abnimmt, wenn die Populationsgröße y anwächst. Der einfachste Fall besteht in der Annahme, dass die momentane Wachstumsrate $\frac{\dot{y}}{y}$ affin linear und fallend in Abhängigkeit von y ist. In diesem Fall entsteht das *Wachstumsmodell von Verhulst (logistisches Modell)*¹:

$$\frac{\dot{y}}{y} = a - b \cdot y \quad \text{oder} \quad \dot{y} = y \cdot (a - b \cdot y) \quad \text{mit } a, b \geq 0$$

□

2.2 Definition

Sei $F : \mathbb{R} \times \mathbb{R}^{m+1} \rightarrow \mathbb{R}$ eine Funktion². Eine Gleichung der Form

$$F(t, y(t), \dot{y}(t), \ddot{y}(t), \dots, y^{(m)}(t)) = 0 , \tag{2.1}$$

die eine Beziehung zwischen den Werten der Funktion y , dem Argument t und den Ableitungen $\dot{y}, \ddot{y}, \dots, y^{(m)}$ herstellt, heißt *implizite Differentialgleichung der Ordnung m*. Eine *Lösung der DGL* ist eine auf einem offenen Intervall I definierte Funktion φ , deren Ableitungen $\dot{\varphi}(t), \dots, \varphi^{(m)}(t)$ existieren und die der DGL (2.1) für alle $t \in I$ genügt, d.h. mit

$$F(t, \varphi(t), \dot{\varphi}(t), \ddot{\varphi}(t), \dots, \varphi^{(m)}(t)) = 0 .$$

¹Zur Lösung siehe Beispiel 2.12 ii).

²Die nachfolgenden Ergebnisse gelten auch, wenn F nur auf Teilmengen des $\mathbb{R} \times \mathbb{R}^{m+1}$ definiert ist.

Existiert eine Funktion $f : \mathbb{D} \times \mathbb{R}^m \rightarrow \mathbb{R}$, so dass die Dgl (2.1) äquivalent ist zur Gleichung³

$$y^{(m)}(t) = f(t, y(t), \dot{y}(t), \ddot{y}(t), \dots, y^{(m-1)}(t)) ,$$

nennt man die DGL (2.1) eine *explizite Differentialgleichung der Ordnung m*. $\square$

2.3 Bemerkung

- i) Analog zu Definition 1.11 wird ein *Differentialgleichungssystem* durch eine Abbildung $\mathbf{F} : \mathbb{R} \times (\mathbb{R}^n)^{m+1} \rightarrow \mathbb{R}^n$ definiert, als eine Gleichung der Form

$$\mathbf{F}(t, \mathbf{y}(t), \dot{\mathbf{y}}(t), \ddot{\mathbf{y}}, \dots, \mathbf{y}^{(m)}(t)) = \mathbf{0} . \quad (2.2)$$

- ii) Man spricht von einer *autonomen* oder *zeitunabhängigen* DGL, wenn $\mathbf{F}$ nicht explizit von t abhängt⁴, d.h. wenn gilt

$$\mathbf{F}(t, \mathbf{y}(t), \dot{\mathbf{y}}(t), \ddot{\mathbf{y}}(t), \dots, \mathbf{y}^{(m)}(t)) = \mathbf{F}(\mathbf{y}(t), \dot{\mathbf{y}}(t), \ddot{\mathbf{y}}(t), \dots, \mathbf{y}^{(m)}(t))$$

- iii) Oft ist es nützlich die gegebene DGL in eine äquivalente Integralgleichung umzuwandeln. Ist $\mathbf{f}$ stetig, so liefern die Regeln über Integration:

Die Funktion $\mathbf{y}(t)$ ist *genau dann* eine Lösung der *Anfangswertaufgabe* (AWA)

$$\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}) \quad \text{mit} \quad \mathbf{y}(t_0) = \mathbf{y}_0 ,$$

wenn sie der folgenden Integralgleichung⁵ genügt:

$$\mathbf{y}(t) = \mathbf{y}_0 + \int_{t_0}^t \mathbf{f}(x, \mathbf{y}(x)) dx .$$

- iv) Mit Hilfe des *Differentialoperators* $D := \frac{d}{dt}$ sind analog zum Differenzenoperator alternative Darstellungen für DGLen möglich, z.B.

$$a\ddot{y} + b\dot{y} + cy = (aD^2 + bD + c)y \quad \square$$

2.4 Beispiel

Betrachten wir zwei Populationen y_1 und y_2 , die um die gleiche Nahrungsquelle oder die gleiche ökologische Nische miteinander konkurrieren, oder noch dramatischer, in der eine

³Dies besagt, dass F nach der letzten Variable auflösbar ist. Bei nicht auflösbaren Gleichungen, liefert der implizite Funktionensatz Kriterien dafür, wann eine implizite DGL wenigstens lokal mit einer expliziten übereinstimmt.

⁴Jede nichtautonome DGL $\dot{y} = f(y, t)$ kann in ein äquivalentes autonomes DGL-System transformiert werden, z.B. $\dot{y} = f(y, r)$, $\dot{r} = 1$ Aus diesem Grund genügt es, autonome DGLen zu betrachten (siehe Bemerkung 3.22).

⁵Das Integral ist immer komponentenweise zu verstehen, wenn $n > 1$ gilt.

Population die Hauptnahrungsquelle für die andere ist. Wir nehmen zunächst an, dass jede Population ihrer eigenen logistischen Gleichung genügt, d.h.

$$\frac{\dot{y}_1}{y_1} = a_1 - b_1 y_1 \quad \text{und} \quad \frac{\dot{y}_2}{y_2} = a_2 - b_2 y_2 .$$

Um eine Wechselwirkung zu modellieren, nehmen wir weiter an, dass die Anwesenheit einer Spezies einen positiven oder negativen Effekt auf die Wachstumsrate der anderen hat. Betrachtet man wieder den einfachsten, d.h. (affin) linearen Fall, so gilt:

$$\frac{\dot{y}_1}{y_1} = a_1 - b_1 y_1 - c_1 y_2 \quad \text{und} \quad \frac{\dot{y}_2}{y_2} = a_2 - b_2 y_2 - c_2 y_1 .$$

Die Interaktion der Populationen wird also durch ein DGL-System 1. Ordnung beschrieben:

$$\begin{aligned} \dot{y}_1 &= y_1(a_1 - b_1 y_1 - c_1 y_2) \\ \dot{y}_2 &= y_2(a_2 - b_2 y_2 - c_2 y_1) \end{aligned}$$

Die Art der Interaktion wird durch das Vorzeichen der Größen c_i bestimmt. Haben c_1 und c_2 z.B. unterschiedliche Vorzeichen, liegt eine parasitäre Situation vor. $\square$

Jede explizite DGL der Ordnung n lässt sich in ein DGL-System erster Ordnung überführen. Diese Aussage ist (samt Beweis) das Analogon zu Lemma 1.13 aus Kapitel 1 über ΔG Len.

2.5 Lemma

Jede explizite DGL m -ter Ordnung lässt sich in ein äquivalentes DGL-System erster Ordnung überführen, d.h. es gilt: Die Lösungen der DGL

$$y^{(m)}(t) = f(t, y(t), \dot{y}(t), \ddot{y}(t), \dots, y^{(m-1)}(t))$$

entsprechen bijektiv den Lösungen des folgenden DGL-Systems 1.-ter Ordnung

$$\begin{aligned} \dot{\phi}_1 &= \phi_2 \\ \dot{\phi}_2 &= \phi_3 \\ \vdots &= \vdots \\ \dot{\phi}_{m-1} &= \phi_m \\ \dot{\phi}_m &= f(t, \phi_1, \phi_2, \dots, \phi_m) \end{aligned}$$

$\square$

2.6 Bemerkung

Der letzte Satz gilt analog für DGLen vektorwertiger Abbildungen $\mathbf{y} : \mathbb{R} \rightarrow \mathbb{R}^n$. Das äquivalente DGL-System besteht dann allerdings aus $m \cdot n$ Gleichungen. $\square$

2.7 Beispiel

Die DGL $y^{(3)} + 2\dot{y} + y = 0$ ist nach Lemma 2.5 äquivalent zu dem DGLs-System

$$\begin{aligned} \dot{\phi}_1 &= \phi_2 \\ \dot{\phi}_2 &= \phi_3 \\ \dot{\phi}_3 &= -2\phi_2 - \phi_1 \end{aligned}$$

$\square$

Wie in der Theorie der Δ GLen untersuchen wir die folgenden Fragen:

- (1) Man gebe Kriterien für die Existenz von Lösungen.
- (2) Man verschaffe sich einen Überblick über die Lösungsgesamtheit einer DGL.
- (3) Man gebe Bedingungen an, die die Eindeutigkeit der Lösung garantieren.

Wegen Lemma 2.5 genügt es, im Folgenden DGL-Systeme erster Ordnung zu betrachten.

2.8 Satz (Existenzsatz von PEANO, lokale Form)

Ist $\mathbf{f} : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ stetig, so besitzt die AWA $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$, $\mathbf{y}(t_0) = \mathbf{y}_0$ wenigstens eine, auf einem offenen Intervall definierte Lösung. $\square$

Beweis:

Auf einen Beweis dieses Satzes, der hochgradig nichttrivial und sehr technisch ist, kann hier verzichtet werden, da er für unsere Zwecke nicht sehr aufschlussreich ist. Er zählt zu den Standardsätzen im Bereich der gewöhnlichen Differentialgleichungen und kann daher in jedem Lehrbuch der Analysis gefunden werden. $\blacksquare$

Wie man an dem folgenden Beispiel erkennen kann, reicht die Stetigkeit von $\mathbf{f}$ noch nicht einmal aus, um die lokale (d.h. in einer Umgebung des Startpunktes) Eindeutigkeit der Lösung einer AWA zu garantieren.

2.9 Beispiel

Für alle $a, b \in \mathbb{R} \cup \{-\infty, +\infty\}$ mit $a \leq 0 \leq b$ sind die Funktionen $\varphi_{a,b}$ mit

$$\varphi_{a,b}(t) := \begin{cases} (t-a)^3 & \text{für } t < a \\ 0 & \text{für } a \leq t \leq b \\ (t-b)^3 & \text{für } b < t \end{cases}$$

Lösungen der AWA

$$\dot{y} = f(t, y) = 3 \cdot y^{\frac{2}{3}}, \quad y(0) = 0.$$

Insbesondere erhält man z.B. die Lösungen

$$\varphi_1(t) \equiv 0 \quad (a = -\infty, b = +\infty) \quad \text{und} \quad \varphi_2(t) = t^3 \quad (a = b = 0).$$

Interessant ist nicht so sehr, dass unendlich viele linear unabhängige Lösungen für diese AWA existieren, sondern dass die beiden angegebenen Lösungen noch nicht einmal in einer kleinen Umgebung des „Startpunktes“ $y(0)=0$ übereinstimmen. Der Grund ist, dass die Funktion $f(t, y) := y^{\frac{2}{3}}$ zwar stetig, aber im Punkt $y=0$ nicht differenzierbar ist⁶. $\square$

⁶Die Steigung ist unendlich, d.h. die Tangente verläuft parallel zur vertikalen Achse.

Der folgende Satz zeigt, dass mit einer kleinen zusätzlichen Forderung, die insbesondere im Fall der stetigen Differenzierbarkeit von f erfüllt ist, zumindest lokal eine Eindeutigkeit der Lösung erreicht werden kann.

2.10 Satz (PICARD-LINDELÖF, lokale Form)

Ist $f : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ stetig und genügt f in einer Umgebung U des Punktes $(t_0, \mathbf{y}_0)$ einer *LIPSCHITZ-Bedingung*, d.h., existiert eine Konstante L mit

$$\|f(t, \mathbf{y}) - f(t, \mathbf{z})\| \leq L \cdot \|\mathbf{y} - \mathbf{z}\| \quad \text{für alle } (t, \mathbf{y}), (t, \mathbf{z}) \text{ in } U,$$

so ist die Lösung der AWA

$$\dot{\mathbf{y}} = f(t, \mathbf{y}), \quad \mathbf{y}(t_0) = \mathbf{y}_0$$

lokal eindeutig, d.h. je zwei Lösungen stimmen in einer Umgebung von t_0 überein. $\square$

Beweis:

Wir betrachten nur den Fall $n=1$, da für $n>1$ der Beweis fast identisch verläuft. Die Existenz einer Lösung ergibt sich aus dem Existenzsatz 2.8. Zum Beweis der lokalen Eindeutigkeit verwenden wir folgenden Hilfssatz:

Ist $h : [a, b] \rightarrow \mathbb{R}$ stetig differenzierbar und $K > 0$ eine Konstante mit der Eigenschaft, dass $h'(x) \leq K \cdot h(x)$ für alle $x \in [a, b]$ ist, so folgt

$$h(x) \leq h(a) \cdot e^{K(x-a)}.$$

Sind $\varphi : I_1 \rightarrow \mathbb{R}$ und $\phi : I_2 \rightarrow \mathbb{R}$ Lösungen der AWA, wobei I_1, I_2 offene und t_0 enthaltende Intervalle sind, so betrachte ein Intervall $[\alpha, \beta]$ mit $t_0 \in]\alpha, \beta[$ und $[\alpha, \beta] \subset I_1 \cap I_2$. Wir definieren nun die Funktion $h : [t_0, \beta] \rightarrow \mathbb{R}$ durch

$$h(t) := (\varphi(t) - \phi(t))^2.$$

Dann gilt:

$$\begin{aligned} |\dot{h}(t)| &= 2|\dot{\varphi}(t) - \dot{\phi}(t)||\varphi(t) - \phi(t)| \\ &= 2|f(t, \varphi) - f(t, \phi)||\varphi(t) - \phi(t)| \\ &\leq 2L|\varphi(t) - \phi(t)|^2 = 2Lh(t) \end{aligned}$$

Wendet man den Hilfssatz mit $K = 2L$, $a = t_0$ und $b = \beta$ an, folgt

$$0 \leq h(t) \leq h(t_0) e^{2L(t-t_0)} = \underbrace{(\varphi(t_0) - \phi(t_0))^2}_{= y_0 - y_0 = 0} e^{2L(t-t_0)} = 0.$$

Daher gilt $h = 0$, d.h. $\varphi = \phi$.

Analog zeigt man, dass auf $[\alpha, t_0]$ die Gleichheit $\varphi = \phi$ gilt, indem man mit der Funktion $h(t) := (\varphi(t_0 - t) - \phi(t_0 - t))^2$ arbeitet. ■

2.11 Bemerkung

- i) Interessanterweise gibt es auch einen Beweis des letzten Satzes, der ein explizites Verfahren zur Konstruktion einer Lösung der AWA $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$, $\mathbf{y}(t_0) = \mathbf{y}_0$ liefert (*successive Approximation*). Man betrachtet hierzu folgende Abbildung T , die auf $C(I, \mathbb{R}^n)$ - den auf dem Intervall I stetigen Abbildungen - definiert ist, durch⁷

$$T(\mathbf{h})(t) := \mathbf{y}_0 + \int_{t_0}^t \mathbf{f}(s, \mathbf{h}(s)) ds .$$

T ist auf einer geeigneten Teilmenge von $C(I, \mathbb{R}^n)$ eine „Kontraktion“, d.h. sind $\mathbf{g}, \mathbf{h} \in C(I, \mathbb{R}^n)$, so gilt mit $\|f\| := \sup_{t \in I} |f(t)|$:

$$\|T(\mathbf{h}) - T(\mathbf{g})\| \leq K \|\mathbf{h} - \mathbf{g}\| \quad (K < 1 \text{ konstant}) .$$

Nach dem Banachschen Fixpunktsatz (siehe Satz B.1 des Anhangs) konvergiert für jede beliebige „Startfunktion“ $\mathbf{h}_0$ die Funktionenfolge $(\mathbf{h}_\ell)_{\ell=0}^\infty$ mit $\mathbf{h}_\ell := T(\mathbf{h}_{\ell-1})$ gegen eine Funktion φ mit der Stationaritätseigenschaft $\varphi = T(\varphi)$ (*Fixpunkt*), d.h. es gilt:

$$\varphi(t) = \mathbf{y}_0 + \int_{t_0}^t \mathbf{f}(s, \varphi(s)) ds .$$

Nach Bemerkung 2.3 iii) ist daher φ eine Lösung der DGL.

- ii) In der Praxis tritt sehr häufig die Frage auf, wie „störungsanfällig“ eine vorgegebene AWA ist, d.h. wie verändert sich die Lösung einer AWA, wenn man den Startwert $\mathbf{y}_0 = \mathbf{y}(t_0)$ und eventuell auch die DGL, d.h. die Funktion f , die die DGL beschreibt, ändert. Vereinfacht kann man sagen, dass sich die AWA gegenüber Veränderungen stetig verhält, wenn f einer Lipschitz-Bedingung genügt. Wir behandeln diese Fragestellung nur hinsichtlich der Störung im Startwert $\mathbf{y}_0$ in Kapitel 3 (Stabilität).
- iii) Die Lipschitz-Bedingung ist automatisch erfüllt, wenn f stetig partiell differenzierbar ist und die Ableitungen beschränkt sind. ■

Die explizite Lösung einer DGL ist nur in den seltensten Fällen möglich. Z.B. ist die DGL

$$\dot{y} = f(t, y) := e^{(-t^2)} \tag{2.3}$$

offensichtlich äquivalent zu

$$y = \int \dot{y} dt = \int e^{(-t^2)} dt .$$

⁷Auf die genaue Angabe von I sei hier verzichtet.

DGL (2.3) kann nicht explizit gelöst werden, da keine Stammfunktion explizit angegeben werden kann. Dieses Beispiel zeigt auch, dass DGLen eine Verallgemeinerung des klassischen Integrationsproblems darstellen⁸. Dennoch gibt es wie bei der Berechnung von Stammfunktionen einige Fälle, in denen eine vorgegebene DGL explizit lösbar ist.

2.1. Spezielle Typen von DGLen erster Ordnung

Wir gehen in diesem Abschnitt von einer DGL der Form $\dot{y} = F(t, y)$ aus und stellen in Abhängigkeit vom Aussehen der Funktion F verschiedene Lösungsmöglichkeiten vor.

2.1.1. DGLen mit getrennten Variablen

DGLen dieses Typs sind von der Form

$$\dot{y} = f(t) \cdot g(y) \quad (2.4)$$

Division durch $g(y)$, Integration nach t und Beachtung der Substitutionsregel liefert

$$\int \frac{1}{g(y)} \dot{y} dt = \int \frac{1}{g(y)} dy .$$

Daher ist die allgemeine Lösung y der DGL (2.4) implizit bestimmt durch die Gleichung

$$\int \frac{1}{g(y)} dy = \int f(t) dt + \text{Konstante}$$

2.12 Beispiel

i) Wir betrachten die DGL

$$\dot{y} = \frac{1}{t} \cdot \frac{-4y^2 + 6y - 7}{4y - 3}$$

mit $f(t) = \frac{1}{t}$ und $g(y) = \frac{-4y^2 + 6y - 7}{4y - 3}$. Einsetzen ergibt für y die Gleichung

$$-\frac{1}{2} \ln |-4y^2 + 6y - 7| = \ln(t) + C \iff 4y^2 - 6y + 7 = \frac{1}{t^2} \cdot \bar{C}$$

ii) Betrachte die DGL $\dot{y} = a \cdot (y-b) \cdot (y-c)$ mit $a \neq 0$. Division durch $(y-b) \cdot (y-c)$ und anschließende Integration mittels Partialbruchzerlegung liefert

$$\int \frac{dy}{(y-b) \cdot (y-c)} = \int a dt \iff \frac{1}{b-c} \cdot \int \left(\frac{1}{y-b} - \frac{1}{y-c} \right) dy = a \cdot t + \text{Konstante},$$

was nach einigen kleineren Umformungen zur Lösung

$$y(t) = \frac{c-b}{e^{(at+K) \cdot (b-c)} - 1} + c$$

führt, wobei K eine Integrationskonstante ist. □

⁸Deshalb wird die Lösung einer DGL häufig auch als Integral bezeichnet.

2.1.2. DGLen mit homogenen Funktionen

DGLen dieses Typs sind von der Form

$$\dot{y} = \frac{P(t, y)}{Q(t, y)}, \quad (2.5)$$

wobei P und Q homogene Funktionen desselben Grades r sind⁹, d.h. es gilt

$$P(\lambda t, \lambda y) = \lambda^r P(t, y) \quad \text{und} \quad Q(\lambda t, \lambda y) = \lambda^r Q(t, y).$$

Mit der Transformation $z = \frac{y}{t}$ folgt

$$y = tz \quad \text{und} \quad \dot{y} = z + t\dot{z}.$$

Setzt man dies in die DGL (2.5) ein, ergibt sich

$$\dot{y} = z + t\dot{z} = \frac{P(t, tz)}{Q(t, tz)} = \frac{t^r P(1, z)}{t^r Q(1, z)},$$

also die äquivalente DGL mit getrennten Variablen

$$\dot{z} = \frac{1}{t} \left[\frac{P(1, z)}{Q(1, z)} - z \right].$$

2.13 Beispiel

Man betrachte die DGL

$$\dot{y} = \frac{3y - 7t}{4y - 3t}.$$

Offensichtlich liegt eine DGL mit homogenen Funktionen vom Grad $r = 1$ vor, mit $P(t, y) = 3y - 7t$ und $Q(t, y) = 4y - 3t$.

Die Transformation $z = \frac{y}{t}$ führt zur äquivalenten DGL aus Beispiel 2.12:

$$\dot{z} = \frac{1}{t} \left[\frac{3z - 7}{4z - 3} - z \right] = \frac{1}{t} \cdot \frac{-4z^2 + 6z - 7}{4z - 3}$$

□

2.1.3. Gebrochen rationale DGLen

DGLen dieses Typs haben folgendes Aussehen:

$$\dot{y} = \frac{at + by + d}{At + By + D}, \quad \text{wobei } a, b, d, A, B, D \text{ konstant.} \quad (2.6)$$

⁹Nicht zu verwechseln mit den homogenen linearen DGLen.

i) Ist $aB - bA \neq 0$, so ist die DGL (2.6) äquivalent zur homogenen DGL

$$\frac{dz}{ds} = \frac{as + bz}{As + Bz}.$$

Man führe hierzu die Transformationen $s := t - \tilde{t}$ und $z := y - \tilde{y}$ durch. Hierbei sind $\tilde{t}, \tilde{y}$ die eindeutig bestimmten Lösungen des linearen Gleichungssystems

$$\begin{aligned} a\tilde{t} + b\tilde{y} + d &= 0 \\ A\tilde{t} + B\tilde{y} + D &= 0 \end{aligned}$$

ii) Ist $aB - bA = 0$, so ist DGL (2.6) äquivalent zu einer separablen DGL. Hierzu führe man die Transformation $z := at + by$ (bzw. $z := At + By$ falls $a=b=0$) und $s := t$ durch.

2.14 Beispiel

Wir betrachten die DGL $\dot{y} = \frac{-7t + 3y - 2}{-3t + 4y - 5}$. Es gilt $aB - bA = -19 \neq 0$ und $\tilde{t} = \frac{7}{19}$, $\tilde{y} = \frac{29}{19}$. Die Transformationen $s := t - \frac{7}{19}$ und $z := y - \frac{29}{19}$ liefern die äquivalente DGL

$$\frac{dz}{ds} = \frac{3z - 7s}{4z - 3s}$$

□

2.1.4. Exakte DGLen

Stellt man eine DGL erster Ordnung $Q(t, y)\dot{y} + P(t, y)=0$ mit Hilfe der Differentiale dt und dy dar, erhält man die „Differentialform“ $P(t, y)dt + Q(t, y)dy=0$. Die Hoffnung ist nun, dass diese Differentialform das totale Differential dF einer Funktion $F(t, y)$ ist, weil dann aus $dF(t, y)=0$ folgt, dass F auf dem Graphen der Funktion y , d.h. auf den Punktepaaen (t, y) , konstant ist, woraus für y die implizite Gleichung $F(t, y)=C$ resultiert.

Zunächst ist die Frage zu beantworten, wie man erkennen kann, ob eine Funktion F mit $dF = P(t, y)dt + Q(t, y)dy$ existiert? Wäre diese Gleichung für eine Funktion F erfüllt, so müsste offensichtlich gelten:

$$\frac{\partial F}{\partial t} = P(t, y) \quad \text{und} \quad \frac{\partial F}{\partial y} = Q(t, y).$$

Nimmt man an, dass F zweimal stetig differenzierbar ist, so gilt bekanntlich

$$\frac{\partial^2 F}{\partial t \partial y} = \frac{\partial^2 F}{\partial y \partial t},$$

woraus sich $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial t}$ in diesem Fall als notwendige Bedingung ergibt. Sofern der Definitionsbereich *einfach zusammenhängend* ist¹⁰, was wir im Folgenden annehmen, ist dies auch hinreichend für die Existenz einer Funktion F mit $dF = P(t, y)dt + Q(t, y)dy$.

¹⁰Dies bedeutet, dass der Rand des Definitionsbereiches zusammenhängend ist.

2.15 Definition

Die DGL

$$Q(t, y)\dot{y} + P(t, y) = 0 \iff Q(t, y)dy + P(t, y)dt = 0$$

heißt *exakt*, wenn $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial t}$ gilt. □

2.16 Beispiel

$$\underbrace{(t - y)}_P dt + \underbrace{(y^{-2} - t)}_Q dy = 0$$

Es gilt $\frac{\partial P}{\partial y} = -1 = \frac{\partial Q}{\partial t}$, so dass eine exakte DGL vorliegt. □

2.17 Satz

Ist die DGL $Q(t, y)\dot{y} + P(t, y) = 0$ exakt, so genügt die gesuchte Funktion y der Gleichung

$$\int_{t_0}^t P(s, y)ds + \int_{y_0}^y Q(t_0, z)dz = \text{Konstante} .$$
□

Beweis:

Aufgrund der Exaktheitsbedingung $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial t}$ existiert eine Funktion $F(t, y)$ mit

$$(*) \quad \frac{\partial F}{\partial t} = P(t, y) \quad \text{und} \quad (**) \quad \frac{\partial F}{\partial y} = Q(t, y) .$$

Integriert man die Gleichung $(*)$ nach der ersten Variablen, so folgt

$$F(t, y) = \int_{t_0}^t P(s, y)ds + F(t_0, y) . \tag{2.7}$$

Differenziert man Gleichung (2.7) partiell nach y , so ergibt sich

$$\begin{aligned} Q(t, y) &\stackrel{(**)}{=} \frac{\partial F}{\partial y}(t, y) = \int_{t_0}^t \frac{\partial P}{\partial y}(s, y)ds + \frac{d}{dy}F(t_0, y) \\ &= \int_{t_0}^t \frac{\partial Q}{\partial s}(s, y)ds + \frac{d}{dy}F(t_0, y) \\ &= Q(t, y) - Q(t_0, y) + \frac{d}{dy}F(t_0, y) . \end{aligned}$$

Damit gilt $\frac{d}{dy}F(t_0, y) = Q(t_0, y)$, d.h.

$$F(t_0, y) - F(t_0, y_0) = \int_{y_0}^y \frac{d}{dz}F(t_0, z)dz = \int_{y_0}^y Q(t_0, z)dz \tag{2.8}$$

Einsetzen von Gleichung (2.8) in Gleichung (2.7) liefert

$$F(t, y) = \int_{t_0}^t P(s, y) ds + \int_{y_0}^y Q(t_0, z) dz + F(t_0, y_0) .$$

Wegen $dF = 0$ ist F konstant und man erhält für y die Bestimmungsgleichung

$$\int_{t_0}^t P(s, y) ds + \int_{y_0}^y Q(t_0, z) dz = \text{konstant} . \quad \blacksquare$$

2.18 Beispiel

Für die exakte DGL

$$(t - y)dt + (y^{-2} - t)dy = 0$$

aus Beispiel 2.16 erhält man mit Hilfe des letzten Satzes

$$\begin{aligned} \int_{t_0}^t P(s, y) ds + \int_{y_0}^y Q(t_0, z) dz &= \int_{t_0}^t (s - y) ds + \int_{y_0}^y (z^{-2} - t_0) dz \\ &= \left[\frac{1}{2} s^2 - y s \right]_{s=t_0}^{s=t} + \left[-z^{-1} - t_0 z \right]_{z=y_0}^{z=y} \\ &= \frac{1}{2} t^2 - yt - \frac{1}{y} + \left(-\frac{1}{2} t_0^2 + \frac{1}{y_0} + t_0 y_0 \right) \end{aligned}$$

Die gesuchte Funktion y genügt daher der Gleichung

$$\frac{1}{2} t^2 - yt - \frac{1}{y} = C = \text{konstant} \quad \square$$

Ist die DGL $Q(t, y)dy + P(t, y)dt = 0$ nicht exakt, d.h. gilt $\frac{\partial P}{\partial y} \neq \frac{\partial Q}{\partial t}$, versucht man eine Funktion $M(t, y)$ zu bestimmen, so dass die DGL

$$M(t, y)Q(t, y)dy + M(t, y)P(t, y)dt = 0$$

exakt ist. Der *integrierende Faktor* M ist also derart zu wählen, dass $\frac{\partial(MP)}{\partial y} = \frac{\partial(MQ)}{\partial t}$ gilt, d.h. M muss der folgenden partiellen DGL genügen:

$$M \left(\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial t} \right) = Q \frac{\partial M}{\partial t} - P \frac{\partial M}{\partial y} \quad (2.9)$$

Häufig kann die partielle DGL (2.9) gelöst werden, wenn man M z.B. nur als Funktion von t , y , $t \cdot y$ oder $\frac{y}{t}$ ansetzt und so die partielle DGL vereinfacht.

2.19 Beispiel

Gegeben sei die DGL

$$\underbrace{(ty^2 - y^3)}_P dt + \underbrace{(1 - ty^2)}_Q dy = 0 \quad (2.10)$$

Dann gilt

$$\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial t} = 2ty - 2y^2 \neq 0 .$$

Setzt man den integrierenden Faktor in der Form $M = M(y)$ an, d.h. gilt $\frac{\partial M}{\partial t} = 0$, erhält man aus Gleichung (2.9):

$$M \cdot \underbrace{\frac{1}{P} \left(\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial t} \right)}_{\frac{2}{y}} = -\frac{\partial M}{\partial y}$$

und damit die gewöhnliche DGL mit getrennten Variablen

$$\frac{2M}{y} = -\frac{dM}{dy} .$$

Eine spezielle Lösung dieser DGL lautet $M(y) = y^{-2}$. Die mit M multiplizierte DGL (2.10) lautet

$$(t - y)dt + (y^{-2} - t)dy = 0 .$$

Sie ist nach Beispiel 2.18 exakt und besitzt die allgemeine Lösung

$$\frac{1}{2}t^2 - yt - \frac{1}{y} = C = \text{konst.} \iff t^2y - 2y^2t - 2 - Cy = 0 .$$

□

2.1.5. Bernoullische Differentialgleichungen

$$\dot{y}(t) + p(t)y(t) = q(t)y^k(t) \quad \text{mit } k \in \mathbb{N}$$

Bernoulli-DGLen lassen sich aus linear inhomogenen DGLen 1.-ter Ordnung ableiten.

Betrachten wir die inhomogene DGL erster Ordnung $\dot{y}(t) + p(t)y(t) = q(t)$, so stellt die Inhomogenität $q(t)$ eine „Störung“ der homogenen DGL $\dot{y}(t) + p(t)y(t) = 0$ dar. Es sei hier angemerkt, dass $q(t)$ nicht etwa eine äußere Vorgabe für die gesuchte Lösungsfunktion $y(t)$ selbst darstellt. Dies kann schon auf Grund der unterschiedlichen „ökonomischen“ oder „physikalischen“ Einheiten von $y(t)$ und $q(t)$ nicht stimmen. Vielmehr stellt $q(t)$ einen äußeren Zwang dar, dem die gesamte linke Seite der DGL „unterliegt“, d.h. die Lösung $y(t)$ wird von $q(t)$ beeinflusst und wird im allgemeinen ihr Aussehen im Vergleich zur Lösung $y_h(t)$ der homogenen DGL signifikant ändern. In vielen ökonomischen und physikalischen Beispielen findet man nun eine Abhängigkeit des äußeren Zwangs von Potenzen von $y(t)$. Man erhält also $\dot{y}(t) + p(t)y(t) = q(t)y^k(t)$ mit $k \in \mathbb{N}$. Formal handelt es sich hier natürlich um eine *inhomogene lineare* ($k = 0$) bzw. *homogene lineare* ($k = 1$) oder eine *nichtlineare* ($k > 1$) DGL erster Ordnung. Für $k > 1$ spricht man von einer Bernoullischen Differentialgleichung.

Um eine Lösung zu finden, substituiert man y durch $z^{\frac{1}{m}}$ derart, dass eine inhomogene lineare DGL (siehe Abschnitt 2.2) entsteht. Ausgehend von

$$\dot{y} + p y = q y^k$$

erhält man nach Substitution $y := z^{\frac{1}{m}}$

$$\frac{d}{dt}(z^{\frac{1}{m}}) + p z^{\frac{1}{m}} = q z^{\frac{k}{m}}$$

und damit

$$\frac{1}{m} \dot{z} z^{\frac{1-m}{m}} + p z^{\frac{1}{m}} = q z^{\frac{k}{m}}.$$

Multiplikation mit $z^{\frac{m-1}{m}}$ liefert

$$\frac{1}{m} \dot{z} + p z = q z^{\frac{k+m-1}{m}}.$$

Wählt man $m = 1 - k$, ist die rechte Seite unabhängig von z und es gilt

$$\frac{1}{1-k} \dot{z} + p z = q.$$

Die Bernoullische Differentialgleichung $\dot{y}(t) + p(t)y(t) = q(t)y^k(t)$ lässt sich also durch die Substitution $y(t) := z^{\frac{1}{1-k}}(t)$ in eine *lineare* inhomogene DGL erster Ordnung überführen.

2.1.6. Riccatische Differentialgleichungen

Die Bernoulli DGL stellt einen recht einfachen Typ von *homogenen nichtlinearen* DGLen dar. Wählt man einen ähnlichen Ansatz der Form $\dot{y}(t) + p(t)y(t) + q(t)y^2(t) = r(t)$, d.h. setzt man in der Bernoulli DGL $k = 2$ und fügt eine Inhomogenität hinzu, so erhält man die sogenannte Riccatische DGL. Findet man nun z.B. durch Raten eine spezielle Lösung $u(t)$, so ergibt sich mit der Substitution $y(t) := u(t) + v(t)$

$$\underbrace{\dot{u} + p u + q u^2}_r + \dot{v} + (p + 2q u) v + q v^2 = r$$

und damit nach Kürzen von r die Bernoullische DGL

$$\dot{v}(t) + (p(t) + 2q(t)u(t))v(t) = -qv^2$$

2.2. Lineare DGLen

Wie im ersten Kapitel wollen wir uns nun auf lineare DGLen einschränken.

2.20 Definition

Ein DGL-System $\mathbf{F}(t, \mathbf{y}, \dot{\mathbf{y}}, \dots, \mathbf{y}^{(m)}) = \mathbf{0}$ der Ordnung m heißt *linear*, wenn $\mathbf{F}$ in den Variablen $\mathbf{y}, \dot{\mathbf{y}}, \dots, \mathbf{y}^{(m)}$ linear ist. Es lässt sich dann in folgende äquivalente Form bringen:

$$\mathbf{A}_m(t)\mathbf{y}^{(m)} + \mathbf{A}_{m-1}(t)\mathbf{y}^{(m-1)} + \dots + \mathbf{A}_0(t)\mathbf{y} = \mathbf{b}(t) \quad (2.11)$$

$\mathbf{y}$ und die Inhomogenität $\mathbf{b}$ seien dabei $\mathbb{R}^n$ -wertige Funktionen und $\mathbf{A}_j(t)$ quadratische $(n \times n)$ -Matrizen mit zeitabhängigen Koeffizienten. Sind die $\mathbf{A}_j$ konstant, d.h. unabhängig

von t , so heißt das DGL-System (2.11) ein lineares DGL-System mit konstanten Koeffizienten. Das System (2.11) heißt *homogen*, falls $\mathbf{b} \equiv \mathbf{0}$ gilt, andernfalls *inhomogen*. Für den Spezialfall $n=1$ erhält man die Definition einer *linearen* DGL m -ter Ordnung. $\square$

Analog zu Satz 1.19 in Abschnitt 1.2 von Kapitel 1 ergibt sich der folgende Existenzsatz, den wir ohne Beweis zitieren. Er ist eine Anwendung der zu Beginn dieses Kapitels gegebenen Existenz- und Eindeigkeitssätze 2.8 und 2.10.

2.21 Satz

Sind $\mathbf{b} : I \rightarrow \mathbb{R}^n$ und $\mathbf{A} : I \rightarrow \mathbb{R}^{n \times n}$ stetige Abbildungen auf dem Intervall I , so besitzt die lineare AWA

$$\dot{\mathbf{y}} - \mathbf{A}(t)\mathbf{y} = \mathbf{b}(t), \quad \mathbf{y}(t_0) = \mathbf{y}_0. \quad (2.12)$$

eine eindeutige auf ganz I definierte Lösung $\varphi : I \rightarrow \mathbb{R}^n$ $\square$

Bevor wir zu Lösungsverfahren kommen, wollen wir uns zunächst überlegen, wie man eine homogene lineare DGL vereinfachen kann, wenn man bereits eine Lösung φ kennt. Mittels Variation der Konstanten wie im Beweis von Satz 1.23 b) entsteht durch die Substitution $y := \varphi \cdot z$ eine lineare DGL, deren Ordnung um 1 reduziert ist¹¹.

2.22 Satz (Reduktionsverfahren von d'Alembert für lineare DGLen)

Seien $a_0, \dots, a_m$ auf einem Intervall stetige Funktionen und φ eine nicht triviale Lösung der homogenen linearen DGL

$$\Lambda(y) := a_m(t)y^{(m)} + a_{m-1}(t)y^{(m-1)} + \dots + a_1(t)\dot{y} + a_0(t)y = 0.$$

Ist ψ eine nicht triviale Lösung der homogenen linearen DGL

$$\Lambda_1(z) := b_{m-1}(t)z^{(m-1)} + \dots + b_1(t)\dot{z} + b_0(t)z = 0,$$

wobei die Koeffizienten $b_j(t)$ für $j = 0, 1, \dots, m-1$ definiert sind durch

$$b_j(t) = \sum_{k=j+1}^m \binom{k}{j+1} \cdot a_k(t) \cdot \varphi^{(k-j-1)}(t),$$

so ist die Funktion $\varphi(t) \cdot \int \psi(t) dt$ eine zu φ linear unabhängige Lösung der DGL $\Lambda(y)=0$. $\square$

Beweis:

Sei ψ eine Lösung der DGL $\Lambda_1(\psi) = 0$. Definiert man $\phi(t) := \int \psi(t) dt$ und beachtet die Leibniz-Regel

$$(f \cdot g)^{(k)} = \sum_{j=0}^k \binom{k}{j} f^{(j)} g^{(k-j)},$$

¹¹Das Reduktionsverfahren gilt entsprechend auch für DGL-Systeme (siehe Satz 2.24).

so folgt:

$$\begin{aligned}
 \Lambda(\varphi \cdot \phi) &= \sum_{k=0}^m a_k (\varphi \cdot \phi)^{(k)} \\
 &= \sum_{k=0}^m a_k \sum_{j=0}^k \binom{k}{j} \phi^{(j)} \varphi^{(k-j)} \\
 &\stackrel{\binom{k}{k+1}=0}{=} \sum_{k=0}^m \sum_{j=0}^m a_k \binom{k}{j} \phi^{(j)} \varphi^{(k-j)} \\
 &= \sum_{j=0}^m \phi^{(j)} \sum_{k=j}^m a_k \binom{k}{j} \varphi^{(k-j)} \\
 &= \underbrace{\phi \sum_{k=0}^m a_k \binom{k}{0} \varphi^{(k)}}_{=\Lambda(\varphi)=0} + \sum_{j=1}^m \underbrace{\phi^{(j)}}_{\psi^{(j-1)}} \underbrace{\sum_{k=j}^m a_k \binom{k}{j} \varphi^{(k-j)}}_{b_{j-1}} \\
 &= \phi \cdot 0 + \Lambda_1(\psi) = 0
 \end{aligned}$$

Die lineare Unabhängigkeit von φ und $\phi := \int \psi(t) dt$ ergibt sich aus folgender Überlegung:

Sind $\alpha, \beta \in \mathbb{R}$ mit

$$\alpha \cdot \varphi(t) + \beta \cdot \varphi(t)\phi(t) = 0 ,$$

so folgt wegen $\varphi \not\equiv 0$ die Gleichung

$$\alpha + \beta \cdot \phi(t) = 0 .$$

Differenziert man die letzte Gleichung nach t und beachtet, dass $\dot{\phi} = \psi$ gilt, so folgt

$$\beta \cdot \psi(t) = 0 .$$

Aus $\psi \not\equiv 0$ folgt $\beta = 0$ und damit auch $\alpha = 0$. ■

2.23 Beispiel

Die homogene lineare DGL zweiter Ordnung

$$t^2(1-t)\ddot{y} + 2t(2-t)\dot{y} + 2(1+t)y = 0$$

besitzt die Lösung $\varphi(t) = t^{-2}$. Die Anwendung des Reduktionsverfahrens liefert mit $a_2(t) = t^2(1-t)$ und $a_1(t) = 2t(2-t)$

$$b_0(t) = \sum_{k=1}^2 \binom{k}{1} a_k(t) \varphi^{(k-1)}(t) = a_1(t) \varphi(t) + 2a_2(t) \dot{\varphi}(t) = 2$$

und

$$b_1(t) = a_2(t) \varphi(t) = 1 - t .$$

Es ergibt sich daher die reduzierte DGL

$$(1-t)\dot{z} + 2z = 0 .$$

Diese ist separabel und besitzt die Lösung $\psi(t) = (1-t)^2$. Damit ergibt sich als zweite Lösung ϕ der ursprünglichen DGL

$$\phi(t) = \varphi(t) \int \psi(t) dt = t^{-2} \int (1-t)^2 dt = -\frac{(1-t)^3}{3t^2} . \quad \square$$

Das Reduktionsverfahren von d'Alembert lässt sich entsprechend auch für lineare DGL-Systeme anwenden. Es hat allerdings den Nachteil, dass das reduzierte System von $n-1$ Gleichungen sich im allgemeinen nicht mehr in eine lineare DGL der Ordnung $n-1$ zurück transformieren lässt.

2.24 Satz (Reduktionsverfahren von d'Alembert für lineare DGL-Systeme)

Sei $\mathbf{A} : I \rightarrow \mathbb{R}^{n \times n}$ mit $\mathbf{A}(t) = (a_{i,j}(t))_{i,j=1}^n$ eine auf dem Intervall I stetige Abbildung (d.h. die Komponentenfunktionen $a_{i,j}$ sind stetig). $\boldsymbol{\varphi}(t) = (\varphi_1(t), \dots, \varphi_n(t))^T$ sei eine bekannte, nicht triviale Lösung des homogenen DGL-Systems

$$\dot{\mathbf{y}} = \mathbf{A}(t)\mathbf{y} .$$

Definiert man die $(n-1) \times (n-1)$ -Matrix

$$\tilde{\mathbf{A}}(t) := \left(a_{k,\ell}(t) - a_{1,\ell}(t) \frac{\varphi_\ell(t)}{\varphi_1(t)} \right)_{k,\ell=2}^n$$

und ist $\boldsymbol{\psi} : I \rightarrow \mathbb{R}^{n-1}$ mit $\boldsymbol{\psi}(t) = (\psi_2(t), \dots, \psi_n(t))^T$ eine Lösung des reduzierten DGL-Systems

$$\dot{\mathbf{z}} = \tilde{\mathbf{A}}(t)\mathbf{z} ,$$

so ist eine weitere von $\boldsymbol{\varphi}$ linear unabhängige Lösung $\boldsymbol{\phi}$ des DGL-Systems $\dot{\mathbf{y}} = \mathbf{A}(t)\mathbf{y}$ gegeben durch

$$\boldsymbol{\phi}(t) := \left(\int_{t_0}^t \frac{1}{\varphi_1(s)} \sum_{\ell=2}^n a_{1,\ell}(s) \psi_\ell(s) ds \right) \cdot \begin{pmatrix} \varphi_1(t) \\ \varphi_2(t) \\ \vdots \\ \varphi_n(t) \end{pmatrix} + \begin{pmatrix} 0 \\ \psi_2(t) \\ \vdots \\ \psi_n(t) \end{pmatrix} \quad \square$$

Beweis:

Wir versuchen eine zweite Lösung $\boldsymbol{\phi}$ der DGL $\dot{\mathbf{y}} = \mathbf{A}(t)\mathbf{y}$ zu finden mit dem Ansatz

$$\boldsymbol{\phi}(t) := \lambda(t) \cdot \boldsymbol{\varphi}(t) + \tilde{\boldsymbol{\psi}}(t) .$$

Hierbei gelte $\tilde{\boldsymbol{\psi}}(t) := (0, \psi_2(t), \dots, \psi_n(t))^T$ mit noch unbekannten Funktionen $\psi_2, \dots, \psi_n$ und der Kurzbezeichnung

$$\lambda(t) := \int_{t_0}^t \frac{1}{\varphi_1(s)} \sum_{\ell=2}^n a_{1,\ell}(s) \psi_\ell(s) ds .$$

ϕ ist genau dann eine Lösung, wenn $\mathbf{A}(t)\phi(t) = \dot{\phi}(t)$ gilt. Einsetzen der speziellen Form von ϕ liefert:

$$\mathbf{A}(t)\phi(t) = \lambda(t)\mathbf{A}(t)\varphi(t) + \mathbf{A}(t)\tilde{\psi}(t) = \dot{\phi}(t) = \dot{\lambda}(t)\varphi(t) + \lambda(t)\dot{\varphi}(t) + \dot{\tilde{\psi}}(t)$$

Da $\dot{\varphi}(t) = \mathbf{A}(t)\varphi(t)$ gilt, ergibt sich, dass $\tilde{\psi}$ der folgenden DGL genügen muss:

$$\mathbf{A}(t)\tilde{\psi}(t) - \dot{\lambda}(t)\varphi(t) = \dot{\tilde{\psi}}(t) = (0, \dot{\psi}_2(t), \dots, \dot{\psi}_n(t))^T$$

Für die erste Komponente dieses Gleichungssystems bedeutet dies

$$\sum_{\ell=2}^n a_{1,\ell}(t)\psi_\ell(t) - \dot{\lambda}(t)\varphi_1(t) = 0 \quad (2.13)$$

und für die Komponenten $i = 2, \dots, n$:

$$\sum_{\ell=2}^n a_{i,\ell}(t)\psi_\ell(t) - \dot{\lambda}(t)\varphi_i(t) = \dot{\psi}_i(t) \quad (2.14)$$

Aus Gleichung (2.13) ergibt sich

$$\dot{\lambda}(t) = \frac{1}{\varphi_1(t)} \sum_{\ell=2}^n a_{1,\ell}(t)\psi_\ell(t) . \quad (2.15)$$

Setzt man dies in Gleichung (2.14) ein, ergibt sich, dass die gesuchten Funktionen $\psi_2, \dots, \psi_n$ den folgenden Gleichungen genügen müssen:

$$\dot{\psi}_i(t) = \sum_{\ell=2}^n \left(a_{i,\ell}(t) - \frac{a_{1,\ell}(t)\varphi_i(t)}{\varphi_1(t)} \right) \psi_\ell(t) \quad (i = 2, \dots, n)$$

Damit ist $\psi := (\psi_2, \dots, \psi_n)^T$ gerade eine Lösung des DGL-Systems

$$\dot{z} = \tilde{\mathbf{A}}(t)z ,$$

Integriert man Gleichung (2.15), erhält man die gewünschte Integralform von λ .

Die Unabhängigkeit der Lösungen ϕ und φ folgt sofort aus der Darstellung von ϕ . ■

Wie in Kapitel 1 (siehe die Sätze 1.21 bzw. 1.25) ergibt sich aus dem Existenzsatz 2.21 für lineare DGL-Systeme folgende Struktur für den Lösungsraum linearer DGLen.

2.25 Satz

Sind $\mathbf{b}(t)$ und $\mathbf{A}(t)$ wie in Satz 2.21, so gelten folgende Aussagen für den homogenen bzw. inhomogenen Lösungsraum $\mathbb{L}_h$ bzw. $\mathbb{L}$ der DGL

$$\dot{\mathbf{y}} - \mathbf{A}(t)\mathbf{y} = \mathbf{b}(t) . \quad (2.16)$$

a) $\mathbb{L}_h$ ist ein Vektorraum der Dimension n . Eine Basis (*Fundamentalsystem*) von $\mathbb{L}_h$ ist gegeben durch die Funktionen $\varphi_1, \dots, \varphi_n$, wobei φ_j die nach Satz 2.21 eindeutigen Lösungen der homogenen AWA $\dot{\mathbf{y}} - \mathbf{A}(t)\mathbf{y} = \mathbf{0}$, $\varphi_j(t_0) = \mathbf{e}_j$ sind.

b) Ist ψ eine spezielle (*partikuläre*) Lösung der inhomogenen DGL, so gilt

$$\mathbb{L} = \{\psi\} + \mathbb{L}_h \quad \square$$

Beweis:

Der Beweis stimmt wörtlich mit den Beweisen der analogen Aussagen für Δ GLen (siehe Satz 1.21) bzw. Δ GL-systeme (siehe Satz 1.25) überein. Wie dort wird für jeden Startpunkt t_0 ein Isomorphismus $\mathfrak{S}_{t_0} : \mathbb{L}_h \rightarrow \mathbb{R}^n$ definiert durch

$$\mathfrak{S}_{t_0}(\varphi) := \varphi(t_0) . \quad \blacksquare$$

Die Untersuchung der linearen Unabhängigkeit von Lösungen linearer DGL-Systeme kann mit Hilfe von Determinanten erfolgen. Wir benötigen hierzu folgenden Begriff.

2.26 Definition

Für n Abbildungen $\varphi_1, \varphi_2, \dots, \varphi_n : I \rightarrow \mathbb{R}^n$ heißt die Funktion $W : I \rightarrow \mathbb{R}$ mit

$$W(t) := \det(\varphi_1(t), \varphi_2(t), \dots, \varphi_n(t))$$

die *Wronski-Determinante* der Funktionen $\varphi_1, \varphi_2, \dots, \varphi_n$. $\square$

Der folgende Satz zeigt, dass mit der Wronski-Determinante Lösungssysteme linearer DGLen mit wenig Aufwand auf lineare Unabhängigkeit untersucht werden können.

2.27 Satz

Mit den Bezeichnungen aus Definition 2.26 sind folgende Aussagen äquivalent:

- a) $\varphi_1, \dots, \varphi_n$ ist ein Fundamentalsystem des linearen DGL-Systems $\dot{\mathbf{y}} = \mathbf{A}(t)\mathbf{y}(t)$.
- b) $\forall t \in I : W(t) \neq 0$, d.h. die Matrix $(\varphi_1(t), \varphi_2(t), \dots, \varphi_n(t))$ ist invertierbar
- c) $\exists s \in I : W(s) \neq 0$, d.h. die Matrix $(\varphi_1(s), \varphi_2(s), \dots, \varphi_n(s))$ ist invertierbar $\square$

Beweis:

- a) $\Rightarrow$ b): Nach Satz 2.25 wird für jedes $t \in I$ durch $\mathfrak{S}_t(\mathbf{y}) := \mathbf{y}(t)$ eine bijektive und lineare Abbildung von $\mathbb{L}_h$ nach $\mathbb{R}^n$ definiert, so dass $\mathfrak{S}_t$ und $\mathfrak{S}_t^{-1}$ jede Basis in eine Basis überführen. Ist nun $(\varphi_1, \dots, \varphi_n)$ ein Fundamentalsystem von $\mathbb{L}_h$, sind die Vektoren $\varphi_1(t) = \mathfrak{S}_t(\varphi_1), \dots, \varphi_n(t) = \mathfrak{S}_t(\varphi_n)$ eine Basis des $\mathbb{R}^n$, so dass gilt

$$0 \neq \det(\varphi_1(t), \dots, \varphi_n(t)) = W(t) .$$

b) $\Rightarrow$ c) Klar.

c) $\Rightarrow$ a): Wegen $W(t_0) = \det(\varphi_1(t_0), \dots, \varphi_n(t_0)) \neq 0$ bilden $\varphi_1(t_0), \dots, \varphi_n(t_0)$ eine Basis des $\mathbb{R}^n$. Damit sind die Urbilder dieser Vektoren, d.h. die Funktionen $\mathfrak{S}_{t_0}^{-1}(\varphi_1(t_0)) = \varphi_1, \dots, \mathfrak{S}_{t_0}^{-1}(\varphi_n(t_0)) = \varphi_n$ linear unabhängig in $\mathbb{L}_h$. ■

2.28 Beispiel

Wir betrachten die DGL $t^2(1-t)\ddot{y} + 2t(2-t)\dot{y} + 2(1+t)y = 0$ aus Beispiel 2.23.

Überführt man die DGL mit Hilfe von Lemma 2.5 in ein System 1. Ordnung, erhält man

$$\dot{\mathbf{y}}(t) = \begin{pmatrix} \dot{y}_1(t) \\ \dot{y}_2(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{2(1+t)}{t^2(1-t)} & -\frac{2(2-t)}{t(1-t)} \end{pmatrix} \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}$$

Um die Stetigkeit der Koeffizienten zu gewährleisten, wählt man als Definitionsbereich eines der Intervalle $] -\infty, 0[,]0, 1[,]1, \infty[$ aus. Dieses homogene DGL-System besitzt wie in Beispiel 2.23 gesehen die Fundamentallösungen

$$\varphi_1(t) = \begin{pmatrix} t^{-2} \\ -2t^{-3} \end{pmatrix}, \quad \varphi_2(t) = \begin{pmatrix} \frac{(t-1)^3}{3t^2} \\ \frac{(t-1)^2(t+2)}{3t^3} \end{pmatrix}$$

Für die Wronski-Determinante $W(t)$ gilt

$$W(t) = \det \begin{pmatrix} t^{-2} & , & \frac{(t-1)^3}{3t^2} \\ -2t^{-3} & , & \frac{(t-1)^2(t+2)}{3t^3} \end{pmatrix} = \frac{(t-1)^2}{t^4},$$

so dass wegen $W(-1) = 4 \neq 0$, $W(\frac{1}{2}) = 4 \neq 0$ und $W(2) = 2^{-4} \neq 0$ die beiden Lösungen in allen uns interessierenden Intervallen ein Fundamentalsystem von $\mathbb{L}_h$ sind. ■

Auch die Lösung von DGLen bzw. DGL-Systemen erster Ordnung erfolgt analog zu den entsprechenden Sätzen über Δ GLen bzw. Δ GL-systeme (siehe Satz 1.23 bzw. 1.25). Es muss nur darauf geachtet werden, dass an die Stelle von Potenzfunktionen und Summen die analogen „stetigen“ Begriffe Exponentialfunktion und Integral treten.

2.29 Satz

Gegeben sei das lineare DGL-System erster Ordnung

$$\dot{\mathbf{y}} - \mathbf{A}(t) \cdot \mathbf{y} = \mathbf{b}(t),$$

mit gegebenen stetigen Funktionen $\mathbf{b} : \mathbb{R} \rightarrow \mathbb{R}^n$ und $\mathbf{A} : \mathbb{R} \rightarrow \mathbb{R}^{n \times n}$. Dann gilt¹²

¹²Die Aussage gilt auch, wenn $\mathbf{b}$ und $\mathbf{A}$ nur auf einem Intervall I definiert sind.

- a) Für den homogenen Lösungsraum gilt $L_h = \left\{ \Phi(t) \cdot c \mid c \in \mathbb{R}^n \right\}$, wobei $\Phi : \mathbb{R} \rightarrow \mathbb{R}^{n \times n}$ definiert ist durch¹³

$$\Phi(t) = \exp \left(\int_{t_0}^t A(s) ds \right)$$

- b) Eine partikuläre Lösung $\psi : \mathbb{R} \rightarrow \mathbb{R}^n$ lautet

$$\psi(t) = \Phi(t) \int_{t_0}^t \Phi(s)^{-1} \cdot b(s) ds . \quad (2.17)$$

- c) Sind b und A konstant und A invertierbar, gilt

$$\Phi(t) = e^{A \cdot t} \quad \text{und} \quad \psi(t) = -A^{-1} \cdot b \quad \square$$

Beweis:

- a) Es gilt gemäß Kettenregel

$$\frac{d}{dt} e^{B(t)} = \dot{B}(t) \cdot e^{B(t)}$$

Wählt man speziell $B(t) = \int_{t_0}^t A(s) ds$ folgt $\dot{B}(t) = A(t)$ und damit die Behauptung

$$\dot{\Phi}(t) = A(t) \cdot \Phi(t) \iff \dot{\Phi}(t) - A(t) \cdot \Phi(t) = 0$$

- b) Wir zeigen mittels Einsetzen, dass ψ der inhomogenen DGL $\dot{y} - A(t) \cdot y = b(t)$ genügt¹⁴. Mit der Produktregel folgt wegen $\dot{\Phi}(t) = A(t) \cdot \Phi(t)$

$$\begin{aligned} \dot{\psi}(t) &= \dot{\Phi}(t) \int_{t_0}^t \Phi(s)^{-1} b(s) ds + \Phi(t) \Phi(t)^{-1} b(t) \\ &= A(t) \Phi(t) \underbrace{\int_{t_0}^t \Phi(s)^{-1} b(s) ds}_{\psi} + b(t) \\ &= A(t) \psi(t) + b(t) \end{aligned}$$

- c) Dies ist eine direkte Konsequenz von Teil b)! ■

¹³Man beachte, dass $\Phi(t_0) = \mathbf{1}$ gilt.

¹⁴Die Formel kann wieder mittels „Variation der Konstanten“, d.h. durch den Ansatz $\psi := \Phi(t) \cdot c(t)$ wie im Beweis von Satz 1.23 b) hergeleitet werden.

2.30 Beispiel

i) Man bestimme die Lösung der linearen AWA

$$\dot{y} - \underbrace{\frac{t-2}{t(1-t)}}_{A(t)} \cdot y = \underbrace{\frac{1}{t^2(1-t)}}_{b(t)}, \quad y(-1) = 3$$

Wir berechnen zunächst

$$\int A(t) dt = \int \frac{t-2}{t(1-t)} dt = \int \left[\frac{1}{t-1} - \frac{2}{t} \right] dt = \ln(t-1) - 2 \ln(t) = \ln\left(\frac{t-1}{t^2}\right)$$

Damit ergibt sich mit Satz 2.29 als homogene Fundamentallösung $\Phi(t)$

$$\Phi(t) = \exp\left(\ln\left(\frac{t-1}{t^2}\right)\right) = \frac{t-1}{t^2}$$

Formel (2.17) liefert als partikuläre Lösung

$$\Psi(t) = \Phi(t) \int \Phi(t)^{-1} \cdot b(t) dt = \frac{t-1}{t^2} \int -(t-1)^{-2} dt = \frac{1}{t^2}$$

Die allgemeine Lösung lautet daher

$$y(t) = \Psi(t) + C \cdot \Phi(t) = \frac{1}{t^2} + C \cdot \frac{t-1}{t^2}$$

und aus der Anfangsbedingung folgt

$$3 \stackrel{!}{=} y(-1) = 1 - 2 \cdot C \implies C = -1.$$

Die gesuchte Lösung lautet also $y(t) = \frac{1}{t^2} - \frac{t-1}{t^2} = \frac{2-t}{t^2}$.

ii) Gegeben sei folgende homogene lineare DGL zweiter Ordnung:

$$\ddot{y} + 8t\dot{y} + 9t^2y = 0$$

Das gemäß Lemma 2.5 zugehörige äquivalente DGL-System lautet

$$\dot{\mathbf{y}}(t) = \begin{pmatrix} \dot{y}_1(t) \\ \dot{y}_2(t) \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 \\ -9t^2 & -8t \end{pmatrix}}_{:=\mathbf{A}(t)} \cdot \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}$$

Dann gilt

$$\int \mathbf{A}(t) dt = \begin{pmatrix} 0 & t \\ -3t^3 & -4t^2 \end{pmatrix} = \mathbf{B} \cdot \underbrace{\begin{pmatrix} -3t^2 & 0 \\ 0 & -t^2 \end{pmatrix}}_D \cdot \mathbf{B}^{-1}$$

mit

$$\mathbf{B} = \begin{pmatrix} -\frac{1}{3t} & , & -\frac{1}{t} \\ 1 & , & 1 \end{pmatrix} \quad , \quad \mathbf{B}^{-1} = \frac{3}{2} \begin{pmatrix} t & , & 1 \\ -t & , & -\frac{1}{3} \end{pmatrix} .$$

Dies liefert wegen $(\mathbf{B} \cdot \mathbf{D} \cdot \mathbf{B}^{-1})^k = \mathbf{B} \cdot \mathbf{D}^k \cdot \mathbf{B}^{-1}$

$$\Phi(t) = \exp \left(\int \mathbf{A}(t) dt \right) = \sum_{k=0}^{\infty} \frac{(\mathbf{B} \cdot \mathbf{D} \cdot \mathbf{B}^{-1})^k}{k!} = \mathbf{B} \cdot \left(\sum_{k=0}^{\infty} \frac{\mathbf{D}^k}{k!} \right) \cdot \mathbf{B}^{-1} = \mathbf{B} \cdot e^{\mathbf{D}} \cdot \mathbf{B}^{-1}$$

Da für die Diagonalmatrix $\mathbf{D}$ offensichtlich

$$e^{\mathbf{D}} = \begin{pmatrix} e^{-3t^2} & , & 0 \\ 0 & , & e^{-t^2} \end{pmatrix} = e^{-t^2} \cdot \begin{pmatrix} e^{-2t^2} & , & 0 \\ 0 & , & 1 \end{pmatrix}$$

gilt, folgt schließlich

$$\Phi(t) = \mathbf{B} \cdot e^{\mathbf{D}} \cdot \mathbf{B}^{-1} = \frac{1}{2} e^{-t^2} \begin{pmatrix} 3 - e^{-2t^2} & , & \frac{1}{t} (1 - e^{-2t^2}) \\ 3t(e^{-2t^2} - 1) & , & 3e^{-2t^2} - 1 \end{pmatrix}$$

- iii) Das vorherige Beispiel ist relativ einfach, weil die entscheidende Matrix $\int \mathbf{A}(t) dt$ diagonalisiert werden kann¹⁵, was die Berechnung der Exponentialmatrix deutlich vereinfacht. Dies ist in vielen Fällen jedoch nicht möglich. In diesen Situationen verwendet man die sog. *Jordan-Normalform* von Matrizen. Diese liefert eine „fast“ diagonale Matrix. Genauer gilt, dass jede $(n \times n)$ -Matrix $\mathbf{M}$ mit einer invertierbaren $(n \times n)$ -Matrix $\mathbf{B}$ auf Jordan-Normalform $\mathbf{J}$ gebracht werden kann (zur genauen Aussage siehe Anhang Satz D.6), d.h.

$$\mathbf{M} = \mathbf{B} \cdot \mathbf{J} \cdot \mathbf{B}^{-1}$$

Damit gilt wie in Teil ii)

$$e^{\mathbf{M}} = \mathbf{B} \cdot e^{\mathbf{J}} \cdot \mathbf{B}^{-1} .$$

Die Berechnung der Matrix $e^{\mathbf{J}}$ ist nun ähnlich einfach wie bei einer Diagonalmatrix (siehe auch Beispiel A.10 iii) des Anhangs). Folgende DGL, in der wir das letzte Beispiel leicht verändert haben, verdeutlicht die Vorgehensweise.

Gegeben sei folgende lineare DGL zweiter Ordnung:

$$\ddot{y} + 4t\dot{y} + 3t^2y = 0$$

¹⁵Zur genauen Vorgehensweise beachte man Beispiel A.10 im mathematischen Anhang auf Seite 174.

Das gemäß Lemma 2.5 zugehörige äquivalente DGL-System hat die Form

$$\dot{\mathbf{y}}(t) = \underbrace{\begin{pmatrix} 0 & , & 1 \\ -3t^2 & , & -4t \end{pmatrix}}_{:=\mathbf{A}(t)} \cdot \mathbf{y}(t)$$

Dann gilt

$$\int \mathbf{A}(t) dt = \begin{pmatrix} 0 & , & t \\ -t^3 & , & -2t^2 \end{pmatrix} = \mathbf{B} \cdot \underbrace{\begin{pmatrix} -t^2 & , & 1 \\ 0 & , & -t^2 \end{pmatrix}}_{\mathbf{J}(t)} \cdot \mathbf{B}^{-1}$$

mit

$$\mathbf{B} = \begin{pmatrix} -\frac{1}{t} & , & 0 \\ 1 & , & -\frac{1}{t^2} \end{pmatrix} \quad , \quad \mathbf{B}^{-1} = \begin{pmatrix} -t & , & 0 \\ -t^3 & , & -t^2 \end{pmatrix} .$$

Wie im vorigen Beispiel folgt dann (mit $\mathbf{J}(t)$ anstelle von $\mathbf{D}$)

$$\Phi(t) = \exp \left(\int \mathbf{A}(t) dt \right) = e^{\mathbf{B} \cdot \mathbf{J}(t) \cdot \mathbf{B}^{-1}} = \mathbf{B} \cdot e^{\mathbf{J}(t)} \cdot \mathbf{B}^{-1}$$

Für die Jordanmatrix $\mathbf{J}$ gilt nun

$$e^{\mathbf{J}(t)} = \begin{pmatrix} e^{-t^2} & , & e^{-t^2} \\ 0 & , & e^{-t^2} \end{pmatrix} = e^{-t^2} \cdot \begin{pmatrix} 1 & , & 1 \\ 0 & , & 1 \end{pmatrix}$$

Zusammen ergibt sich

$$\Phi(t) = \mathbf{B} \cdot e^{\mathbf{J}} \cdot \mathbf{B}^{-1} = e^{-t^2} \cdot \mathbf{B} \cdot \begin{pmatrix} 1 & , & 1 \\ 0 & , & 1 \end{pmatrix} \cdot \mathbf{B}^{-1} = e^{-t^2} \begin{pmatrix} 1+t^2 & , & t \\ -t^3 & , & 1-t^2 \end{pmatrix}$$

iv) Wir betrachten das DGL-System $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$, mit der konstanten Matrix

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 3 & 1 & -1 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & -1 & -2 & 2 & 2 \end{pmatrix}$$

Es gilt

$$\mathbf{A} = \mathbf{B}\mathbf{J}\mathbf{B}^{-1}$$

mit den Matrizen

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 & 0 \end{pmatrix}, \quad \mathbf{B}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 \end{pmatrix}$$

und der Jordanmatrix

$$J = \left(\begin{array}{cc|ccc} 3 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ \hline 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{array} \right)$$

Dies liefert

$$\Phi(t) = e^{A \cdot t} = B \cdot e^{J \cdot t} \cdot B^{-1} = B \cdot \left(\begin{array}{c|ccc} e^{J_1 \cdot t} & \mathcal{O} & & \\ \hline \mathcal{O} & e^{J_2 \cdot t} & & \end{array} \right) \cdot B^{-1}.$$

Nun gilt (zur Rechnung siehe Beispiel A.10 iii) im Anhang)

$$e^{J \cdot t} = \left(\begin{array}{cc|ccc} e^{3t} & 0 & 0 & 0 & 0 \\ 0 & e^{3t} & 0 & 0 & 0 \\ \hline 0 & 0 & e^{2t} & t e^{2t} & \frac{1}{2} t^2 e^{2t} \\ 0 & 0 & 0 & e^{2t} & t e^{2t} \\ 0 & 0 & 0 & 0 & e^{2t} \end{array} \right)$$

und es folgt letztendlich

$$\Phi(t) = B \cdot e^{J \cdot t} \cdot B^{-1} = \begin{pmatrix} e^{3t} & 0 & 0 & 0 & 0 \\ 0 & e^{3t} & e^{3t} - e^{2t} & -t e^{2t} & 0 \\ 0 & 0 & e^{2t} & t e^{2t} & 0 \\ 0 & 0 & 0 & e^{2t} & 0 \\ 0 & -e^{3t} + e^{2t} & -e^{3t} + e^{2t} - t e^{2t} & -\frac{1}{2} t^2 e^{2t} + 2t e^{2t} & e^{2t} \end{pmatrix} \quad \square$$

Liegt eine lineare DGL höherer Ordnung mit konstanten Koeffizienten vor, kann statt der Umwandlung in ein lineares DGL-System der Ordnung 1 (wie in den Beispielen 2.30ii), iii), iv) alternativ auch die schon bei Δ GL-en eingeführte Nullstellenmethode verwendet werden. Führt man formal den Differentialoperator $D := \frac{d}{dt}$ ein, so nimmt die lineare DGL mit konstanten Koeffizienten

$$y^{(n)} + a_{n-1} y^{(n-1)} + \dots + a_1 \dot{y} + a_0 y = 0$$

folgende Form an:

$$(D^n + a_{n-1} D^{n-1} + \dots + a_1 D + a_0) y(t) = 0.$$

Ist λ eine Nullstelle der charakteristischen Gleichung $D^n + a_{n-1} D^{n-1} + \dots + a_1 D + a_0 = 0$ mit der Vielfachheit k , so sind die Funktionen

$$\varphi^{(j)}(t) := t^j e^{\lambda t} \quad (j = 0, 1, \dots, k-1)$$

Lösungen der homogenen DGL. Eine spezielle Lösung findet man wiederum mittels Variation der Konstanten. Man erhält (inklusive Beweis) das stetige Analogon zu Satz 1.27.

2.31 Satz (Lösung linearer DGLen mit konstanten Koeffizienten)

Sei z eine Nullstelle der Vielfachheit k der charakteristischen Gleichung der homogenen linearen DGL mit konstanten Koeffizienten

$$y^{(m)} + a_{n-1}y^{(m-1)} + \dots + a_1\dot{y} + a_0y = 0 \quad (2.18)$$

Dann ist für jedes Polynom P mit $\text{grad}(P) \leq k-1$ die Funktion $\varphi(t) := P(t) \cdot e^{zt}$ ebenfalls eine Lösung der DGL (2.18). $\square$

2.32 Beispiel

Gegeben sei die homogene lineare DGL

$$y^{(6)} + y^{(5)} - 4y^{(4)} + 2y^{(3)} - 11\ddot{y} + \dot{y} - 6y = 0$$

Die zugehörige charakteristische Gleichung

$$x^6 + x^5 - 4x^4 + 2x^3 - 11x^2 + x - 6 = 0$$

liefert die Nullstellen (inklusive Vielfachheit)

$$-i, -i, i, i, -3, 2$$

und damit gemäß Satz 2.31 die Lösungsfunktionen

$$e^{-it}, t e^{-it}, e^{it}, t e^{it}, e^{-3t}, e^{2t}$$

Um reelle Lösungen zu erhalten, wählen wir von den komplexen Lösungsfunktionen den Real- bzw. Imaginärteil. Damit ergeben sich als reelle Fundamentalebasis die Funktionen

$$\cos(t), \sin(t), t \cdot \cos(t), t \cdot \sin(t), e^{-3t}, e^{2t} \quad \square$$

Wie Satz 2.29 zeigt, ist die Lösung homogener linearer DGL-Systeme 1. Ordnung besonders einfach, wenn die das System beschreibende Matrix konstant ist, da dann alle homogenen Lösungen von der Form $\mathbf{y}(t) = e^{At} \cdot \mathbf{c}$ sind.

Die Beispiele 2.30 ii) - iv) zeigen jedoch, dass die Berechnung von e^{At} einen hohen Aufwand erfordert, so dass Satz 2.29 nur eingeschränkt verwendet werden kann. Kennt man jedoch Eigenwerte und Eigenvektoren von $\mathbf{A}$, so stellt er ein äußerst nützliches Ergebnis dar. Ist nämlich $\mathbf{v}$ ein Eigenvektor von $\mathbf{A}$ zum Eigenwert λ von $\mathbf{A}$, d.h. gilt $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, so ist nach dem letzten Satz $\mathbf{y}(t) := e^{At} \mathbf{v}$ eine Lösung des homogenen DGL-Systems, die sich explizit berechnen lässt. Man erhält nämlich:

$$\mathbf{y}(t) = e^{At} \mathbf{v} = \sum_{j=0}^{\infty} \frac{\mathbf{A}^j \cdot t^j}{j!} \mathbf{v} = \sum_{j=0}^{\infty} \frac{\mathbf{A}^j \cdot \mathbf{v}}{j!} t^j = \sum_{j=0}^{\infty} \frac{\lambda^j \mathbf{v}}{j!} t^j = e^{\lambda t} \mathbf{v}$$

Genauere Informationen über diesen Sachverhalt liefert der folgende Satz.

2.33 Satz

Sei $\mathbf{A} \in \mathbb{R}^{n \times n}$ konstant.

- Für $\lambda \in \mathbb{C}$ und $\mathbf{0} \neq \mathbf{v} \in \mathbb{C}^n$ ist die Abbildung φ mit $\varphi(t) := e^{\lambda t} \mathbf{v}$ genau dann eine Lösung des homogenen DGL-Systems $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$, wenn λ ein Eigenwert der Matrix $\mathbf{A}$ zum Eigenvektor $\mathbf{v}$ ist.
- Die Lösungen $\varphi_j(t) := e^{\lambda_j t} \mathbf{v}_j$ ($j=1, \dots, k$) des DGL-Systems $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$ sind genau dann linear unabhängig, wenn die Vektoren $\mathbf{v}_1, \dots, \mathbf{v}_k$ linear unabhängig sind. $\square$

Beweis:

- Ist $\varphi(t) := e^{\lambda t} \mathbf{v}$ eine Lösung des DGL-Systems $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$, so gilt für alle $t \in \mathbb{R}$:

$$\lambda e^{\lambda t} \mathbf{v} = \dot{\varphi}(t) = \mathbf{A}\varphi(t) = \mathbf{A} \cdot (e^{\lambda t} \mathbf{v}) = e^{\lambda t} \mathbf{A} \cdot \mathbf{v} .$$

Speziell für $t = 0$ folgt $\lambda \mathbf{v} = \mathbf{A}\mathbf{v}$.

Ist umgekehrt λ ein Eigenwert von $\mathbf{A}$ zum Eigenvektor $\mathbf{v}$, so folgt für alle $t \in \mathbb{R}$:

$$\frac{d}{dt} e^{\lambda t} \mathbf{v} = e^{\lambda t} \lambda \mathbf{v} = e^{\lambda t} \mathbf{A}\mathbf{v} = \mathbf{A} \cdot (e^{\lambda t} \mathbf{v}) .$$

Daher genügt $\varphi(t) := e^{\lambda t} \mathbf{v}$ dem DGL-System $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$.

- Sind die Abbildungen $\varphi_j(t) := e^{\lambda_j t} \mathbf{v}_j$ linear unabhängig, so folgt die lineare Unabhängigkeit der Vektoren $\mathbf{v}_j$ folgendermaßen. Seien $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ mit

$$\mathbf{0} = \sum_{j=1}^k \alpha_j \mathbf{v}_j = \sum_{j=1}^k \alpha_j \varphi_j(0) .$$

Dann ist die Abbildung $\sum_{j=1}^k \alpha_j \varphi_j$ Lösung der AWA $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$, $\mathbf{y}(0) = \mathbf{0}$. Da $\mathbf{y} \equiv \mathbf{0}$ diese AWA ebenfalls löst, folgt wegen der Eindeutigkeitsaussage in Satz 2.21, dass beide Abbildungen übereinstimmen, d.h. es gilt

$$\sum_{j=1}^k \alpha_j \varphi_j \equiv \mathbf{0} .$$

Wegen der linearen Unabhängigkeit der φ_j muss daher $\alpha_1 = \dots = \alpha_k = 0$ gelten. Sind umgekehrt die Vektoren $\mathbf{v}_1, \dots, \mathbf{v}_k$ linear unabhängig, so folgt die lineare Unabhängigkeit der φ_j , denn

$$\sum_{j=1}^k \alpha_j \varphi_j = \mathbf{0} \implies \sum_{j=1}^k \alpha_j \varphi_j(0) = \sum_{j=1}^k \alpha_j \mathbf{v}_j = \mathbf{0} \implies \alpha_1 = \dots = \alpha_k = 0 \quad \blacksquare$$

2.34 Bemerkung

- i) Liegt eine Basis von Eigenvektoren vor, liefert Satz 2.33 ein Fundamentalsystem des DGL-Systems $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$. Dies ist insbesondere dann erfüllt, wenn die $(n \times n)$ -Matrix $\mathbf{n}$ verschiedene Eigenwerte besitzt, da die Eigenvektoren zu verschiedenen Eigenwerten linear unabhängig sind¹⁶.
- ii) Das Problem komplexer Eigenwerte und Eigenvektoren lässt sich wie früher beheben. Ist λ ein komplexer Eigenwert der reellen Matrix $\mathbf{A}$ mit dem Eigenvektor $\mathbf{v} = \mathbf{a} + i\mathbf{b}$, so ist auch $\bar{\lambda}$ ein Eigenwert zum Eigenvektor $\bar{\mathbf{v}} = \mathbf{a} - i\mathbf{b}$, denn wegen $\bar{\mathbf{A}} = \mathbf{A}$ gilt:

$$\bar{\lambda}\bar{\mathbf{v}} = \overline{\lambda\mathbf{v}} = \overline{\mathbf{A}\mathbf{v}} = \bar{\mathbf{A}}\bar{\mathbf{v}} = \mathbf{A}\bar{\mathbf{v}}$$

Also sind $\boldsymbol{\psi}(t) := e^{\lambda t} \mathbf{v}$ und $\bar{\boldsymbol{\psi}}(t) := e^{\bar{\lambda}t} \bar{\mathbf{v}}$ Lösungen des linearen DGL-Systems. Aufgrund der Linearität ergeben sich dann linear unabhängige reellwertige Lösungen durch:

$$\operatorname{Re}(\boldsymbol{\psi}) = \frac{1}{2}(\boldsymbol{\psi} + \bar{\boldsymbol{\psi}}), \quad \operatorname{Im}(\boldsymbol{\psi}) = \frac{1}{2i}(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}}) \quad \square$$

Probleme ergeben sich bei Eigenwerten mit einer *algebraischen Vielfachheit* $m \geq 2$, d.h. mit mehrfachen Nullstellen des charakteristischen Polynoms, da die Dimension des zugehörigen Eigenraumes (*geometrische Vielfachheit*) i.a. ungleich der algebraischen Vielfachheit m ist, wie man etwa in Beispiel D.5, Anhang D sehen kann.

Der folgende Satz ist das Analogon zu Satz 1.27 bzw. Satz 2.31, wo gezeigt wurde, dass für einen Eigenwert λ der Vielfachheit k alle Funktionen vom Typ $P(t)e^{\lambda t}$ Lösungen der homogenen DGL $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$ sind, sofern P ein Polynom mit $\operatorname{grad}(P) \leq k-1$ ist. Da ein DGL-System vorliegt, ist P entsprechend ein vektorwertiges Polynom, dessen Koeffizienten Eigenvektoren (evtl. höherer Stufen) von $\mathbf{A}$ sind. Dies soll zunächst durch eine Plausibilitätsüberlegung motiviert werden.

Nehmen wir hierzu an, dass Vektoren $\mathbf{a}$, $\mathbf{b}$ und $\mathbf{c}$ ($\mathbf{a}, \mathbf{b} \neq \mathbf{0}$) existieren, so dass

$$\boldsymbol{\varphi}(t) = \underbrace{(\mathbf{a}t^2 + \mathbf{b}t + \mathbf{c})}_{P(t)} e^{\lambda t}$$

eine Lösung des homogenen DGL-Systems $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$ ist.

Setzt man $\boldsymbol{\varphi}$ in das DGL-System ein, ergibt sich die Forderung

$$\underbrace{[\lambda \mathbf{a}t^2 + (2\mathbf{a} + \lambda \mathbf{b})t + \mathbf{b} + \lambda \mathbf{c}]}_{=\dot{\boldsymbol{\varphi}}(t)} e^{\lambda t} \stackrel{!}{=} \underbrace{\mathbf{A}(\mathbf{a}t^2 + \mathbf{b}t + \mathbf{c})}_{\mathbf{A}\boldsymbol{\varphi}} e^{\lambda t}$$

Kürzen von $e^{\lambda t}$ und anschließender Koeffizientenvergleich liefert

$$(\mathbf{A} - \lambda \mathbf{1})\mathbf{c} \stackrel{!}{=} \mathbf{b}, \quad \frac{1}{2}(\mathbf{A} - \lambda \mathbf{1})\mathbf{b} \stackrel{!}{=} \mathbf{a}, \quad (\mathbf{A} - \lambda \mathbf{1})\mathbf{a} \stackrel{!}{=} \mathbf{0}$$

¹⁶Siehe Satz D.4 b) des Anhangs.

und damit

$$(\mathbf{A} - \lambda \mathbf{1})^3 \mathbf{c} = (\mathbf{A} - \lambda \mathbf{1})^2 \mathbf{b} = (\mathbf{A} - \lambda \mathbf{1}) \mathbf{a} = \mathbf{0}.$$

Also sind $\mathbf{a}$, $\mathbf{b}$ und $\mathbf{c}$ Eigenvektoren der Stufe 1, 2 bzw. 3 und die gefundene Lösung lautet

$$\varphi(t) = (\mathbf{a}t^2 + \mathbf{b}t + \mathbf{c}) e^{\lambda t} = \left(\underbrace{(\mathbf{A} - \lambda \mathbf{1})^2 \mathbf{c}_2^1}_{=\mathbf{a}} t^2 + \underbrace{(\mathbf{A} - \lambda \mathbf{1}) \mathbf{c} t + \mathbf{c}}_{=\mathbf{b}} \right) e^{\lambda t}$$

Diese Aussage kann auf Eigenvektoren beliebiger Stufe verallgemeinert werden und liefert einen alternativen Weg zur Bestimmung der Lösung ohne Verwendung der Jordanform von Matrizen wie in Beispiel 2.30 iii) und iv). Genauer gilt:

2.35 Satz

Sei $\mathbf{A}$ eine $(n \times n)$ -Matrix mit dem Eigenwert λ und $\mathbf{v}$ ein zugehöriger Eigenvektor der Stufe k . Dann ist eine Lösung der DGL $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$ gegeben durch die Abbildung $\varphi : \mathbb{R} \rightarrow \mathbb{R}^n$ mit

$$\varphi(t) := e^{\lambda t} \sum_{j=0}^{k-1} \frac{t^j}{j!} (\mathbf{A} - \lambda \mathbf{1})^j \mathbf{v}.$$

□

Beweis:

Wegen $(\mathbf{A} - \lambda \mathbf{1})\mathbf{A} = \mathbf{A}(\mathbf{A} - \lambda \mathbf{1})$ und wegen

$$\begin{aligned} (*) \quad (\mathbf{A} - \lambda \mathbf{1})^{k-1}(\lambda \mathbf{v}) &= (\mathbf{A} - \lambda \mathbf{1})^{k-1}(\mathbf{A}\mathbf{v} - (\mathbf{A} - \lambda \mathbf{1})\mathbf{v}) \\ &= (\mathbf{A} - \lambda \mathbf{1})^{k-1}\mathbf{A}\mathbf{v} - \underbrace{(\mathbf{A} - \lambda \mathbf{1})^k \mathbf{v}}_{=\mathbf{0}} \\ &= (\mathbf{A} - \lambda \mathbf{1})^{k-1}\mathbf{A}\mathbf{v} \end{aligned}$$

folgt:

$$\begin{aligned} \dot{\varphi}(t) &= \lambda e^{\lambda t} \sum_{j=0}^{k-1} \frac{t^j}{j!} (\mathbf{A} - \lambda \mathbf{1})^j \mathbf{v} + e^{\lambda t} \sum_{j=1}^{k-1} \frac{t^{j-1}}{(j-1)!} (\mathbf{A} - \lambda \mathbf{1})^j \mathbf{v} \\ &= e^{\lambda t} \left[\sum_{j=0}^{k-1} \frac{t^j}{j!} (\mathbf{A} - \lambda \mathbf{1})^j (\lambda \mathbf{v}) + \sum_{j=0}^{k-2} \frac{t^j}{j!} (\mathbf{A} - \lambda \mathbf{1})^{j+1} \mathbf{v} \right] \\ &= e^{\lambda t} \left[\sum_{j=0}^{k-2} \frac{t^j}{j!} (\mathbf{A} - \lambda \mathbf{1})^j [\lambda \mathbf{v} + (\mathbf{A} - \lambda \mathbf{1})\mathbf{v}] + \frac{t^{k-1}}{(k-1)!} (\mathbf{A} - \lambda \mathbf{1})^{k-1} (\lambda \mathbf{v}) \right] \\ &\stackrel{(*)}{=} e^{\lambda t} \left[\sum_{j=0}^{k-2} \frac{t^j}{j!} (\mathbf{A} - \lambda \mathbf{1})^j \mathbf{A}\mathbf{v} + \frac{t^{k-1}}{(k-1)!} (\mathbf{A} - \lambda \mathbf{1})^{k-1} \mathbf{A}\mathbf{v} \right] \\ &= e^{\lambda t} \sum_{j=0}^{k-1} \frac{t^j}{j!} (\mathbf{A} - \lambda \mathbf{1})^j \mathbf{A}\mathbf{v} \\ &= \mathbf{A} e^{\lambda t} \sum_{j=0}^{k-1} \frac{t^j}{j!} (\mathbf{A} - \lambda \mathbf{1})^j \mathbf{v} = \mathbf{A} \varphi(t) \end{aligned}$$

■

2.36 Beispiel

Wir betrachten das DGL-System $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$, wobei $\mathbf{A}$ wie in Beispiel D.5 a) bzw. b) sei¹⁷. $\mathbf{A}$ besitzt die Eigenwerte $\lambda_1 = 0$ (algebraische Vielfachheit 1) und $\lambda_2 = -1$ (algebraische Vielfachheit 3). $\mathbf{v}_0 = (1, 1, 0, 0)^T$ ist Eigenvektor zum Eigenwert λ_1 . $\mathbf{v}_1 = (1, 0, -1, 0)^T$, $\mathbf{v}_2 = (0, 1, 1, -1)^T$ und $\mathbf{v}_3 = (0, 1, 0, 0)^T$ sind Eigenvektoren der Stufe 1, 2 und 3 zum Eigenwert λ_2 . Nach dem letzten Satz ist daher ein Fundamentalsystem des DGL-Systems $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$ gegeben durch

$$\begin{aligned}\varphi_0(t) &= e^{\lambda_1 t} \mathbf{v}_0 = \mathbf{v}_0 \\ \varphi_1(t) &= e^{\lambda_2 t} \mathbf{v}_1 = e^{-t} \mathbf{v}_1 \\ \varphi_2(t) &= e^{\lambda_2 t} \left(\mathbf{v}_2 + t(\mathbf{A} - \lambda_2 \mathbf{1})\mathbf{v}_2 \right) = e^{-t} \left(\begin{pmatrix} 0 \\ 1 \\ 1 \\ -1 \end{pmatrix} + t \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} \right) \\ \varphi_3(t) &= e^{\lambda_2 t} \left(\mathbf{v}_3 + t(\mathbf{A} - \lambda_2 \mathbf{1})\mathbf{v}_3 + \frac{t^2}{2}(\mathbf{A} - \lambda_2 \mathbf{1})^2 \mathbf{v}_3 \right) \\ &= e^{-t} \left(\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 0 \\ 1 \\ 1 \\ -1 \end{pmatrix} + \frac{t^2}{2} \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} \right)\end{aligned}$$

□

2.3. Laplace-Transformation

Häufig ist es nützlich, mittels einer Integraltransformation von einer DGL zu einer transformierten Gleichung überzugehen, die rein algebraisch und daher einfacher zu lösen ist, als die ursprüngliche DGL. Dennoch ist die Schwierigkeit, eine DGL zu lösen, damit nicht beseitigt worden. Im allgemeinen steckt sie in der Rücktransformation der Lösung der transformierten Gleichung. Als Integraltransformation benutzt man oft die *Laplace-Transformation*:

$$\mathcal{L}\{f\}(s) := \int_0^\infty e^{-st} f(t) dt$$

Der Grund für das Arbeiten mit der Laplace-Transformation liegt in der Rechenregel

$$\mathcal{L}\{\dot{f}\}(s) = s \cdot \mathcal{L}\{f\}(s) - f(0).$$

Sie ergibt sich folgendermaßen mittels partieller Integration:

$$\begin{aligned}\mathcal{L}\{\dot{f}\}(s) &= \int_0^\infty e^{-st} \dot{f}(t) dt \stackrel{\text{part. Int.}}{=} \left[e^{-st} f(t) \right]_{t=0}^{t=\infty} - \int_0^\infty \left(\frac{d}{dt} e^{-st} \right) f(t) dt \\ &= -f(0) + s \int_0^\infty e^{-st} f(t) dt \\ &= s \cdot \mathcal{L}\{f\}(s) - f(0)\end{aligned}$$

¹⁷Dieses DGL-System beschreibt die gedämpfte Schwingung zweier Pendel, die durch eine Feder miteinander verbunden sind

Diese Eigenschaft, dass die Ableitung bzgl. der Variablen t bei der Laplace-Transformation in eine Multiplikation mit der Variablen s überführt wird, liefert eine rein algebraische Gleichung. Deren Lösung enthält die Startwerte $y(0)$, $\dot{y}(0)$, $\dots$ (siehe den Differentiationssatz auf Seite 68), was die Berechnung der sonst vorkommenden Integrationskonstanten in der allgemeinen Lösung erspart. Man betrachte hierzu folgendes Beispiel.

2.37 Beispiel

Bei gegebener konstanter $(n \times n)$ -Matrix $\mathbf{A}$ betrachte man die lineare AWA

$$\dot{\mathbf{y}}(t) = \mathbf{A}\mathbf{y}(t) + \mathbf{b}(t), \quad \mathbf{y}(0) = \mathbf{y}_0,$$

Wendet man den Laplace-Operator an und bezeichnet $\mathcal{L}\{\mathbf{y}\}$ mit $\mathbf{Y}$, so erhält man unter Beachtung der Linearität von $\mathcal{L}$ und der gerade hergeleiteten Rechenregel

$$\mathcal{L}\{\dot{\mathbf{y}}\}(s) = s \cdot \mathcal{L}\{\mathbf{y}\}(s) - \mathbf{y}(0) = s \cdot \mathbf{Y}(s) - \mathbf{y}_0.$$

Es ergibt sich damit die äquivalente (rein algebraische Gleichung)

$$s \cdot \mathbf{Y}(s) - \mathbf{y}_0 = \mathbf{A} \cdot \mathbf{Y}(s) + \mathcal{L}\{\mathbf{b}\}(s),$$

mit der Lösung

$$\mathbf{Y}(s) = (s \cdot \mathbf{1} - \mathbf{A})^{-1} (\mathcal{L}\{\mathbf{b}\}(s) + \mathbf{y}_0).$$

Die gesuchte Lösung der AWA ist dann

$$\mathbf{y}(t) = \mathcal{L}^{-1}\{\mathbf{Y}(s)\}(t),$$

d.h. die Abbildung $\mathbf{y}$, deren Laplacetransformierte gerade $\mathbf{Y}$ liefert. Die genaue Lösung und weitere Beispiele werden in Beispiel 2.40 vorgestellt. $\square$

2.38 Bemerkung

- i) Für die Existenz der Laplacetransformierten $\mathcal{L}\{f\}$ ist die exponentielle Wachstumsbeschränkung $|f(t)| \leq K e^{ct}$ erforderlich. $\mathcal{L}\{f\}(s)$ ist dann eine analytische Funktion auf der Menge der komplexen Zahlen s mit $\operatorname{Re}(s) > c$. Alle Laplacefunktionen $g(s)$, die auf solche Weise entstehen, haben die Eigenschaft, dass sie nebst ihren Ableitungen beim Grenzübergang $s \rightarrow \infty$ verschwinden, genauer:

$$\lim_{\operatorname{Re}(s) \rightarrow \infty} \mathcal{L}\{f\}(s) = 0$$

- ii) Ist die Funktion $g(s)$ vorgegeben und das Urbild $f(t) := \mathcal{L}^{-1}\{g\}(t)$ gesucht, so gilt:

$$f(t) = \mathcal{L}^{-1}\{g\}(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{st} g(s) ds$$

- iii) Unter der *Faltung* zweier Funktionen f_1 und f_2 versteht man die folgendermaßen definierte Funktion $f_1 * f_2$:

$$(f_1 * f_2)(t) := \int_0^t f_1(t - \tau) f_2(\tau) d\tau$$

Es gilt $f_1 * f_2 = f_2 * f_1$. □

Die Laplacetransformation $\mathcal{L}$ und ihre Inverse $\mathcal{L}^{-1}$ haben folgende Eigenschaften¹⁸:

1. Additionssatz

$$\mathcal{L}\{af_1 + bf_2\} = a\mathcal{L}\{f_1\} + b\mathcal{L}\{f_2\} \quad (a, b \in \mathbb{R} \text{ konstant})$$

$$\mathcal{L}^{-1}\{ag_1 + bg_2\} = a\mathcal{L}^{-1}\{g_1\} + b\mathcal{L}^{-1}\{g_2\} \quad (a, b \in \mathbb{R} \text{ konstant})$$

2. Differentiationssatz

$$\mathcal{L}\{f^{(n)}\}(s) = s^n \mathcal{L}\{f\}(s) - \sum_{j=0}^{n-1} s^{n-1-j} f^{(j)}(0)$$

$$\mathcal{L}^{-1}\left\{\frac{d^n}{ds^n} g\right\}(t) = (-1)^n t^n \mathcal{L}^{-1}\{g\}(t)$$

3. Integrationssatz

$$\mathcal{L}\left\{\int_0^t f(\tau) d\tau\right\}(s) = \frac{1}{s} \mathcal{L}\{f\}(s)$$

$$\mathcal{L}^{-1}\left\{\int_s^\infty g(\sigma) d\sigma\right\}(t) = \frac{1}{t} \mathcal{L}^{-1}\{g\}(t)$$

4. Verschiebungssatz

$$\mathcal{L}\{f(t - a)\}(s) = e^{-as} \mathcal{L}\{f(t)\}(s)$$

$$\mathcal{L}^{-1}\{g(s + a)\}(t) = e^{-at} \mathcal{L}^{-1}\{g(s)\}(t)$$

5. Ähnlichkeitssatz

$$\mathcal{L}\{f(at)\}(s) = \frac{1}{a} \mathcal{L}\{f\}\left(\frac{s}{a}\right)$$

$$\mathcal{L}^{-1}\{g(as)\}(t) = \frac{1}{a} \mathcal{L}^{-1}\{g(s)\}\left(\frac{t}{a}\right)$$

¹⁸Wir setzen stillschweigend voraus, dass die auftretenden Funktionen transformierbar sind.

6. Dämpfungssatz

$$\mathcal{L}\{e^{-at} f(t)\}(s) = \mathcal{L}\{f\}(s + a)$$

$$\mathcal{L}^{-1}\{e^{-as} g(s)\}(t) = \mathcal{L}^{-1}\{g\}(t - a)$$

7. Multiplikationssatz

$$\mathcal{L}\{t^n f(t)\}(s) = (-1)^n \frac{d^n}{ds^n} \mathcal{L}\{f\}(s)$$

8. Divisionssatz

$$\mathcal{L}\left\{\frac{1}{t} f(t)\right\}(s) = \int_s^\infty \mathcal{L}\{f\}(\sigma) d\sigma$$

$$\mathcal{L}^{-1}\left\{\frac{1}{s} g(s)\right\}(t) = \int_0^t \mathcal{L}^{-1}\{g\}(\sigma) d\sigma$$

9. Faltungssatz

$$\mathcal{L}\{f_1 * f_2\} = \mathcal{L}\{f_1\} \cdot \mathcal{L}\{f_2\}$$

$$\mathcal{L}^{-1}\{g_1 \cdot g_2\} = \mathcal{L}^{-1}\{g_1\} * \mathcal{L}^{-1}\{g_2\}$$

2.39 Beispiel

Offensichtlich gilt

$$\mathcal{L}\{1\}(s) = \int_0^\infty e^{-st} dt = \left[-\frac{1}{s} e^{-st} \right]_{t=0}^{t=\infty} = \frac{1}{s}.$$

Hieraus folgt mit Hilfe des Multiplikationssatzes:

$$\mathcal{L}\{t^n\}(s) = \mathcal{L}\{t^n \cdot 1\}(s) = (-1)^n \frac{d^n}{ds^n} \left(\frac{1}{s} \right) = \frac{n!}{s^{n+1}}.$$

Mit dem Dämpfungssatz folgt hieraus wiederum

$$\mathcal{L}\{e^{-at} t^n\}(s) = \frac{n!}{(s+a)^{n+1}}.$$

□

Die folgende Tabelle gibt ein paar Beispiele von Laplacetransformierten.

$f(t)$	1	t^n	$e^{\alpha t}$	$t^n e^{\alpha t}$	$\cos(\alpha t)$	$\sin(\alpha t)$	$\cosh(\alpha t)$	$\sinh(\alpha t)$
$\mathcal{L}\{f\}(s)$	$\frac{1}{s}$	$\frac{n!}{s^{n+1}}$	$\frac{1}{s-\alpha}$	$\frac{n!}{(s-\alpha)^{n+1}}$	$\frac{s}{s^2+\alpha^2}$	$\frac{\alpha}{s^2+\alpha^2}$	$\frac{s}{s^2-\alpha^2}$	$\frac{\alpha}{s^2-\alpha^2}$

Die entsprechende Tabelle für die inverse Laplacetransformation hat folgendes Aussehen.

$g(s)$	$\frac{1}{s}$	$\frac{1}{s^{n+1}}$	$\frac{1}{s-\alpha}$	$\frac{1}{(s-\alpha)^{n+1}}$	$\frac{s}{s^2+\alpha^2}$	$\frac{1}{s^2+\alpha^2}$	$\frac{s}{s^2-\alpha^2}$	$\frac{1}{s^2-\alpha^2}$
$\mathcal{L}^{-1}\{g\}(t)$	1	$\frac{t^n}{n!}$	$e^{\alpha t}$	$\frac{t^n e^{\alpha t}}{n!}$	$\cos(\alpha t)$	$\frac{\sin(\alpha t)}{\alpha}$	$\cosh(\alpha t)$	$\frac{\sinh(\alpha t)}{\alpha}$

Das folgende Beispiel zeigt die früher schon erwähnte Anwendung der Laplacetransformation bei der Lösung von Differentialgleichungen.

2.40 Beispiel

i) Wir betrachten die lineare AWA

$$\ddot{y}(t) - 3\ddot{y}(t) + \dot{y}(t) - 3y(t) = 6e^{3t}; \quad y(0) = 1, \quad \dot{y}(0) = 0, \quad \ddot{y}(0) = 1$$

Anwendung des Laplaceoperators auf die DGL liefert:

$$\mathcal{L}\{\ddot{y}(t) - 3\ddot{y}(t) + \dot{y}(t) - 3y(t)\}(s) = 6\mathcal{L}\{e^{3t}\}(s)$$

Bezeichnen wir $\mathcal{L}\{y\}$ mit Y , so liefert die Anwendung der Rechenregeln für den Laplaceoperator:

$$s^3Y(s) - [s^2y(0) + s\dot{y}(0) + \ddot{y}(0)] - 3\{s^2Y(s) - [sy(0) + \dot{y}(0)]\} + sY(s) - y(0) - 3Y(s) = \frac{6}{s-3}$$

Einsetzen der Anfangsbedingungen und Zusammenfassen liefert:

$$(s^2 + 1)(s - 3)Y(s) = \frac{6}{s-3} + s^2 - 3s + 2$$

Löst man diese Gleichung nach Y auf, ergibt sich nach einigen algebraischen Umformungen:

$$Y(s) = \frac{1}{25} \left(\frac{29s}{s^2+1} - \frac{3}{s^2+1} + \frac{15}{(s-3)^2} - \frac{4}{s-3} \right)$$

Die Rücktransformation (siehe vorherige Tabelle) ergibt

$$y(t) = \mathcal{L}^{-1}\{Y(s)\}(t) = \frac{1}{25} (29 \cos(t) - 3 \sin(t) + 15t e^{3t} - 4 e^{3t})$$

ii) Auch DGL-Systeme lassen sich auf diese Art lösen. Betrachten wir hierzu das lineare DGL-System

$$\begin{aligned} \dot{y}_1 &= y_1 - y_2 + t \\ \dot{y}_2 &= 4y_1 - 3y_2 + 2 \end{aligned}$$

Da Anfangsbedingungen fehlen, definieren wir $y_1(0) := c_1$, $y_2(0) := c_2$ und wenden auf beide Gleichungen des DGL-Systems die Laplacetransformation an. Es ergibt sich das folgende Gleichungssystem für die Laplacetransformierten $Y_i := \mathcal{L}\{y_i\}$

$$\begin{aligned}(s-1)Y_1 + Y_2 &= \frac{1}{s^2} + c_1 \\ -4Y_1 + (s+3)Y_2 &= \frac{2}{s} + c_2\end{aligned}$$

Die Lösung dieses Gleichungssystems lautet

$$\begin{aligned}Y_1(s) &= \frac{c_1 s^3 + (3c_1 - c_2)s^2 - s + 3}{s^2(s+1)^2} \\ Y_2(s) &= \frac{c_2 s^3 + (2 + 4c_1 - c_2)s^2 - 2s + 4}{s^2(s+1)^2}\end{aligned}$$

Die Rücktransformation ergibt dann die gesuchte Lösung

$$\begin{aligned}y_1(t) &= \mathcal{L}^{-1}\{Y_1(s)\}(t) = 3t - 7 + e^{-t}(C_1 t + C_2) \\ y_2(t) &= \mathcal{L}^{-1}\{Y_2(s)\}(t) = 4t - 10 + e^{-t}(2C_1 t + 2C_2 - C_1)\end{aligned}$$

mit den Kurzbezeichnungen $C_1 := 4 + 2c_1 - c_2$ und $C_2 := 7 + c_1$.

iii) Betrachtet man die lineare AWA

$$\dot{\mathbf{y}}(t) = \mathbf{A}\mathbf{y}(t) + \mathbf{b}(t), \quad \mathbf{y}(0) = \mathbf{y}_0$$

mit der $(n \times n)$ -Matrix $\mathbf{A}$, so liefert die Laplacetransformation mit $\mathbf{Y} := \mathcal{L}\{\mathbf{y}\}$ wie in Beispiel 2.37 gezeigt die Lösung

$$\mathbf{Y}(s) = (s\mathbf{I} - \mathbf{A})^{-1} \cdot (\mathcal{L}\{\mathbf{b}\}(s) + \mathbf{y}_0)$$

Beachtet man, dass - genau wie im eindimensionalen - gilt

$$\mathcal{L}^{-1}\{(s\mathbf{I} - \mathbf{A})^{-1}\}(t) = e^{\mathbf{A}t},$$

so ergibt sich die gesuchte Lösung der AWA mit Hilfe der Faltungsregel als

$$\begin{aligned}
 \mathbf{y}(t) &= \mathcal{L}^{-1}\{\mathbf{Y}(s)\}(t) \\
 &= \mathcal{L}^{-1}\{(s-\mathbf{A})^{-1} \cdot \mathcal{L}\{\mathbf{b}(t)\}(s)\}(t) + \mathcal{L}^{-1}\{(s-\mathbf{A})^{-1} \cdot \mathbf{y}_0\}(t) \\
 &= \mathcal{L}^{-1}\{(s-\mathbf{A})^{-1}\} * \mathcal{L}^{-1}\{\mathcal{L}\{\mathbf{b}(t)\}(s)\}(t) + e^{\mathbf{A}t} \mathbf{y}_0 \\
 &= e^{\mathbf{A}t} * \mathbf{b}(t) + e^{\mathbf{A}t} \mathbf{y}_0 \\
 &= \int_0^t e^{\mathbf{A}(t-\tau)} \mathbf{b}(\tau) d\tau + e^{\mathbf{A}t} \mathbf{y}_0 \\
 &= e^{\mathbf{A}t} \int_0^t e^{-\mathbf{A}\tau} \mathbf{b}(\tau) d\tau + e^{\mathbf{A}t} \mathbf{y}_0
 \end{aligned}$$

Dies ist gerade die Aussage von Satz 2.29 mit der Lösungsmatrix $\Phi(t) = e^{\mathbf{A}t}$.

- iv) Wendet man die Laplacetransformation auf eine partielle DGL mit n unabhängigen Veränderlichen an, entsteht eine partielle DGL mit „nur“ noch $n-1$ unabhängigen Veränderlichen. Vorausgesetzt ist dabei, dass die Anfangswerte hinsichtlich der Variablen, die man der Laplacetransformation unterwerfen will, gegeben sind. Insbesondere kann man aus einer partiellen DGL mit zwei unabhängigen Veränderlichen eine gewöhnliche DGL machen, wie das folgende Beispiel zeigt.

Der Wert $C = C(x, t)$ eines europäischen Call in Abhängigkeit vom Aktienkurs x und der Laufzeit t des Calls ist bestimmt durch die partielle DGL von Black/Scholes:

$$\frac{1}{2}\sigma^2 x^2 \frac{\partial^2 C}{\partial x^2} + bx \frac{\partial C}{\partial x} - iC = \frac{\partial C}{\partial t}$$

Hierbei bezeichnet b bzw. i die stetigen Bestandshaltekosten bzw. den stetigen Zinssatz und σ die Volatilität des Aktienkurses. Die Randbedingung dieser DGL ist durch den Wert $C(x, 0)$ des Call zum Verfallszeitpunkt $t=0$ gegeben.

Unterwirft man diese DGL der Laplacetransformation $\mathcal{L}_t$ bzgl. der Variablen t , setzt $c(x, s) := \mathcal{L}_t\{C(x, t)\}(s)$ und beachtet, dass gilt

$$\mathcal{L}_t\left\{\frac{\partial^n C}{\partial t^n}\right\}(s) = \frac{d^n}{ds^n} \mathcal{L}_t\{C(x, t)\}(s) = \frac{d^n c(x, s)}{ds^n},$$

so erhält man folgende gewöhnliche lineare DGL in der Variablen x :

$$(*) \quad \frac{1}{2}\sigma^2 x^2 \frac{d^2 c}{dx^2} + bx \frac{dc}{dx} - ic(x, s) = sc(x, s) - C(x, 0)$$

Wie früher gesehen lautet die allgemeine Lösung dieser DGL

$$c(x, s) = \lambda_1(s)\varphi_1(x, s) + \lambda_2(s)\varphi_2(x, s) + \psi(x, s).$$

Hierbei sei φ_1, φ_2 ein Fundamentalsystem der homogenen und ψ eine partikuläre Lösung der DGL (*).

Etwas Vorsicht ist bei den Integrationskonstanten λ_j angebracht. Diese sind nämlich Funktionen in Abhängigkeit von s , da bei der Integration der DGL (*) (bzgl. der Variablen x !!!) die Größe s als Konstante behandelt wird. λ_1 und λ_2 sind so zu bestimmen, dass die Anfangsbedingung(-en) für C bzw. die Laplacetransformierte c erfüllt sind. Aufgrund dieser s -Abhängigkeit der λ_j wird die Rücktransformation deutlich erschwert.

□

3

Stabilität dynamischer Systeme

Wir wollen nun die „Stabilität“ der Lösungen eines DGL-Systems $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$ untersuchen, wobei die Abbildung $\mathbf{f} : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ stetig differenzierbar sei. Wir fragen dabei, wie sich Störungen bzw. Änderungen der Anfangsbedingung $\mathbf{y}(t_0)$ auf die Lösung des Systems auswirken. Man wird das System dann als „stabil“ empfinden, wenn „kleine“ Störungen des Startwertes nur „kleine“ Änderungen bei der Lösung hervorrufen, d.h. wenn die Lösungen als Funktion des Startwertes stetig sind. Vor einer genauen Definition betrachten wir zunächst ein Beispiel.

3.1 Beispiel

Betrachte die DGL $\dot{y} = y \cdot (2 - y)$. Die Lösung dieser separablen DGL mit der Anfangsbedingung $y(0)=a$ wurde in Beispiel 2.12 ii) bestimmt und lautet:

$$y(t) = \frac{2a e^{2t}}{2 - a + a e^{2t}}$$

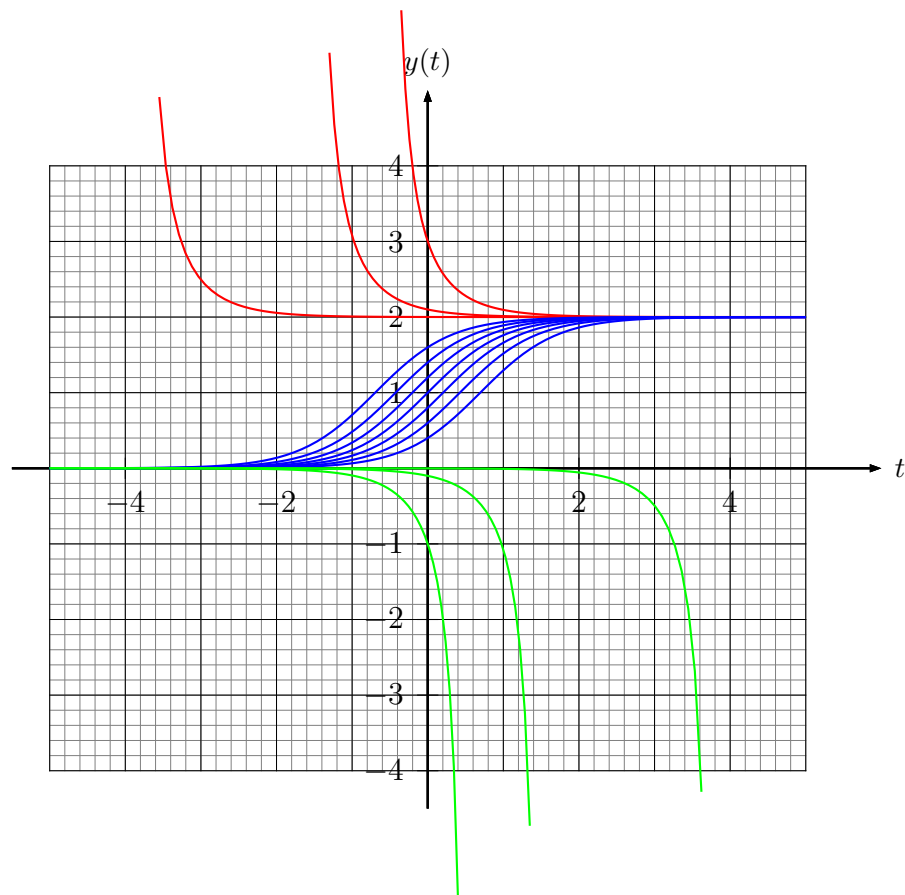
Zwei Lösungen dieser DGL, nämlich $y \equiv 0$ und $y \equiv 2$, die durch die Startwerte $a = 0$ und $a = 2$ entstehen, haben die Eigenschaft, zeitlich konstant (*stationär*) zu sein.

Die stationäre Lösung $y \equiv 2$ wird asymptotisch stabil genannt, da jede Lösung y mit dem Startwert $y(0) > 0$ mit zunehmender Zeit gegen sie konvergiert.

Die zweite stationäre Lösung $y \equiv 0$ zeigt dagegen fast das gegenteilige Verhalten. Jede Lösung y mit $y(0) \neq 0$ läuft mit zunehmender Zeit von ihr weg¹ (siehe hierzu auch die nachfolgende Graphik). Man nennt sie daher instabil.

Die übrigen Lösungen zeigen das erwartete „normale“ Verhalten, d.h. kleine Änderungen im Startwert a führen zu kleinen Änderungen in der Lösung. Man spricht in diesem Fall von neutral stabilen Lösungen. Z.B. unterscheiden sich die Lösungen zu den Startwerten $a = 0.8$, $a = 1.0$, $a = 1.2$, $a = 1.4$ kaum. (siehe auch nachfolgende Skizze)

¹Zu beachten ist, dass die Lösungsfunktionen eine Polstelle bei $t = \frac{1}{2} \ln(\frac{a-2}{a})$ besitzen. Für $a < 0$ liefert dies eine Polstelle im positiven Bereich.



Lösungen der DGL $\dot{y} = y(2 - y)$ mit den Startwerten
 $y(0) = -1.0, -0.1, -0.001, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 2.001, 2.1, 3.0$. $\square$

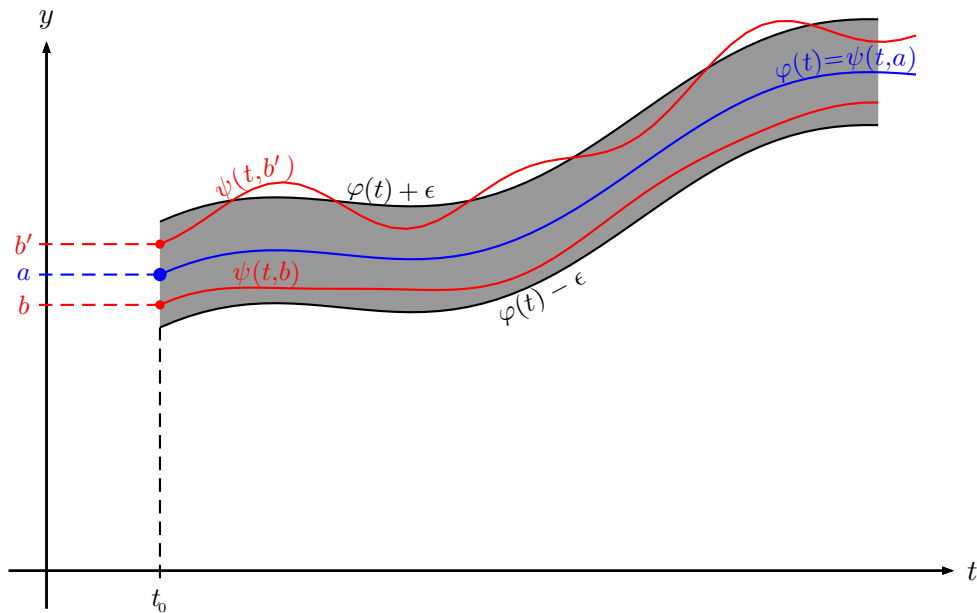
3.2 Definition

$\psi(t, \mathbf{z})$ bezeichne die Lösung der AWA $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$, $\mathbf{y}(t_0) = \mathbf{z}$.

- a) Die Lösung $\psi(t, \mathbf{a})$ heißt *stabil*, wenn zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert, mit

$$\|\mathbf{a} - \mathbf{b}\| < \delta \implies \forall t \geq t_0 : \|\psi(t, \mathbf{a}) - \psi(t, \mathbf{b})\| < \varepsilon$$

- b) $\psi(t, \mathbf{a})$ heißt *asymptotisch stabil*, wenn $\psi(t, \mathbf{b})$ mit wachsendem t gegen $\psi(t, \mathbf{a})$ konvergiert. Je nachdem, ob dies für alle $\mathbf{b}$ oder nur für Startwerte $\mathbf{b}$ in einer Umgebung von $\mathbf{a}$ gilt, spricht man von *globaler* oder *lokaler asymptotischer Stabilität*.
- c) Eine stabile, jedoch nicht asymptotisch stabile Lösung heißt *neutral stabil*. Eine weder neutral noch asymptotisch stabile Lösung heißt *instabil*. $\square$



ε -Umgebung einer Lösung $\varphi(t) = \psi(t, \mathbf{a})$ der DGL $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$.

3.3 Bemerkung

Im Prinzip genügt es, Stabilitätsuntersuchungen auf *stationäre*, d.h. *zeitlich konstante Lösungen* φ einzuschränken, für die also $\mathbf{f}(t, \varphi) = \dot{\varphi} = \mathbf{0}$ gilt. Untersucht man nämlich die Stabilität einer vorgegebenen Lösung φ , liegt es nahe, mit den neuen Koordinaten

$$\bar{\mathbf{y}} := \mathbf{y} - \varphi$$

zu arbeiten, da $\bar{\mathbf{y}}$ die Abweichung der Lösung $\mathbf{y}$ von φ angibt. Wegen $\dot{\bar{\mathbf{y}}}(t) = \dot{\mathbf{y}}(t) - \dot{\varphi}(t)$ führt dies zur transformierten DGL

$$\dot{\bar{\mathbf{y}}} = \underbrace{\mathbf{f}(t, \underbrace{\bar{\mathbf{y}} + \varphi}_{\mathbf{y}})}_{\mathbf{y}} - \dot{\varphi}$$

Die „ungestörte“ Lösung $\mathbf{y} = \varphi$ entspricht in der neuen DGL der stationären Lösung $\bar{\mathbf{y}} \equiv \mathbf{0}$, so dass im Prinzip sogar Stabilitätsuntersuchungen für die stationäre Lösung $\mathbf{y} \equiv \mathbf{0}$ genügen (siehe hierzu auch Satz 3.5 im Fall linearer DGLen). Entsprechend kann durch die Zeittransformation $\tau := t - t_0$ die Fragestellung auch auf den Startzeitpunkt $t = 0$ reduziert werden. $\square$

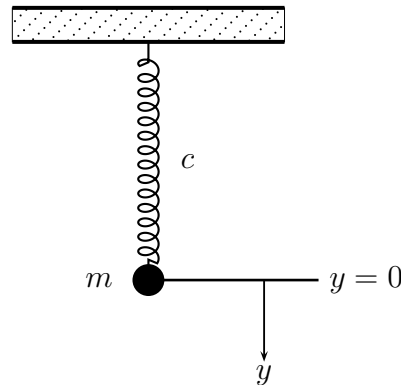
Um quantitative Stabilitätsaussagen machen zu können, werden wir uns - wie schon bei der Lösung von DGLen - zunächst auf lineare DGLen zurückziehen. Das folgende Beispiel zeigt die unterschiedlichen Fälle, die sich bei der Stabilitätsuntersuchung linearer Systeme ergeben.

3.4 Beispiel

Wir betrachten die folgende lineare DGL zweiter Ordnung:

$$m\ddot{y} + r\dot{y} + cy = 0, \quad (m, c > 0, r \geq 0)$$

Hierdurch wird z.B. die Bewegung einer an einer Feder aufgehängten Masse m beschrieben, die durch eine Auslenkung zum Zeitpunkt 0 hervorgerufen wird (*Federpendel*). $y(t)$ gibt den Ort der Masse zum Zeitpunkt t an und wird gemessen durch den Abstand zur Ruheposition, die als Nullpunkt des y -Koordinatensystems gewählt wird² (siehe auch nachfolgende Skizze). Der Term ry beschreibt eine proportional zur Geschwindigkeit wirkende Reibung. Der Term cy gibt außerdem an, dass die Rückstellkraft proportional zur Auslenkung der Feder ist.



Die Eigenwerte dieser DGL, d.h. die Nullstellen der charakteristischen Gleichung lauten

$$\lambda_{1/2} = \alpha \pm \sqrt{\beta} \quad \text{mit} \quad \alpha = -\frac{r}{2m}, \quad \beta = \frac{r^2 - 4mc}{4m^2}.$$

Die allgemeine Lösung y der DGL lautet daher gemäß Satz 2.31:

$$y(t) = \begin{cases} C_1 e^{\alpha t} + C_2 t e^{\alpha t} & \text{falls } \beta = 0 \\ C_1 e^{(\alpha - \sqrt{\beta})t} + C_2 e^{(\alpha + \sqrt{\beta})t} & \text{falls } \beta > 0 \\ C_1 e^{\alpha t} \sin(t\sqrt{|\beta|}) + C_2 e^{\alpha t} \cos(t\sqrt{|\beta|}) & \text{falls } \beta < 0 \end{cases}$$

Zur Stabilitätsuntersuchung betrachten wir die Fälle mit ($r > 0$) bzw. ohne Reibung ($r = 0$).

1. Fall: $r > 0$

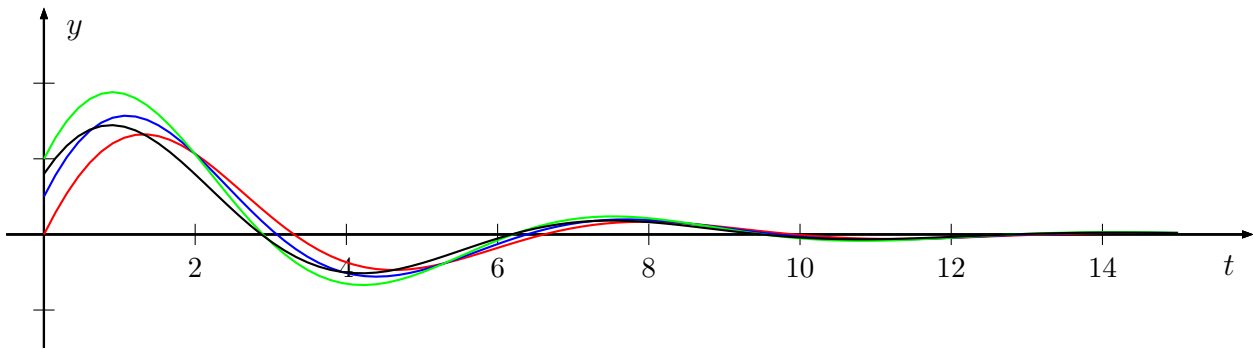
In diesem Fall gilt $\alpha = -\frac{r}{2m} < 0$ bzw. $\alpha - \sqrt{\beta} \leq \alpha + \sqrt{\beta} < 0$ (sofern $\beta > 0$). Daher folgt

$$\lim_{t \rightarrow \infty} e^{\alpha t} = \lim_{t \rightarrow \infty} e^{(\alpha \pm \sqrt{\beta})t} = \lim_{t \rightarrow \infty} e^{\alpha t} \sin(t\sqrt{|\beta|}) = \lim_{t \rightarrow \infty} e^{\alpha t} \cos(t\sqrt{|\beta|}) = 0,$$

so dass $\lim_{t \rightarrow \infty} y(t) = 0$ für jede Lösung $y(t)$ gilt. Damit ist die stationäre Lösung $y \equiv 0$ stabil und global asymptotisch stabil, da jede andere Lösung gegen sie konvergiert. Nachfolgende Graphik beschreibt im Fall³ $r > 0$ Lösungen mit den Startwerten $y(0) = 0$, $y(0) = \frac{1}{2}$, $y(0) = 0.5$ und $y(0) = 0.8$. Die zur Eindeutigkeit der Lösung erforderliche zweite Startbedingung (die Anfangsgeschwindigkeit $\dot{y}(0)$) wurde in den ersten drei Fällen gleich 2, im vierten Fall gleich 1.5 gewählt. Zu beachten ist, dass die Richtung der positiven y -Achse im Gegensatz zur vorigen Zeichnung nach oben zeigt.

²Die ursprüngliche Gleichung mit der Gravitationskonstante g lautet $m\ddot{z} + r\dot{z} + cz = -mg$. Die angegebene Gleichung entsteht erst durch die Transformation $y := z + \frac{mg}{c}$. Hierdurch geht die stationäre Lösung $z^* \equiv -\frac{mg}{c}$ (Ruhelage) gemäß Bemerkung 3.3 in die stationäre Lösung $y \equiv 0$ über.

³Ist $\beta \geq 0$ liegt eine Überdämpfung vor, d.h. die Reibung ist so stark, dass keine Schwingung mehr entsteht. Ist $\beta < 0$, entsteht eine gedämpfte Schwingung, die gegen Null konvergiert (siehe Graphik).



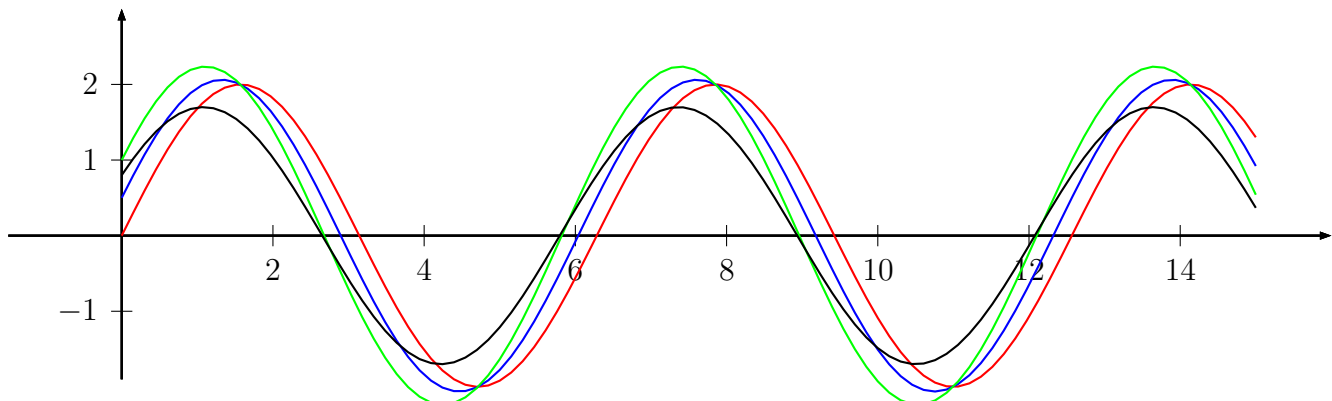
Lösungen der DGL $m\ddot{y} + r\dot{y} + cy = 0$ ($m=1, r=0.625, c=1$) für die Startwerte
 $y(0)=0, \dot{y}(0)=2$, $y(0)=0.5, \dot{y}(0)=2$, $y(0)=1, \dot{y}(0)=2$, $y(0)=0.8, \dot{y}(0)=1.5$

2. Fall: $r = 0$

In diesem Fall gilt $\alpha=0, \beta<0$ und die allgemeine Lösung lautet in diesem Fall:

$$y(t) = C_1 \sin(t\sqrt{\frac{c}{m}}) + C_2 \cos(t\sqrt{\frac{c}{m}})$$

Wegen der fehlenden Reibung entsteht eine unendlich lang andauernde Schwingung. Jede Lösung ist nun neutral stabil, d.h. zwei Lösungen, deren Startwerte sich um einen gewissen Betrag unterscheiden, werden für alle Zeiten um höchstens diesen Betrag differieren (siehe nachfolgende Skizze anstelle einer Rechnung. Für einen Beweis sei auf Satz 3.6 verwiesen.).



Lösungen der DGL $m\ddot{y} + r\dot{y} + cy = 0$ ($m=1, r=0, c=1$) für die Startwerte
 $y(0)=0, \dot{y}(0)=2$, $y(0)=0.5, \dot{y}(0)=2$, $y(0)=1, \dot{y}(0)=2$, $y(0)=0.8, \dot{y}(0)=1.5$ $\square$

Die Überlegungen des letzten Beispiels lassen sich auf beliebige lineare DGL-Systeme übertragen. Vor dem entscheidenden Stabilitätssatz soll zunächst gezeigt werden, dass es genügt homogene DGLen zu betrachten. Dies ist ein Spezialfall von Bemerkung 3.3, wo gezeigt wurde, dass es ausreicht die Stabilität der Lösung $\mathbf{y} \equiv \mathbf{0}$ zu untersuchen.

3.5 Satz

Jede partikuläre Lösung des inhomogenen linearen DGL-Systems $\dot{\mathbf{y}}(t) = \mathbf{A}(t)\mathbf{y}(t) + \mathbf{b}(t)$ mit $\mathbf{A}(t) \in \mathbb{R}^{n \times n}$ hat die gleichen Stabilitätseigenschaften wie die stationäre Lösung $\mathbf{y} \equiv \mathbf{0}$ des homogenen Systems $\dot{\mathbf{y}}(t) = \mathbf{A}(t)\mathbf{y}(t)$ und umgekehrt. $\square$

Beweis:

Seien $\varphi(t)$ und $\psi(t)$ zwei Lösungen der inhomogenen DGL. Dann ist aufgrund der Linearität die Abbildung $\vartheta(t) := \psi(t) - \varphi(t)$ eine Lösung der homogenen DGL. Wegen

$$\|\vartheta - \mathbf{0}\| = \|\vartheta\| = \|\psi - \varphi\|$$

ergibt sich daher, dass φ die gleichen Stabilitätseigenschaften wie $\mathbf{0}$ hat.

Ist umgekehrt φ eine Lösung der inhomogenen DGL und ϑ eine homogene Lösung, so ist $\psi := \varphi - \vartheta$ eine weitere inhomogene Lösung. Wegen

$$\|\vartheta - \mathbf{0}\| = \|\psi - \varphi\|$$

hat dann $\mathbf{0}$ die gleichen Stabilitätseigenschaften wie φ . ■

Beispiel 3.4 lässt vermuten, dass sich die Stabilitätseigenschaften eines homogenen DGL-Systems $\dot{\mathbf{y}}(t) = \mathbf{A}\mathbf{y}(t)$ mit konstanter Systemmatrix $\mathbf{A}$ an dem Verhalten der Eigenwerte ablesen lassen. Dies beruht darauf, dass gemäß den Sätzen 2.33 und 2.35 ein Fundamentalsystem von Lösungen gegeben ist durch

$$\varphi(t) := e^{\lambda t}(\mathbf{a} + \mathbf{b}t + \mathbf{c}t^2 + \dots),$$

wobei λ ein Eigenwert der Matrix $\mathbf{A}$ und $\mathbf{a}$, $\mathbf{b}$ und $\mathbf{c}$ Eigenvektoren entsprechender Stufe sind. Man erhält folgendes Ergebnis.

3.6 Satz

Sei $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$ ein homogenes lineares DGL-System mit der reellen invertierbaren⁴ $(n \times n)$ -Matrix $\mathbf{A}$. Dann gelten folgende Aussagen:

- a) Sind die Realteile aller Eigenwerte negativ, so ist die stationäre Lösung $\mathbf{y} \equiv \mathbf{0}$ global asymptotisch stabil.
- b) Existiert ein Eigenwert mit positivem Realteil, so ist $\mathbf{y} \equiv \mathbf{0}$ eine instabile Lösung.
- c) Besitzt $\mathbf{A}$ nur Eigenwerte, deren Realteil kleiner oder gleich 0 ist und besitzen die rein komplexen Eigenwerte (d.h. mit Realteil 0) eine vollständige Menge von Eigenvektoren (algebraische = geometrische Vielfachheit des Eigenwertes), so ist $\mathbf{y} \equiv \mathbf{0}$ eine neutral stabile stationäre Lösung.
- d) Besitzt $\mathbf{A}$ mindestens einen rein komplexen Eigenwert ohne vollständige Menge von Eigenvektoren, so ist $\mathbf{y} \equiv \mathbf{0}$ eine instabile stationäre Lösung. □

⁴Ist $\mathbf{A}$ singular, so existiert ein Vektor $\mathbf{v} \neq \mathbf{0}$, der die Gleichung $\mathbf{A}\mathbf{v} = \mathbf{0}$ löst. Wegen $\mathbf{0} = \dot{\mathbf{v}} = \mathbf{A}\mathbf{v} = \mathbf{0}\mathbf{v}$ ist $\mathbf{v}$ eine stationäre, d.h. zeitunabhängige Lösung der DGL $\dot{\mathbf{y}}(t) = \mathbf{A}\mathbf{y}(t)$ und ein Eigenvektor von $\mathbf{A}$ zum Eigenwert 0. In diesem Fall kann die Lösung $\mathbf{y} = \mathbf{0}$ höchstens noch lokal jedoch nicht mehr global stabil sein, da für jedes $r \in \mathbb{R}$ auch $r \cdot \mathbf{v}$ eine Lösung ist.

Beweis:

- a) Gemäß den Sätzen 2.33 und 2.35 existiert eine Basis von Lösungen der Form⁵

$$\varphi(t) := e^{\lambda t} \mathbf{v}(t) .$$

Hierbei ist λ ein Eigenwert und $\mathbf{v}(t)$ eine vektorwertige Abbildung, deren Komponentenfunktionen aus Polynomen bestehen. Nun gilt nach Voraussetzung $\lambda = a + i \cdot b$, mit $a < 0$ und daher

$$\varphi(t) = e^{at} e^{ibt} \mathbf{v}(t) .$$

Weil für jedes Polynom $P(t)$

$$\lim_{t \rightarrow \infty} e^{at} P(t) = 0$$

gilt, folgt daher wegen $|e^{ibt}| = 1$

$$\lim_{t \rightarrow \infty} \varphi(t) = \mathbf{0} .$$

Da sich jede Lösung als Linearkombination dieser Basislösungen φ darstellen lässt, muss $\mathbf{0}$ global stabil sein.

- b) Ist $\lambda = a + i \cdot b$ ein Eigenwert mit $a > 0$, so ist - wie schon in Teil a) gesehen - die Abbildung $\varphi(t) := e^{at} e^{ibt} \mathbf{v}(t)$ eine Lösung des DGL-Systems. Wegen $a > 0$ und $|e^{ibt}| = 1$ gilt diesmal jedoch

$$\lim_{t \rightarrow \infty} \|\varphi(t)\| = \infty .$$

Damit ist $\delta \cdot \varphi$ eine Lösung des DGL-Systems $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$, die - je nach Wahl von δ - beliebig nahe bei $\mathbf{y} \equiv \mathbf{0}$ startet, sich wegen $\lim_{t \rightarrow \infty} \|\delta \cdot \varphi(t)\| = |\delta| \cdot \lim_{t \rightarrow \infty} \|\varphi(t)\| = \infty$ mit zunehmender Zeit jedoch immer weiter von $\mathbf{y} \equiv \mathbf{0}$ entfernt. Daher ist $\mathbf{y} \equiv \mathbf{0}$ instabil.

- c) Da die Fundamentallösungen, deren Eigenwerte negative Realteile besitzen, gemäß Teil a) für $t \rightarrow \infty$ gegen $\mathbf{0}$ konvergieren, können wir o.B.d.A. annehmen, dass nur rein komplexe Eigenwerte existieren. Weil nach Voraussetzung zu jedem Eigenwert eine vollständige Menge von Eigenvektoren existiert, kann eine Basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ von Eigenvektoren gefunden werden. Die zugehörigen Lösungen

$$\varphi_k(t) := e^{ib_k t} \mathbf{v}_k \quad (k = 1, \dots, n)$$

bilden gemäß Satz 2.33 eine Basis⁶ für den Lösungsraum der DGL $\dot{\mathbf{y}} = \mathbf{A} \cdot \mathbf{y}$.⁷ Außer-

⁵Da $\mathbf{A}$ eine reelle Matrix ist, treten komplexe Eigenwerte immer konjugiert komplex auf, was gemäß Bemerkung 2.34 die Konstruktion einer reellen Lösungsbasis gestattet. Dies ist hier nicht erforderlich, da es für die Konvergenzuntersuchungen ausreicht, dass jede Lösung eine Linearkombination der komplexen Lösungen ist.

⁶Bei der Darstellung ist zu beachten, dass die Eigenwerte $\lambda_k = ib_k$ in ihrer Vielfachheit auftreten.

⁷Aus den Real- bzw. Imaginärteilen dieser komplexen Basis kann nach Bemerkung 2.34 eine Basis von reellen

dem wird wegen der linearen Unabhängigkeit durch

$$\left\| \sum_{j=1}^n x_j \mathbf{v}_j \right\| := \max_{j=1, \dots, n} |x_j|$$

eine Norm auf $\mathbb{R}^n$ definiert⁸.

Wir zeigen zunächst, dass $\mathbf{0}$ stabil ist und wählen hierzu $\varepsilon > 0$ beliebig.

$\mathbf{y}_0 = y_1 \mathbf{v}_1 + \dots + y_n \mathbf{v}_n$ sei nun ein Startwert mit der Eigenschaft

$$\|\mathbf{y}_0 - \mathbf{0}\| = \|\mathbf{y}_0\| < \delta := \varepsilon.$$

Die Lösung $\mathbf{y}$ der DGL mit $\mathbf{y}(0) = \mathbf{y}_0$ lautet dann:

$$\mathbf{y}(t) = \sum_{j=1}^n y_j \cdot \boldsymbol{\varphi}_j(t) = \sum_{j=1}^n y_j \cdot e^{ib_j t} \mathbf{v}_j$$

Wegen $|y_j| \leq \|\mathbf{y}_0\|$ und $|e^{ib_j t}| = 1$ folgt dann jedoch für alle $t \geq 0$:

$$\|\mathbf{y}(t) - \mathbf{0}\| = \|\mathbf{y}(t)\| = \max_{j=1, \dots, n} |y_j \cdot e^{ib_j t}| \leq \|\mathbf{y}_0\| < \varepsilon$$

Daher bleibt $\mathbf{y}(t)$ in der vorgegebenen ε -Umgebung von $\mathbf{0}$, wenn nur $\mathbf{y}(t)$ in der δ -Umgebung von $\mathbf{0}$ startet. Damit ist $\mathbf{0}$ stabil.

Da $\mathbf{y}(t)$ als periodische Abbildung *nicht* gegen $\mathbf{0}$ konvergieren kann, ist $\mathbf{0}$ ein neutral stabiles Gleichgewicht.

- d) Hat $\mathbf{A}$ einen Eigenwert $\lambda = ib$ ($b \in \mathbb{R}$) der Vielfachheit $k \geq 2$ ohne vollständige Menge von Eigenvektoren, so existiert gemäß Satz 2.33 eine Lösung $\boldsymbol{\varphi}$ der Form

$$\boldsymbol{\varphi}(t) = e^{ibt} (\mathbf{v}_1 + t\mathbf{v}_2 + \dots + t^{k-1} \mathbf{v}_k).$$

Wegen $|e^{ibt}| = 1$ gilt $\lim_{t \rightarrow \infty} \|\boldsymbol{\varphi}(t)\| = \infty$. Wie in Teil b) folgt dann, dass $\mathbf{y} \equiv \mathbf{0}$ eine instabile stationäre Lösung ist. ■

3.7 Beispiel

Wandelt man die lineare DGL $m\ddot{y} + r\dot{y} + cy = 0$ aus Beispiel 3.4 in ein äquivalentes DGL-System $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$ um, erhält man die Matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -\frac{c}{m} & -\frac{r}{m} \end{pmatrix}$$

Lösungen konstruiert werden. Diese sind von der Form $\boldsymbol{\psi}_k(t) := f_k(t) \mathbf{u}_k$. Die reellwertigen Funktionen f_k sind dabei periodisch wegen der Periodizität von $e^{ibt} = \cos(bt) + i \sin(bt)$ und die reellen Vektoren $\mathbf{u}_k$ bilden eine Basis des $\mathbb{R}^n$.

⁸Da alle Normen auf dem $\mathbb{R}^n$ äquivalent sind, ist der Beweis von der Wahl der Norm unabhängig

mit dem charakteristischen Polynom

$$p(\lambda) = \det(\lambda E - A) = \lambda^2 + \frac{r}{m}\lambda + \frac{c}{m}.$$

Die Eigenwerte λ_1, λ_2 sind die Nullstellen dieses Polynoms und lauten

$$\lambda_{1/2} = -\frac{r}{2m} \pm \frac{1}{2m}\sqrt{r^2 - 4mc}.$$

Für $r > 0$ gilt $\operatorname{Re}(\lambda_2) \leq \operatorname{Re}(\lambda_1) < 0$ und mit Satz 3.6 a) folgt die globale Stabilität von $\mathbf{y} \equiv \mathbf{0}$.

Im Fall $r = 0$ erhält man zwei verschiedene rein komplexe Eigenwerte und damit nach Satz 3.6 c), dass $\mathbf{0}$ eine neutral stabile Lösung ist. $\square$

3.8 Bemerkung

Es ist nach Satz 3.6 wichtig, möglichst einfach entscheiden zu können, ob die charakteristische Gleichung ausschließlich Nullstellen mit negativen Realteilen besitzt. Da die Nullstellenberechnung i.A. recht mühsam ist, wird man versuchen, direkt aus den Koeffizienten der DGL auf ihre Stabilität zu schließen. Am bekanntesten ist das *Kriterium von Routh-Hurwitz*:

Alle Nullstellen der Gleichung

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0 \quad (a_i \text{ reell, } a_n > 0)$$

haben genau dann nur negative Realteile, wenn alle Koeffizienten a_i und zugleich alle n Hauptabschnittsdeterminanten der Matrix

$$\begin{pmatrix} a_1 & a_0 & 0 & 0 & \dots & 0 \\ a_3 & a_2 & a_1 & 0 & \dots & 0 \\ a_5 & a_4 & a_3 & a_2 & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ a_{2n-1} & a_{2n-2} & a_{2n-3} & a_{2n-4} & \dots & a_n \end{pmatrix} = \begin{pmatrix} a_1 & a_0 & 0 & 0 & \dots & 0 \\ a_3 & a_2 & a_1 & 0 & \dots & 0 \\ a_5 & a_4 & a_3 & a_2 & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & 0 & 0 & \dots & a_n \end{pmatrix}$$

positiv sind. Hierbei gelte die Konvention $a_m := 0$ für $m > n$. $\square$

Das Ergebnis aus Satz 3.6 gestattet zumindest lokale Stabilitätsaussagen für nichtlineare autonome DGL-Systeme $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$, wenn man F mit Hilfe des Satzes von Taylor (bzw. mit Hilfe des Differentials) linearisiert. Wie früher schon bemerkt, genügt es dabei, stationäre Lösungen zu betrachten.

3.9 Satz

Sei $\mathbf{y}^*$ eine stationäre Lösung des autonomen DGL-Systems $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$, mit der zweimal stetig differenzierbaren Abbildung $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Dann gilt

- Hat jeder Eigenwert der Ableitungsmatrix $\mathbf{f}'(\mathbf{y}^*)$ einen negativen Realteil, so ist $\mathbf{y}^*$ eine lokal asymptotisch stabile stationäre Lösung von $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$.

- b) Hat mindestens ein Eigenwert der Ableitungsmatrix $\mathbf{f}'(\mathbf{y}^*)$ einen positiven Realteil, so ist $\mathbf{y}^*$ eine instabile stationäre Lösung von $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$. $\square$

Beweis:

- a) Nach dem Satz von Taylor (bzw. der Definition der totalen Ableitung) gilt

$$\mathbf{f}(\mathbf{y}) = \underbrace{\mathbf{f}(\mathbf{y}^*)}_{=0} + \mathbf{f}'(\mathbf{y}^*) \cdot (\mathbf{y} - \mathbf{y}^*) + \mathbf{g}(\mathbf{y} - \mathbf{y}^*) .$$

Die Abbildung $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ hat dabei die Eigenschaft (siehe z.B. Anhang Satz A.8)

$$\lim_{\mathbf{y} \rightarrow \mathbf{y}^*} \frac{|\mathbf{g}(\mathbf{y} - \mathbf{y}^*)|}{|\mathbf{y} - \mathbf{y}^*|} = 0 .$$

Setzt man $\mathbf{A} := \mathbf{f}'(\mathbf{y}^*)$ und definiert die neue Variable $\mathbf{x} := \mathbf{y} - \mathbf{y}^*$, so erhält man wegen $\dot{\mathbf{x}} = \dot{\mathbf{y}}$ und $\mathbf{f}(\mathbf{y}^*) = \mathbf{0}$ das zu $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ äquivalente System

$$(*) \quad \dot{\mathbf{x}} = \mathbf{A} \cdot \mathbf{x} + \mathbf{g}(\mathbf{x}) , \quad \text{mit} \quad \lim_{\mathbf{x} \rightarrow \mathbf{0}} \frac{|\mathbf{g}(\mathbf{x})|}{|\mathbf{x}|} = 0 .$$

Um die lokal asymptotische Stabilität von $\mathbf{y}^*$ nachzuweisen, ist $\lim_{t \rightarrow \infty} \mathbf{y}(t) = \mathbf{y}^*$ oder äquivalent hierzu $\lim_{t \rightarrow \infty} \mathbf{x}(t) = \lim_{t \rightarrow \infty} (\mathbf{y}(t) - \mathbf{y}^*) = \mathbf{0}$ zu zeigen. Da für betragsmäßig kleine $\mathbf{x}$ das DGL-System $(*)$ wegen $\mathbf{g}(\mathbf{x}) \approx \mathbf{0}$ durch $\dot{\mathbf{x}} = \mathbf{A} \cdot \mathbf{x}$ approximiert wird, genügt es daher zu zeigen, dass $\mathbf{0}$ eine global asymptotische Lösung von $\dot{\mathbf{x}} = \mathbf{A} \cdot \mathbf{x}$ ist. Dies folgt mit Satz 3.6 aus der Voraussetzung, dass $\mathbf{A} = \mathbf{f}'(\mathbf{y}^*)$ nur Eigenwerte mit negativem Realteil besitzt.

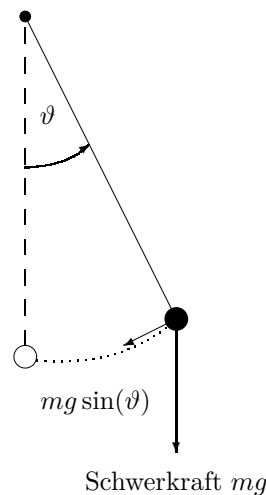
- b) Analog zu Teil a). $\blacksquare$

3.10 Bemerkung

Besitzt $\mathbf{f}'(\mathbf{y}^*)$ Eigenwerte mit Realteil 0, d.h. liegen die Fälle c) und d) von Satz 3.6 vor, kann keine genauere Aussage über das Stabilitätsverhalten von $\mathbf{y}^*$ gemacht werden. Dies liegt daran, dass das System $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ nur für betragsmäßig kleine $\mathbf{x}$ eine gute Approximation des ursprünglichen Systems $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{g}(\mathbf{x})$ ist und die Eigenschaften von $\mathbf{A}$ und $\mathbf{f}$ daher nur in diesem „guten“ Approximationsbereich die gleichen sind. $\square$

3.11 Beispiel

- i) Wir betrachten die Bewegung eines Pendels der Länge ℓ . Die Beschreibung erfolge mit Hilfe des Auslenkwinkels ϑ von der Ruheposition (siehe Skizze).



Die Bewegungsgleichung lautet:

$$\ddot{\vartheta} + r\dot{\vartheta} + \frac{g}{\ell} \sin(\vartheta) = 0$$

Hierbei ist $r > 0$ ein die Reibung beschreibender Parameter, m die an dem Pendel befestigte Masse und g die Erdbeschleunigung. Das zugehörige äquivalente Gleichungssystem mit $y_1 := \vartheta$ und $y_2 := \dot{\vartheta}$ lautet:

$$\dot{\mathbf{y}} = \begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \mathbf{f}(y_1, y_2) = \begin{pmatrix} y_2 \\ -\frac{g}{\ell} \sin(y_1) - r y_2 \end{pmatrix}$$

Die stationären Lösungen $\mathbf{y}^*$ sind gegeben durch die Gleichung $\mathbf{f}(\mathbf{y}^*) = \mathbf{0}$. Man erhält wegen der Periodizität des Systems unendlich viele Lösungen

$$\begin{pmatrix} y_1^* \\ y_2^* \end{pmatrix} = \begin{pmatrix} n\pi \\ 0 \end{pmatrix} \quad (n \in \mathbb{Z}).$$

Von den Ortskoordinaten her gesehen, ergeben sich nur zwei unterschiedliche stationäre Lösungen, die sich jedoch je nach Startgeschwindigkeit erst nach k Umdrehungen ($n = 2k$ oder $n = 2k+1$) einstellen. Es ist das in Ruhe hängende Pendel, wenn n gerade ist und das in Ruhe stehende Pendel, wenn n ungerade ist.

Die Stationarität (Ruhe) der Lösungen lässt sich auch an der zweiten Koordinate $y_2^* = 0$ ablesen. Wegen $y_2 = \dot{\vartheta}$ bedeutet dies nämlich, dass in den stationären Lösungen die Geschwindigkeit 0 ist.

Nun gilt

$$\mathbf{f}'(\mathbf{y}) = \begin{pmatrix} 0 & 1 \\ -\frac{g}{\ell} \cos(y_1) & -r \end{pmatrix}$$

und daher

$$\mathbf{f}'(0, 0) = \begin{pmatrix} 0 & 1 \\ -\frac{g}{\ell} & -r \end{pmatrix}, \quad \mathbf{f}'(\pi, 0) = \begin{pmatrix} 0 & 1 \\ \frac{g}{\ell} & -r \end{pmatrix}.$$

Die charakteristischen Polynome von $\mathbf{f}'(0, 0)$ und $\mathbf{f}'(\pi, 0)$ lauten

$$p(\lambda) = \lambda^2 + r\lambda + \frac{g}{\ell} \quad \text{und} \quad p(\lambda) = \lambda^2 + r\lambda - \frac{g}{\ell}.$$

Nach dem Kriterium von Routh-Hurwitz besitzen alle Eigenwerte von $\mathbf{f}'(0, 0)$ negative Realteile. Daher ist $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ nach Satz 3.9 eine lokal asymptotisch stabile Lösung.

Da $\mathbf{f}'(\pi, 0)$ einen positiven Eigenwert besitzt, ist $\begin{pmatrix} \pi \\ 0 \end{pmatrix}$ eine instabile stationäre Lösung. Dies entspricht der Erfahrung, dass es fast unmöglich ist, das Pendel so anzustoßen (Anfangsbedingung), dass es aufrecht stehen bleibt, da schon bei geringsten Änderungen der Anfangsbedingung diese Position verfehlt wird.

ii) Wir betrachten ein konkretes Beispiel des Populationsmodells aus Beispiel 2.4, nämlich

$$\dot{\mathbf{y}} = \begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \mathbf{f}(y_1, y_2) = \begin{pmatrix} y_1(4 - y_1 - y_2) \\ y_2(6 - y_2 - 3y_1) \end{pmatrix}.$$

Dieses System besitzt die vier stationären Lösungen $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 6 \end{pmatrix}$, $\begin{pmatrix} 4 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$. Wegen

$$\mathbf{f}'(\mathbf{y}) = \begin{pmatrix} 4 - 2y_1 - y_2 & -y_1 \\ -3y_2 & 6 - 2y_2 - 3y_1 \end{pmatrix}$$

besitzen die stationären Lösungen folgende Stabilitätseigenschaften:

$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$: Die Ableitungsmatrix ist $\mathbf{f}'(0, 0) = \begin{pmatrix} 4 & 0 \\ 0 & 6 \end{pmatrix}$.

Da zwei positive Eigenwerte vorliegen, nämlich 4 und 6, ist $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ eine instabile stationäre Lösung. Ökologisch bedeutet dies folgendes:

Sind beide Populationsgrößen nahe bei 0, führt eine leichte Vergrößerung von y_1 oder y_2 dazu, dass beide Populationen anwachsen, besonders dann, wenn beide so klein sind, dass sie sich nicht gegenseitig bedrohen.

$\begin{pmatrix} 0 \\ 6 \end{pmatrix}$: Die Ableitungsmatrix ist $\mathbf{f}'(0, 6) = \begin{pmatrix} -2 & 0 \\ -18 & -6 \end{pmatrix}$.

Da zwei negative Eigenwerte vorliegen, nämlich -2 und -6 , ist $\begin{pmatrix} 0 \\ 6 \end{pmatrix}$ eine lokal asymptotisch stabile stationäre Lösung. Ökologisch bedeutet dies folgendes:

Liegt y_2 nahe bei 6 und y_1 nahe bei 0, stabilisiert sich y_2 bei 6, während die erste Population ausstirbt.

$\begin{pmatrix} 4 \\ 0 \end{pmatrix}$: Die Ableitungsmatrix ist $\mathbf{f}'(4, 0) = \begin{pmatrix} -4 & -4 \\ 0 & -6 \end{pmatrix}$.

Es liegen wiederum zwei negative Eigenwerte vor, nämlich -4 und -6 . $\begin{pmatrix} 4 \\ 0 \end{pmatrix}$ ist ebenfalls

eine lokal asymptotisch stabile stationäre Lösung. Ökologisch bedeutet dies analog zur vorigen Überlegung:

Liegt y_1 nahe bei 4 und y_2 nahe bei 0, stabilisiert sich y_1 bei 4, während die zweite Population ausstirbt.

$\begin{pmatrix} 1 \\ 3 \end{pmatrix}$: Die Ableitungsmatrix ist $\mathbf{f}'(1, 3) = \begin{pmatrix} -1 & -1 \\ -9 & -3 \end{pmatrix}$.

Das charakteristische Polynom lautet $p(\lambda) = \lambda^2 + 4\lambda - 6$, mit einer positiven und einer negativen Nullstelle. $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ ist daher eine instabile stationäre Lösung.

Das DGL-System beschreibt also zwei Populationen, die nicht auf einem konstanten positiven Level (=stationärer Zustand) koexistieren können. $\square$

3.1. Ljapunov-Funktionen

Eine zweite Methode der Stabilitätsuntersuchung arbeitet ohne direkte Bestimmung der Lösung. Sie ist daher - ähnlich wie die im nächsten Abschnitt behandelten Phasendiagramme - hervorragend dazu geeignet, qualitative Stabilitätsuntersuchungen anzustellen.

Die Idee besteht darin, zu einer vorgegebenen stationären Lösung $\mathbf{y}^*$ des autonomen DGL-Systems $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ eine Funktion V (*Ljapunovfunktion*) zu bestimmen, die den „Abstand“ eines beliebigen Punktes $\mathbf{y}$ zu $\mathbf{y}^*$ misst. Ist $\mathbf{y}(t)$ eine Lösung der DGL $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ und $\mathbf{y}^*$ z.B. asymptotisch stabil, so muss $\lim_{t \rightarrow \infty} V(\mathbf{y}(t)) = 0$ gelten.

Der naheliegende Ansatz

$$V(\mathbf{y}) := \|\mathbf{y} - \mathbf{y}^*\|$$

scheitert in den meisten Fällen, da die euklidische Norm zwar den Abstand zwischen $\mathbf{y}$ und $\mathbf{y}^*$ misst, diese Abstandsmessung jedoch nicht die spezielle Struktur des vorgegebenen DGL-Systems berücksichtigt. Wäre z.B. $\mathbf{y}^*$ eine neutral stabile Lösung, deren „Umgebungstrajektorien“ Ellipsen der Form

$$\frac{(y_1 - y_1^*)^2}{a^2} + \frac{(y_2 - y_2^*)^2}{b^2} = r^2$$

bilden (siehe etwa Beispiel 3.16 iii)), so würde die euklidische Norm einen ständig wechselnden (periodischen) Abstand der Trajektorien zu $\mathbf{y}^*$ melden, obwohl man bei neutraler Stabilität intuitiv einen konstanten Abstand erwarten würde. Diese Abstandsmessung, die den (im euklidischen Sinne) „tatsächlichen“ Abstand misst, ist also nicht geeignet, die neutral stabile Bewegungsstruktur zu erkennen.

Verwendet man jedoch eine passend zur geometrischen Struktur der Lösungen gewichtete euklidische Norm

$$V(\mathbf{x}) = V(x_1, x_2) := \frac{(x_1 - y_1^*)^2}{a^2} + \frac{(x_2 - y_2^*)^2}{b^2},$$

so gilt für jede Lösung $\mathbf{y}(t)$ der DGL:

$$V(\mathbf{y}(t)) = \frac{(y_1(t) - y_1^*)^2}{a^2} + \frac{(y_2(t) - y_2^*)^2}{b^2} = \text{konstant}$$

Entscheidend für die Stabilitätsuntersuchung mittels Ljapunovfunktionen ist also eine an das System angepasste Abstandsfunktion. Aus diesem Grund ist es auch nicht möglich, feste Regeln zur Bestimmung von Ljapunovfunktionen anzugeben.

Der Vorteil von Ljapunovfunktionen ist, dass keine weiteren Informationen über den Bahnverlauf der Lösungsfunktionen notwendig sind. Vielmehr ist dabei ein intuitives Gefühl für das durch die DGL beschriebene System erforderlich. Ist $\mathbf{y}^*$ z.B. asymptotisch stabil, so sollte sich dies für jede in der Nähe von $\mathbf{y}^*$ startende Lösung $\mathbf{y}(t)$ dadurch äußern, dass die Funktion $t \mapsto V(\mathbf{y}(t))$ eine im Verlauf der Zeit monoton fallende Funktion ist. Das zeitliche Monotonieverhalten von $V(\mathbf{y}(t))$ lässt sich am Vorzeichen der Ableitung $\frac{d}{dt}V(\mathbf{y}(t))$ erkennen. Wegen

$$\frac{d}{dt}V(\mathbf{y}(t)) = V'(\mathbf{y}(t))\dot{\mathbf{y}}(t) = V'(\mathbf{y}(t))\mathbf{f}(\mathbf{y}(t))$$

muss daher nur das Vorzeichen von $V'(\mathbf{x})\mathbf{f}(\mathbf{x})$ untersucht werden. Insbesondere gibt dann der Bereich der $\mathbf{y}$ mit $\frac{d}{dt}V(\mathbf{y}(t)) < 0$ gewissermaßen den „Anziehungsbereich“ von $\mathbf{y}^*$ an, da in diesem Fall der durch V gemessene Abstand zwischen $\mathbf{y}(t)$ und $\mathbf{y}^*$ im Lauf der Zeit abnimmt, bis schließlich $\mathbf{y}^*$ erreicht wird. Diese Überlegungen führen zu

3.12 Definition

Sei $\mathbf{y}^*$ eine stationäre Lösung des DGL-Systems $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$, mit der stetig differenzierbaren Abbildung $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $\mathbf{f}(\mathbf{y}) = (f_1(\mathbf{y}), \dots, f_n(\mathbf{y}))$. Eine Funktion $V : \Omega \rightarrow \mathbb{R}$ heißt *Ljapunovfunktion* für $\mathbf{y}^*$, wenn gilt:

- a) Ω ist eine offene Umgebung von $\mathbf{y}^*$.
- b) V besitzt in $\mathbf{y}^*$ ein strenges lokales Minimum.
- c) Die Funktion $\mathbf{x} \mapsto V'(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) = \sum_{k=1}^n \frac{\partial V}{\partial x_k} \cdot f_k(\mathbf{x})$ nimmt in $\Omega \setminus \{\mathbf{y}^*\}$ entweder nur positive, nur negative Werte oder nur den Wert 0 an. □

3.13 Bemerkung

Da mit V auch $V + \text{Konstante}$ eine Ljapunov-Funktion ist, wird häufig $V(\mathbf{y}^*) = 0$ in Definition 3.12 gefordert. Dies ist eine von vielen Ähnlichkeiten zu Potentialfunktionen in der Physik (siehe hierzu auch Bemerkung 3.15). □

3.14 Satz

Sei $\mathbf{y}^*$ eine isolierte stationäre Lösung der DGL $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ und V eine Ljapunov-Funktion für $\mathbf{y}^*$. Dann gilt:

- a) Ist $V'(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) < 0$, so ist $\mathbf{y}^*$ eine lokal asymptotisch stabile stationäre Lösung und Ω gehört zum Anziehungsbereich von $\mathbf{y}^*$.
- b) Ist $V'(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) = 0$, so ist $\mathbf{y}^*$ eine neutral stabile stationäre Lösung und die Lösungsbahnen von $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ sind die Niveauflächen von V .
- c) Ist $V'(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) > 0$, so ist $\mathbf{y}^*$ eine instabile stationäre Lösung. $\square$

Beweis:

Da Aussage c) die Umkehrung von Teil a) ist und Aussage b) eine analoge Argumentation verwendet, beweisen wir beispielhaft Teil a).

Sei also $V : \Omega \rightarrow \mathbb{R}$ eine Ljapunovfunktion für $\mathbf{y}^*$ mit $V'(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) < 0$, die im Punkt $\mathbf{y}^*$ ein strenges lokales Minimum besitzt. O.B.d.A. kann nach Bemerkung 3.13 ii) angenommen werden, dass $V(\mathbf{y}^*) = 0$ gilt.

Sei $\varepsilon > 0$ beliebig. Da Ω offen ist und in $\mathbf{y}^*$ ein strenges lokales Minimum von V liegt, kann ε so klein gewählt werden, dass $U_\varepsilon(\mathbf{y}^*) = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{y}^*\| \leq \varepsilon\}$ ganz in Ω liegt und V dort mit Ausnahme von $\mathbf{y}^*$ nur positive Werte annimmt. M bezeichne das Minimum von V auf dem Rand von $U_\varepsilon(\mathbf{y}^*)$. Nach den Eigenschaften von V und $U_\varepsilon(\mathbf{y}^*)$ gilt $M > 0$. Ω_0 bezeichne die Menge der Punkte $\mathbf{x}$ in $U_\varepsilon(\mathbf{y}^*)$ mit $V(\mathbf{x}) < M$. Die Menge Ω_0 ist nicht leer, da sie $\mathbf{y}^*$ enthält und offen aufgrund der Stetigkeit von V .

Wir zeigen nun, dass jede in Ω_0 startende Lösung in Ω_0 bleibt. Sei hierzu $\tilde{\mathbf{y}}(t)$ eine Lösung der DGL, die in Ω_0 startet, d.h. die $V(\tilde{\mathbf{y}}(0)) < M$ erfüllt. Nach Voraussetzung gilt jedoch

$$\frac{d}{dt}V(\tilde{\mathbf{y}}(t)) = V'(\tilde{\mathbf{y}}(t)) \cdot \dot{\tilde{\mathbf{y}}}(t) = V'(\tilde{\mathbf{y}}(t)) \cdot \mathbf{f}(\tilde{\mathbf{y}}(t)) < 0,$$

so dass die Werte von V auf der Trajektorie von $\tilde{\mathbf{y}}$ (streng) monoton abnehmen. Da außerhalb von Ω_0 die Werte von V größer als M sind, muss der Orbit von $\tilde{\mathbf{y}}$ in Ω_0 und damit in $U_\varepsilon(\mathbf{y}^*)$ verlaufen. Zum Überschreiten des Randes von Ω_0 müssten die Werte von $V(\tilde{\mathbf{y}}(t))$ nämlich mindestens den Wert M annehmen, was nur dadurch zu erreichen ist, dass $V(\tilde{\mathbf{y}}(t))$ zumindest lokal anwächst. Damit ist gezeigt, dass $\mathbf{y}^*$ stabil ist. $\blacksquare$

3.15 Bemerkung

- i) Aussage a) von Satz 3.14 kann noch verbessert werden. Gilt nämlich $V'(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) \leq 0$ und enthält die Menge $R := \{\mathbf{x} \in \Omega \mid V'(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) = 0\}$ mit Ausnahme von $\{\mathbf{y}^*\}$ keine weitere *invariante Teilmenge*⁹, so ist $\mathbf{y}^*$ asymptotisch stabil.
- ii) Leider gibt es keine allgemeine Methode zur Konstruktion von Ljapunovfunktionen, weshalb sich das Auffinden solcher Funktionen unter Umständen als sehr schwierig, wenn nicht sogar als unmöglich erweisen kann. Im Prinzip erfordert die Bestimmung

⁹Eine Menge $M \subset \mathbb{R}^n$ heißt invariant, wenn jede Lösung $\mathbf{y}$ der DGL $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$, die in M startet, für alle Zeiten in M bleibt, d.h. wenn gilt: $\mathbf{y}(0) \in M \implies \forall t : \mathbf{y}(t) \in M$

der Ljapunovfunktion sehr viel Gefühl für das Verhalten des durch die DGL beschriebenen Systems (darin liegt natürlich umgekehrt auch der Vorteil einer Ljapunovfunktion). Es gibt jedoch einige Fälle in denen das Auffinden einer Ljapunovfunktion relativ einfach ist.

Ist z.B. $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ mit $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ein *Gradientensystem*, d.h. existiert eine Funktion $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}$ mit $\mathbf{g}' = \mathbf{f}$, so erhält man eine Ljapunovfunktion V , indem man $V := \pm \mathbf{g}$ definiert. Es gilt dann nämlich:

$$V'(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) = \pm \mathbf{g}'(\mathbf{x}) \cdot \mathbf{g}'(\mathbf{x}) = \pm \|\mathbf{g}'(\mathbf{x})\|^2$$

In diesem Fall sind also die strengen lokalen Minima von $\mathbf{g}$ die stabilen bzw. instabilen stationären Lösungen von $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$. Bei mechanischen Systemen mit Energieerhaltung (siehe nachfolgendes Beispiel 3.16 ii)) kann man die Energiefunktion als Ljapunovfunktion wählen. Bei thermodynamischen oder Informationssystemen ist die negative Entropiefunktion eine Ljapunovfunktion.

iii) Oft wird in der Definition von Ljapunovfunktionen nur $V'(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) \leq 0$ gefordert. $\square$

3.16 Beispiel

- i) $\dot{y} = py - y^3 =: \mathbf{f}(y)$ ist ein Gradientensystem ($p \geq 0$ konstant). Wählt man nämlich $\mathbf{g}(y) := \frac{1}{2}py^2 - \frac{1}{4}y^4$, so gilt $\mathbf{g}' = \mathbf{f}$. Die Menge aller strengen Minima der Ljapunovfunktion $V := -G$ sind die stabilen Lösungen von $\dot{y} = \mathbf{f}(y)$.

Für $p \leq 0$ ist nur $y^* = 0$ stabil. Für $p > 0$ wird $y^* = 0$ instabil und es entstehen stattdessen zwei neue stabile stationäre Zustände,

$$y^* = \pm \sqrt{p}.$$

Man spricht von einer *Bifurkation* oder *Elementarkatastrophe*.

- ii) Wir betrachten die Bewegung eines Pendels der Länge ℓ wie in Beispiel 3.11 i). Die Bewegungsgleichung lautet:

$$\ddot{\vartheta} + r\dot{\vartheta} + \frac{g}{\ell} \sin(\vartheta) = 0$$

Das zugehörige äquivalente Gleichungssystem mit $y_1 := \vartheta$ und $y_2 := \dot{\vartheta}$ lautet:

$$\dot{\mathbf{y}} = \begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \mathbf{f}(y_1, y_2) = \begin{pmatrix} y_2 \\ -\frac{g}{\ell} \sin(y_1) - ry_2 \end{pmatrix}$$

Als Ljapunovfunktion V wählen wir die Gesamtenergie (kinetische + potentielle), d.h.

$$V : \Omega = \{(y_1, y_2) \mid |y_1| < \pi\} \rightarrow \mathbb{R} \quad \text{mit} \quad V(y_1, y_2) = \frac{y_2^2}{2} + \frac{g}{\ell}(1 - \cos(y_1)) .$$

Ist $\mathbf{y}$ eine Lösung, so gilt:

$$\frac{d}{dt}V(\mathbf{y}(t)) = V'(\mathbf{y}(t))\dot{\mathbf{y}}(t) = \begin{pmatrix} \frac{g}{\ell} \sin(y_1) \\ y_2 \end{pmatrix} \cdot \begin{pmatrix} y_2 \\ -ry_2 - \frac{g}{\ell} \sin(y_1) \end{pmatrix} = -ry_2^2 \leq 0$$

Im Fall $r = 0$ ist daher $\mathbf{y} \equiv \mathbf{0}$ (d.h. $\vartheta = \dot{\vartheta} = 0$) ein neutral stabiles Gleichgewicht. Es liegt ein konservatives System vor. Startet man mit einer bestimmten Anfangsenergie E_0 , so bleibt diese im Verlauf der gesamten Bewegung erhalten. Es findet nur eine periodische Verschiebung zwischen potentieller und kinetischer Energie statt. Die Bahn der zugehörigen Bewegung (Lösung) ist gegeben durch

$$V(y_1, y_2) = E_0 .$$

Im Fall $r > 0$ ist das System *dissipativ*, da durch die Reibung ein ständiger Energieverlust stattfindet. Nun wird $\mathbf{y} \equiv \mathbf{0}$ zur lokal asymptotisch stabilen Lösung, weil $\frac{d}{dt}V(\mathbf{y}(t)) \leq 0$ für jede Lösung $\mathbf{y}$ gilt. Nach Bemerkung 3.15 i) müssen wir hierzu allerdings noch nachweisen, dass die Menge $R := \{x \in \Omega \mid V'(x) \cdot \mathbf{f}(x) = 0\}$ mit Ausnahme von $\{\mathbf{0}\}$ keine weiteren invarianten Teilmengen enthält. Sei hierzu $\mathbf{y} = (y_1, y_2) \in R$. Wegen $0 = V'(\mathbf{y}) \cdot \dot{\mathbf{y}} = V'(\mathbf{y}) \cdot \mathbf{f}(\mathbf{y}) = -ry_2^2$ folgt dann jedoch, dass die zweite Komponente von $\mathbf{y}$, nämlich y_2 gleich Null sein muss. Ist nun $\mathbf{y}(t)$ eine Lösung der DGL, die in R startet, also mit $y_2(0) = 0$, so folgt aus der zweiten Komponente der DGL, dass $\dot{y}_2(t) = -\frac{g}{\ell} \sin(y_1(t))$ gilt. Damit unterliegt die zweite Komponente der Lösung $\mathbf{y}$ einer zeitlichen Veränderung, kann also nicht den Wert Null beibehalten und damit auch nicht in R bleiben. $\mathbf{0}$ ist nur lokal asymptotisch stabil, weil ein Startwert mit großer Anfangsgeschwindigkeit y_2 dazu führen kann, dass das Pendel erst nach mehreren Umdrehungen die Ruheposition einnimmt. Diese entspricht dann nicht den Koordinaten $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ sondern $\begin{pmatrix} 2k\pi \\ 0 \end{pmatrix}$ mit $k \in \mathbb{Z}$.

iii) Gegeben sei die ein Federpendel beschreibende lineare DGL (siehe Beispiel 3.4)

$$m\ddot{y} + r\dot{y} + cy = 0 , \quad (m, c > 0, \quad r \geq 0)$$

Das äquivalente lineare DGL-System $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}) = \mathbf{A}\mathbf{y}$ mit $y_1 := y$ und $y_2 := \dot{y}$ besitzt die Matrix

$$A = \begin{pmatrix} 0 & 1 \\ -\frac{c}{m} & -\frac{r}{m} \end{pmatrix} .$$

Als Ljapunovfunktion wählen wir wie in Teil ii) dieses Beispiels die Gesamtenergie

$$V(y_1, y_2) := \frac{1}{2} \left(y_2^2 + \frac{c}{m} y_1^2 \right) .$$

Wie in Teil ii) dieses Beispiels folgt wegen

$$V'(y_1, y_2) \cdot \mathbf{f}(y_1, y_2) = -\frac{r}{m} y_2^2,$$

dass $\mathbf{y} \equiv \mathbf{0}$ eine neutral bzw. asymptotisch stabile stationäre Lösung ist, je nachdem, ob Reibung vorhanden ($r > 0$) oder nicht vorhanden ($r = 0$) ist.

iv) Gegeben sei das DGL-System ($A, B, C, D, E, F > 0$)

$$\begin{aligned}\dot{y}_1 &= y_1(A - By_2 - Ey_1) \\ \dot{y}_2 &= y_2(Dy_1 - C - Fy_2) .\end{aligned}$$

Dies ist die allgemeine Form des Populationsmodells wie in Beispiel 2.4 bzw. 3.11ii). E und F seien so klein gewählt, dass eine stationäre Lösung $\mathbf{y}^* := (y_1^*, y_2^*)$ mit positiven Komponenten existiert. Wir untersuchen $\mathbf{y}^*$ auf Stabilität im Bereich $\Omega := \{(y_1, y_2) \mid y_1 > 0, y_2 > 0\}$. Hierzu wählen wir die Ljapunov-Funktion

$$V(x_1, x_2) := D(x_1 - y_1^* \ln(x_1)) + B(x_2 - y_2^* \ln(x_2))$$

Mittels Differentialrechnung lässt sich zunächst zeigen, dass V im Punkt $\mathbf{y}^*$ ein absolutes Minimum für den Bereich Ω annimmt.

Eine kleine Rechnung zeigt außerdem, dass gilt:

$$V'(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) = -DE(y_1^* - x_1)^2 - BF(y_2^* - x_2)^2 < 0$$

Daher ist $\mathbf{y}^*$ für den Bereich Ω global asymptotisch stabil. $\square$

3.2. Richtungsfeld, Phasendiagramm

Da die meisten nichtlinearen DGLen keine explizite Lösung gestatten, wird man versuchen, zumindest qualitative Aussagen über die Lösung einer DGL bzw. eines DGL-Systems und deren Stabilitätsverhalten zu machen. Hierzu müssen wir eine gegebene DGL geometrisch interpretieren. Betrachten wir zunächst eine DGL erster Ordnung:

$$\dot{y} = \mathbf{f}(t, y)$$

Diese gibt uns in jedem Punkt (t, y) die Steigung $\dot{y}$ der Funktion $y=y(t)$. Die DGL fragt daher nach der Funktion, deren Graph in jedem Punkt (t, y) das vorgegebene Steigungsverhalten $\dot{y}=\mathbf{f}(t, y)$ besitzt. Ist φ eine Lösung, gilt $\dot{\varphi}=\mathbf{f}(t, \varphi)$. Interpretiert man $\dot{\varphi}(t)$ als Steigung der Tangente an den Graphen von φ im Punkt $(t, \varphi(t))$, muss diese mit dem vorgegebenen

Wert $\mathbf{f}(t, \varphi)$ übereinstimmen. Zeichnet man in jedem Punkt (t, y) den Vektor $(1, \mathbf{f}(t, y))^T$ (angedeutet durch kleine Linien oder Vektoren), nennt man die Funktion $R : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ mit

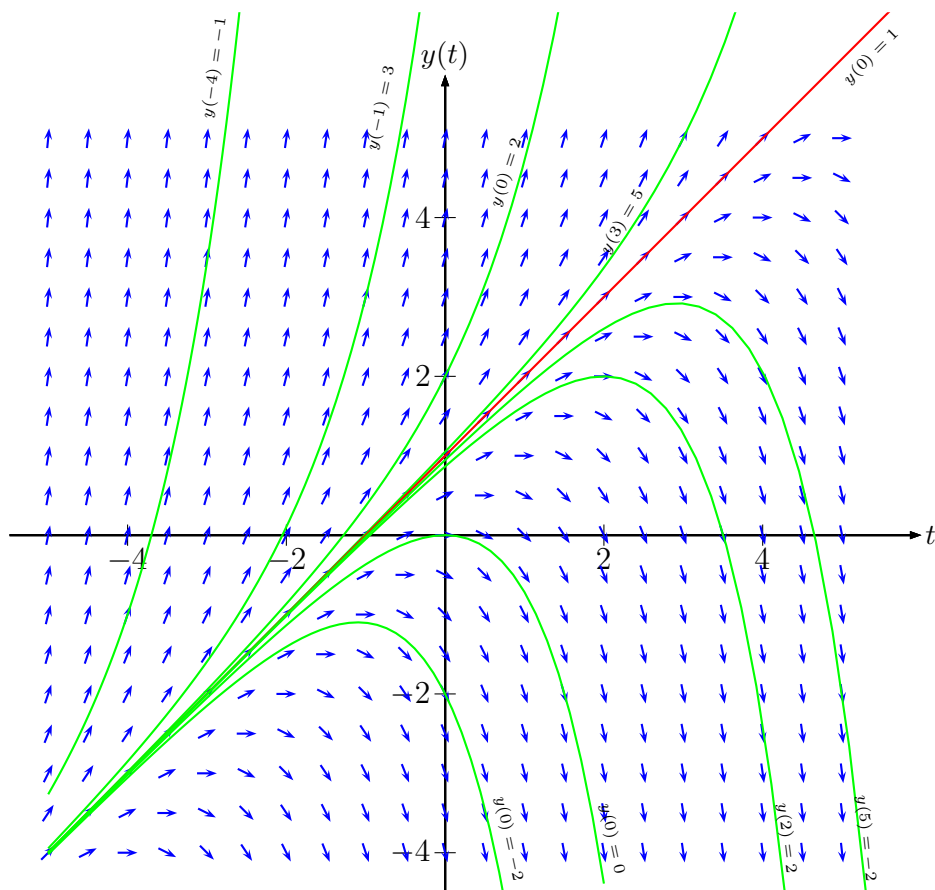
$$R(t, y) := (1, \mathbf{f}(t, y))^T$$

das *Richtungsfeld* der DGL $\dot{y} = \mathbf{f}(t, y)$, was zu folgender anschaulichen Vorstellung über die Lösbarkeit führt:

Die differenzierbare Funktion φ ist genau dann Lösung der DGL $\dot{y} = \mathbf{f}(t, y)$, wenn für jedes t die Richtung $(1, \dot{\varphi}(t))$ der Tangente an den Graphen von φ in $(t, \varphi(t))$ mit dem vorgegebenen Richtungsfeld $R(t, y)$ übereinstimmt.

3.17 Beispiel

Wir betrachten die DGL $\dot{y} = y - t$ mit der allgemeinen Lösung $y(t) = C e^t + t + 1$. Folgende Graphik zeigt das Richtungsfeld dieser DGL und einzelne Lösungen mit den Anfangsbedingungen $y(5) = -2$, $y(3) = 5$, $y(2) = 2$, $y(0) = 2$, $y(0) = -2$, $y(-1) = 3$, $y(-4) = 5$ bzw. $y(0) = 1$.



Richtungsfeld¹⁰ und ausgewählte Lösungen der DGL $\dot{y} = y - t$.

¹⁰Die Verwendung von Vektoren ist nicht unbedingt erforderlich. Es können auch kleine Linienelemente verwendet werden, da die Vektoren immer in Richtung der positiven t -Achse zeigen (siehe Beispiel 3.18).

Hervorzuheben ist an dieser Stelle besonders das unterschiedliche asymptotische Verhalten der Lösungen mit den Anfangsbedingungen $y(0) = 2$ und $y(2) = 2$. Diese lauten

$$y_1(t) = e^t + t + 1 \quad \text{und} \quad y_2(t) = -e^{t-2} + t + 1.$$

Es gilt nämlich

$$\lim_{t \rightarrow +\infty} y_1(t) = +\infty \quad \text{und} \quad \lim_{t \rightarrow +\infty} y_2(t) = -\infty.$$

□

Eine systematische Vorgehensweise bei der Konstruktion von Richtungsfeldern besteht darin, *Isoklinen* zu bestimmen, d.h. die Menge aller Punkte (t, y) , in denen das Richtungsfeld dieselbe Steigung aufweist. Bei gegebener DGL $\dot{y} = f(t, y)$ liefert die konstante Steigung $\dot{y} = C$ die implizite Isoklinengleichung

$$f(t, y) = C. \quad (3.1)$$

Löst man Gleichung (3.1), sofern dies möglich ist, nach y in Abhängigkeit von t auf, erhält man die explizite Isoklinengleichung (Man beachte hierzu nachfolgendes Beispiel 3.18.).

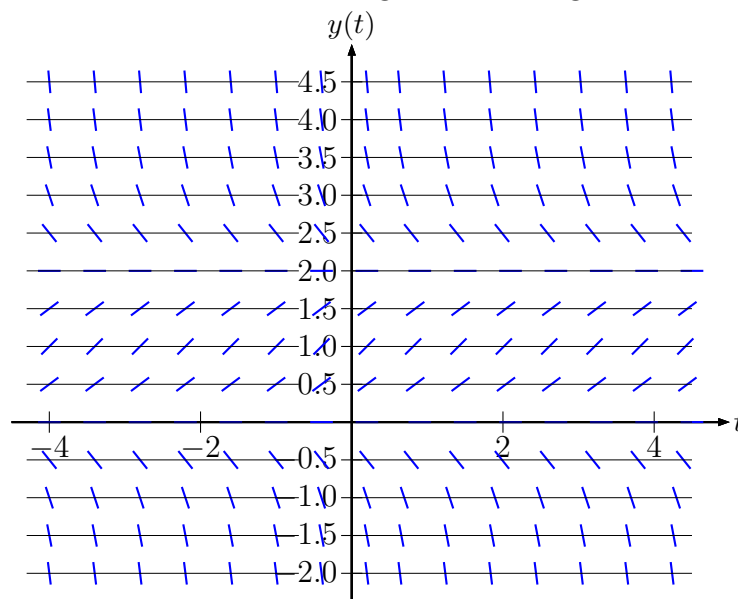
3.18 Beispiel

Wir skizzieren mit der „Isoklinenmethode“ das Richtungsfeld der DGL $\dot{y} = y \cdot (2 - y)$ (Was hier natürlich überflüssig ist, da wir die Lösung kennen.). Hierzu berechnen wir zunächst die Isoklinen mit der konstanten Steigung $\dot{y} = C$. Diese sind gegeben durch die Gleichung

$$\dot{y} = y \cdot (2 - y) = C \iff y = 1 \pm \sqrt{1 - C}.$$

Dies sind zwei Parallelen zur t -Achse im Abstand $1 \pm \sqrt{1 - C}$.

Die beiden folgenden Graphiken geben zuerst die Isoklinen für unterschiedliche Werte von C und dann das daraus resultierende Richtungsfeld mit ausgewählten Lösungen der DGL.

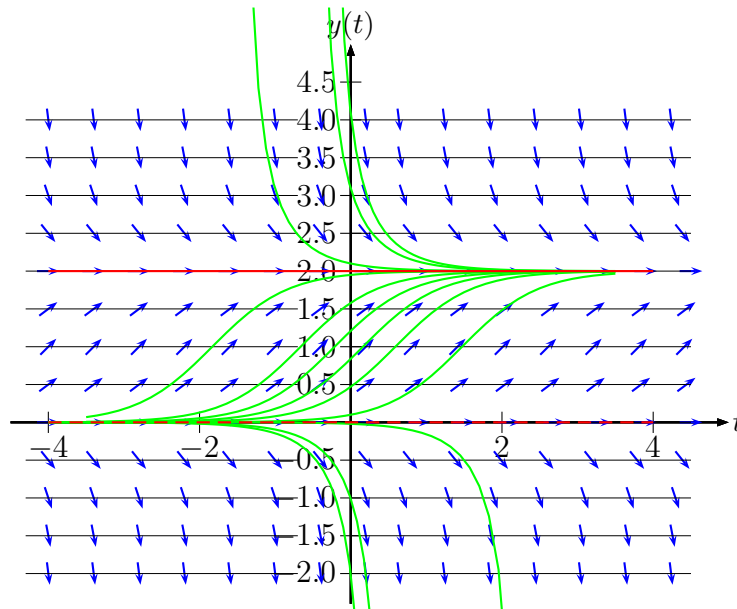


Ausgewählte Isoklinen der DGL $\dot{y} = y \cdot (2 - y)$ für $\dot{y} = C \in \{-8, -\frac{21}{4}, -3, -\frac{5}{4}, 0, \frac{3}{4}, 1\}$.

Die allgemeine Lösung der AWA $\dot{y} = y \cdot (2-y)$, $y(0)=a$ lautet (siehe Beispiel 2.12):

$$y(t) = \frac{2a e^{2t}}{2-a+a e^{2t}}$$

Für $y(0)=0$ und $y(0)=2$ ergeben sich die rot gezeichneten stationären Lösungen $y \equiv 0$, $y \equiv 2$. Das Richtungsfeld mit ausgewählten Lösungen hat damit folgendes Aussehen:



Richtungsfeld und ausgewählte Lösungen der DGL $\dot{y} = y \cdot (2-y)$. □

Beispiel 3.18 illustriert das Hauptverhalten autonomer DGLen $\dot{y} = f(y)$:

Die zeitliche Entwicklung eines autonomen Systems wird nur von der Größe des Startwertes y_0 nicht jedoch vom Startzeitpunkt t_0 des Systems beeinflusst¹¹.

Dies ergibt sich aus folgender Überlegung. Bezeichnen wir mit φ die Lösung der DGL $\dot{y} = f(y)$ mit $y(t_0) = y_0$, so gilt für $\psi(t) := \varphi(t + t_0 - t_1)$:

$$\dot{\psi}(t) = \dot{\varphi}(t + t_0 - t_1) = f(\varphi(t + t_0 - t_1)) = f(\psi(t))$$

und

$$\psi(t_1) = \varphi(t_1 + t_0 - t_1) = \varphi(t_0) = y_0.$$

Daher ist ψ die Lösung der DGL $\dot{y} = f(y)$ mit $y(t_1) = y_0$. Sie ergibt sich also aus φ durch einfache Translation.

Damit ist auch das asymptotische Verhalten nur vom Startwert y_0 nicht jedoch vom Startzeitpunkt t_0 abhängig. Daher kann man die Information, die das Richtungsfeld liefert, effizienter vermitteln, wenn man nur auf die y -Achse (*Phasenraum*) schaut. Hierzu wird eine

¹¹Dies steht im Gegensatz zu nichtautonomen DGLen, wie etwa in Beispiel 3.17. Hier zeigen die Lösungen mit $y(0) = 2$ und $y(2) = 2$ für $t \rightarrow +\infty$ ein völlig unterschiedliches Verhalten. Dies resultiert daraus, dass zwar die Startwerte, nicht jedoch die Startzeitpunkte gleich sind.

gegebene Lösungsfunktion auf die y -Achse projiziert. Man konzentriert sich damit unter Vernachlässigung der Änderungen in t auf das zeitliche Verhalten von $y(t)$.

Betrachten wir hierzu nochmals Beispiel 3.18. Bei einem Startwert $y_0 > 2$, driftet man nach unten auf 2 zu. Startet man genau in 2, bleibt man ewig dort. Wählt man einen Startpunkt zwischen 0 und 2 driftet man nach oben auf 2 zu. Startet man genau in 0 bleibt man immer in diesem Punkt. Startet man jedoch nur einen Bruchteil entfernt von 0, treibt man von 0 weg (entweder nach $-\infty$ oder auf 2 zu).



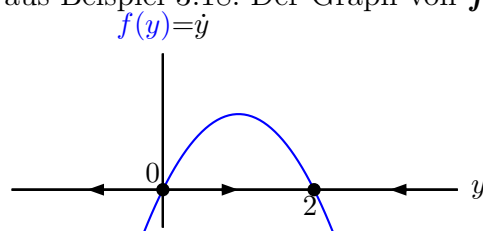
Phasenportrait der DGL $\dot{y} = y \cdot (2-y)$

Dieses *Phasenportrait* oder *Phasendiagramm* gibt uns somit die wesentlichen Informationen, die wir auch aus dem Richtungsfeld ablesen können¹². Zu beachten ist, dass es i.a. nicht möglich ist, nichtautonome DGLen mit Hilfe solcher eindimensionalen Phasenportraits zu beschreiben, da ihr zeitliches Verhalten davon abhängt, zu welchem Zeitpunkt t_0 der Prozess startet. D.h. die Lösung mit $y(t_0) = y_0$ wird i.a. eine andere Zeitdynamik zeigen als die Lösung mit $y(t_1) = y_0$ (siehe Beispiel 3.17).

Zu klären bleibt noch, wie man das Phasenportrait einer autonomen DGL $\dot{y} = \mathbf{f}(y)$ erhalten kann, ohne auf das Richtungsfeld zurückzugreifen. Dies ist relativ einfach, da wir nur wissen müssen, wo $y(t)$ in seinen Werten zu- bzw. abnimmt, d.h. wo $\dot{y} > 0$ bzw. $\dot{y} < 0$ gilt. Wegen $\dot{y} = \mathbf{f}(y)$ genügt es, den Graphen von $\mathbf{f}$ zu kennen. Bereiche, in denen $\mathbf{f}(y)$ positiv ist, zeigen anwachsendes Verhalten von y und werden mit einem nach rechts zeigenden Pfeil gekennzeichnet. Intervalle, in denen $\mathbf{f}$ negativ ist, werden mit einem nach links zeigenden Pfeil gekennzeichnet.

3.19 Beispiel

Betrachte $\dot{y} = \mathbf{f}(y) = y \cdot (2-y)$ aus Beispiel 3.18. Der Graph von $\mathbf{f}$ hat folgendes Aussehen:



¹²Die Bewegungsrichtung der y -Werte muss durch Pfeile angegeben werden, da im Gegensatz zum Richtungsfeld die Zeitachse nicht vorliegt.

Bis auf eine Drehung entspricht die y -Achse dem Phasenportrait auf Seite 96. $\square$

3.20 Bemerkung

Der Linearisierungssatz 3.9 liefert in diesem Spezialfall Stabilitätsaussagen über die stationären Lösungen der DGL $\dot{y} = \mathbf{f}(y)$:

Sei z eine stationäre Lösung der DGL $\dot{y} = \mathbf{f}(y)$, d.h. gelte $\mathbf{f}(z) = 0$. Ist $\mathbf{f}'(z) < 0$, ist z eine lokal asymptotisch stabile stationäre Lösung. Ist $\mathbf{f}'(z) > 0$, ist z eine instabile stationäre Lösung.

Wie in Satz 3.9 sind im Fall $\mathbf{f}'(z) = 0$ keine genaueren Aussagen ohne weitere Untersuchungen möglich. Problematisch sind z.B. folgende Typen von stationären Lösungen: $\square$

Wie bei autonomen DGLen in einer Veränderlichen gibt es auch bei DGL-Systemen in zwei Veränderlichen die Möglichkeit, das Verhalten des Systems graphisch zu untersuchen. Wir gehen hierzu im Folgenden von einem autonomen DGL-System der folgenden Form aus:

$$\dot{\mathbf{y}}(t) = \begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \mathbf{f}(\mathbf{y}(t)) = \begin{pmatrix} f_1(y_1, y_2) \\ f_2(y_1, y_2) \end{pmatrix}.$$

Dieses System können wir als Bewegung eines Teilchens in der y_1 - y_2 -Ebene interpretieren. $\dot{\mathbf{y}}(t)$ gibt dann die Geschwindigkeit des Teilchens im Punkt $\mathbf{y}(t)$ an.

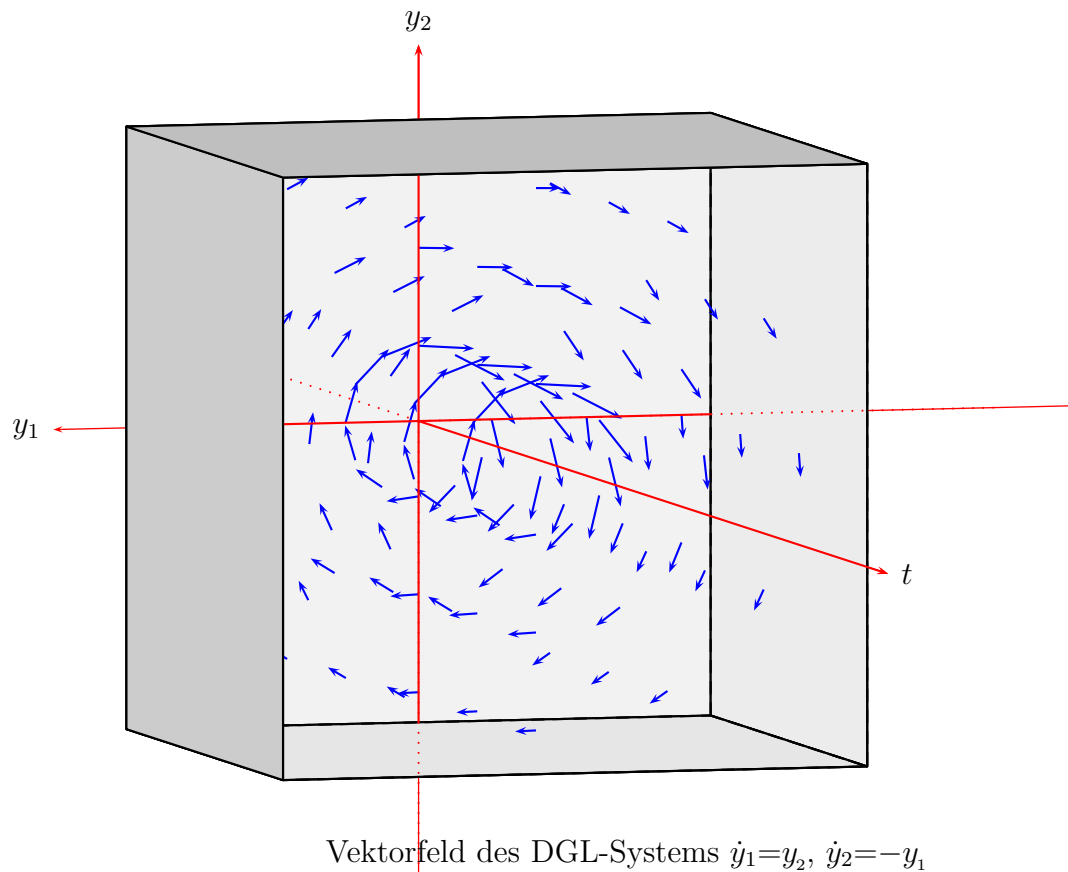
Die erste Möglichkeit ist die Analogie zum Richtungsfeld. Wir zeichnen hierzu in jedem Punkt des 3-dimensionalen t - y_1 - y_2 -Koordinatensystems den „Geschwindigkeitsvektor“ $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$. Zur besseren Übersicht wird man - wie früher - nur eine kleine Zahl von repräsentativen Punkten auswählen und die Länge der Vektoren ignorieren, d.h. nur die Geschwindigkeitsrichtung durch kurze Linien oder Vektoren andeuten. Die zugehörige Graphik ist das 3-dimensionale Analogon zum Richtungsfeld und wird als *Vektorfeld* der DGL bezeichnet. Die Pfade der Lösungsfunktionen (*Orbits*, *Trajektorien*) müssen so verlaufen, dass in jedem Punkt der zugehörige Vektor des Vektorfeldes tangential verläuft. Das folgende Beispiel zeigt ein solches Vektorfeld und beschreibt, wie - analog zum Übergang vom Richtungsfeld zum Phasenportrait - durch Projektion das Phasendiagramm entsteht.

3.21 Beispiel

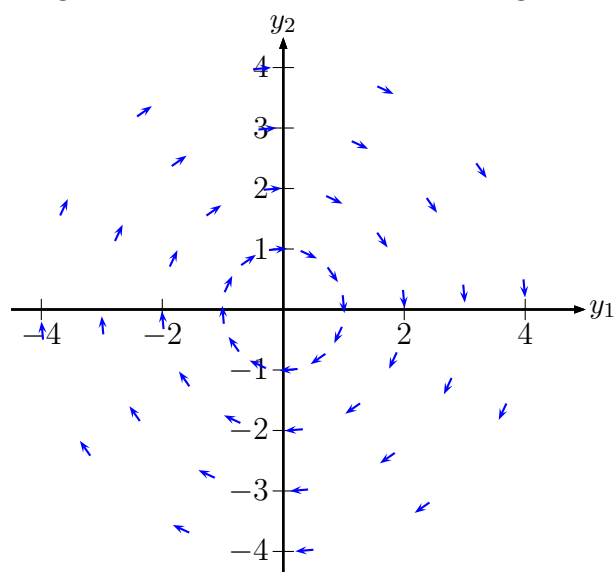
Wir betrachten das DGL-System

$$\dot{\mathbf{y}} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \cdot \mathbf{y} \iff \begin{aligned} \dot{y}_1 &= y_2 \\ \dot{y}_2 &= -y_1 \end{aligned}$$

Die allgemeine Lösung lautet $\mathbf{y}(t) = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = C \begin{pmatrix} \sin(t) \\ \cos(t) \end{pmatrix}$. In einem t - y_1 - y_2 -Koordinatensystem wird hierdurch eine kreisförmige Spiralbewegung um die t -Achse beschrieben. Daher hat das 3-dimensionale *Vektorfeld* folgendes Aussehen:



Da die entstehende Graphik schwer zu erstellen und zudem sehr unübersichtlich ist, bietet sich wie vorher beim Richtungsfeld die Möglichkeit, die Information effizienter zu vermitteln, indem die t -Achse durch Projektion auf die y_1 - y_2 -Ebene, den sog. *Phasenraum*, ausgeblendet wird. In dieser 2-dimensionalen Projektion des Vektorfeldes wird - wegen des Fehlens der Zeitachse - die Bewegungsrichtung der y -Werte durch Vektorpfeile angedeutet. Führt man dies mit dem vorigen Vektorfeld durch, entsteht folgende Graphik:



□

3.22 Bemerkung

Eine nichtautonome DGL $\dot{y} = f(t, y)$ kann durch die Substitution $y_1 := t, y_2 := y$ künstlich in ein autonomes System umgewandelt werden. Man erhält:

$$\begin{aligned}\dot{y}_1 &= 1 \\ \dot{y}_2 &= f(y_1, y_2)\end{aligned}$$

Projiziert man das Vektorfeld dieses Systems auf die y_1 - y_2 -Ebene liefert gerade das Richtungsfeld der ursprünglichen DGL. $\square$

Vervollständigt wird die graphische Darstellung des Vektorfeldes, indem man zusätzlich eine repräsentative Zahl von dreidimensionalen Lösungsfunktionen $(t, y_1(t), y_2(t))$ in den *Phasenraum* ($=y_1$ - y_2 -Ebene) mitprojiziert. Die entstehende Graphik im Phasenraum wird wie im 1-dimensionalen Fall als *Phasendiagramm* oder Phasenportrait bezeichnet. Wegen der fehlenden Zeitachse muss die zeitabhängige Bewegungsrichtung mit Pfeilen gekennzeichnet werden. Folgendes Beispiel zeigt, wie man ein vollständiges Phasendiagramm erstellt.

3.23 Beispiel

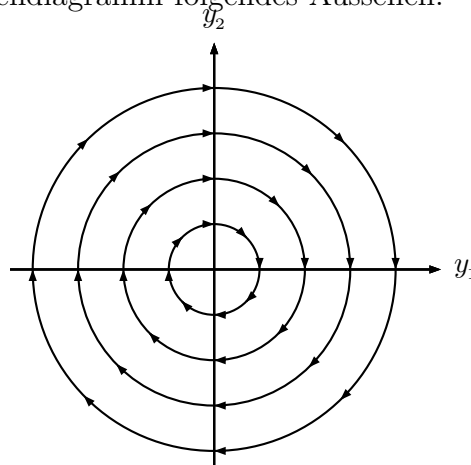
- i) Betrachten wir das DGL-System $\begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \begin{pmatrix} \dot{y}_2 \\ -y_1 \end{pmatrix}$ aus Beispiel 3.21. Division der beiden Gleichungen liefert folgende DGL für die Lösungsfunktionen im Phasenraum:

$$\frac{\dot{y}_2}{\dot{y}_1} = \frac{dy_2}{dy_1} = -\frac{y_1}{y_2}$$

Diese DGL ist separabel und man erhält als Lösung:

$$\int y_2 dy_2 = - \int y_1 dy_1 + \text{Konst.} \iff y_2^2 + y_1^2 = C$$

Die in die y_1 - y_2 -Ebene projizierten Lösungen, bestehen also aus konzentrischen Kreisen. Das zugehörige Vektorfeld zeigt, dass diese Orbits im Uhrzeigersinn durchlaufen werden. Also hat das Phasendiagramm folgendes Aussehen:



Phasendiagramm des DGL-Systems $\dot{y}_1 = y_2, \dot{y}_2 = -y_1$

Es liegen hier also periodische Orbits vor, da die Lösungskurven geschlossen sind¹³.

ii) Betrachten wir das DGL-System

$$\begin{aligned}\dot{y}_1 &= 2y_1 \\ \dot{y}_2 &= -2y_2\end{aligned}$$

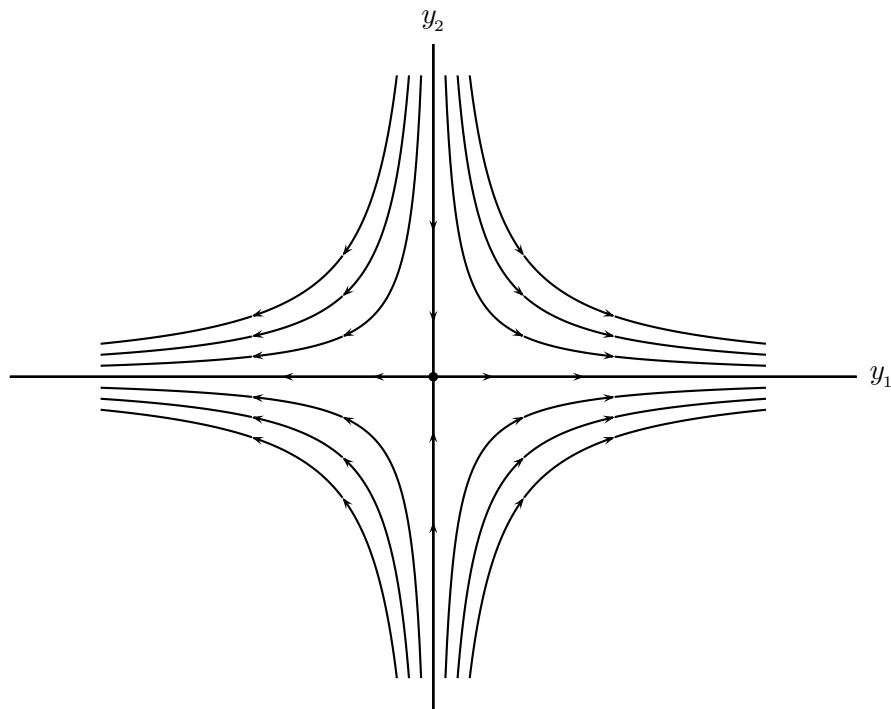
Dann gilt:

$$\frac{\dot{y}_2}{\dot{y}_1} = \frac{dy_2}{dy_1} = -\frac{y_2}{y_1}$$

Die allgemeine Lösung lautet

$$y_2 \cdot y_1 = C.$$

Die Lösungskurven im Phasenraum haben daher die Form $y_2 = \frac{C}{y_1}$, falls $C \neq 0$. Im Fall $C=0$ gilt $y_2=0$ oder $y_1=0$. Die Richtung, in der diese Orbits durchlaufen werden, kann anhand der Vorzeichen von y_1 und y_2 ermittelt werden. So gilt im 1. Quadranten $y_1 > 0$ und $y_2 > 0$. Aus der DGL folgt dann $\dot{y}_1 = 2y_1 > 0$ und $\dot{y}_2 = -2y_2 < 0$ im 1. Quadranten, d.h. die y_1 -Werte nehmen zu, wohingegen die y_2 -Werte im 1. Quadranten abnehmen. Das Phasendiagramm hat daher folgendes Aussehen:



Phasendiagramm des DGL-Systems $\dot{y}_1 = 2y_1$, $\dot{y}_2 = -2y_2$

□

¹³Die kleinste Zahl $T > 0$ mit $\mathbf{y}(t) = \mathbf{y}(t+T)$ heißt Periode des Orbits. Hier sind alle nicht konstanten Lösungen periodisch mit der Periode 2π .

Da die projizierten Lösungsfunktionen i.a. nicht so einfach wie in Beispiel 3.23 als Lösung der DGL $\frac{\dot{y}_2}{\dot{y}_1} = \frac{dy_2}{dy_1}$ berechnet werden können, soll nun noch eine einfache aber effektive Methode zur Bestimmung des Phasendiagramms beschrieben werden. Wir greifen hierzu das früher schon verwendete Beispiel 3.11 ii) erneut auf.

3.24 Beispiel

Wir betrachten das DGL-System $\dot{\mathbf{y}} = \begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \begin{pmatrix} y_1 \cdot (4 - y_1 - y_2) \\ y_2 \cdot (6 - y_2 - 3y_1) \end{pmatrix} = \mathbf{f}(\mathbf{y})$.

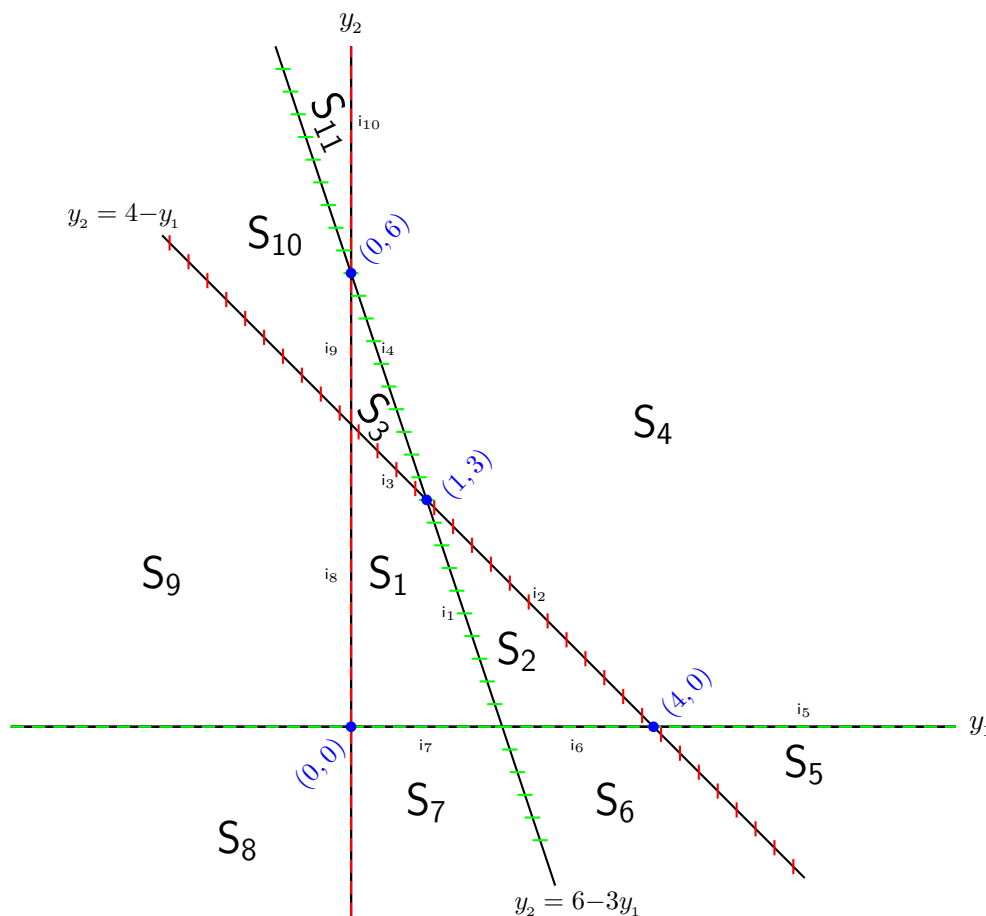
1. Schritt: Bestimmung der Isoklinen $\dot{y}_1 = 0$ und $\dot{y}_2 = 0$.

Wir bestimmen zunächst die *Isoklinen* (d.h. den geometrischen Ort aller Punkte gleicher Steigung), in denen das Vektorfeld eine horizontale oder vertikale Richtung besitzt. Diese berechnen sich durch $\dot{y}_2 = 0$ bzw. $\dot{y}_1 = 0$. Die Schnittpunkte dieser Isoklinen liefern gleichzeitig die stationären Lösungen des DGL-Systems. Es gilt

$$\dot{y}_1 = 0 \iff (y_1 = 0) \vee (4 - y_1 - y_2 = 0)$$

$$\dot{y}_2 = 0 \iff (y_2 = 0) \vee (6 - y_2 - 3y_1 = 0)$$

Die stationären Punkte sind daher $(0, 0)$, $(0, 6)$, $(4, 0)$, $(1, 3)$.

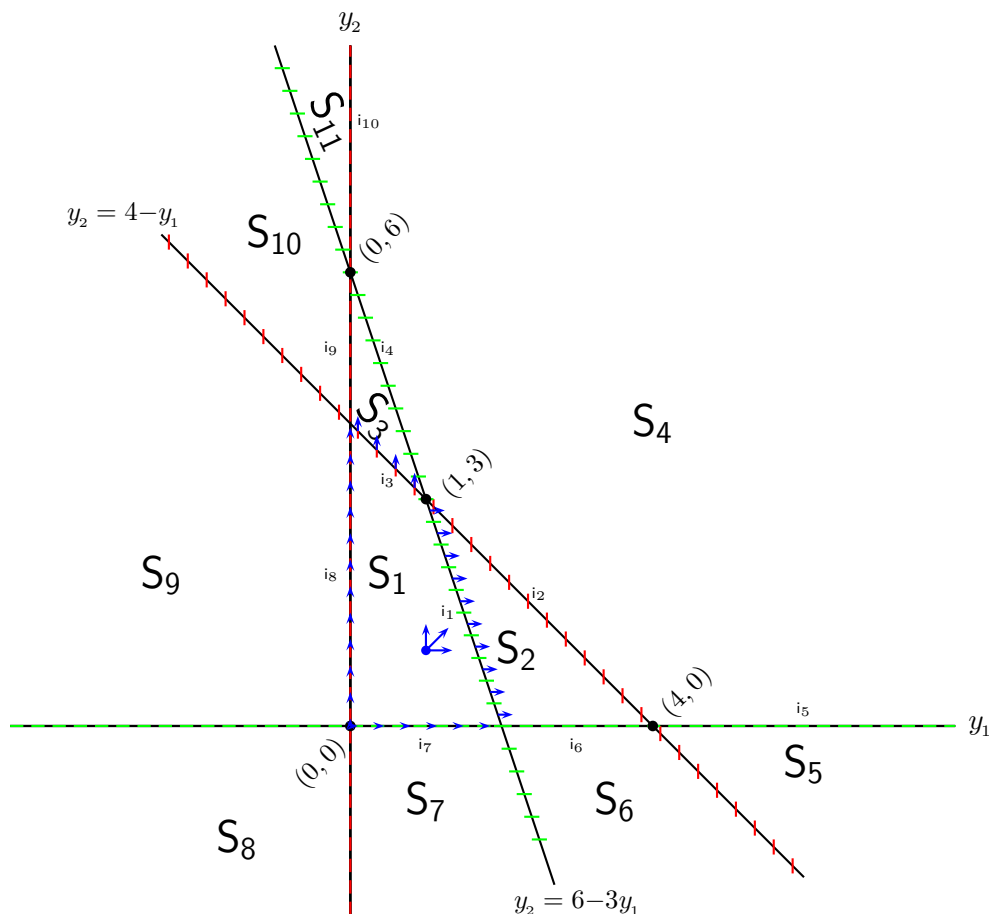


Isoklinen $\dot{y}_1=0$, $\dot{y}_2=0$ des DGL-Systems $\dot{y}_1 = y_1(4 - y_1 - y_2)$, $\dot{y}_2 = y_2(6 - y_2 - 3y_1)$

Wir tragen die Isoklinen in einem y_1 - y_2 -Koordinatensystem ein (inklusive kleiner horizontaler bzw. vertikaler Steigungslinien). Die Steigungsrichtung - nach oben (Norden) oder unten (Süden), rechts (Osten) oder links (Westen) - ist an dieser Stelle noch unklar. Sie wird erst im zweiten Schritt bestimmt. Die Isoklinen¹⁴ zerlegen die gesamte Ebene in einzelne Sektoren (hier die Sektoren S_1 bis S_{11}).

2. Schritt: Skizze des Vektorfeldes

Wir füllen die Sektoren und Isoklinen mit Richtungsvektoren. Hierzu beachte man, dass im Inneren eines jeden Sektors, die Vektoren *nicht* nach Norden, Süden (vertikal), Westen oder Osten (horizontal) zeigen, da dies nur *auf* den in Schritt 1 bestimmten Isoklinen der Fall ist. Für jeden Sektor bleiben daher nur die Richtungen NO, NW, SO, SW.



Bestimmung des Vektorfeldes von Sektor S_1 mittels Punkt $(1|1)$.

Aufgrund der geforderten Stetigkeit der DGL (f ist stetig), kann das Vektorfeld in einem Sektor außerdem nur in eine dieser Richtungen zeigen, sonst erhielte man einen Widerspruch zum Zwischenwertsatz. Würde *innerhalb* eines Sektors ein Vektor z.B. nach NO und ein anderer nach NW zeigen, müsste nach dem Zwischenwertsatz aufgrund der Stetigkeit des Vektorfeldes (f ist stetig) im Innern auch die Richtung N angenommen werden. Dies ist

¹⁴Aus rechnerischen Gründen sind die Isoklinen in diesem Beispiel Geraden. Im allgemeinen muss mit nicht-linearen Isoklinen gerechnet werden (siehe Beispiel 3.25 iii)).

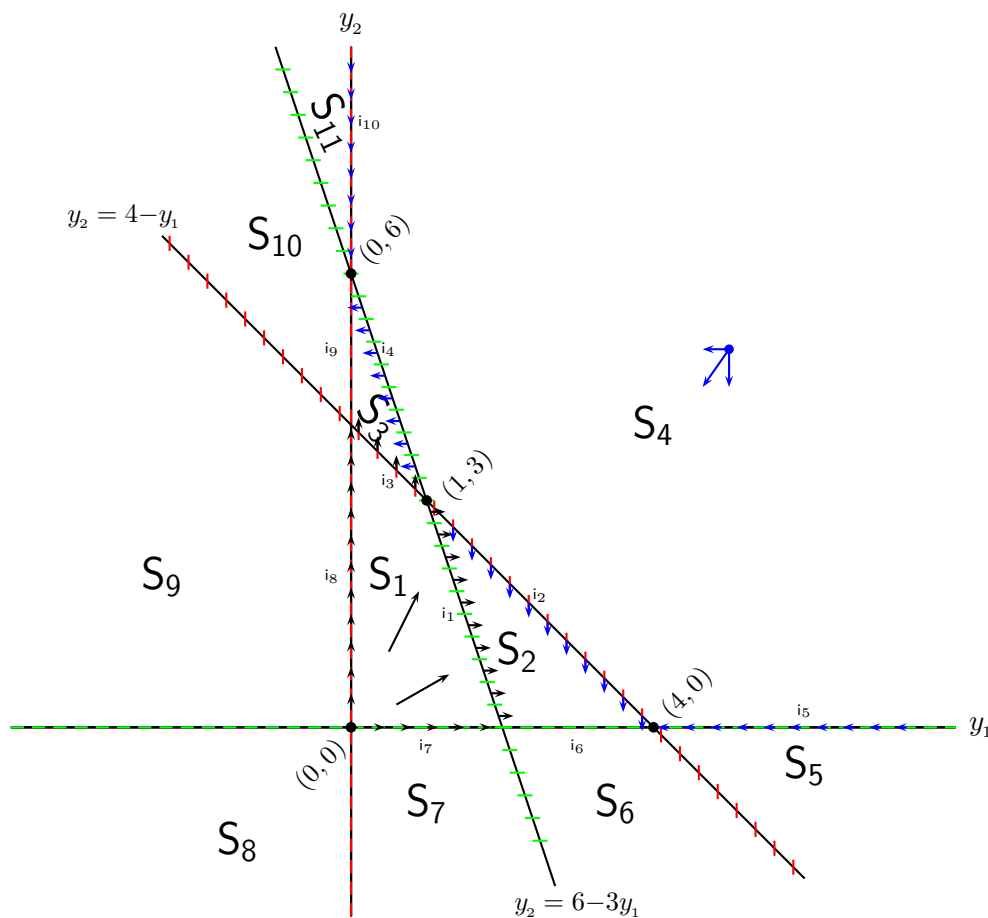
jedoch nur auf der Isoklinen $\dot{y}_1 = 0$ möglich. Es genügt daher, in jedem Sektor einen Punkt auszuwählen und die Richtung des Vektorfeldes in diesem Punkt zu bestimmen. Jeder Vektor in diesem Sektor muss dann in die „gleiche“ Richtung zeigen.

Wegen des stetigen Übergangs von einem Sektor in den benachbarten ist dann auch die Vektorrichtung *auf* den zugehörigen Teilabschnitten der Isoklinen bestimmt, d.h. ob die Vektoren auf den Isoklinen in Richtung N oder S bzw. W oder O zeigen.

Im vorliegenden Beispiel starten wir mit Sektor S_1 , da dessen zentrale Lage Schlüsse über das Steigungsverhalten besonders vieler Isoklinenabschnitte zulässt, nämlich die Abschnitte i_1, i_3, i_7 und i_8 . Wir betrachten hierzu den Punkt $y=(1, 1)$ in S_1 . Wegen

$$\dot{y} = f(1, 1) = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

gilt offensichtlich $\dot{y}_1 = \dot{y}_2 = 2$, d.h. der Geschwindigkeitsvektor zeigt in Richtung (2,2), also in Richtung NO. Daher zeigen alle Vektoren dieses Sektors in Richtung NO. Damit muss die Richtung der Vektoren auf den Isoklinenabschnitten i_1, i_3, i_7, i_8 , die Sektor S_1 begrenzen, entweder N oder O sein. Da auf i_8 und i_3 eine vertikale Steigungsrichtung vorliegt, muss hier die Bewegungsrichtung N sein. Auf i_7 und i_1 erhält man analog die Bewegungsrichtung O.

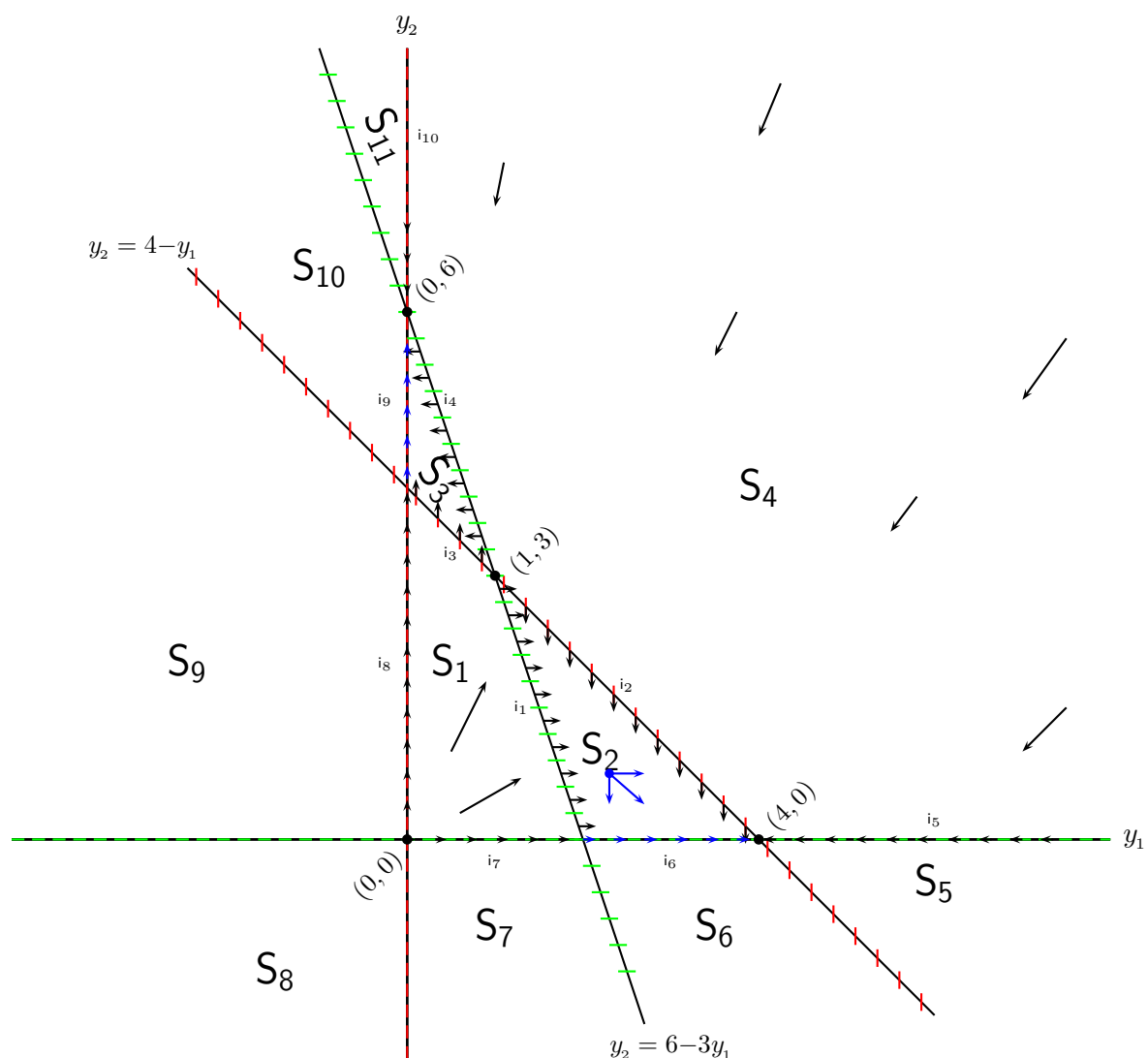


Bestimmung des Vektorfeldes von Sektor S_4 mittels Punkt (5|5).

Im Sektor S_4 wählen wir den Punkt $(5,5)$. Der Geschwindigkeitsvektor in diesem Punkt ist $(-30, -70)$ und zeigt damit Richtung SW. Auf den begrenzenden Randisoklinen i_2, i_4, i_5, i_{10} müssen daher die Vektoren entweder in Richtung S oder W zeigen. Da auf i_{10} und i_2 eine vertikale Bewegung stattfindet, erhält man dort Steigungsvektoren Richtung S. Auf i_4 und i_5 ergibt sich analog eine Bewegung Richtung W.

Aufgrund des nun bekannten Steigungsverhaltens der Isoklinenabschnitte i_3 und i_4 muss im Sektor S_3 die Bewegungsrichtung NW vorliegen. Hierdurch ergibt sich auf i_9 eine Bewegung Richtung N.

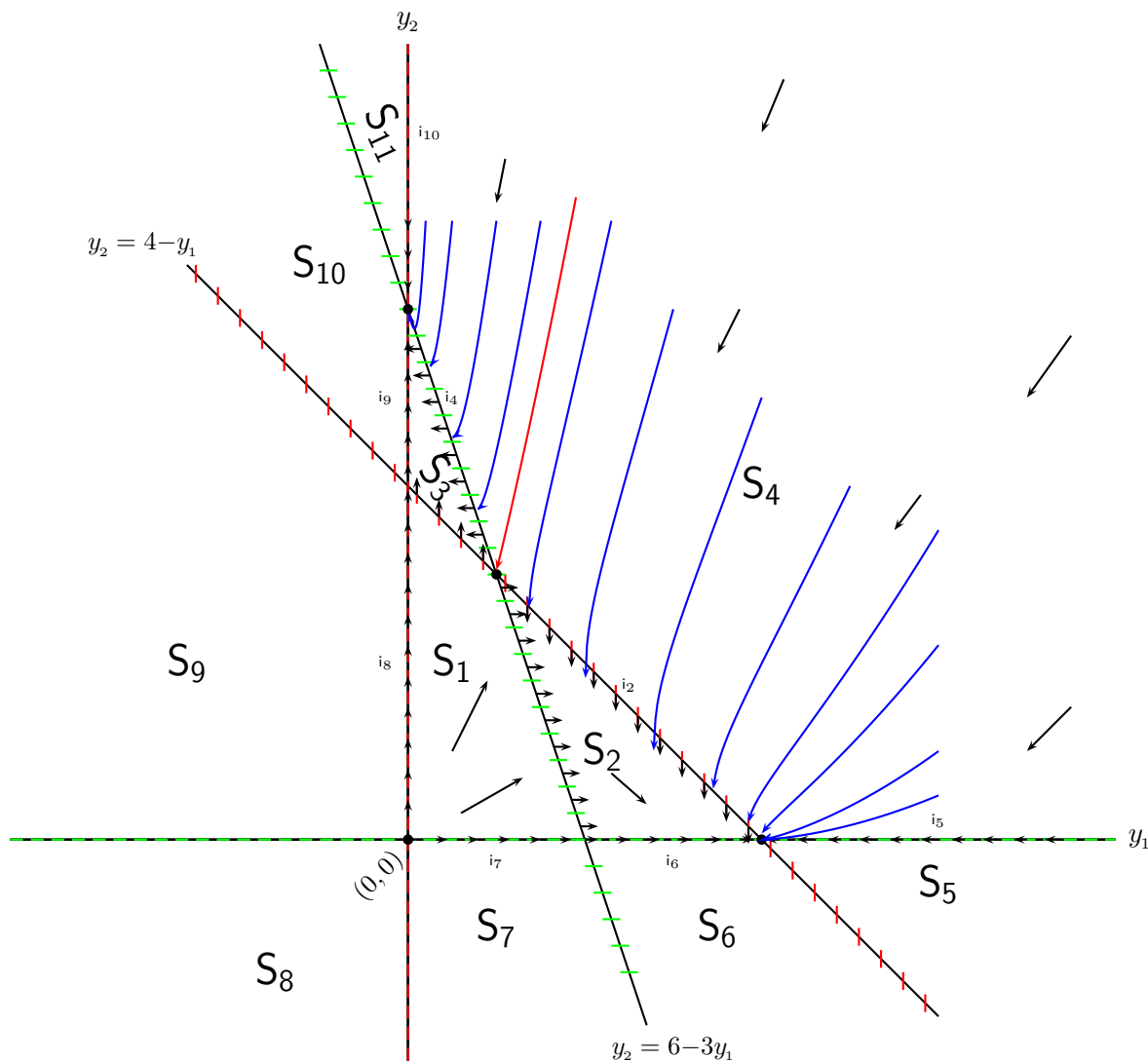
Analog ergibt sich im Sektor S_2 die Bewegungsrichtung SO und damit auf i_6 eine Bewegung Richtung O. Daher hat das Vektorfeld (genauer: die Projektion des Vektorfeldes in den Phasenraum) im 1. Quadranten folgendes Aussehen:



Das Vektorfeld im 1. Quadranten.

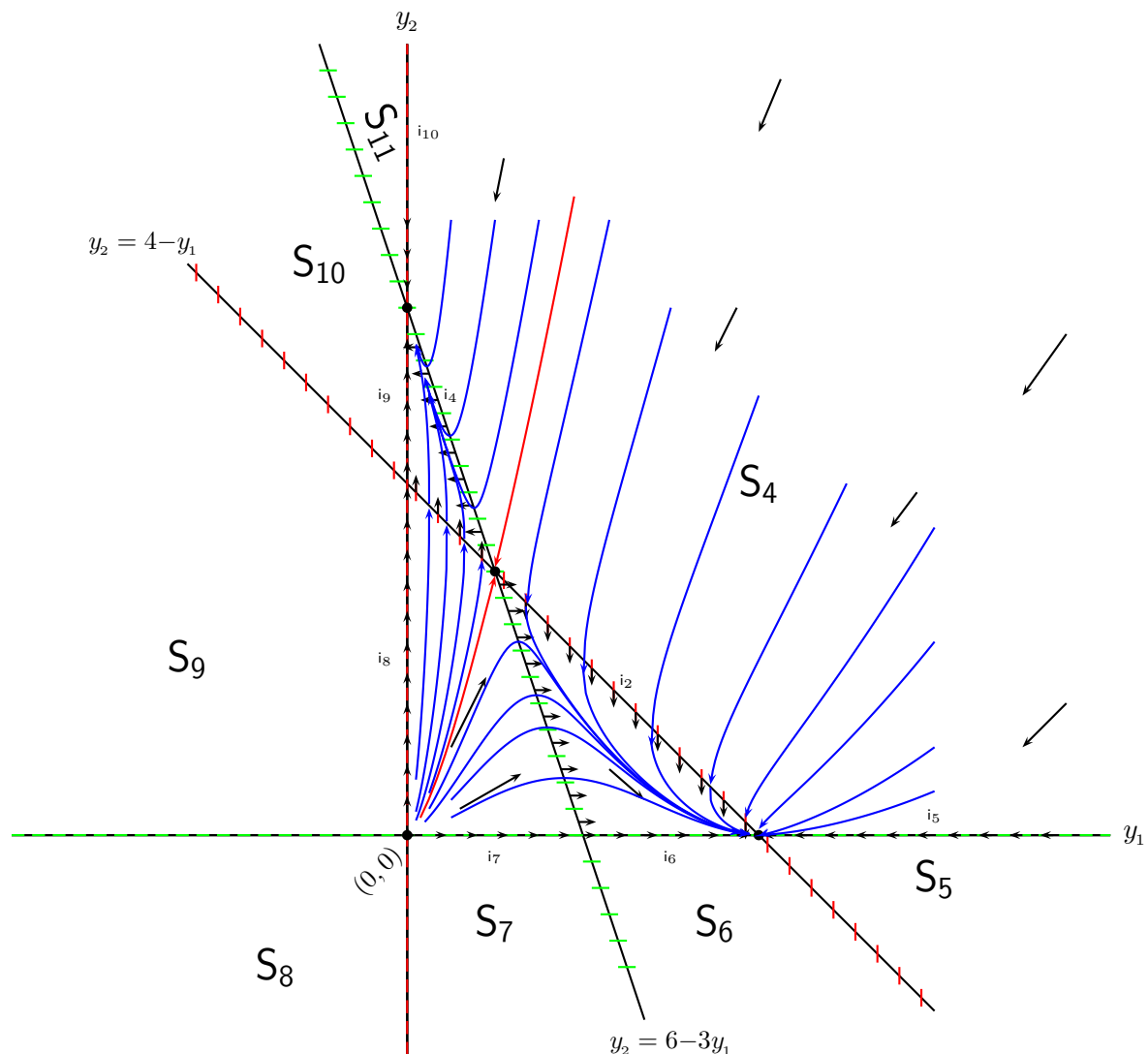
3. Schritt: Skizze des Phasendiagramms

Zum gefundenen (repräsentativen) Vektorfeld (genauer dessen Projektion auf die y_1 - y_2 -Ebene) skizzieren wir repräsentative Lösungskurven. Wir starten beispielsweise in Sektor S_4 . Jede Lösungskurve im Innern des Sektors bewegt sich Richtung SW. Isoklinenabschnitt i_4 wird horizontal überquert, Isoklinenabschnitt i_2 vertikal. Eine spezielle Lösung läuft genau in die stationäre Lösung (1,3) und bleibt dort stehen.



Das Phasendiagramm von $\dot{y}_1 = y_1(4 - y_1 - y_2)$, $\dot{y}_2 = y_2(6 - y_2 - 3y_1)$ im Sektor S_4

Die Lösungskurven, die Abschnitt i_4 queren, landen in Sektor S_3 , wo sie nach NW drehen. Da sie Isoklinenabschnitt i_9 nicht kreuzen dürfen, konvergieren sie in den stationärer Punkt (0,6). Das Phasendiagramm hat damit im 1. Quadranten folgendes Aussehen:



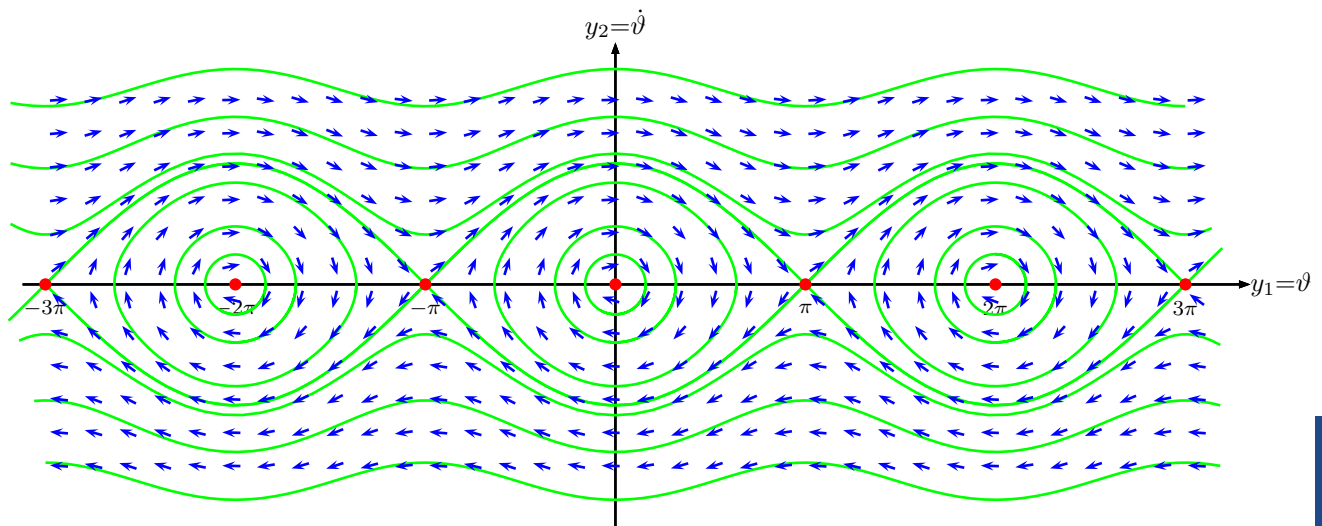
Das Phasendiagramm von $\dot{y}_1 = y_1(4 - y_1 - y_2)$, $\dot{y}_2 = y_2(6 - y_2 - 3y_1)$ im 1. Quadranten

Das entstandene Bild ist wertvoller, als die explizite Lösung (sofern diese existiert), weil wir die zeitliche Entwicklung verfolgen können. Z.B. sieht man, dass Startwerte im Sektor S_4 zu Lösungskurven führen, die mit $t \rightarrow +\infty$ gegen die stationären Lösungen $(4,0)$ oder $(0,6)$ konvergieren, je nachdem, ob man im unteren oder oberen Bereich von S_4 startet. Diese zwei Phänomene werden durch eine spezielle Lösungskurve (*Separatrix*) getrennt. Dies ist die einzige Trajektorie, die zum instabilen Gleichgewicht $(1,3)$ führt. Auch die Stabilitätseigenschaften der stationären Lösungen lassen sich sehr gut ablesen. $\square$

Zum Abschluss dieses Kapitels sollen noch einige weitere Beispiele für Phasenportraits gegeben und eine qualitative Analyse des zugehörigen Systems durchgeführt werden.

3.25 Beispiel

- i) Betrachten wir das Pendel aus Beispiel 3.16 ii). Das System besitzt die stationären Lösungen $(\ell\pi \mid 0)$ mit $\ell \in \mathbb{Z}$. Liegt keine Reibung vor, d.h. gilt $r=0$ ist das Phasendiagramm periodisch und hat folgendes Aussehen:



Das Phasendiagramm von $\dot{y}_1 = y_2, \dot{y}_2 = -\frac{g}{\ell} \sin(y_1) - r y_2$ im Fall $r = 0$.

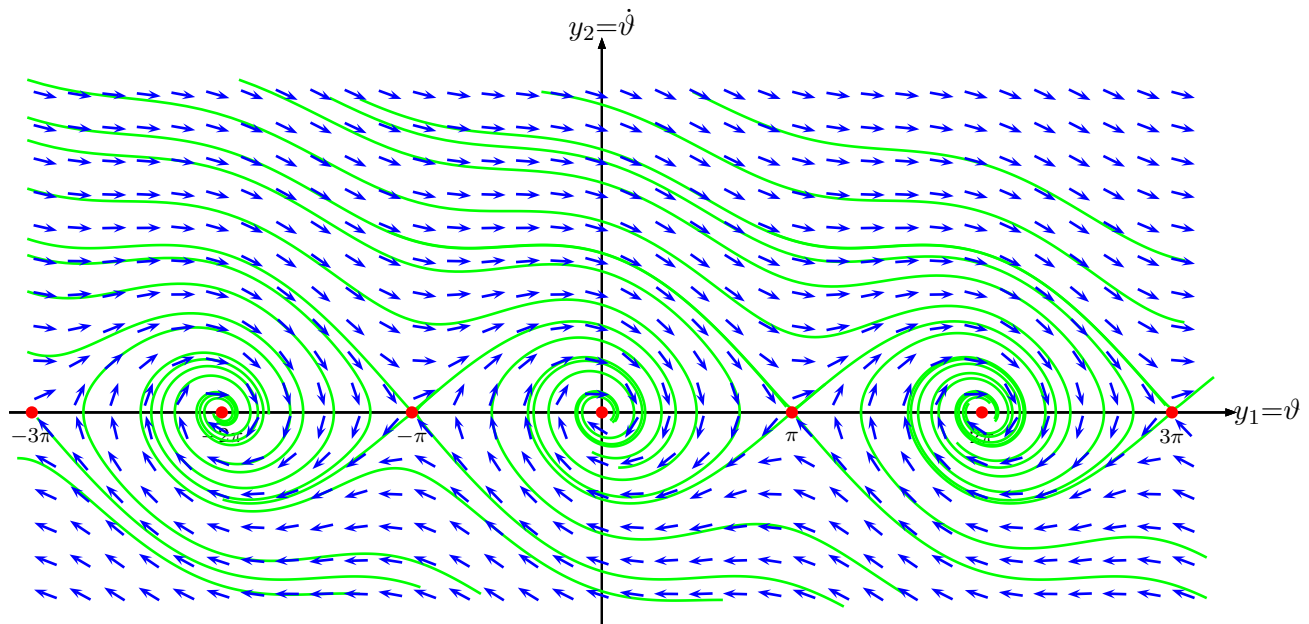
Die stationären Lösungen $(2k\pi | 0)$ sind neutral stabil, wie die in sich geschlossenen Lösungstrajektorien in der Nähe dieser Punkte zeigen. Auf diesen geschlossenen Lösungsbahnen bleibt die Gesamtenergie konstant. Es findet jedoch ein ständiger Austausch zwischen potentieller Energie (abhängig von $y_1 = \vartheta$) und kinetischer Energie (abhängig von $y_2 = \dot{\vartheta}$) statt. Hat die potentielle Energie den Wert 0, d.h. gilt $y_1 = \vartheta = 2k\pi$, so hat das Pendel eine maximale kinetische Energie, d.h. $y_2 = \dot{\vartheta}$ ist maximal und umgekehrt. Die stationären Lösungen $y_1 = (2k+1)\pi$ sind instabil.

Auf den nicht geschlossenen Lösungsbahnen besitzt das Pendel eine so hohe Energie, dass die Geschwindigkeit $y_2 = \dot{\vartheta}$ niemals 0 wird und eine ständige Rotation um den Aufhängungspunkt stattfindet. Hierdurch wächst auch der Auslenkungswinkel $y_1 = \vartheta$ auf diesen Bahnen ständig an.

Die stationären Lösungen $((2k+1)\pi | 0)$ sind instabil. Sie werden nur bei ganz bestimmten Anfangsbedingungen erreicht und entsprechen einem aufrecht stehenden Pendel. Geringste Störungen in den Anfangsbedingungen führen sofort von dieser Lösung weg (*Sattelpunktstabilität*).

Ist Reibung vorhanden, d.h. gilt $r > 0$, liegt eine ähnliche Struktur vor. Aufgrund der Reibung findet jedoch ein ständiger Energieverlust statt, der dazu führt, dass nach einiger Zeit das Pendel - nach unten hängend - zur Ruhe kommt. Je nach der Anzahl der vorher stattgefundenen Umdrehungen besitzt diese stationäre Lösung eine andere Winkelkoordinate. Daher sind aufgrund der Reibung die stationären Lösungen $(2k\pi | 0)$ lokal asymptotisch stabil.

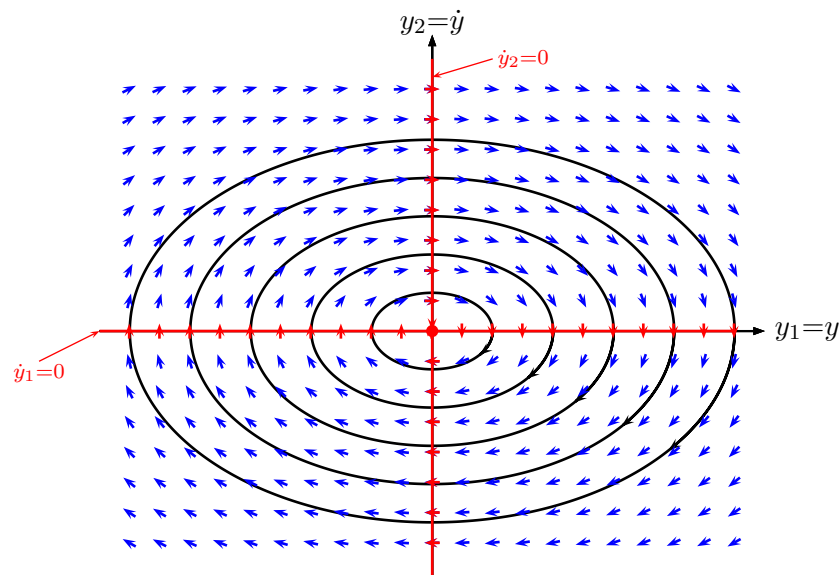
Wie vorher sind die stationären Lösungen $(2(k+1)\pi | 0)$ instabil (sattelpunktstabil). Rein theoretisch kann die Startbedingung so gewählt werden, dass das Pendel nach mehreren Umdrehungen aufrecht stehend zur Ruhe kommt.



Das Phasendiagramm von $\dot{y}_1 = y_2, \dot{y}_2 = -\frac{g}{\ell} \sin(y_1) - r y_2$ im Fall $r > 0$.

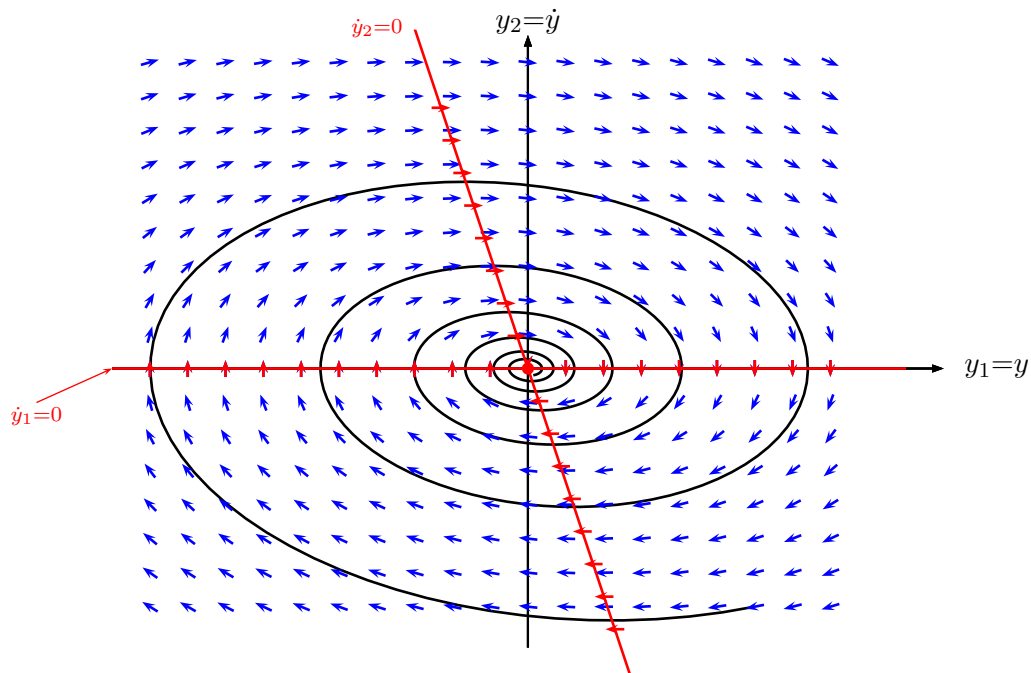
- ii) Ein von der Struktur her ähnliches Verhalten zeigt das Federpendel aus Beispiel 3.16 iii). Da hier jedoch explizit die Ortskoordinate y zur Beschreibung verwendet wird, bleiben bei den Lösungsbahnen die Werte von y_1 (= Ortskoordinate y) und y_2 (= Geschwindigkeitskoordinate $\dot{y}$) beschränkt.

Im Gegensatz zu Teil i) dieses Beispiels entsteht daher kein periodisches Phasendiagramm. Im Fall $r=0$ ist die stationäre Lösung $(0 \mid 0)$ neutral stabil, wobei jetzt aufgrund der Verwendung der Ortskoordinate ausschließlich geschlossene Lösungstrajektorien auftreten.



Das Phasendiagramm von $m\ddot{y} + r\dot{y} + cy = 0$ im Fall $r = 0$.

Im Fall $r > 0$ ist die stationäre Lösung $(0 \mid 0)$ global asymptotisch stabil. Die Verwendung der Ortskoordinate führt hier zu spiralförmigen Lösungstrajektorien.

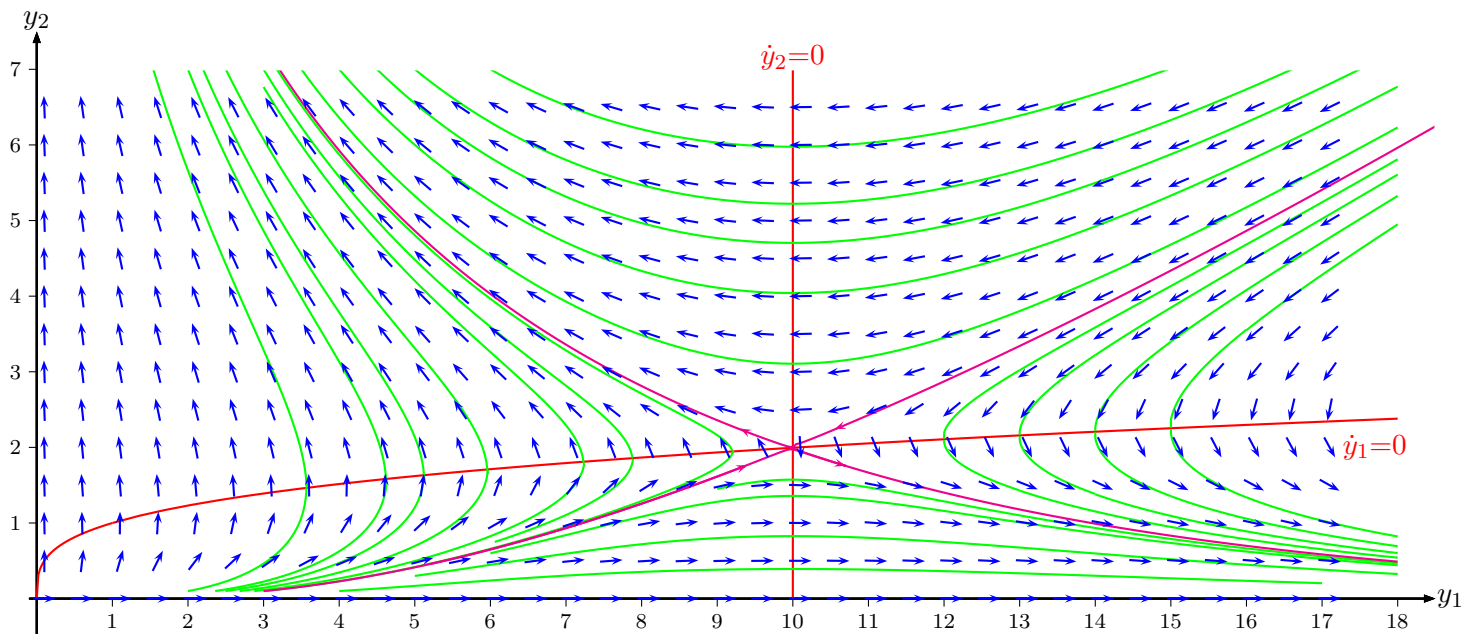


Das Phasendiagramm von $m\ddot{y} + r\dot{y} + cy = 0$ im Fall $r > 0$.

iii) Wir betrachten das DGL-System

$$\dot{\mathbf{y}} = \begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \mathbf{f}(\mathbf{y}) := \begin{pmatrix} y_1^{0,3} - y_2 \\ y_2 \cdot [5y_1^{-0,7} - 1] \end{pmatrix}$$

Aufgrund der Nichtlinearität des Systems ist die Isoquante $\dot{y}_1=0 \iff y_2=y_1^{0,3}$ nichtlinear. Das Phasenportrait hat im ersten Quadranten folgendes Aussehen:



Das Phasendiagramm von $\dot{y}_1 = y_1^{0,3} - y_2$, $\dot{y}_2 = y_2 [5y_1^{-0,7} - 1]$ im 1. Quadranten.

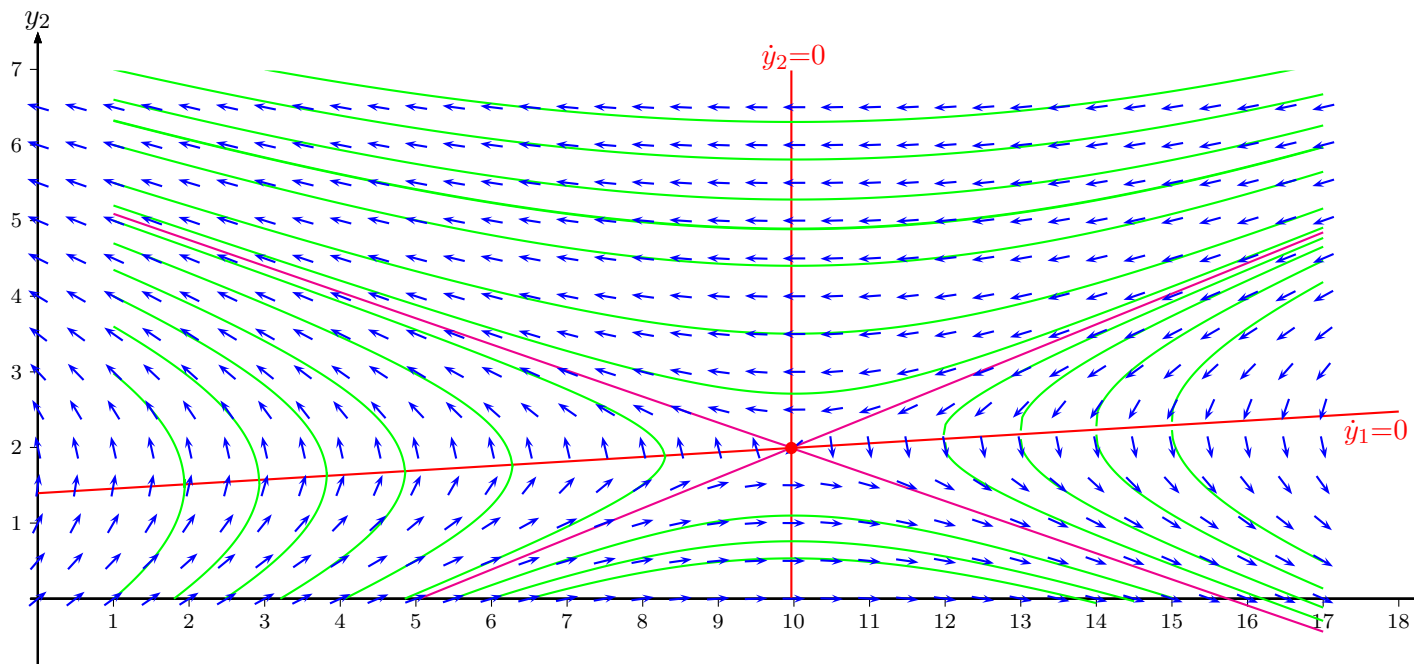
Offensichtlich ist die stationäre Lösung $\mathbf{y}^* = (5^{\frac{10}{7}} \mid 5^{\frac{3}{7}}) \approx (10 \mid 2)$ Sattelpunktstabil. Das Phasendiagramm zeigt in der Umgebung dieses Punktes (bis auf eine Koordinatenrotation) das annähernd gleiche Aussehen wie das Phasendiagramm der linearen DGL in Beispiel 3.23 ii). Dies entspricht gerade der Aussage von Satz 3.9 b). Das Phasendiagramm der linearisierten DGL ist daher in der Nähe von $\mathbf{y}^*$ eine gute Näherung für das Original. Die Linearisierung des DGL-System $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ durch $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}^*) + \mathbf{f}'(\mathbf{y}^*) \cdot (\mathbf{y} - \mathbf{y}^*)$ liefert (mit $\alpha=0.3$, $\beta=-0.7$)

$$\mathbf{f}'(\mathbf{y}^*) = \begin{pmatrix} \frac{\alpha}{5} & -1 \\ \frac{\beta}{5} & 0 \end{pmatrix}$$

und daher

$$\dot{\mathbf{y}} = \begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \begin{pmatrix} \frac{\alpha}{5} \cdot y_1 - \beta \cdot 5^{-\frac{\alpha}{\beta}} - y_2 \\ \frac{\beta}{5} \cdot y_1 - \beta \cdot 5^{-\frac{\alpha}{\beta}} \end{pmatrix}$$

Das linearisierte DGL-System liefert folgendes Phasendiagramm:



□

4

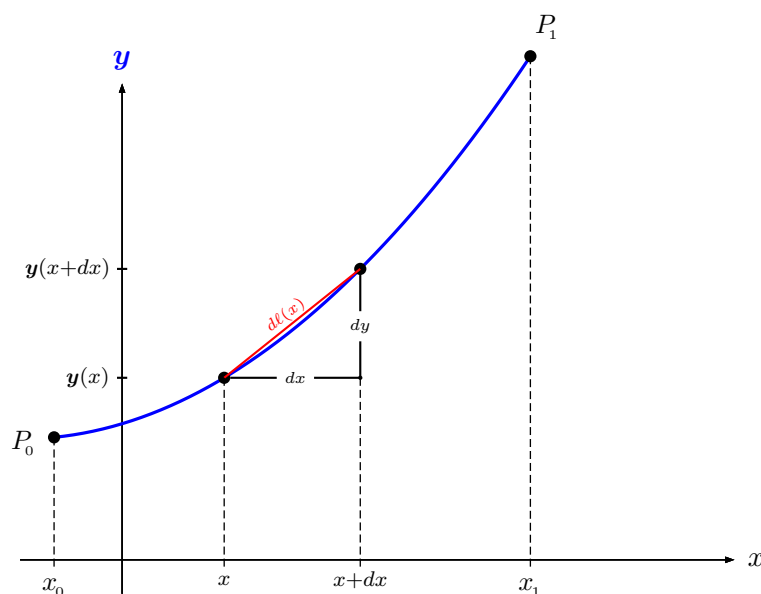
Dynamische Optimierung

Im Unterschied zur Extremwertbestimmung reeller Funktionen $f : \mathbb{D} \rightarrow \mathbb{R}$, mit $\mathbb{D} \subset \mathbb{R}^N$, befasst sich die dynamische Optimierung mit der Ermittlung von Minima und Maxima von *Funktionalen* $V : \mathbb{D}(V) \rightarrow \mathbb{R}$, wobei $\mathbb{D}(V)$ eine Menge von Funktionen (Wegen) ist. Analog zum klassischen Fall werden dabei nur relative Extrema bestimmt, d.h. Wege, die V in Bezug auf „benachbarte“ Wege einen größten oder kleinsten Wert erteilen. Die folgenden Beispiele zeigen, dass es vernünftig ist, von vornherein anzunehmen dass V ein Integral ist. Dies stellt keine Einschränkung dar, denn mit $I := \frac{d}{dt} V(g(t))$ gilt $V(g(t)) = \int I dt$. Plausibel erscheint dies auch, wenn man bekannte zeitdiskrete Probleme betrachtet, wo das zu optimierende Gesamtergebnis (Gewinn, Kosten o.ä.) die Summe der Teilergebnisse zu den betrachteten Zeitpunkten ist. Im kontinuierlichen Fall geht diese Summe in ein Integral über.

4.1 Beispiel

- i) Gesucht sei in der (x, y) -Ebene die kürzeste Verbindungsstrecke $y=y(x)$ zwischen zwei vorgegebenen Punkten $P_0=(x_0, y_0)$ und $P_1=(x_1, y_1)$. Nach Pythagoras gilt für die Länge $d\ell$ der Verbindungsstrecke zwischen x und $x+dx$ (siehe folgende Skizze)

$$(d\ell)^2 = (dx)^2 + (dy)^2 = (dx)^2 + (y' \cdot dx)^2 = (1 + y'^2) \cdot (dx)^2 .$$



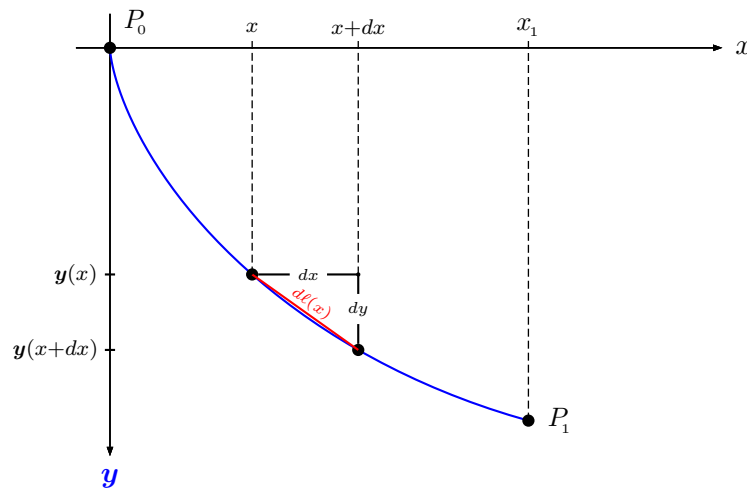
Daher lautet die mathematische Forderung für die Länge L der gesuchten Strecke:

$$L[y] := \int_{x_0}^{x_1} d\ell(x) = \int_{x_0}^{x_1} \sqrt{1 + y'^2} dx \rightarrow \min$$

$$y(x_0) = y_0, \quad y(x_1) = y_1$$

ii) (*Brachistochronenproblem*)

Zwischen zwei in verschiedener Höhe gelegenen Punkten P_0 und P_1 , ist eine Verbindungskurve so zu bestimmen, dass die Fallzeit eines Teilchens, das sich unter dem Einfluss der Schwerkraft reibungsfrei längs dieser Kurve bewegt, minimal wird. Wir verwenden ein kartesisches Koordinatensystem, dessen Ursprung in P_0 liegt und dessen y -Achse in Richtung des tiefer gelegenen Punktes P_1 zeigt (siehe nachfolgende Skizze).



Bezeichnet $y = y(x)$ die gesuchte Bahnkurve und war das Teilchen im Punkt P_0 in Ruhe, so folgt wegen der Energieerhaltung (mit der Erdbeschleunigung g)

$$\text{potentielle Energie} = mgy = \frac{1}{2}mv^2 = \text{kinetische Energie} ,$$

woraus sich für die Bahngeschwindigkeit v des Teilchens ergibt:

$$v = \sqrt{2gy}$$

Bezeichnet $d\ell$ das im Zeitraum dt zurückgelegte Bogenstück, so gilt andererseits für die Geschwindigkeit des Teilchens

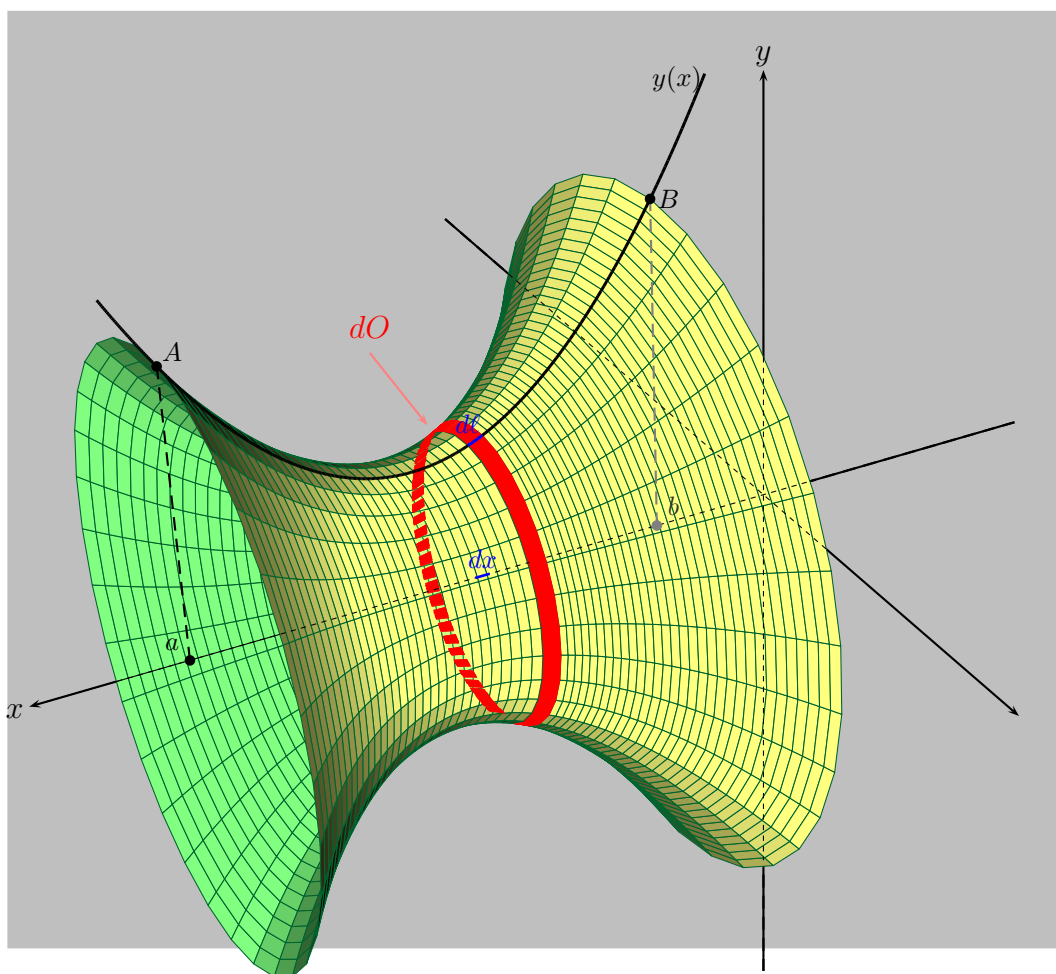
$$v = \frac{d\ell}{dt} \iff dt = \frac{d\ell}{v} = \frac{d\ell}{\sqrt{2gy}}$$

Beachtet man, dass (wie in Teil i) dieses Beispiels) $d\ell = \sqrt{1+y'^2} \cdot dx$ gilt, so ist die für die gesamte Bahn erforderliche zu minimierende Laufzeit:

$$t = \int_{P_0}^{P_1} dt = \int_{P_0}^{P_1} \frac{d\ell}{v} = \frac{1}{\sqrt{2 \cdot g}} \int_0^{x_1} \frac{\sqrt{1+y'^2}}{\sqrt{y}} \cdot dx \rightarrow \min$$

$$y(0) = 0, \quad y(x_1) = y_1$$

- iii) Gesucht sei die Verbindungsstrecke zwischen zwei Punkten der $(x-y)$ -Ebene, so dass die Oberfläche O des zugehörigen Rotationskörpers (ohne Deckel) minimal ist.



Die Länge $d\ell$ eines Bogenstückes der gesuchten Kurve $y(x)$ ist nach Teil i):

$$d\ell = (1 + y'^2)^{\frac{1}{2}} \cdot dx$$

Das Bogenstück $d\ell$ erzeugt bei der Rotation einen Zylinder der Höhe $d\ell$ und Oberfläche

$$dO = 2\pi y \cdot d\ell = 2\pi y \cdot (1 + y'^2)^{\frac{1}{2}} \cdot dx$$

Das Problem für die zu minimierende Gesamtoberfläche O lautet dann

$$O(y) = \int_a^b dO = 2\pi \int_a^b y(1 + y'^2)^{\frac{1}{2}} dx \rightarrow \min \quad \text{mit } y(a) = A, y(b) = B$$

- iv) Wir betrachten einen Monopolisten, der ein Gut produziert. Die Kostenfunktion in Abhängigkeit von der produzierten Menge q laute

$$K(q) = \alpha q^2 + \beta q + \gamma \quad (\alpha, \beta, \gamma > 0),$$

mit der Annahme, dass der Output q gleich der nachgefragten Menge ist. Nimmt man an, dass die nachgefragte Menge nicht nur vom Preis p , sondern auch von der Preisänderungsrate $\dot{p}(t) := \frac{dp}{dt}$ abhängt, gilt

$$q = a - b \cdot p(t) + c \cdot \dot{p}(t) \quad (a, b > 0, c \neq 0).$$

Der Gewinn g zum Zeitpunkt t ist dann

$$\begin{aligned} g(p, \dot{p}) &= p(t) \cdot q(t) - K(t) \\ &= -b(1+\alpha b) \cdot p^2 + (a+2\alpha ab+\beta b)p \\ &\quad -\alpha c^2 \dot{p}^2 - c(2\alpha a+\beta) \dot{p} + c(1+2\alpha b) p \dot{p} \\ &\quad -(\alpha a^2 + \beta a + \gamma) \end{aligned}$$

Ziel ist es nun einen optimalen Preispfad $p(t)$ zu bestimmen, so dass der Gesamtgewinn G über eine endliche Zeitperiode $[0, T]$ maximiert wird. Hierbei seien der Anfangspreis p_0 und der Endpreis p_T gegeben. Damit liegt folgendes Problem vor:

$$\begin{aligned} G(p) &= \int_0^T g(p, \dot{p}) e^{-it} dt \rightarrow \max \\ \text{u.d.N.} \quad p(0) &= p_0 \\ p(T) &= p_T \quad (p_0, p_T, T \text{ gegeben}) \end{aligned}$$

Die betrachtete Zeitperiode sei dabei so klein, dass die Annahme einer festen Nachfrage und Kostenfunktion plausibel erscheint. U.U. kann dann auch auf den Abzinsungsfaktor verzichtet werden. $\square$

Das allgemeine Kontroll-Problem

Das statische Optimierungsproblem in der Ökonomie besteht in der optimalen Verteilung von Ressourcen zu einem vorgegebenen Zeitpunkt. Mathematisch bedeutet dies, den Wert

der systembeschreibenden Zustandsvariablen oder Instrumente (in der allgemeinen Definition 4.3 mit $\mathbf{y}$ bezeichnet) innerhalb eines Zulässigkeitsbereiches so zu wählen, dass eine vorgegebene Zielfunktion optimiert wird.

Bei der dynamischen Optimierung muss dagegen eine optimale Allokation der Ressourcen für ein ganzes Zeitintervall vorgenommen werden, so dass der Gesamtnutzen für den gesamten betrachteten Zeitraum optimal ist. Mathematisch formuliert heißt dies, die Zeitpfade der Zustandsvariablen innerhalb einer gegebenen Klasse von Zeitpfaden zu bestimmen. Diese Optimierung wird im allgemeinen Fall indirekt durch eine optimale Wahl von systemsteuernden Kontrollgrößen (in der allgemeinen Definition 4.3 mit $\mathbf{u}$ bezeichnet) vorgenommen und *nicht* durch direkte Bestimmung der optimalen Zustandsvariablen $\mathbf{y}$.

Die Idee hinter diesem *allgemeinen Kontrollansatz* verdeutlicht folgendes Beispiel.

4.2 Beispiel

Wir betrachten den Abbau (Förderung) einer endlichen Ressource S (= Zustandsvariable). Der Bestand der Ressource vermindert sich gemäß folgender Vorschrift

$$(*) \quad \dot{S} = -E(t),$$

wobei E die Förderungsrate zum Zeitpunkt t bezeichnet. Statt mit S wird man bevorzugt mit der Variablen E (= Steuervariable) arbeiten. Diese gestattet auf der einen Seite über die DGL (*) die Beschreibung von S , auf der anderen Seite steuert sie den Abbau der Ressource und ist damit als Kontrollvariable geeignet. Will man zum Beispiel den Nutzen U , der durch den Abbau der Ressource entsteht, über eine Zeitperiode $[0, T]$ optimieren, könnte folgendes Problem entstehen:

$$\int_0^T U(E) e^{-\rho t} dt \rightarrow \max$$

$$\text{u.d.N.} \quad \dot{S} = -E(t) \quad \text{mit} \quad S(0) = S_0, \quad S_0 \geq S(T) \geq 0 \quad \text{frei}$$

□

4.3 Definition

Das *allgemeine Kontroll-Problem* der dynamischen Optimierung lautet:

$$V = V(\mathbf{u}(t)) := \int_{t_0}^{t_1} I(\mathbf{y}(t), \mathbf{u}(t), t) dt + F(\mathbf{y}(t_1), t_1) \rightarrow \max$$

$$\text{u.d.N.} \quad \dot{\mathbf{y}}(t) = \mathbf{f}(\mathbf{y}(t), \mathbf{u}(t), t)$$

$$\mathbf{y}(t_0) = \mathbf{y}_0 \quad (t_0, \mathbf{y}_0 \text{ gegeben})$$

$$(\mathbf{y}(t_1), t_1) \in \mathcal{T} \subset \mathbb{R}^n \times \mathbb{R} \quad (\mathcal{T} \text{ gegeben})$$

$$\mathbf{u}(t) \in \Omega \subset \mathbb{R}^k \quad \text{für} \quad t_0 \leq t \leq t_1 \quad (\Omega \text{ gegeben})$$

Dabei heißen

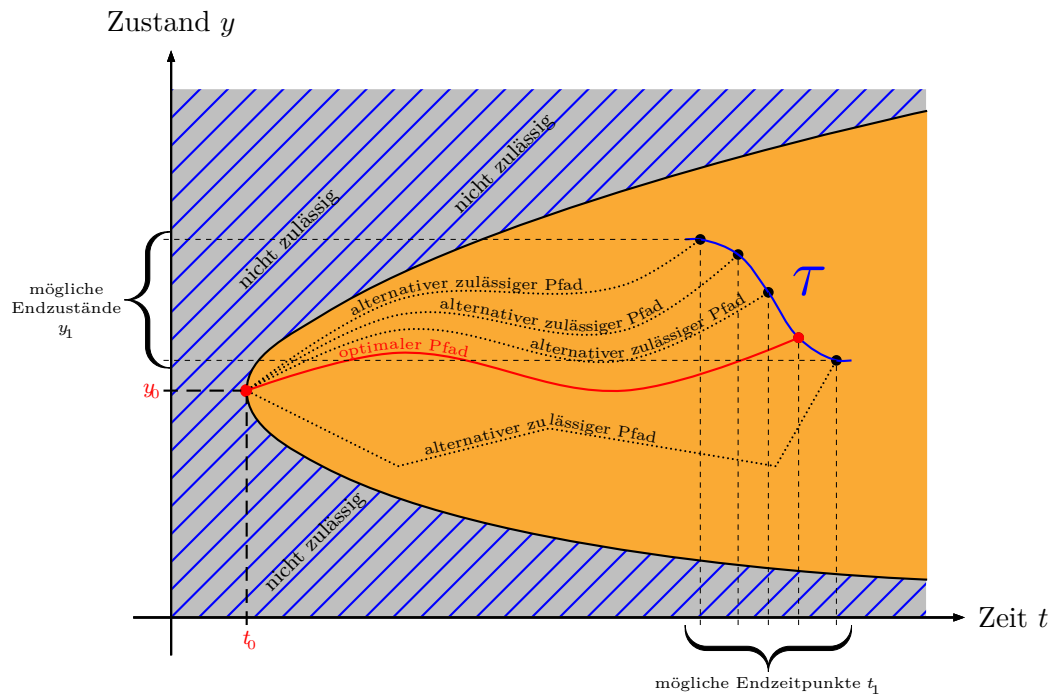
- V Zielfunktional, I intermediäre, F finale Funktion, $\mathcal{T}$ Transversalitätsbedingung
- $\dot{\mathbf{y}}(t) := \frac{d\mathbf{y}}{dt} = \mathbf{f}(\mathbf{y}(t), \mathbf{u}(t), t)$ die Bewegungsgleichung des Systems
- t_0 der gegebene Anfangszeitpunkt
- t_1 der je nach Aussehen von $\mathcal{T}$ gegebene oder zu bestimmende Endzeitpunkt
- $\mathbf{u}(t) = (u_1(t), \dots, u_k(t))$ der Kontroll- oder Steuervektor (Regelgröße) zum Zeitpunkt t . Die Menge Ω aller zulässigen $\mathbf{u}$ heißt Steuerbereich oder Kontrollmenge.
- $\mathbf{y}(t) = (y_1(t), \dots, y_n(t)) \in \mathbb{R}^n$ der Zustands- oder Phasenvektor zum Zeitpunkt t
- $\mathbf{y}_0$ der gegebene Anfangszustand
- $\mathbf{y}_1$ der je nach Aussehen von $\mathcal{T}$ gegebene oder zu bestimmende Endzustand. $\square$

Weitere Voraussetzungen sind:

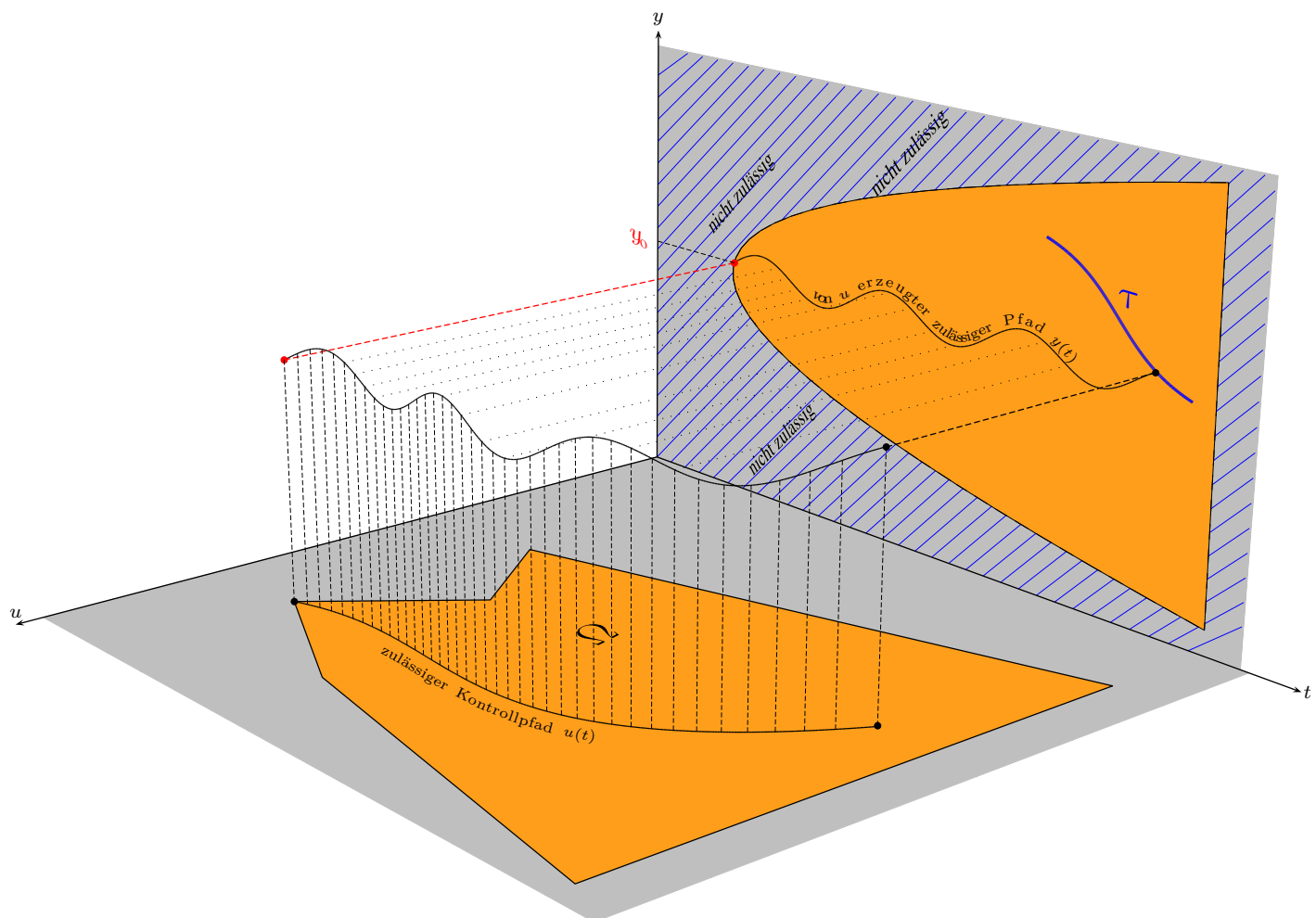
- a) Die auftretenden Funktionen I , F und $\mathbf{f}$ sind stetig differenzierbar
- b) Die Zustandstrajektorien $t \mapsto \mathbf{y}(t)$ sind stückweise stetig differenzierbar
- c) Die Kontrolltrajektorien $t \mapsto \mathbf{u}(t)$ sind stückweise stetig

4.4 Bemerkung

- i) Die Bezeichnung intermediär für die Funktion I rührt daher, dass diese die Abhängigkeit der Zielfunktion V vom Zeitpfad der Zustands- und Kontrollvariablen und von der Zeit *innerhalb* des maßgebenden Zeitintervalls beschreibt.
- ii) Die Funktion F beschreibt die Abhängigkeit der Zielfunktion V vom Endzustand und Endzeitpunkt (daher auch die Bezeichnung final) und gibt gewissermaßen eine Zusatzvergütung bzw. eine Konventionalstrafe in Abhängigkeit vom erreichten Endzustand und Endzeitpunkt an.
- iii) Der Begriff Steuer- oder Kontrollvariable rührt daher, dass diese über ein DGL-System (die Bewegungsgleichung) die systembeschreibenden Zustandsvariablen festlegen (siehe etwa Beispiel 4.2). Die möglichen Zustandstrajektorien heißen *zulässig*. Der Zulässigkeitsbereich enthält alle Trajektorien $\mathbf{y}(t)$, die sich aus der Bewegungsgleichung ergeben, wenn sich $\mathbf{u}(t)$ im Kontrollbereich Ω „bewegt“. Er ist also abhängig von $\mathbf{f}$ und Ω (und t_0 , falls $\mathbf{f}$ nicht autonom ist). Folgendes Bild beschreibt die Geometrie des allgemeinen Kontrollproblems für eine Zustandsvariable ($n=1$):



Bezieht man im letzten Bild die Wirkung der Kontrollvariablen u mit ein, entsteht folgende 3-dimensionale Graphik. Sie zeigt, wie durch Wahl eines zulässigen „Kontrollpfades“ $u(t)$ in Ω der zugehörige zulässige „Zustandspfad“ $y(t)$ als Lösung der Bewegungsgleichung $\dot{y} = f(y(t), u(t), t)$ entsteht.



- iv) Besitzt die Zielfunktion V die in Definition 4.3 vorgegebene Form, spricht man von dem *Problem von Bolza*. Ist die finale Funktion $F \equiv 0$ (die klassische Form der dynamischen Optimierung), spricht man von dem *Problem von Lagrange*. Ist die intermediäre Funktion $I \equiv 0$, so spricht man von dem *Problem von Mayer*.

Das Problem von Bolza ist dabei nur scheinbar die allgemeinste Form. In Wirklichkeit sind alle drei Probleme äquivalent.

Das Problem von Bolza kann durch die zusätzliche Zustandsvariable $z(t)$ mit

$$\dot{z}(t) = I(\mathbf{y}, \mathbf{u}, t), \quad z(t_0) = 0$$

in das Problem von Mayer überführt werden, weil dann nämlich gilt:

$$z(t) = \int_{t_0}^t I(\mathbf{y}, \mathbf{u}, t) dt \quad \text{d.h.} \quad V = z(t_1) + F(\mathbf{y}(t_1), t_1)$$

Andererseits kann das Problem von Bolza in das Problem von Lagrange transformiert werden. Hierzu beachte man, dass gilt:

$$F(\mathbf{y}(t_1), t_1) = \int_{t_0}^{t_1} \left[\frac{d}{dt} F(\mathbf{y}(t), t) \right] dt + F(\mathbf{y}(t_0), t_0)$$

Damit hat V die Form

$$V = \int_{t_0}^{t_1} \left[I(\mathbf{y}, \mathbf{u}, t) + \frac{d}{dt} F(\mathbf{y}, t) \right] dt + F(\mathbf{y}_0, t_0) .$$

Weil $\mathbf{y}_0$ und t_0 vorgegeben sind, ist $F(\mathbf{y}_0, t_0)$ konstant. Da additiv konstante Größen keinen Einfluss auf die Bestimmung des Optimums haben, sind alle drei Optimierungsprobleme äquivalent.

- v) Im folgenden werden die auftretenden Funktionen des öfteren nur nach einem Teil der Variablen differenziert. Ist $f(y_1, \dots, y_n, u_1, \dots, u_k, t)$ eine Funktion in den Variablen $y_1, \dots, y_n, u_1, \dots, u_k, t$, so verwenden wir folgende Schreibweisen:

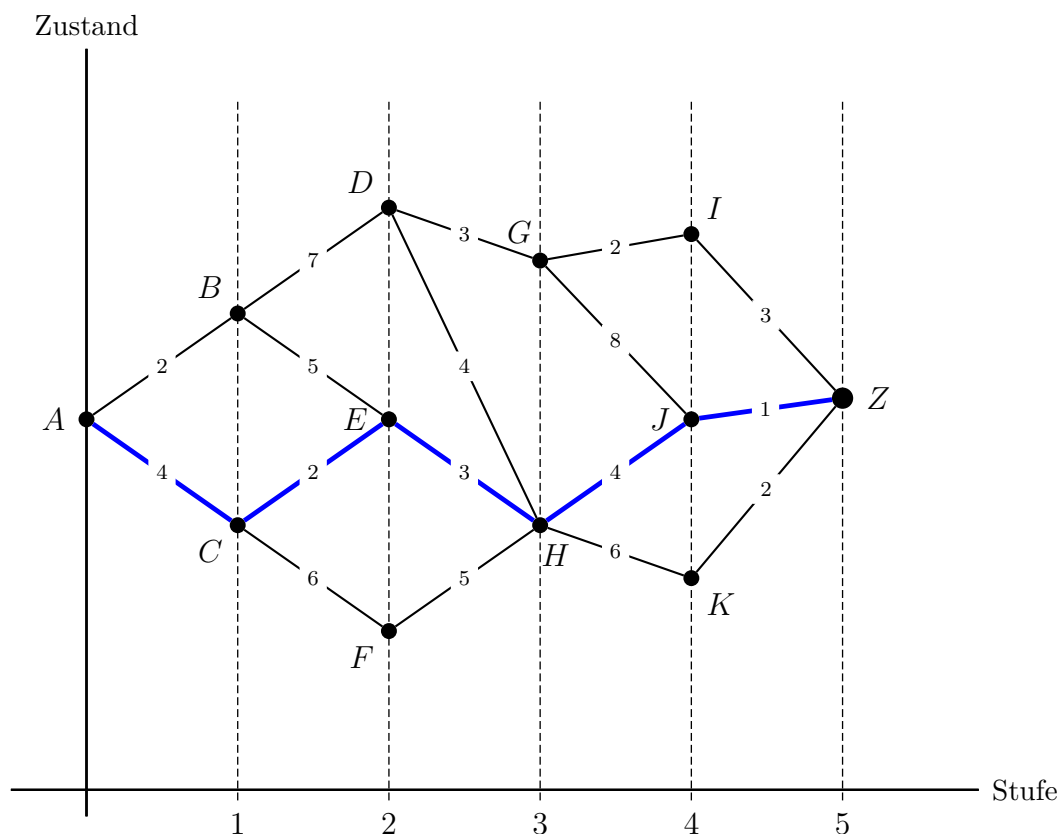
$$D_{\mathbf{y}} f = \frac{\partial f}{\partial \mathbf{y}} := \left(\frac{\partial f}{\partial y_1}, \dots, \frac{\partial f}{\partial y_n} \right), \quad D_{\mathbf{u}} f = \frac{\partial f}{\partial \mathbf{u}} := \left(\frac{\partial f}{\partial u_1}, \dots, \frac{\partial f}{\partial u_k} \right)$$

Analog steht $D_{\mathbf{y}}^2 f = \frac{\partial^2 f}{\partial \mathbf{y}^2}$ bzw. $D_{\mathbf{u}}^2 f = \frac{\partial^2 f}{\partial \mathbf{u}^2}$ für die zweite Ableitung (smatrix) nach den Variablen $y_1, \dots, y_n$ bzw. $u_1, \dots, u_k$. $\square$

Die Schwierigkeit der dynamischen Optimierung lässt sich schon im Fall von (zeit)diskreten Vorgängen, wo man von dynamischer Programmierung spricht, beobachten. Betrachten wir hierzu folgendes Beispiel.

4.5 Beispiel

Der Transport eines Gutes von einem Punkt A zu einem Punkt Z verursacht, je nachdem, welcher Weg gewählt wird, unterschiedliche Transportkosten (abhängig von Topographie, Straßengebühren, Weglänge usw.). Der Transportweg lasse sich in 5 Stufen unterteilen, wobei in jeder Stufe je nach aktuellem Standort Weggabelungen vorliegen, die eine Entscheidung (Kontrollvariable!) über die weitere Wegwahl erforderlich machen. Im Startpunkt A stehen zwei alternative Wege zur Auswahl: Ein Weg nach B mit Kosten 2 GE und ein Weg nach C mit Kosten 4 GE. Nach dieser ersten Wegentscheidung müssen weitere Entscheidungen getroffen werden, bis man schließlich am Zielpunkt Z ankommt.

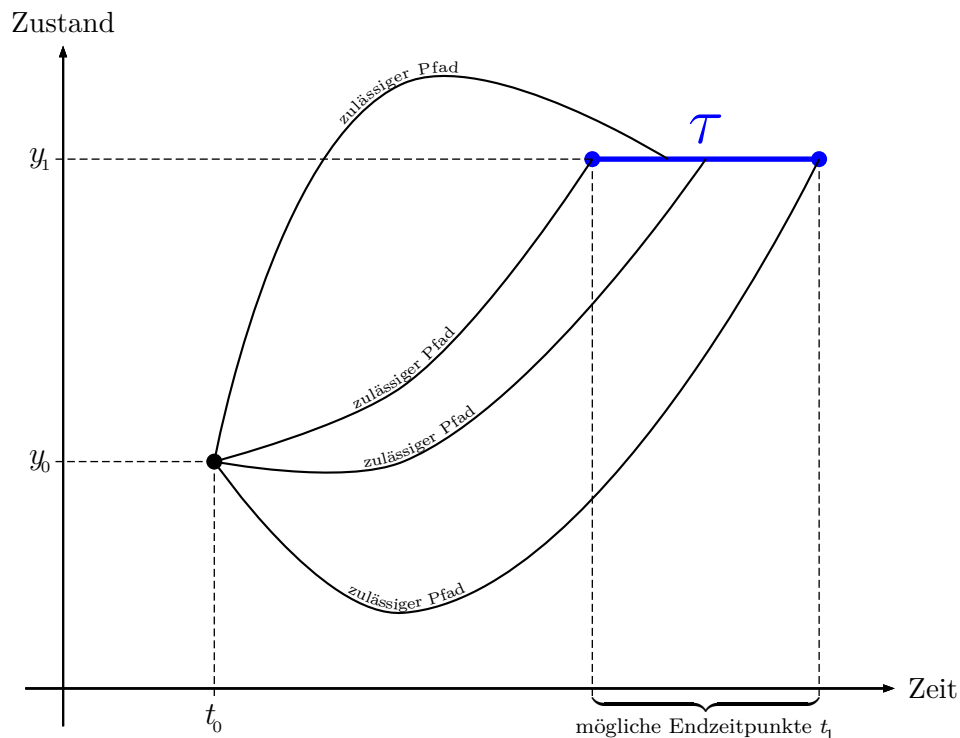


Das Beispiel zeigt, dass ein einstufiger Optimierungsprozess, bei dem man die Kosten von Stufe zu Stufe minimiert, nicht zum gesuchten kostenminimalen Gesamtweg führt. Diese Vorgehensweise würde nämlich zum Weg $ABEHJZ$ führen, dessen Gesamtkosten 15 GE betragen, im Gegensatz zum optimalen Weg $ACEHJZ$ mit den Gesamtkosten 14 GE. Dieser einstufige Optimierungsalgorithmus führt deswegen zu einer Fehlentscheidung, weil hierbei nicht berücksichtigt wird, welchen Einfluss eine Entscheidung auf die späteren Entscheidungsmöglichkeiten ausübt. Entscheidet man sich z.B. auf der ersten Stufe für den augenblicklich günstigsten Weg nach B , führt dies auf der nächsten Stufe nur zu relativ ungünstigen Alternativen, mit Kosten 5 bzw. 7 GE. $\square$

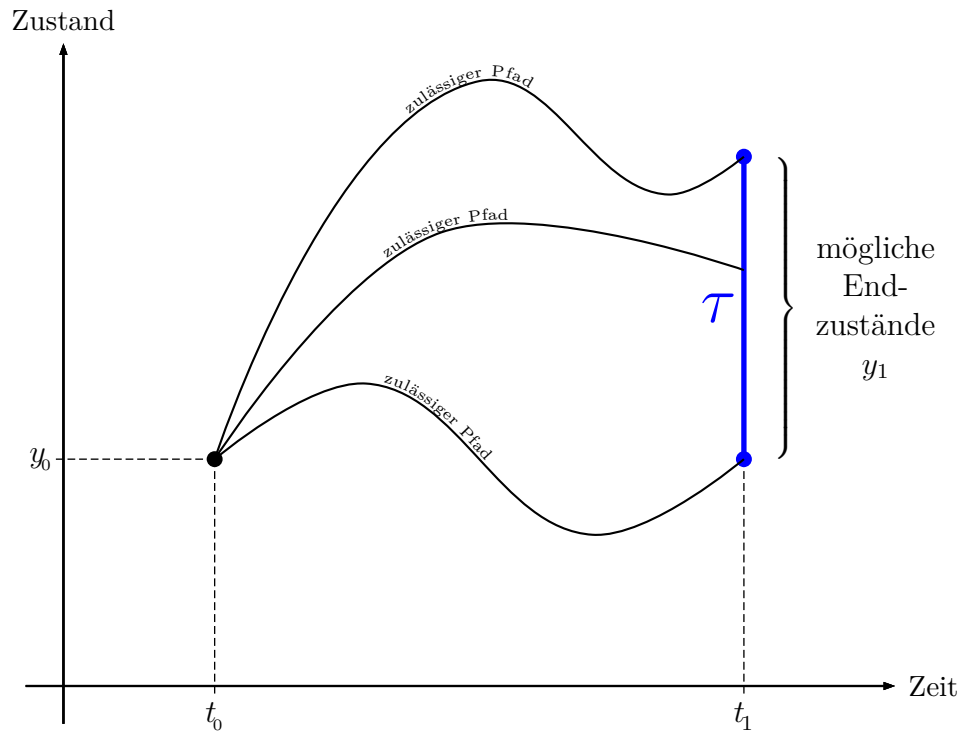
4.6 Bemerkung

Obwohl wir bei der Formulierung des Optimierungsproblems nur variable Endzustände (bzw. Zeiten) betrachtet haben, ist es auf die gleiche Art und Weise möglich, mit variablen Anfangsgrößen zu arbeiten. Die Vorgabe eines bestimmten Anfangswertes $\mathbf{y}(t_0) = \mathbf{y}_0$ stellt jedoch keine große Restriktion dar, da man bei den üblichen Optimierungsplänen meistens von der aktuellen Ausgangsposition ausgeht. Dahingegen muss beim Endzustand i.a. damit gerechnet werden, dass nur ein *Zielbereich* vorgegeben ist. Dies gibt zwar einen zusätzlichen Freiheitsgrad und führt dementsprechend zu besseren Optimierungsergebnissen, erhöht jedoch den Rechenaufwand. Zwei Extremfälle sind hier erwähnenswert:

- 1) Der Endzustand $\mathbf{y}_1$ ist fest vorgegeben, der Endzeitpunkt ist innerhalb eines vorgegebenen (evtl. unbeschränkten) Zeitintervalls frei. So könnte z.B. bei einer vorgegebenen Produktion eine genau bestimmte Qualitätsstufe $\mathbf{y}_1$ gefordert werden, wobei man jedoch zur Kostenoptimierung eine evtl. längere Produktionszeit akzeptiert.



- 2) Der Endzeitpunkt t_1 ist fest vorgegeben, $\mathbf{y}(t_1) = \mathbf{y}_1$ darf jedoch Werte innerhalb eines vorgegebenen Bereichs (z.B. ein Intervall) annehmen. Ein solches Problem kann sich etwa dadurch ergeben, dass bei der Suche nach einem optimalen Preispfad von vornherein negative Preise ausgeschlossen sind und zum Zeitpunkt t_1 eine offizielle Preisobergrenze vorgegeben ist.



Im allgemeinen Fall wird weder der Endzeitpunkt t_1 , noch der Endzustand y_1 festliegen, sondern es wird ein funktionaler Zusammenhang der Form $\mathcal{T}(y_1, t_1) = 0$ zwischen diesen beiden Größen bestehen. Obwohl in diesem Fall t_1 und y_1 frei sind, liegt wegen des funktionalen Zusammenhangs nur ein Freiheitsgrad vor. So könnte z.B. bei einer gegebenen Produktion ein Kunde an zwei Zielen interessiert sein:

- möglichst schnelle Fertigung
- möglichst gute Qualität

Da unter kostenminimalen Gesichtspunkten beide Ziele nicht gleichzeitig erreicht werden können, wird man einen Kompromiss in Form eines funktionalen Zusammenhangs akzeptieren, z.B. nach dem Motto: „je länger die Produktionszeit, desto höher die Qualität“. $\square$

Es gibt drei im Prinzip gleichwertige alternative Möglichkeiten, das allgemeine Kontrollproblem zu lösen, nämlich die

- 1.) Variationsrechnung
- 2.) Dynamische Programmierung
- 3.) Optimale Kontrolltheorie (Maximum Prinzip)

Zu bemerken ist vorab, dass die optimale Kontrolltheorie vom praktischen Standpunkt wohl die vorteilhafteste Methode darstellt, weil Restriktionen direkt verarbeitet werden können und weil sie im Gegensatz zur dynamischen Programmierung meistens Vermutungen

über die Natur der gesuchten Lösungen mitliefert. Die natürlichste, weil der statischen Optimierung nachempfundene und daher auch die mit Abstand älteste Lösungstheorie, ist die *Variationsrechnung*. Da die Ähnlichkeit zur statischen Optimierung (klassische Differentialrechnung) auch das Verständnis wesentlich erleichtert, soll diese Methode zuerst vorgestellt werden. Der Einfachheit halber sei dabei vorausgesetzt, dass die intermediäre Funktion I zweimal differenzierbar ist.

4.1. Variationsrechnung

4.7 Definition

Das *klassische Variationsproblem* lautet:

$$V(\mathbf{y}) = \int_{t_0}^{t_1} I(\mathbf{y}(t), \dot{\mathbf{y}}(t), t) dt \rightarrow \begin{matrix} \max \\ \min \end{matrix}$$

u.d.N. $\mathbf{y}(t_0) = \mathbf{y}_0, \quad \mathbf{y}(t_1) = \mathbf{y}_1$

Hierbei sind der Anfangs- und Endzeitpunkt t_0 und t_1 , sowie der Anfangs- und Endzustand $\mathbf{y}_0$ bzw. $\mathbf{y}_1$ vorgegeben. I sei zweimal differenzierbar. $\mathbb{D}(V)$ - der Definitionsbereich von V - bestehe aus allen stetigen Funktionen $\mathbf{y}$, die stückweise stetig differenzierbar sind. $\square$

4.8 Bemerkung

- i) Das klassische Variationsproblem ist ein Spezialfall des allgemeinen Kontrollproblems. Es ist ein Problem von Lagrange, bei dem die finale Funktion 0 und die Kontrollvariable $\mathbf{u}$ gleich $\dot{\mathbf{y}}$ ist, wobei $\mathbf{u}=\dot{\mathbf{y}}$ alle Werte annehmen darf (d.h. $\Omega=\mathbb{R}^n$). Die Bewegungsgleichung lautet trivialerweise $\dot{\mathbf{y}} = f(\mathbf{y}, \mathbf{u}, t) := \mathbf{u}$.
- ii) Wie im klassischen Fall muss zwischen lokalen und absoluten Extremwerten unterschieden werden. So ist $\mathbf{y}^*$ ein relatives Maximum (bzw. Minimum) von V , wenn für alle $\mathbf{y}$ in einer Umgebung von $\mathbf{y}^*$ gilt:

$$V(\mathbf{y}^*) \geq V(\mathbf{y}) \quad \text{bzw.} \quad V(\mathbf{y}^*) \leq V(\mathbf{y})$$

Ein absolutes Maximum bzw. Minimum liegt vor, wenn für alle $\mathbf{y}$ im Definitionsbereich $\mathbb{D}(V)$ von V gilt:

$$V(\mathbf{y}^*) \geq V(\mathbf{y}) \quad \text{bzw.} \quad V(\mathbf{y}^*) \leq V(\mathbf{y}).$$

$\square$

Ist $\mathbf{y}^*$ ein lokales Maximum bzw. Minimum von V und $\mathbf{h}(t)$ eine beliebige Funktion, so muss die Funktion $\tilde{V}(\varepsilon) := V(\mathbf{y}^* + \varepsilon \mathbf{h})$ in $\varepsilon = 0$ ein lokales Maximum bzw. Minimum

besitzen. Dies ist genau dann der Fall, wenn gilt:

$$\begin{aligned}\partial V(\mathbf{y}^*, \mathbf{h}) &:= \tilde{V}'(0) = \left. \frac{d\tilde{V}}{d\varepsilon} \right|_{\varepsilon=0} = 0 \quad \text{und} \\ \partial^2 V(\mathbf{y}^*, \mathbf{h}) &:= \tilde{V}''(0) = \left. \frac{d^2\tilde{V}}{d\varepsilon^2} \right|_{\varepsilon=0} \leq 0 \quad (\text{bzw. } \geq 0)\end{aligned}\tag{4.1}$$

$\partial V(\mathbf{y}^*, \mathbf{h})$ bzw. $\partial^2 V(\mathbf{y}^*, \mathbf{h})$ heißen die *erste* und *zweite Variation* von V .

Gleichung (4.1) ist aus praktischen Gründen nicht sehr nützlich, da diese Bedingung für alle Störfunktionen $\mathbf{h}$ zu überprüfen ist. Der folgende Satz dient in der Variationsrechnung dazu, aus den Eigenschaften eines Integrals Rückschlüsse auf den Integranden zu ziehen. Er gestattet es, die Bedingung (4.1) so zu formulieren, dass die Variablen ε und $\mathbf{h}$ nicht mehr auftauchen.

4.9 Satz (Fundamentallemma der Variationsrechnung)

Sei $g : [t_0, t_1] \rightarrow \mathbb{R}$ eine stetige Funktion. Gilt $\int_{t_0}^{t_1} g(t)h(t)dt \leq 0$ für jede nichtnegative unendlich oft differenzierbare Funktion $h : [t_0, t_1] \rightarrow \mathbb{R}$ mit $h^{(n)}(t_0) = h^{(n)}(t_1) = 0$ für alle $n \in \mathbb{N}$, so folgt $g \leq 0$. Die analoge Aussage gilt in den Fällen ≥ 0 bzw. $= 0$. $\square$

Beweis:

Wäre $g(s) > 0$ für ein $s \in]t_0, t_1[$, so existiert wegen der Stetigkeit von g ein Intervall $[a, b] \subset]t_0, t_1[$, das s enthält und in dem $g > 0$ gilt. Definiert man nun die unendlich oft differenzierbare Funktion h durch

$$h(t) = \begin{cases} 0 & \text{falls } t \notin [a, b] \\ e^{-\frac{1}{(t-a)^2(b-t)^2}} & \text{falls } t \in]a, b[\end{cases}$$

so folgt mit dem Mittelwertsatz der Integralrechnung

$$\int_{t_0}^{t_1} g(t)h(t)dt = \int_a^b g(t)h(t)dt \stackrel{\text{MWS}}{=} g(x) \cdot h(x) \cdot (b-a) \quad \text{für ein } x \in]a, b[.$$

Wegen $g(x) > 0$, $h(x) > 0$ und $b-a > 0$ ist dies ein Widerspruch zu $\int_{t_0}^{t_1} g(t)h(t)dt \leq 0$. $\blacksquare$

4.10 Satz

Ist $\mathbf{y}^*(t)$ ein lokales Extremum der klassischen Variationsaufgabe 4.7

$$\begin{aligned}V(\mathbf{y}) &= \int_{t_0}^{t_1} I(\mathbf{y}(t), \dot{\mathbf{y}}(t), t)dt \rightarrow \max_{\min} \\ \text{u.d.N.} \quad \mathbf{y}(t_0) &= \mathbf{y}_0, \quad \mathbf{y}(t_1) = \mathbf{y}_1\end{aligned}$$

so muss $\mathbf{y}^*(t)$ der *Euler-Lagrange-Gleichung* genügen, d.h.

$$\frac{\partial I}{\partial \mathbf{y}} - \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{\mathbf{y}}} \right) = \mathbf{0}.$$

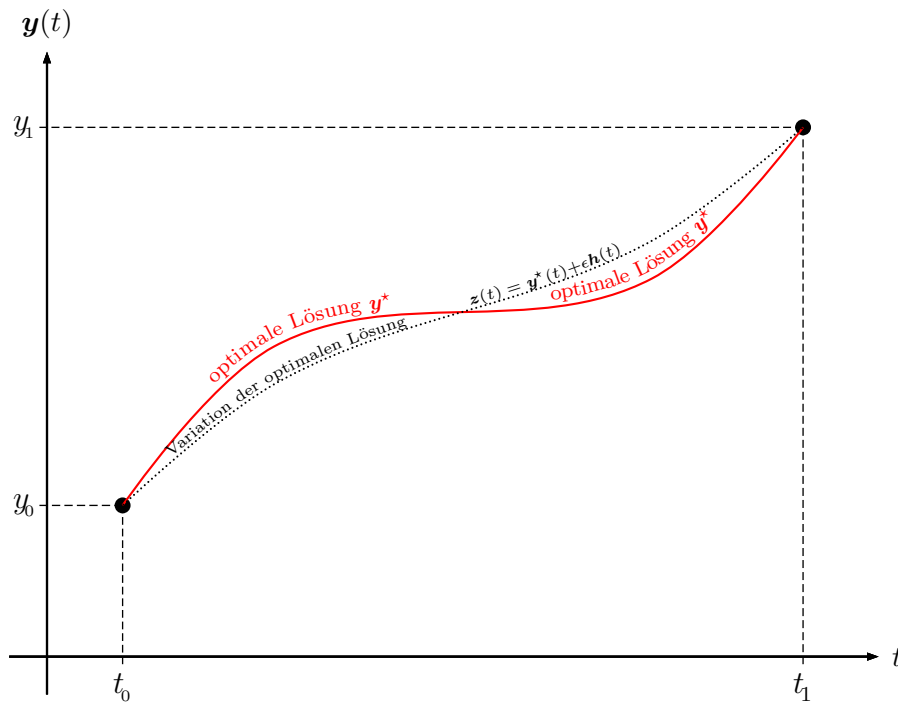
$\square$

Beweis:

Es genügt den Fall $n = 1$ zu betrachten, da der Fall $n > 1$ völlig analog läuft.

Sei $y^*(t)$ eine stetig differenzierbare (lokal) optimale Lösung des Problems und $h : [t_0, t_1] \rightarrow \mathbb{R}$ eine beliebige stetig differenzierbare Funktion mit der Eigenschaft $h(t_0) = h(t_1) = 0$. Dann ist $z := y^* + \varepsilon \cdot h$ für jedes ε eine stetig differenzierbare Funktion mit

$$z(t_i) = y^*(t_i) + \varepsilon h(t_i) = y_i + 0 = y_i \quad \text{für } i = 0, 1$$



Da y^* ein lokales Extremum von V ist, muss $\tilde{V}(\varepsilon) := V(y^* + \varepsilon h)$ für jedes h ein lokales Extremum an der Stelle $\varepsilon^* = 0$ besitzen. Nach der klassischen Differentialrechnung muss daher $\tilde{V}'(0) = 0$ gelten. (d.h. ∂V , die erste Variation von V , ist 0). Dies liefert:

$$\begin{aligned} 0 &= \tilde{V}'(0) \\ &= \left. \frac{dV(y^* + \varepsilon h)}{d\varepsilon} \right|_{\varepsilon=0} \\ &= \left. \frac{d}{d\varepsilon} \int_{t_0}^{t_1} I(z, \dot{z}, t) dt \right|_{\varepsilon=0} \\ &= \int_{t_0}^{t_1} \left. \frac{d}{d\varepsilon} I(z, \dot{z}, t) \right|_{\varepsilon=0} dt \\ &= \int_{t_0}^{t_1} \left[\frac{\partial I}{\partial y}(y, \dot{y}, t) \cdot h(t) + \frac{\partial I}{\partial \dot{y}}(y, \dot{y}, t) \cdot \dot{h}(t) \right]_{(y, \dot{y})=(y^*, \dot{y}^*)} \Big|_{\varepsilon=0} dt \\ &= \int_{t_0}^{t_1} \left[\frac{\partial I}{\partial y}(y^*, \dot{y}^*, t) \cdot h(t) + \frac{\partial I}{\partial \dot{y}}(y^*, \dot{y}^*, t) \cdot \dot{h}(t) \right] dt \end{aligned}$$

Integriert man den zweiten Term mittels partieller Integration, so ergibt sich

$$\begin{aligned} \int_{t_0}^{t_1} \frac{\partial I}{\partial \dot{y}}(y^*, \dot{y}^*, t) \cdot \dot{h}(t) dt &= \left[\frac{\partial I}{\partial \dot{y}}(y^*, \dot{y}^*, t) \cdot h(t) \right]_{t_0}^{t_1} - \int_{t_0}^{t_1} \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{y}} \right) \cdot h(t) dt \\ &\stackrel{h(t_0)=h(t_1)=0}{=} - \int_{t_0}^{t_1} \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{y}} \right) \cdot h(t) dt \end{aligned}$$

Einsetzen liefert:

$$0 = \int_{t_0}^{t_1} \left[\frac{\partial I}{\partial y}(y^*, \dot{y}^*, t) - \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{y}} \right) \right] h(t) dt .$$

Da h beliebig war, muss dann nach dem Fundamentallemma der Variationsrechnung gelten:

$$\frac{\partial I}{\partial y}(y^*, \dot{y}^*, t) - \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{y}} \right)(y^*, \dot{y}^*, t) = 0 \quad \blacksquare$$

4.11 Bemerkung

i) Die Euler-Gleichung ist eine DGL (bzw. DGL-System) zweiter Ordnung, denn es gilt:

$$\frac{\partial I}{\partial y} - \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{y}} \right) = \frac{\partial I}{\partial y} - \frac{\partial^2 I}{\partial y \partial \dot{y}} \cdot \dot{y} - \frac{\partial^2 I}{\partial \dot{y}^2} \cdot \ddot{y} - \frac{\partial^2 I}{\partial t \partial \dot{y}}$$

Nach Satz 4.10 muss die optimale Lösung also der folgenden AWA genügen:

$$\begin{aligned} \frac{\partial^2 I}{\partial \dot{y}^2} \cdot \ddot{y} + \frac{\partial^2 I}{\partial y \partial \dot{y}} \cdot \dot{y} + \left(\frac{\partial^2 I}{\partial t \partial \dot{y}} - \frac{\partial I}{\partial y} \right) &= 0 \\ y(t_0) = y_0, \quad y(t_1) &= y_1 \end{aligned}$$

ii) Die Lösungen der Eulerschen DGL werden *Extremale* genannt. Je nach Größe des Definitionsbereiches $\mathbb{D}(V)$, d.h., ob $\mathbb{D}(V)$ eine Teilmenge der stetigen oder differenzierbaren Funktionen ist, spricht man von *starken* oder *schwachen Lösungen*.

iii) Eine leichte Rechnung zeigt, dass die Euler-Lagrange-Gleichung im eindimensionalen Fall äquivalent ist zu

$$\frac{d}{dt} \left(I - \left(\frac{\partial I}{\partial \dot{y}} \right) \dot{y} \right) - \frac{\partial I}{\partial t} = 0.$$

iv) Es lohnt sich einige Spezialfälle der Euler-DGL zu betrachten, nämlich die Fälle, in denen die intermediäre Funktion I eine oder mehrere der drei Variablen $t, y, \dot{y}$ nicht explizit enthält.

1. Spezialfall

Die intermediäre Funktion I hängt nicht explizit von $\mathbf{y}$ ab, d.h. es gilt $I = I(\dot{\mathbf{y}}, t)$ oder $\frac{\partial I}{\partial \mathbf{y}} = \mathbf{0}$. Die Euler-Gleichung lautet dann:

$$-\frac{d}{dt} \left(\frac{\partial I}{\partial \dot{\mathbf{y}}} \right) = \mathbf{0}$$

Diese kann direkt integriert werden und ist daher äquivalent zu

$$\frac{\partial I}{\partial \dot{\mathbf{y}}} = \text{konstant.}$$

2. Spezialfall

I ist nicht explizit von $\dot{\mathbf{y}}$ abhängig d.h. es gilt $\frac{\partial I}{\partial \dot{\mathbf{y}}} = \mathbf{0}$. Die Euler-Gleichung lautet dann:

$$\frac{\partial I}{\partial \mathbf{y}} = \mathbf{0}.$$

Dies ist gerade die klassische Bedingung für relative Extrema. Das Problem ist in Wirklichkeit ein statisches Problem, das nur künstlich mit einer Zeitvariable indiziert worden ist.

3. Spezialfall

I ist nicht explizit zeitabhängig, d.h. es gilt $I = I(y, \dot{y})$ bzw. $\frac{\partial I}{\partial t} = 0$. Ist die Zustandsvariable eindimensional ergibt sich nach Teil iii) der Bemerkung für die Euler-DGL:

$$\frac{d}{dt} \left(I - \left(\frac{\partial I}{\partial \dot{y}} \right) \dot{y} \right) = 0.$$

Diese kann integriert werden und man erhält dann:

$$I - \left(\frac{\partial I}{\partial \dot{y}} \right) \dot{y} = \text{konstant.}$$

4. Spezialfall

I ist weder von $\mathbf{y}$ noch von t abhängig, d.h. $I = I(\dot{\mathbf{y}})$. Nach dem ersten Spezialfall gilt dann $\frac{\partial I}{\partial \dot{\mathbf{y}}} = \text{konstant}$. Da $\frac{\partial I}{\partial \dot{\mathbf{y}}}$ die Größen t und $\mathbf{y}$ nicht explizit enthält, ist jede affin lineare Funktion $\mathbf{y}$ eine Lösung dieser DGL (im Gegensatz zu Spezialfall 1!).

Da hierdurch jede Anfangsbedingung erfüllt werden kann und jede Lösung in einer Umgebung daher mit einer Geraden übereinstimmt, sind dies auch alle Extremale.

- v) Manchmal spricht man nur von der *Euler-Gleichung*. Die Bezeichnung Euler-Lagrange-Gleichung wird dann ausschließlich bei dynamischen Optimierungsproblemen mit zusätzlichen Nebenbedingungen verwendet. $\square$

4.12 Beispiel

Wir versuchen die zu Beginn in Beispiel 4.1 gegebenen Probleme zu lösen.

$$\text{i) } \int_{x_0}^{x_1} \sqrt{1+y'^2} dx \rightarrow \min$$

$$y(x_0) = y_0, \quad y(x_1) = y_1$$

Die Intermediäre Funktion I lautet $I(y, y', x) = I(y')$. Nach Spezialfall 4 aus Bemerkung 4.11 iv) muss $y^*(x) = ax + b$ gelten. Die Anfangsbedingungen liefern:

$$y^*(x) = \frac{y_1 - y_0}{x_1 - x_0} \cdot x + \frac{y_0 x_1 - y_1 x_0}{x_1 - x_0}.$$

$$\text{ii) } V = \int_0^{x_1} \frac{\sqrt{1+y'^2}}{\sqrt{y}} dx \rightarrow \min \quad \text{mit} \quad y(0) = 0, \quad y(x_1) = y_1$$

Da hier die intermediäre Funktion $I = \frac{\sqrt{1+y'^2}}{\sqrt{y}}$ nicht explizit von x abhängt, muss nach Fall 3 aus Bemerkung 4.11 iv) gelten:

$$I - \left(\frac{\partial I}{\partial y'} \right) \cdot y' = \frac{\sqrt{1+y'^2}}{\sqrt{y}} - \frac{y'}{\sqrt{y(1+y'^2)}} \cdot y' = \text{konst}$$

Bildung des Hauptnenners, Quadrieren und Bildung des Kehrwertes liefert¹:

$$y(1+y'^2) = 2K$$

Mit der Substitution

$$y'(x) := \cot \left(\frac{z(x)}{2} \right) = \frac{\cos \left(\frac{z(x)}{2} \right)}{\sin \left(\frac{z(x)}{2} \right)}$$

ergibt sich wegen $1 + \cot^2(\alpha) = \frac{1}{\sin^2(\alpha)}$:

$$y(1+y'^2) = y \cdot \frac{1}{\sin^2 \left(\frac{z}{2} \right)} = 2K \quad \text{also} \quad y = 2K \sin^2 \left(\frac{z}{2} \right).$$

Differenzieren nach x liefert

$$y' = 2K \sin \left(\frac{z}{2} \right) \cos \left(\frac{z}{2} \right) \cdot z'$$

Vergleicht man dies mit der Substitution $y' = \frac{\cos \left(\frac{z}{2} \right)}{\sin \left(\frac{z}{2} \right)}$, so folgt

$$2K \sin^2 \left(\frac{z}{2} \right) \cdot z' = 1 \quad \text{oder} \quad 2K \sin^2 \left(\frac{z}{2} \right) dz = dx$$

¹Der Faktor 2 im konstanten Term wird aus rechnerischen Gründen eingeführt.

Integration unter Beachtung von $2 \sin^2\left(\frac{z}{2}\right) = 1 - \cos(z)$ liefert:

$$x = K \cdot (z - \sin(z)) + K_2$$

Zusammen mit $y = 2K \sin^2\left(\frac{z}{2}\right) = K \cdot (1 - \cos(z))$ hat man eine parametrische Darstellung für $y(x)$.

Es ist die parametrische Darstellung einer *Zykloide*. Dies sind *Rollkurven*, die dadurch entstehen, dass ein Kreis auf einer Leitkurve abrollt, ohne dabei zu „gleiten“. Bezeichnet P einen Punkt auf der Peripherie des Kreises, so beschreibt P bei dieser Bewegung eine Zykloide. Die gerade beschriebene Zykloide entsteht, wenn die oben genannte Leitkurve eine Gerade ist. Ihre Darstellung in kartesischen Koordinaten lautet:

$$K \cdot \cos\left(\frac{x - K_2 + \sqrt{y(2K - y)}}{K}\right) = K - y$$

$$\text{iii) } \quad V(y) = \int_a^b y(1 + y'^2)^{\frac{1}{2}} dx \rightarrow \min$$

$$y(a) = A, \quad y(b) = B$$

Analog zu Teil ii) resultiert

$$y(1 + y'^2)^{\frac{1}{2}} - yy''(1 + y'^2)^{-\frac{1}{2}} = c$$

Nach einigen naheliegenden Umformungen erhält man

$$y' = \frac{1}{c} \sqrt{y^2 - c^2}$$

Diese DGL ist separabel. Es gilt

$$c \int \frac{dy}{\sqrt{y^2 - c^2}} = x + k \quad (k \text{ konst.})$$

wegen $\int \frac{dy}{\sqrt{y^2 - c^2}} = \ln\left(\frac{y + \sqrt{y^2 - c^2}}{c}\right)$ erhält man

$$e^{\frac{(x+k)}{c}} = \frac{\sqrt{y^2 - c^2} + y}{c}.$$

Auflösen nach y liefert

$$y(x) = \frac{c}{2} \left[e^{\frac{(x+k)}{c}} + e^{-\frac{(x+k)}{c}} \right].$$

Die Nebenbedingungen liefern c und k . Dies ist eine modifizierte Form der *Kettenkurve*

$$y(x) = \frac{1}{2} (e^x + e^{-x}) = \cosh(x).$$

$$\begin{aligned}
\text{iv)} \quad G(p) &= \int_0^T g(p, \dot{p}) dt \rightarrow \max \\
\text{u.d.N.} \quad p(0) &= p_0, \quad p(T) = p_T \\
\text{mit} \quad g(p, \dot{p}) &= -b(1 + \alpha b)p^2 + (a + 2\alpha ab + \beta b)p \\
&\quad - \alpha c^2 \dot{p}^2 - c(2\alpha a + \beta)\dot{p} + c(1 + 2\alpha b)p\dot{p} \\
&\quad - (\alpha a^2 + \beta a + \gamma).
\end{aligned}$$

Die Euler-Lagrange-Gleichung (Es kann auch der 3. Spezialfall aus Bemerkung 4.11 iv) verwendet werden.)

$$0 = \frac{\partial g}{\partial p} - \frac{d}{dt} \left(\frac{\partial g}{\partial \dot{p}} \right) = -\frac{\partial^2 g}{\partial \dot{p}^2} \cdot \ddot{p} - \frac{\partial^2 g}{\partial p \partial \dot{p}} \cdot \dot{p} - \frac{\partial^2 g}{\partial t \partial \dot{p}} + \frac{\partial g}{\partial p}$$

liefert eine DGL zweiter Ordnung:

$$\ddot{p} - \frac{b(1 + \alpha b)}{\alpha c^2} p = -\frac{a + 2\alpha ab + \beta b}{2\alpha c^2}$$

Eine spezielle Lösung $\bar{p}$ lautet

$$\bar{p} = \frac{a + 2\alpha ab + \beta b}{2b(1 + \alpha b)}$$

und ist offensichtlich konstant. Die Nullstellen $\lambda_{1/2} = \pm \sqrt{\frac{b(1 + \alpha b)}{\alpha c^2}}$ der charakteristischen Gleichung $\lambda^2 - \frac{b(1 + \alpha b)}{\alpha c^2} = 0$ liefern die allgemeine Lösung:

$$p^*(t) = c_1 e^{\lambda_1 t} + c_2 e^{-\lambda_1 t} + \bar{p}$$

c_1 und c_2 ergeben sich über die Nebenbedingung

$$\begin{aligned}
p_0 &= c_1 + c_2 + \bar{p} \\
p_T &= c_1 e^{\lambda_1 T} + c_2 e^{-\lambda_1 T} + \bar{p}.
\end{aligned}$$

Man erhält:

$$\begin{aligned}
c_1 &= \frac{p_0 - \bar{p} - (p_T - \bar{p}) e^{\lambda_1 T}}{1 - e^{2\lambda_1 T}} \\
c_2 &= \frac{p_0 - \bar{p} - (p_T - \bar{p}) e^{-\lambda_1 T}}{1 - e^{-2\lambda_1 T}}
\end{aligned}$$

Zu beachten ist, dass der Parameter c , der den Einfluss von $\dot{p}$ auf die Nachfrage beschreibt nur über λ_1 in die Lösung eingeht. Da c in λ_1 jedoch nur in der Potenz 2 auftaucht, spielt nur der Betrag von c eine Rolle.

- v) Das letzte Beispiel kann allgemeiner untersucht werden, indem man für den Gewinn g den allgemeinen Ansatz $g = g(p, \dot{p})$ verwendet. Da g nicht zeitabhängig ist, muss nach Fall 3 von Bemerkung 4.11 iv) der optimale Preispfad folgende Gleichung erfüllen:

$$g - \left(\frac{\partial g}{\partial \dot{p}} \right) \dot{p} = C = \text{konst.}$$

Da diese Gleichung auch im statischen Fall gelten muss, wo g von $\dot{p}$ unabhängig ist, muss die Konstante C wegen $\frac{\partial g}{\partial \dot{p}} = 0$ gerade gleich dem maximalen statischen Gewinn g_s sein. Damit lautet die Gleichung

$$g - \dot{p} \left(\frac{\partial g}{\partial \dot{p}} \right) = g_s.$$

Division durch g und Umstellen liefert

$$\epsilon_{g,\dot{p}} = \dot{p} \cdot \frac{\frac{\partial g}{\partial \dot{p}}}{g} = 1 - \frac{g_s}{g}$$

Dies besagt also, dass der „dynamische Monopolist“ den Preis so wählen wird, dass die Elastizität des Gewinns bezüglich der Preisänderung $\dot{p}$ gleich $1 - \frac{g_s}{g}$ ist.

Im Gegensatz hierzu wird der „statische Monopolist“, bei dem der Gewinn nur vom Preis p abhängt, die notwendige Bedingung $\frac{dg}{dp} = 0$ erfüllen, was (nach Multiplikation mit $\frac{p}{g}$) äquivalent ist zu $\epsilon_{g,p} = 0$. $\square$

4.13 Bemerkung

- i) Ist die intermediäre Funktion I abhängig von $t, \mathbf{y}, \dot{\mathbf{y}}, \dots, \mathbf{y}^{(n)}$, kann dies durch die Einführung künstlicher Zustandsvariablen auf den ursprünglichen Fall reduziert werden. Ist zum Beispiel

$$I(t, y, \dot{y}, \ddot{y}) = ty^2 + y\dot{y} + \ddot{y}^2,$$

so führt man die neue Variable $z := \dot{y}$ ein und man erhält

$$I = ty^2 + yz + \dot{z}^2.$$

Alternativ hierzu kann mit der gleichen Methode wie in Satz 4.10 eine notwendige Bedingung für ein Extremal hergeleitet werden, die sog. *Euler-Poisson-Gleichung*

$$\frac{\partial I}{\partial \mathbf{y}} - \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{\mathbf{y}}} \right) + \frac{d^2}{dt^2} \left(\frac{\partial I}{\partial \ddot{\mathbf{y}}} \right) + \dots + (-1)^n \frac{d^n}{dt^n} \left(\frac{\partial I}{\partial \mathbf{y}^{(n)}} \right) = \mathbf{0}$$

Für

$$I(t, y, \dot{y}, \ddot{y}) = ty^2 + y\dot{y} + \ddot{y}^2$$

erhält man z.B. die Euler-Poisson-Gleichung

$$(2ty + \dot{y}) - \frac{d}{dt}(y) + \frac{d^2}{dt^2}(2\ddot{y}) = 2ty + 2y^{(4)} = 0$$

- ii) Der optimale Pfad $\mathbf{y}^*$ muss der Euler-Lagrange-Gleichung nur in dem Bereich genügen, in dem stetige Differenzierbarkeit vorliegt, da dies beim Beweis von Satz 4.10 benötigt wurde. Ist die Lösung $\mathbf{y}^*$ nur stückweise stetig differenzierbar, d.h. liegen Eckpunkte vor, in denen $\dot{\mathbf{y}}^*$ unstetig ist, so muss die *Eckenbedingung von Weierstraß/Erdmann* als weitere notwendige Bedingung erfüllt sein. Sie besitzt kein Analogon im Bereich der statischen Optimierung, da sie wesentlich vom zeitlichen Verhalten der Lösung abhängt und besagt, dass in den Eckpunkten die Funktionen $\frac{\partial I}{\partial \dot{\mathbf{y}}}$ und $I - \frac{\partial I}{\partial \dot{\mathbf{y}}} \cdot \dot{\mathbf{y}}$ stetig sein müssen. $\square$

4.14 Lemma

Sei a konstant, $I(x, t)$ stetig differenzierbar und $b(x)$ differenzierbar. Dann gilt:

$$\frac{d}{dx} \int_a^{b(x)} I(x, t) dt = \int_a^{b(x)} \frac{\partial I}{\partial x}(x, t) dt + I(x, b(x)) \cdot b'(x)$$

Beweis:

Sei $H(x, t)$ eine Stammfunktion von I bezüglich t , d.h. $\frac{\partial H}{\partial t} = I$. Dann gilt wegen der 2-mal stetigen Differenzierbarkeit von H :

$$\frac{\partial I}{\partial x} = \frac{\partial^2 H}{\partial x \partial t} = \frac{\partial}{\partial t} \left(\frac{\partial H}{\partial x} \right).$$

Daher ist $\frac{\partial H}{\partial x}$ eine Stammfunktion von $\frac{\partial I}{\partial x}$ bezüglich der Variablen t , d.h. es gilt

$$\int_u^v \frac{\partial I}{\partial x} dt = \int_u^v \frac{\partial^2 H}{\partial x \partial t} dt = \left[\frac{\partial H}{\partial x} \right]_{t=u}^{t=v}$$

Es folgt

$$\begin{aligned} \frac{d}{dx} \int_a^{b(x)} I(x, t) dt &= \frac{d}{dx} [H(x, b(x)) - H(x, a)] \\ &= \frac{\partial H}{\partial t}(x, b(x)) b'(x) + \frac{\partial H}{\partial x}(x, b(x)) - \frac{\partial H}{\partial x}(x, a) \\ &= I(x, b(x)) b'(x) + \left[\frac{\partial H}{\partial x}(x, t) \right]_{t=a}^{t=b(x)} \\ &= I(x, b(x)) b'(x) + \int_a^{b(x)} \frac{\partial I}{\partial x} dt \end{aligned} \quad \blacksquare$$

4.15 Satz

Ist $\mathbf{y}^*$ ein lokales Extremum der klassischen Variationsaufgabe

$$V(\mathbf{y}) = \int_{t_0}^{t_1} I(\mathbf{y}, \dot{\mathbf{y}}, t) dt \rightarrow \text{Opt.}$$

$$\text{u.d.N } \mathbf{y}(t_0) = \mathbf{y}_0, \quad \mathcal{T}(\mathbf{y}(t_1), t_1) = 0$$

wobei $\mathcal{T} : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$ differenzierbar sei, so muss $\mathbf{y}^*$ folgenden Gleichungen genügen:

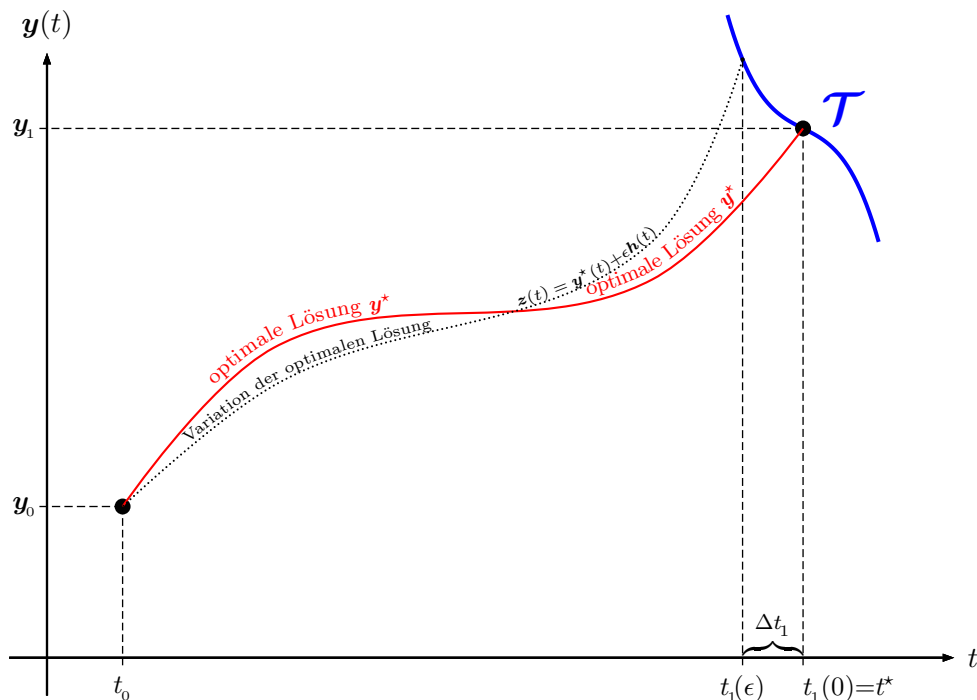
$$\frac{\partial I}{\partial \mathbf{y}} - \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{\mathbf{y}}} \right) = \mathbf{0} \quad (4.2)$$

$$\left[\left(I - \frac{\partial I}{\partial \dot{\mathbf{y}}} \cdot \dot{\mathbf{y}} \right) \frac{\partial \mathcal{T}}{\partial \mathbf{y}} - \frac{\partial I}{\partial \dot{\mathbf{y}}} \cdot \frac{\partial \mathcal{T}}{\partial t} \right]_{t=t^*} = 0 \quad (4.3) \quad \square$$

Beweis:

Ohne Einschränkung kann wieder $n = 1$ angenommen werden.

Sei t^* der zum optimalen Weg y^* gehörende optimale Endzeitpunkt. Für eine beliebige Störfunktion h mit $h(t_0) = 0$, $h(t_1)$ frei², definieren wir³ $z(t) := y^*(t) + \varepsilon h(t)$. Mit $t_1 = t_1(\varepsilon)$ bezeichnen wir den Endzeitpunkt der gestörten Bewegung z . Insbesondere gilt $t_1(0) = t^*$.



Wir definieren wie früher $\tilde{V}(\varepsilon) := V(z(t)) = V(y^*(t) + \varepsilon h(t))$. Wegen der Optimalität muss

² $h(t_1)$ ist nur bedingt frei, da $z(t_1(\varepsilon))$ auf $\mathcal{T}$ liegen muss, d.h. dass $\mathcal{T}(z(t_1(\varepsilon)), t_1) = 0$ gelten muss.

³Für $t \geq t^*$ wird $y^*(t)$ linear fortgesetzt.

$\frac{d\tilde{V}}{d\varepsilon}\Big|_{\varepsilon=0} = 0$ gelten.

$$\begin{aligned}
 \frac{d\tilde{V}}{d\varepsilon} &= \frac{d}{d\varepsilon} \int_{t_0}^{t_1(\varepsilon)} I(z(t), \dot{z}(t), t) dt \\
 &\stackrel{4.14}{=} \int_{t_0}^{t_1(\varepsilon)} \left[\frac{\partial I}{\partial y} h + \frac{\partial I}{\partial \dot{y}} \dot{h} \right] dt + I(z(t), \dot{z}(t), t_1(\varepsilon)) \cdot \frac{dt_1}{d\varepsilon} \\
 &\stackrel{\text{part. Int.}}{=} \int_{t_0}^{t_1(\varepsilon)} \left[\frac{\partial I}{\partial y} - \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{y}} \right) \right] h(t) dt + \left[\frac{\partial I}{\partial \dot{y}} h(t) \right]_{t_0}^{t_1(\varepsilon)} + I(z, \dot{z}, t_1(\varepsilon)) \cdot \frac{dt_1}{d\varepsilon} \\
 &\stackrel{h(t_0)=0}{=} \int_{t_0}^{t_1(\varepsilon)} \left[\frac{\partial I}{\partial y} - \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{y}} \right) \right] h(t) dt + \frac{\partial I}{\partial \dot{y}}(z, \dot{z}, t_1(\varepsilon)) h(t_1(\varepsilon)) + I(z, \dot{z}, t_1(\varepsilon)) \cdot \frac{dt_1}{d\varepsilon}
 \end{aligned}$$

Die notwendige Bedingung $\frac{d\tilde{V}}{d\varepsilon}\Big|_{\varepsilon=0} = 0$ liefert dann mit $\Delta t := \frac{dt_1}{d\varepsilon}\Big|_{\varepsilon=0}$

$$\int_{t_0}^{t^*} \left[\frac{\partial I}{\partial y} - \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{y}} \right) \right] h(t) dt + \frac{\partial I}{\partial \dot{y}}(y^*, \dot{y}^*, t^*) \cdot h(t^*) + I(y^*, \dot{y}^*, t^*) \cdot \Delta t = 0$$

Da die zwei letzten Terme auf der linken Seite unabhängig vom Verlauf von h sind, müssen folgende Gleichungen gelten:

$$\int_{t_0}^{t^*} \left[\frac{\partial I}{\partial y} - \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{y}} \right) \right] h(t) dt = 0 \quad (4.4)$$

$$\frac{\partial I}{\partial \dot{y}}(y^*, \dot{y}^*, t^*) h(t^*) + I(y^*, \dot{y}^*, t^*) \Delta t = 0 \quad (4.5)$$

Damit ist wieder die Euler-Lagrange-Gleichung erfüllt:

$$\frac{\partial I}{\partial y} - \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{y}} \right) = 0.$$

Gleichung (4.5) lässt sich weiter vereinfachen, wenn man beachtet, dass eine Kopplung zwischen $h(t)$ und Δt besteht, da die gestörte Bewegung auf $\mathcal{T}$ enden muss. Nach Voraussetzung muss nämlich gelten:

$$\mathcal{T}(z(t_1(\varepsilon)), t_1(\varepsilon)) = 0.$$

Differenzieren dieser Gleichung nach ε liefert wegen $z(t_1(\varepsilon)) = y^*(t_1(\varepsilon)) + \varepsilon \cdot h(t_1(\varepsilon))$:

$$\begin{aligned}
 0 &= \frac{\partial \mathcal{T}}{\partial y} \cdot \frac{dz(t_1(\varepsilon))}{d\varepsilon} + \frac{\partial \mathcal{T}}{\partial t} \cdot \frac{dt_1}{d\varepsilon} \\
 &= \frac{\partial \mathcal{T}}{\partial y} \left(\frac{dy^*}{dt} \cdot \frac{dt_1}{d\varepsilon} + h(t_1) + \varepsilon \cdot \frac{dh}{dt} \cdot \frac{dt_1}{d\varepsilon} \right) + \frac{\partial \mathcal{T}}{\partial t} \cdot \frac{dt_1}{d\varepsilon}
 \end{aligned}$$

Für $\varepsilon = 0$ ergibt sich mit der oben eingeführten Kurzbezeichnung $\Delta t = \frac{dt_1}{d\varepsilon}\Big|_{\varepsilon=0}$:

$$0 = \frac{\partial \mathcal{T}}{\partial y}(y^*, t^*) \cdot \left(\dot{y}^*(t^*) \Delta t + h(t^*) \right) + \frac{\partial \mathcal{T}}{\partial t}(y^*, t^*) \cdot \Delta t$$

Multipliziert man diese Gleichung mit $I(y^*, \dot{y}^*, t)$ und beachtet, dass nach Gleichung (4.5) die Beziehung

$$I(y^*, \dot{y}^*) \Delta t = -\frac{\partial I}{\partial \dot{y}}(y^*, \dot{y}^*) h(t^*),$$

gilt, so erhält man

$$\left[\left(I - \frac{\partial I}{\partial \dot{y}} \cdot \dot{y} \right) \frac{\partial \mathcal{T}}{\partial y} - \frac{\partial \mathcal{T}}{\partial t} \cdot \frac{\partial I}{\partial \dot{y}} \right]_{t=t^*} \cdot h(t^*) = 0$$

Da h beliebig war, folgt die Behauptung

$$\left[\left(I - \frac{\partial I}{\partial \dot{y}} \cdot \dot{y} \right) \frac{\partial \mathcal{T}}{\partial y} - \frac{\partial \mathcal{T}}{\partial t} \cdot \frac{\partial I}{\partial \dot{y}} \right]_{t=t^*} = 0 \quad \blacksquare$$

4.16 Bemerkung

- i) Bedingung (4.3) in Satz 4.15 ersetzt die ursprüngliche Nebenbedingung $\mathbf{y}(t_1) = \mathbf{y}_1$.
- ii) Es ist nützlich - wie schon in Bemerkung 4.6 erwähnt - ein paar spezielle Formen von $\mathcal{T}$ und deren Wirkung auf die zweite Bedingung in Satz 4.15 zu betrachten.

1. Fall (1. Fall in Bemerkung 4.6):

$t_1 = t^*$ liegt fest, $\mathbf{y}(t^*)$ ist frei, d.h. es gilt $\mathcal{T}(\mathbf{y}, t) = t - t^*$. Wegen $\frac{\partial \mathcal{T}}{\partial \mathbf{y}} = \mathbf{0}$, $\frac{\partial \mathcal{T}}{\partial t} = 1$ lautet die zweite Bedingung

$$\left[\frac{\partial I}{\partial \dot{\mathbf{y}}} \right]_{t=t^*} = \mathbf{0}$$

Diese wird häufig als *natürliche Randbedingung* bezeichnet, da sie bei freiem $\mathbf{y}$ die Größe von $y^* = y(t^*)$ festlegt.

2. Fall:

$\mathcal{T}(\mathbf{y}, t) = \mathbf{0}$ kann nach $\mathbf{y}$ aufgelöst werden, d.h. die Transversalitätsbedingung kann in die Form $\mathcal{T}(\mathbf{y}, t) = \mathbf{y} - \mathbf{g}(t)$ gebracht werden. Wegen $\frac{\partial \mathcal{T}}{\partial \mathbf{y}} = \mathbf{1}$ (= Einheitsmatrix) und $\frac{\partial \mathcal{T}}{\partial t} = -\dot{\mathbf{g}}(t)$ erhält man die Bedingung:

$$\left[I + (\dot{\mathbf{g}} - \dot{\mathbf{y}}) \frac{\partial I}{\partial \dot{\mathbf{y}}} \right]_{t=t^*} = \mathbf{0}$$

Ein Sonderfall ergibt sich, wenn t_1 frei und $\mathbf{y}(t_1) := \mathbf{y}_1$ vorgegeben ist (2. Fall in Bemerkung 4.6). In dieser Situation gilt $\mathcal{T}(\mathbf{y}, t) = \mathbf{y} - \mathbf{y}_1$ d.h. $\mathbf{g}(t) = \mathbf{y}_1 = \text{konstant}$. Wegen $\dot{\mathbf{g}} = \mathbf{0}$ lautet die zweite Bedingung

$$\left[I - \frac{\partial I}{\partial \dot{\mathbf{y}}} \dot{\mathbf{y}} \right]_{t=t^*} = \mathbf{0} \quad \square$$

4.17 Beispiel

- i) Man bestimme den Weg y kürzesten Abstandes, den der Punkt (x_0, y_0) von der Kurve $x \mapsto g(x)$ besitzt. Dieses Problem führt auf die Aufgabe

$$\begin{aligned} V(y) &= \int_{x_0}^{x_1} \sqrt{1 + y'^2} dx \rightarrow \min \\ y(x_0) &= y_0, \quad y(x_1) = g(x_1), \quad x_1 \text{ unbekannt, d.h. } \mathcal{T}(y, x) = y - g(x) \end{aligned}$$

Da I nur von y' abhängt, sind die Extremalen Geraden. Es gilt also $y^*(x) = a \cdot (x - x_0) + y_0$. Nach dem 2. Fall von Bemerkung 4.16 bestimmt sich x_1 durch die Transversalitätsbedingung

$$\left[\sqrt{1 + y'^2} + (g' - y') \cdot \frac{y'}{\sqrt{1 + y'^2}} \right]_{x=x_1} = 0.$$

Nach Multiplikation mit $\sqrt{1 + y'^2}$ ergibt sich

$$y^{*'}(x_1) \cdot g'(x_1) = -1.$$

Im optimalen Punkt x_1 steht der optimale Weg y^* (eine Gerade) also senkrecht auf der durch g beschriebenen Kurve.

- ii) Wir betrachten folgendes Problem:

$$\begin{aligned} V(y) &:= \int_0^1 \left(\frac{1}{2} \dot{y}^2 + y\dot{y} + \dot{y} + y \right) dt \rightarrow \text{Opt.} \\ y(0) &= \frac{1}{2}, \quad y(1) \text{ frei (d.h. } \mathcal{T}(y, t) = t - 1) \end{aligned}$$

Die Euler-Lagrange-Gleichung lautet $\ddot{y} = 1$, woraus folgt

$$y^*(t) = \frac{1}{2}t^2 + c_1t + c_2.$$

Aus $y^*(0) = \frac{1}{2}$ folgt

$$c_2 = \frac{1}{2}.$$

Da der 1. Fall von Bemerkung 4.16 vorliegt, muss gelten:

$$0 = \left[\frac{\partial I}{\partial \dot{y}} \right]_{t=1} = \dot{y}^*(1) + y^*(1) + 1 = 2c_1 + 3$$

Dies liefert

$$c_1 = -\frac{3}{2}, \text{ also } y^*(t) = \frac{1}{2}(t^2 - 3t + 1), \quad y^*(1) = -\frac{1}{2}. \quad \square$$

4.1.1. Nebenbedingungen

Die Variationsrechnung gestattet auch die Einbindung von Nebenbedingungen. Wie bei der statischen Optimierung spielt dabei die Methode der Lagrange-Multiplikatoren eine entscheidende Rolle. Wir betrachten drei Typen von Nebenbedingungen:

- 1.) Nebenbedingungen in Gleichungsform
- 2.) Nebenbedingungen in Ungleichungsform
- 3.) Nebenbedingungen in Integralform.

1. Nebenbedingungen in Gleichungsform

Das allgemeine Problem lautet in diesem Fall

$$\begin{aligned} V(\mathbf{y}) &= \int_{t_0}^{t_1} I(\mathbf{y}, \dot{\mathbf{y}}, t) dt \rightarrow \max \\ \text{u.d.N.} \quad \mathbf{y}(t_0) &= \mathbf{y}_0, \quad \mathbf{y}(t_1) = \mathbf{y}_1 \\ \mathbf{g}(\mathbf{y}, \dot{\mathbf{y}}, t) &= \mathbf{0} \end{aligned} \quad (4.6)$$

Hierbei sei $\mathbf{g}$ eine $\mathbb{R}^m$ -wertige Abbildung, d.h. $\mathbf{g}$ ist vektorwertig und besteht aus m Komponentenfunktionen, wobei wir - um Trivialitäten zu vermeiden - fordern, dass die Anzahl n der Zustandsvariablen größer ist als m . Um die lineare Unabhängigkeit des DGL-Systems zu garantieren, verlangen wir außerdem, dass $(\frac{\partial \mathbf{g}}{\partial \dot{\mathbf{y}}})$ maximalen Rang m besitzt. Ist $\mathbf{g}$ von $\dot{\mathbf{y}}$ unabhängig, d.h. gilt $\mathbf{g} = \mathbf{g}(\mathbf{y}, t)$, ist entsprechend $\text{Rg}(\frac{\partial \mathbf{g}}{\partial \dot{\mathbf{y}}}) = m$ zu fordern.

Analog zum statischen Fall bilden wir eine „intermediäre Lagrangefunktion“ $\mathcal{J}$ mit den Lagrangemultiplikatoren $\boldsymbol{\lambda}(t) := (\lambda_1(t), \dots, \lambda_m(t))$, nämlich

$$\mathcal{J}(\mathbf{y}, \dot{\mathbf{y}}, t, \boldsymbol{\lambda}) := I(\mathbf{y}, \dot{\mathbf{y}}, t) - \boldsymbol{\lambda}(t) \cdot \mathbf{g}(\mathbf{y}, \dot{\mathbf{y}}, t)$$

und die neue Zielfunktion:

$$\mathcal{V}(\mathbf{y}, \dot{\mathbf{y}}, \boldsymbol{\lambda}) := \int_{t_0}^{t_1} \mathcal{J}(\mathbf{y}, \dot{\mathbf{y}}, t, \boldsymbol{\lambda}) dt$$

Wie im statischen Fall ergibt sich die Lösung als Sattelpunkt, d.h. durch Maximierung bzgl. $\mathbf{y}$ und Minimierung bzgl. $\boldsymbol{\lambda}$. Es gilt:

Ist $\mathbf{y}^*$ eine Lösung des Optimierungsproblems (4.6), existiert ein Lagrange-Multiplikator $\boldsymbol{\lambda}^*(t)$, so dass die Euler-Lagrange-Gleichungen für $\mathcal{V}$ gelten:

$$\frac{\partial \mathcal{J}}{\partial \mathbf{y}} - \frac{d}{dt} \left(\frac{\partial \mathcal{J}}{\partial \dot{\mathbf{y}}} \right) = \mathbf{0} \quad (4.7)$$

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{\lambda}} - \frac{d}{dt} \left(\frac{\partial \mathcal{J}}{\partial \dot{\boldsymbol{\lambda}}} \right) = \mathbf{0} \quad (4.8)$$

4.18 Bemerkung

- i) 1) $\lambda * g$ wird *nicht* wie im statischen Fall zu V , sondern zu I addiert.
 2) $\lambda = \lambda(t)$ ist zeitabhängig. Dies ist erforderlich, weil die Nebenbedingung zu jedem Zeitpunkt erfüllt werden muss.
- ii) Da $\mathcal{J}$ von $\dot{\lambda}$ unabhängig ist, gilt $\frac{\partial \mathcal{J}}{\partial \dot{\lambda}} = \mathbf{0}$ und man erhält aus (4.8)

$$\frac{\partial \mathcal{J}}{\partial \lambda} - \frac{d}{dt} \left(\frac{\partial \mathcal{J}}{\partial \dot{\lambda}} \right) = \frac{\partial \mathcal{J}}{\partial \lambda} = -g(\mathbf{y}, \dot{\mathbf{y}}, t) = \mathbf{0},$$

d.h. die Euler-Lagrange-Gleichungen bezüglich der Variablen λ liefern analog zum eindimensionalen Fall gerade die Nebenbedingung $g = \mathbf{0}$.

- iii) Gleichung (4.7) liefert n DGLen zweiter Ordnung. Zusammen mit den m Nebenbedingungen $g(\mathbf{y}, \dot{\mathbf{y}}, t) = \mathbf{0}$ können diese dazu verwendet werden, $\mathbf{y}$ und λ zu bestimmen. Die noch fehlenden $2 \cdot n$ Konstanten ergeben sich aus den Anfangsbedingungen $\mathbf{y}(t_0) = \mathbf{y}_0$ und $\mathbf{y}(t_1) = \mathbf{y}_1$. $\square$

4.19 Beispiel

Man bestimme die Kurve kleinster Länge, die zwei vorgegebene Punkte $\mathbf{y}_0, \mathbf{y}_1 \in \mathbb{R}^3$ auf der durch ϕ definierten Fläche $\{\mathbf{x} \in \mathbb{R}^3 \mid \phi(\mathbf{x}) = 0\}$ verbindet.

Bezeichnet $\mathbf{y}(t) := (y_1(t), y_2(t), y_3(t))$ die gesuchte Kurve in parametrisierter Form mit $\mathbf{y}(0) = \mathbf{y}_0$ und $\mathbf{y}(1) = \mathbf{y}_1$ und beachtet man, dass die Länge L der Kurve $\mathbf{y}(t)$ gleich

$$\int_0^1 \sqrt{\dot{y}_1^2 + \dot{y}_2^2 + \dot{y}_3^2} dt = \int_0^1 \|\dot{\mathbf{y}}\| dt$$

ist, so lautet das Problem:

$$\begin{aligned} L &= \int_0^1 \|\dot{\mathbf{y}}\| dt \rightarrow \min \\ \text{u.d.N.} \quad \mathbf{y}(0) &= \mathbf{y}_0, \quad \mathbf{y}(1) = \mathbf{y}_1 \\ \phi(\mathbf{y}) &= 0 \end{aligned}$$

Mit der intermediären Lagrangefunktion $\mathcal{J} = I + \lambda(t) \cdot \phi$ erhält man wegen $\frac{\partial I}{\partial \mathbf{y}_j} = 0$ die Euler-Lagrange-Gleichungen:

$$\begin{aligned} \lambda(t) \cdot \frac{\partial \phi}{\partial y_1} - \frac{d}{dt} \left[\dot{y}_1 (\dot{y}_1^2 + \dot{y}_2^2 + \dot{y}_3^2)^{-\frac{1}{2}} \right] &= 0 \\ \lambda(t) \cdot \frac{\partial \phi}{\partial y_2} - \frac{d}{dt} \left[\dot{y}_2 (\dot{y}_1^2 + \dot{y}_2^2 + \dot{y}_3^2)^{-\frac{1}{2}} \right] &= 0 \\ \lambda(t) \cdot \frac{\partial \phi}{\partial y_3} - \frac{d}{dt} \left[\dot{y}_3 (\dot{y}_1^2 + \dot{y}_2^2 + \dot{y}_3^2)^{-\frac{1}{2}} \right] &= 0 \\ \phi(y_1, y_2, y_3) &= 0 \end{aligned}$$

Hiermit bestimmt man die optimalen Pfade $y_1(t)$, $y_2(t)$, $y_3(t)$ und $\lambda(t)$. Die dann noch fehlenden 6 Konstanten ergeben sich aus den Randbedingungen $\mathbf{y}(0)=\mathbf{y}_0$, $\mathbf{y}(1)=\mathbf{y}_1$. $\square$

2. Nebenbedingungen in Ungleichungsform

Das allgemeine Problem lautet:

$$\begin{aligned} V(\mathbf{y}) &= \int_{t_0}^{t_1} I(\mathbf{y}, \dot{\mathbf{y}}, t) dt \rightarrow \max \\ \mathbf{y}(t_0) &= \mathbf{y}_0, \quad \mathbf{y}(t_1) = \mathbf{y}_1 \\ \mathbf{g}(\mathbf{y}, \dot{\mathbf{y}}, t) &\leq \mathbf{0} \end{aligned} \quad (4.9)$$

Wie im 1. Fall (4.6) ist $\mathbf{g}$ eine $\mathbb{R}^m$ -wertige Funktion, bestehend aus m Komponenten.

Da Ungleichungen wesentlich schwächere Forderungen darstellen als Gleichungen, muss nicht unbedingt $m < n$ gelten. Es ist nur darauf zu achten, dass die Ungleichungen miteinander verträglich sind. Wir bilden wie im ersten Fall die „intermediäre Lagrange-Funktion“

$$\mathcal{J}(\boldsymbol{\lambda}, \mathbf{y}, \dot{\mathbf{y}}, t) := I(\mathbf{y}, \dot{\mathbf{y}}, t) - \boldsymbol{\lambda}(t) \cdot \mathbf{g}(\mathbf{y}, \dot{\mathbf{y}}, t).$$

Als notwendige Bedingung für die gesuchte Lösung muss gelten:

$$\frac{\partial \mathcal{J}}{\partial \mathbf{y}} - \frac{d}{dt} \left(\frac{\partial \mathcal{J}}{\partial \dot{\mathbf{y}}} \right) = \mathbf{0} \quad (4.10)$$

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{\lambda}} = -\mathbf{g}(\mathbf{y}, \dot{\mathbf{y}}, t) \geq \mathbf{0} \quad (4.11)$$

$$\boldsymbol{\lambda}(t) \geq \mathbf{0} \quad (4.12)$$

$$\boldsymbol{\lambda}(t) \cdot \frac{\partial \mathcal{J}}{\partial \boldsymbol{\lambda}} = -\boldsymbol{\lambda}(t) \cdot \mathbf{g}(\mathbf{y}, \dot{\mathbf{y}}, t) = 0 \quad (4.13)$$

Wie im statischen Fall stellen (4.11) - (4.13) die Kuhn-Tucker-Bedingungen dar (siehe Satz F.3 des Anhangs). Es gilt wie üblich der *Satz vom komplementären Schlupf* im Optimum:

Ist die i -te Restriktion streng erfüllt, d.h. gilt $g_i(\mathbf{y}, \dot{\mathbf{y}}, t) < 0$, so folgt $\lambda_i = 0$.

Ist der i -te Lagrange-Multiplikator $\lambda_i > 0$, so ist die i -te Restriktion als Gleichheit erfüllt. Hierdurch wird garantiert, dass im Optimum

$$\mathcal{J} = I - \boldsymbol{\lambda} \cdot \mathbf{g}(t) = I$$

gilt, so dass der optimale Wert von $\mathcal{V} = \int \mathcal{J}(\mathbf{y}, \dot{\mathbf{y}}, t) dt$ mit dem optimalen Wert von V übereinstimmt.

3. Nebenbedingungen in Integralform

Dies ist einer der wichtigsten Nebenbedingungstypen. Eine der frühesten Aufgaben, in denen solche Integralbedingungen auftauchten, war das *isoperimetrische Problem*. Hierbei ist eine geometrische Figur maximaler Fläche zu bestimmen, so dass die einhüllende Kurve eine vorgegebene Länge besitzt (siehe Beispiel 4.20). Die allgemeine Form lautet:

$$\begin{aligned} V(\mathbf{y}) &= \int_{t_0}^{t_1} I(\mathbf{y}, \dot{\mathbf{y}}, t) dt \rightarrow \max \\ \text{u.d.N} \quad &\mathbf{y}(t_0) = \mathbf{y}_0, \quad \mathbf{y}(t_1) = \mathbf{y}_1 \\ &\int_{t_0}^{t_1} \mathbf{g}(\mathbf{y}, \dot{\mathbf{y}}, t) dt = \mathbf{k} \end{aligned}$$

$\mathbf{g}$ ist hierbei eine stetig differenzierbare $\mathbb{R}^m$ -wertige Abbildung und $\mathbf{k}$ ein konstanter vorgegebener Vektor. Wie bei Ungleichungen wird der *Freiheitsgrad* des Problems durch die Integralnebenbedingung nicht weiter eingeschränkt, so dass auch $m \geq n$ zulässig ist.

Bildet man mit den (konstanten) Lagrangemultiplikatoren $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)$ die „intermediäre Lagrangefunktion“

$$\mathcal{J}(\mathbf{y}, \dot{\mathbf{y}}, \boldsymbol{\lambda}, t) := I(\mathbf{y}, \dot{\mathbf{y}}, t) - \boldsymbol{\lambda} * \mathbf{g}(\mathbf{y}, \dot{\mathbf{y}}, t),$$

so lautet die notwendige Euler-Lagrange-Gleichung:

$$\frac{\partial \mathcal{J}}{\partial \mathbf{y}} - \frac{d}{dt} \left(\frac{\partial \mathcal{J}}{\partial \dot{\mathbf{y}}} \right) = \mathbf{0}$$

Heuristisch lässt sich die Konstanz der Lagrangemultiplikatoren dadurch erklären, dass die Nebenbedingung in Integralform vorliegt, also schon die gesamte Wegabhängigkeit erfasst. Wir wollen dies für den Fall $m = n = 1$ auch direkt mit Hilfe des ersten Falles von Nebenbedingungen in Gleichungsform beweisen. Wir betrachten also das Problem

$$\begin{aligned} V(y) &= \int_{t_0}^{t_1} I(y, \dot{y}, t) dt \rightarrow \max \\ \text{u.d.N} \quad &y(t_0) = y_0, \quad y(t_1) = y_1 \\ &\int_{t_0}^{t_1} g(y, \dot{y}, t) dt = k \end{aligned}$$

Wir definieren die Funktion $\gamma(t) := \int_{t_0}^t g(y(s), \dot{y}(s), s) ds$, für die offensichtlich gilt

$$\gamma(t_0) = 0, \quad \gamma(t_1) = k, \quad \dot{\gamma}(t) = g(y, \dot{y}, t).$$

Führt man formal γ als zusätzliche Variable ein, so lautet das ursprüngliche Problem:

$$\begin{aligned} V(y, \gamma) = V(y) &= \int_{t_0}^{t_1} I(y, \dot{y}, t) dt \rightarrow \max \\ \text{u.d.N.} \quad &\begin{pmatrix} y \\ \gamma \end{pmatrix}(t_0) = \begin{pmatrix} y_0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} y \\ \gamma \end{pmatrix}(t_1) = \begin{pmatrix} y_1 \\ k \end{pmatrix} \\ &-\dot{\gamma}(t) + g(y, \dot{y}, t) = 0 \end{aligned}$$

Es liegt damit ein Optimierungsproblem mit einer Nebenbedingung in Gleichungsform wie im 1. Fall vor. Mit der intermediären Lagrangefunktion

$$\mathcal{J} := I - \lambda(t) \cdot (g - \dot{\gamma})$$

lautet die Euler-Lagrange-Gleichung:

$$\begin{aligned} 0 &= \frac{\partial \mathcal{J}}{\partial y} - \frac{d}{dt} \left(\frac{\partial \mathcal{J}}{\partial \dot{y}} \right) \\ 0 &= \frac{\partial \mathcal{J}}{\partial \gamma} - \frac{d}{dt} \left(\frac{\partial \mathcal{J}}{\partial \dot{\gamma}} \right) = 0 - \frac{d}{dt} \lambda(t) \\ 0 &= \frac{\partial \mathcal{J}}{\partial \lambda} - \frac{d}{dt} \left(\frac{\partial \mathcal{J}}{\partial \dot{\lambda}} \right) = 0 - \frac{d}{dt} \lambda(t) = \frac{\partial \mathcal{J}}{\partial \lambda} = g(y, \dot{y}, t) - \dot{\gamma}(t) \end{aligned}$$

Die zweite Bedingung liefert daher gerade $\lambda(t) = \text{konst.}$

4.20 Beispiel

- i) Man bestimme eine Kurve vorgegebener Länge ℓ in der $(t-y)$ -Ebene⁴ zwischen zwei vorgegebenen Punkten (t_0, y_0) und (t_1, y_1) , so dass die unter der Kurve gelegene Fläche maximal ist. Die mathematische Forderung lautet:

$$\begin{aligned} V(y) &= \int_{t_0}^{t_1} y(t) dt \rightarrow \max \\ \text{u.d.N.} \quad &y(t_0) = y_0, \quad y(t_1) = y_1 \\ &\int_{t_0}^{t_1} (1 + \dot{y}^2)^{\frac{1}{2}} dt = \ell \end{aligned}$$

Da $\mathcal{J} = y - \lambda(1 + \dot{y}^2)^{\frac{1}{2}}$ unabhängig von t ist, erhält man nach Bemerkung 4.11 iv) die notwendige Forderung:

$$\mathcal{J} - \dot{y} \cdot \frac{\partial \mathcal{J}}{\partial \dot{y}} = c_1 = \text{konstant} \iff y - c_1 = \frac{\lambda}{(1 + \dot{y}^2)^{\frac{1}{2}}}$$

⁴Damit das Problem sinnvoll ist, sollte ℓ größer als die Länge der Gerade zwischen den beiden Punkten sein.

Auflösen nach $\dot{y}$ liefert:

$$\dot{y} = \frac{\sqrt{\lambda^2 - (y-c_1)^2}}{y-c_1}$$

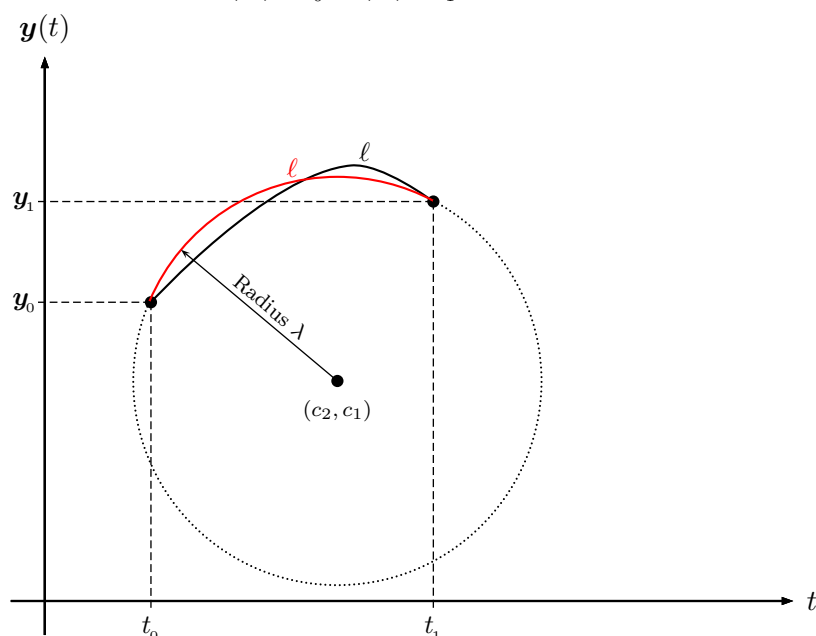
Diese DGL ist separabel und man erhält

$$\int \frac{y-c_1}{\sqrt{\lambda^2 - (y-c_1)^2}} dy = \int dt - c_2 \iff -\sqrt{\lambda^2 - (y-c_1)^2} = t - c_2,$$

Was äquivalent ist zu

$$(y-c_1)^2 + (t-c_2)^2 = \lambda^2.$$

Dies ist ein Kreis mit Mittelpunkt (c_2, c_1) und Radius λ . c_2, c_1 und λ ergeben sich aus den Anfangsbedingungen $y(t_0)=y_0$, $y(t_1)=y_1$ und der Integralbedingung.



- ii) Gegeben sei eine Resource mit endlichem Bestand S_0 zum Zeitpunkt t_0 . Bezeichnet $q(t)$ die zum Zeitpunkt t abgebaute Menge der Resource und $N(q)$ den zugehörigen Wert (Nutzen), so lautet das entstehende Problem:

$$V(q) = \int_{t_0}^{t_1} N(q) e^{-\rho t} dt \rightarrow \max$$

u.d.N $\int_{t_0}^{t_1} q(t) dt = S_0$

□

4.21 Bemerkung

Bei isoperimetrischen Problemen gilt das *Prinzip der Reziprozität*. Es besagt:

Ist $y(t)$ ein Maximum von V u.d.N. $W = \text{konst}$, so
 ist $y(t)$ ein Minimum von W u.d.N. $V = \text{konst}$.

So ist im letzten Beispiel 4.21 i) die Lösungskurve, die bei gegebener Länge die darunter liegende Fläche maximiert auch die Kurve, die bei gegebener Fläche die Kurvenlänge minimiert, da man den gleichen Typ von DGL für die Extremalen erhält. $\square$

4.1.2. Bedingungen zweiter Ordnung

Ähnlich wie im statischen Fall bietet die zweite Variation von V die Möglichkeit, über den Typ des vorliegenden Extremums zu entscheiden.

Mit den Bezeichnungen von Satz 4.10 muss nämlich im Fall eines lokalen Minimums

$$\partial^2 V = \left. \frac{d^2 \tilde{V}}{d\varepsilon^2} \right|_{\varepsilon=0} \geq 0$$

und im Fall eines lokalen Maximums

$$\partial^2 V = \left. \frac{d^2 \tilde{V}}{d\varepsilon^2} \right|_{\varepsilon=0} \leq 0$$

gelten. Entsprechend erhält man auch eine hinreichende Bedingung für ein lokales Minimum bzw. Maximum durch

$$\partial^2 V > 0 \quad \text{bzw.} \quad \partial^2 V < 0$$

Diese Aussagen sind jedoch erst dann nützlich, wenn man es - wie bei der 1. Variation - schafft, die explizite Abhängigkeit des Terms $\partial^2 V$ von der Störfunktion h zu beseitigen. Bei der notwendigen Bedingung 1. Ordnung $\partial V = 0$ wurde hierzu das Fundamentallemma der Variationsrechnung verwendet. Wir versuchen dies mit der zweiten Variation ebenfalls und berechnen zunächst $\partial^2 V$. Mit $z = y + \varepsilon h$ gilt

$$\begin{aligned} \frac{d}{d\varepsilon} \left[\frac{\partial I}{\partial y}(z, \dot{z}, t) \right]_{\varepsilon=0} &= \frac{\partial^2 I}{\partial y^2} \cdot h + \frac{\partial^2 I}{\partial y \partial \dot{y}} \cdot \dot{h} \quad \text{und} \\ \frac{d}{d\varepsilon} \left[\frac{\partial I}{\partial \dot{y}}(z, \dot{z}, t) \right]_{\varepsilon=0} &= \frac{\partial^2 I}{\partial y \partial \dot{y}} \cdot h + \frac{\partial^2 I}{\partial \dot{y}^2} \cdot \dot{h} \end{aligned}$$

und daher

$$\begin{aligned} \left. \frac{d^2 \tilde{V}}{d\varepsilon^2} \right|_{\varepsilon=0} &= \left. \frac{d}{d\varepsilon} \int_{t_0}^{t_1} \left[\frac{\partial I}{\partial y} \cdot h + \frac{\partial I}{\partial \dot{y}} \cdot \dot{h} \right] dt \right|_{\varepsilon=0} \\ &= \int_{t_0}^{t_1} \left[\frac{\partial^2 I}{\partial y^2} \cdot h^2(t) + 2 \frac{\partial^2 I}{\partial y \partial \dot{y}} \cdot \dot{h}(t)h(t) + \frac{\partial^2 I}{\partial \dot{y}^2} \cdot \dot{h}^2(t) \right] dt \end{aligned}$$

Integriert man den zweiten Term unter Beachtung von $2 \cdot \dot{h} \cdot h = \frac{d}{dt} [h^2]$ partiell, folgt:

$$\left. \frac{d^2 \tilde{V}}{d\varepsilon^2} \right|_{\varepsilon=0} = \int_{t_0}^{t_1} \left[\left\{ \frac{\partial^2 I}{\partial y^2} - \frac{d}{dt} \left(\frac{\partial^2 I}{\partial y \partial \dot{y}} \right) \right\} h^2(t) + \frac{\partial^2 I}{\partial \dot{y}^2} \cdot \dot{h}^2(t) \right] dt$$

Hier zeigt sich, dass eine Anwendung des Fundamentallemmas der Variationsrechnung prinzipiell nicht möglich ist, da die beiden „störenden“ Terme h^2 und $\dot{h}^2$ nicht unabhängig voneinander sind. Es lässt sich jedoch zeigen, dass aufgrund der Eigenschaften der Ableitung einer Funktion der zweite Term unter dem Integral dominiert, woraus folgt:

$$\partial^2 V = \left. \frac{d^2 \tilde{V}}{d\varepsilon^2} \right|_{\varepsilon=0} \geq 0 \implies \frac{\partial^2 I}{\partial \dot{y}^2} \geq 0$$

$$\partial^2 V = \left. \frac{d^2 \tilde{V}}{d\varepsilon^2} \right|_{\varepsilon=0} \leq 0 \implies \frac{\partial^2 I}{\partial \dot{y}^2} \leq 0$$

Unglücklicherweise kann bei dieser Argumentation jedoch aus $\frac{\partial^2 I}{\partial \dot{y}^2} > 0$ nicht geschlossen werden, dass $\partial^2 V > 0$ gilt. Hier geht also die Analogie zum statischen Fall verloren. Man erhält damit folgende Aussage:

4.22 Satz (Legendre-Bedingung)

Besitzt das klassische Variationsproblem ein lokales Maximum bzw. Minimum in $\mathbf{y}^*$, so gilt neben der Euler-Lagrange-Gleichung auch die *Legendre-Bedingung*

$$\frac{\partial^2 I}{\partial \dot{\mathbf{y}}^2}(\mathbf{y}^*, \dot{\mathbf{y}}^*, t) \leq 0 \quad \text{bzw.} \quad \frac{\partial^2 I}{\partial \dot{\mathbf{y}}^2}(\mathbf{y}^*, \dot{\mathbf{y}}^*, t) \geq 0.$$

□

Im folgenden Beispiel soll diese weitere notwendige Bedingung u.a. für die in 4.1 vorgestellten Probleme überprüft werden. Wegen der bisher verwendeten Schreibweise werden wir die Variable x mit t und entsprechend y' mit $\dot{y}$ bezeichnen.

4.23 Beispiel

i) Beispiel 4.1 i):

$$I(y, \dot{y}, t) = \sqrt{1 + \dot{y}^2} \implies \frac{\partial^2 I}{\partial \dot{y}^2} = (1 + \dot{y}^2)^{-\frac{3}{2}} > 0$$

Damit ist die notwendige Legendre-Bedingung für ein lokales Minimum erfüllt. Insbesondere ist ein lokales Maximum ausgeschlossen.

ii) Beispiel 4.1 ii):

$$I(y, \dot{y}, t) = \frac{\sqrt{1 + \dot{y}^2}}{\sqrt{y}} \implies \frac{\partial^2 I}{\partial \dot{y}^2} = \frac{1}{\sqrt{y}} \cdot \frac{\partial^2}{\partial \dot{y}^2} (\sqrt{1 + \dot{y}^2}) = \frac{(1 + \dot{y}^2)^{-\frac{3}{2}}}{\sqrt{y}}$$

Da $y > 0$ vorausgesetzt ist, gilt $\frac{\partial^2 I}{\partial \dot{y}^2} > 0$ und die notwendige Bedingung für ein lokales Minimum ist erfüllt.

iii) Beispiel 4.1 iii):

$$I(y, \dot{y}, t) = y\sqrt{1 + \dot{y}^2} \implies \frac{\partial^2 I}{\partial \dot{y}^2} = y \cdot (1 + \dot{y}^2)^{-\frac{3}{2}}$$

Daher gilt $\operatorname{sgn}\left(\frac{\partial^2 I}{\partial \dot{y}^2}\right) = \operatorname{sgn}(y)$. Die Extremalen des Problems lauten (wie gesehen)

$$y^*(t) = \frac{c}{2} \left[e^{\frac{t+k}{c}} + e^{-\frac{t+k}{c}} \right],$$

so dass

$$\operatorname{sgn}\left(\frac{\partial^2 I}{\partial \dot{y}^2}(y^*, \dot{y}^*, t)\right) = \operatorname{sgn}(y^*) = \operatorname{sgn}(c)$$

gilt. Damit hat $\frac{\partial^2 I}{\partial \dot{y}^2}(y^*, \dot{y}^*, t)$ für das gesamte betrachtete „Zeitintervall“ $[a, b]$ ein einheitliches Vorzeichen. Da y außerdem die Höhe der Kurve angibt und $y(a) > 0$ gilt, ist $c > 0$. Daher ist die Legendre-Bedingung für ein lokales Minimum erfüllt.

iv) Beispiel 4.1 iv):

$$\begin{aligned} I(p, \dot{p}, t) = & -b(1 + \alpha b)p^2 + (a + 2\alpha ab + \beta b)p - \alpha c^2 \dot{p}^2 \\ & - c(2\alpha a + \beta)\dot{p} + c(1 + 2\alpha b)p\dot{p} - (\alpha a^2 + \beta + \gamma) \end{aligned}$$

$$\implies \frac{\partial^2 I}{\partial \dot{p}^2} = -\alpha c^2 < 0, \text{ wegen } \alpha > 0.$$

Die Legendre-Bedingung für ein lokales Maximum ist erfüllt.

v) Das folgende Beispiel beschreibt einen Fall, in dem die Legendre-Bedingung nicht erfüllt ist.

$$\begin{aligned} T(y) &= \int_0^1 (\dot{y}^2 - 13\sqrt{1 + \dot{y}^2}) dt \rightarrow \min \\ \text{u.d.N.} \quad & y(0) = 0, \quad y(1) = 1 \end{aligned}$$

Da I nur von $\dot{y}$ abhängt, sind die Extremalen Geraden. Aufgrund der Randbedingungen ergibt sich $y^*(t) = t$. Wegen

$$\frac{\partial^2 I}{\partial \dot{y}^2} = 2 - 13(1 + \dot{y}^2)^{-\frac{3}{2}} = 2 - \frac{13}{\sqrt{8}} < 0$$

liegt in y^* kein (schwaches) relatives Minimum vor. □

Wie vorher beschrieben eignet sich die zweite Variation nicht, um hinreichende Bedingungen für lokale Extremwerte zu geben. Dahingegen lässt sich die klassische Aussage aus der statischen Optimierung, dass jede stationäre Stelle (Nullstelle der ersten Ableitung), einer auf dem gesamten Definitionsbereich konkaven (bzw. konvexen) Funktion, schon ein globales Maximum (bzw. Minimum) ist, auf die dynamische Optimierung übertragen.

4.24 Satz

Für das klassische Variationsproblem wie in Satz 4.10 (mit festem Anfangs- und Endzustand und festem Anfangs- und Endzeitpunkt) sei die intermediäre Funktion I als Funktion der Variablen $\mathbf{y}$ und $\dot{\mathbf{y}}$ konkav (bzw. konvex). Dann ist jede Lösung der Euler-Lagrange-Gleichung ein absolutes Maximum (bzw. Minimum) des Variationsproblems. $\square$

Beweis:

Wir verwenden folgende Eigenschaft differenzierbarer Funktionen $f : \mathbb{R}^N \rightarrow \mathbb{R}$ (Satz 7.41 Schindler, Mathematik für Ökonomen):

$$f \text{ ist konkav} \iff \forall \mathbf{x}, \mathbf{x}^* \in \mathbb{R}^N : f(\mathbf{x}) - f(\mathbf{x}^*) \leq f'(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*)$$

Sei $\mathbf{y}^*$ eine Lösung der Euler-Lagrange-Gleichung

$$\frac{\partial I}{\partial \mathbf{y}} - \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{\mathbf{y}}} \right) = \mathbf{0},$$

mit der in $\mathbf{y}, \dot{\mathbf{y}}$ konkaven intermediären Funktion I . Sind nun $\mathbf{y}, \mathbf{z}$ beliebige zulässige Pfade. (d.h. mit $\mathbf{y}(t_0) = \mathbf{z}(t_0) = \mathbf{y}_0$ und $\mathbf{y}(t_1) = \mathbf{z}(t_1) = \mathbf{y}_1$), so gilt:

$$\begin{aligned} I(\mathbf{y}, \dot{\mathbf{y}}, t) - I(\mathbf{z}, \dot{\mathbf{z}}, t) &\leq \left(\frac{\partial I}{\partial \mathbf{y}}(\mathbf{z}), \frac{\partial I}{\partial \dot{\mathbf{y}}}(\mathbf{z}) \right) \cdot \begin{pmatrix} \mathbf{y} - \mathbf{z} \\ \dot{\mathbf{y}} - \dot{\mathbf{z}} \end{pmatrix} \\ &= \frac{\partial I}{\partial \mathbf{y}}(\mathbf{z}) \cdot (\mathbf{y} - \mathbf{z}) + \frac{\partial I}{\partial \dot{\mathbf{y}}}(\mathbf{z}) \cdot (\dot{\mathbf{y}} - \dot{\mathbf{z}}) \end{aligned}$$

Integriert man auf beiden Seiten nach t , folgt

$$\begin{aligned} V(\mathbf{y}) - V(\mathbf{z}) &\leq \int_{t_0}^{t_1} \left[\frac{\partial I}{\partial \mathbf{y}}(\mathbf{z}) \cdot (\mathbf{y} - \mathbf{z}) + \frac{\partial I}{\partial \dot{\mathbf{y}}}(\mathbf{z}) \cdot (\dot{\mathbf{y}} - \dot{\mathbf{z}}) \right] dt \\ &\stackrel{\text{part. Int.}}{=} \int_{t_0}^{t_1} \underbrace{\left[\frac{\partial I}{\partial \mathbf{y}}(\mathbf{z}) - \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{\mathbf{y}}}(\mathbf{z}) \right) \right]}_{= \mathbf{0} \text{ für } \mathbf{z}=\mathbf{y}^*} \cdot (\mathbf{y} - \mathbf{z}) dt, \end{aligned}$$

Für $\mathbf{z} = \mathbf{y}^*$ ist die rechte Seite 0 und es folgt $V(\mathbf{y}) \leq V(\mathbf{y}^*)$ für alle zulässigen $\mathbf{y}$. $\blacksquare$

4.25 Beispiel

i) Beispiel 4.1 i):

$$I(y, \dot{y}, t) = \sqrt{1 + \dot{y}^2}.$$

I enthält nur die Variable $\dot{y}$. Wegen $\frac{\partial^2 I}{\partial \dot{y}^2} = (1 + \dot{y}^2)^{-\frac{3}{2}} > 0$ ist daher I als Funktion der Variablen y und $\dot{y}$ konvex. Die gefundene Lösung ist damit das absolute Minimum.

ii) Beispiel 4.1 ii):

$$I(y, \dot{y}, t) = \frac{\sqrt{1 + \dot{y}^2}}{\sqrt{y}}.$$

Wir untersuchen die Definitheitseigenschaften der zweiten Ableitung (bzgl. $y, \dot{y}$):

$$I''(y, \dot{y}) = \begin{pmatrix} \frac{\partial^2 I}{\partial y^2} & \frac{\partial^2 I}{\partial y \partial \dot{y}} \\ \frac{\partial^2 I}{\partial \dot{y} \partial y} & \frac{\partial^2 I}{\partial \dot{y}^2} \end{pmatrix} = \begin{pmatrix} \frac{3}{4} y^{-\frac{5}{2}} (1 + \dot{y}^2)^{\frac{1}{2}} & -\frac{1}{2} y^{-\frac{3}{2}} \dot{y} (1 + \dot{y}^2)^{-\frac{1}{2}} \\ -\frac{1}{2} y^{-\frac{3}{2}} \dot{y} (1 + \dot{y}^2)^{-\frac{1}{2}} & (1 + \dot{y}^2)^{-\frac{3}{2}} y^{-\frac{1}{2}} \end{pmatrix}$$

$$\implies \det I''(y, \dot{y}) = \frac{3 - \dot{y}^2}{4y^3(1 + \dot{y}^2)}$$

Da diese Determinante negative Werte annehmen kann, ist $D_{y\dot{y}}^2$ nicht überall positiv definit, so dass I nicht global konvex ist. Damit ist die hinreichende Bedingung für ein absolutes Minimum nicht erfüllt. $\square$

4.2. Dynamische Programmierung

Die Methode der dynamischen Programmierung kann direkt auf das allgemeine Kontrollproblem mit festem Anfangs- und Endzustand angewendet werden (siehe Definition 4.3):

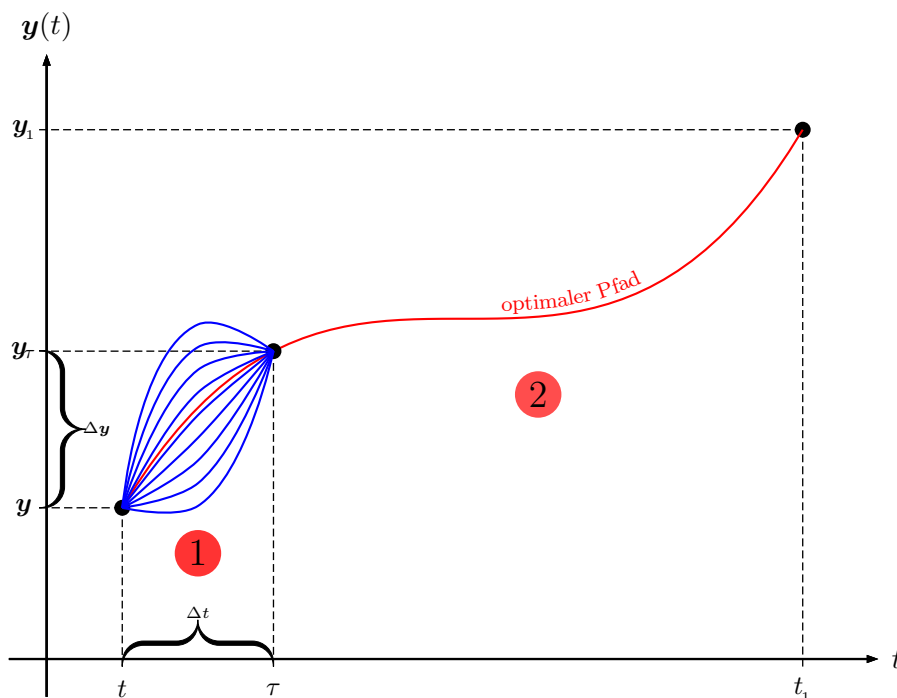
$$\begin{aligned} V(\mathbf{u}) &= \int_{t_0}^{t_1} I(\mathbf{y}, \mathbf{u}, t) dt + F(\mathbf{y}_1, t_1) \rightarrow \max \\ \dot{\mathbf{y}} &= \mathbf{f}(\mathbf{y}, \mathbf{u}, t) \\ \mathbf{y}(t_0) &= \mathbf{y}_0, \quad \mathbf{y}(t_1) = \mathbf{y}_1 \\ \mathbf{u} &\in \Omega \end{aligned}$$

In der dynamischen Programmierung bettet man das Kontrollproblem zunächst in eine größere Problemklasse ein, indem man bei vorgegebenem Endzustand $(\mathbf{y}_1, t_1)$ einen beliebigen freien Anfangszustand $(\mathbf{y}, t)$ zulässt, statt nur einen Anfangszustand $(\mathbf{y}_0, t_0)$ zu betrachten. Durch diesen Ansatz hängt der optimale Zielfunktionswert V^* vom jeweiligen Anfangszustand ab, ist also eine Funktion $V^*(\mathbf{y}, t)$ des Startzustands $(\mathbf{y}, t)$ ⁵, während bei der ursprünglichen Problemstellung V^* eine feste reelle Zahl liefert. Speziell für $(\mathbf{y}, t) := (\mathbf{y}_0, t_0)$ erhält man die Lösung des Ausgangsproblems.

⁵Beachte, dass V^* eine Funktion von $\mathbb{R}^{n+1}$ nach $\mathbb{R}$ ist, die implizit, jedoch nicht explizit von $\mathbf{u}$ abhängt.

Mit Hilfe des sog. *Optimalitätsprinzips* leitet man aus dieser allgemeineren Problemstellung eine fundamentale Rekursionsformel her, die zu einer partiellen (!) DGL für $V^*(\mathbf{y}, t)$ führt (*Bellman-Gleichung*), die das Änderungsverhalten von $V^*(\mathbf{y}, t)$ in Abhängigkeit seiner Variablen t und $\mathbf{y}$ beschreibt.

Das folgende Bild zeigt den optimalen Pfad des Kontrollproblems mit Startzustand $(\mathbf{y}, t)$ und Zielzustand $(\mathbf{y}_1, t_1)$. Das Optimalitätsprinzip besagt, dass jeder Teil des optimalen Pfades für sich betrachtet optimal sein muss.⁶ Zerlegt man diesen Pfad zu einem beliebigen Zeitpunkt τ in zwei Teile, so stellt das Optimalitätsprinzip fest, dass Teil (2) der Kurve eine optimale Lösung des gleichen Kontrollproblems jedoch mit Startwert $(\mathbf{y}_\tau, \tau)$ sein muss, unabhängig davon, wie das System in den Zustand $\mathbf{y}_\tau$ gelangt ist.



Zur Lösung dieses verallgemeinerten Optimierungsproblems, d.h. zur Bestimmung von $V^*(\mathbf{y}, t)$ beschreiben wir das Änderungsverhalten von $V^*(\mathbf{y}, t)$. Hierzu betrachten wir die infinitesimale Änderung $(\mathbf{y}_\tau, \tau) := (\mathbf{y} + \Delta \mathbf{y}, t + \Delta t)$ des Anfangszustandes $(\mathbf{y}, t)$.

Der Unterschied zwischen $V^*(\mathbf{y}, t)$ und $V^*(\mathbf{y} + \Delta \mathbf{y}, t + \Delta t)$ wird ausschließlich durch die intermediäre Funktion I unter dem Integral hervorgerufen, da F nur von $(\mathbf{y}_1, t_1)$ abhängt und damit von einer Änderung im Startwert nicht beeinflusst wird. Der Übergang vom Zustand $(\mathbf{y}, t)$ zum Zustand $(\mathbf{y}_\tau, \tau)$ muss so erfolgen, dass die intermediäre Funktion I im Zeitraum Δt für V einen größtmöglichen Beitrag leistet. Der Anteil $I(\mathbf{y}, \mathbf{u}, t)\Delta t$, den I im Zeitraum Δt zum Wert von V beiträgt, hängt davon ab, welcher Weg gewählt wird, um von $\mathbf{y}$ nach $\mathbf{y} + \Delta \mathbf{y}$ zu gelangen. Die genaue Wegauswahl erfolgt mit Hilfe der Steuergröße $\mathbf{u}$ über die Bewegungsgleichung $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}, \mathbf{u}, t)$. Da der maximale Wert von V beim Startpunkt

⁶Eine auf Aris zurückgehende Beschreibung des Optimalitätsprinzips lautet: *Wenn man mit dem Wenigen, was man gerade hat, nicht das Beste tut, hätte man auch mit dem, was man hätte haben können, nicht das Beste getan.*

$(\mathbf{y}+\Delta\mathbf{y}, t+\Delta t)$ bekannt ist - nämlich $V^*(\mathbf{y}+\Delta\mathbf{y}, t+\Delta t)$ - ist zur Berechnung des Maximums $V^*(\mathbf{y}, t)$ die Steuervariable $\mathbf{u}$ so zu wählen, dass für den zugehörigen Weg von $\mathbf{y}$ nach $\mathbf{y}+\Delta\mathbf{y}$ der Beitrag $I(\mathbf{y}, \mathbf{u}, t)\Delta t$ maximal (bei einem Min-Problem möglichst klein) wird. Man erhält damit die fundamentale Rekursionsformel⁷:

$$V^*(\mathbf{y}, t) = \max_{\mathbf{u}} \left[I(\mathbf{y}, \mathbf{u}, t)\Delta t + V^*(\mathbf{y}+\Delta\mathbf{y}, t+\Delta t) \right] \quad (4.14)$$

Nimmt man nun an, dass $V^*(\mathbf{y}, t)$ in den Größen $\mathbf{y}$ und t stetig differenzierbar ist, so kann $V^*(\mathbf{y}+\Delta\mathbf{y}, t+\Delta t)$ im Punkt $(\mathbf{y}, t)$ mit Hilfe des Satzes von Taylor entwickelt werden:

$$V^*(\mathbf{y}+\Delta\mathbf{y}, t+\Delta t) = V^*(\mathbf{y}, t) + \frac{\partial V^*}{\partial \mathbf{y}}(\mathbf{y}, t) \cdot \Delta\mathbf{y} + \frac{\partial V^*}{\partial t}(\mathbf{y}, t) \cdot \Delta t + \dots$$

Setzt man dies in (4.14) ein, kürzt $V^*(\mathbf{y}, t)$ und dividiert durch Δt , so erhält man

$$0 = \max_{\mathbf{u}} \left[I(\mathbf{y}, \mathbf{u}, t) + \frac{\partial V^*}{\partial \mathbf{y}}(\mathbf{y}, t) \cdot \frac{\Delta\mathbf{y}}{\Delta t} + \frac{\partial V^*}{\partial t}(\mathbf{y}, t) + \dots \text{hot} \right]$$

Bildet man den Grenzwert $\Delta t \rightarrow 0$ und beachtet

$$\lim_{\Delta t \rightarrow 0} \frac{\Delta\mathbf{y}}{\Delta t} = \dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}, \mathbf{u}, t),$$

so erhält man, da V^* nicht explizit von $\mathbf{u}$ abhängt, folgende Aussage.

4.26 Satz

$V^*(\mathbf{y}, t)$ genügt der *Gleichung von Bellman*

$$-\frac{\partial V^*}{\partial t} = \max_{\mathbf{u}} \left[I(\mathbf{y}, \mathbf{u}, t) + \frac{\partial V^*}{\partial \mathbf{y}}(\mathbf{y}, t) \cdot \mathbf{f}(\mathbf{y}, \mathbf{u}, t) \right]$$

mit der Randbedingung

$$V^*(\mathbf{y}(t_1), t_1) = F(\mathbf{y}(t_1), t_1).$$

□

4.27 Bemerkung

- i) Die Randbedingung in Satz 4.26 besagt: Startet man bei der dynamischen Programmierung erst im Endzustand $(\mathbf{y}_1, t_1)$, so ist der optimale Wert von V^* gerade gleich dem Wert der finalen Funktion $F(\mathbf{y}_1, t_1)$.
- ii) Die Bellman-Gleichung ist eine partielle DGL, deren Lösung die Funktion V^* liefert. Wertet man diese an der Stelle $(\mathbf{y}_0, t_0)$ aus, so erhält man als Spezialfall den uns eigentlich interessierenden Wert $V^*(\mathbf{y}_0, t_0)$. □

⁷ $V^*(\mathbf{y}+\Delta\mathbf{y}, t+\Delta t)$ kann nicht aus dem Maximum gezogen werden, da $\Delta\mathbf{y}=\mathbf{f}(\mathbf{y}, \mathbf{u}, t)\Delta t$ explizit von $\mathbf{u}$ abhängt

4.28 Beispiel

Wir betrachten folgendes Problem:

$$\begin{aligned} V(u) &= \int_{t_0}^{t_1} (y - u) dt \rightarrow \max \\ \dot{y} &= \sqrt{u} \\ y(t_0) &= y_0, \quad y(t_1) = y_1, \quad u > 0 \text{ beliebig} \end{aligned}$$

Die Bellman-Gleichung lautet:

$$-\frac{\partial V^*}{\partial t} = \max_u \underbrace{\left[(y - u) + \frac{\partial V^*}{\partial y} \cdot \sqrt{u} \right]}_{T(u)}$$

Die notwendige Bedingung $\frac{\partial T}{\partial u} = 0$ für das Maximum von $T = T(u)$ liefert

$$\sqrt{u^*} = \frac{1}{2} \cdot \frac{\partial V^*}{\partial y} \iff u^* = \frac{1}{4} \left(\frac{\partial V^*}{\partial y} \right)^2.$$

Setzt man dieses u in die Bellman-Gleichung ein, erhält man die partielle DGL:

$$-\frac{\partial V^*}{\partial t} = y + \frac{1}{4} \left(\frac{\partial V^*}{\partial y} \right)^2$$

□

4.29 Bemerkung

- i) Da der dynamischen Programmierung ein allgemeinerer Ansatz zugrunde liegt als der klassischen Variationsrechnung, ist es möglich die Ergebnisse der Variationsrechnung aus der Bellman-Gleichung herzuleiten. Beispielhaft soll dies für die Euler-Lagrange-Gleichung und die Legendre-Bedingung geschehen.

Das klassische Variationsproblem ist der Spezialfall des dynamischen Programmierproblems, in dem gilt:

$$\dot{\mathbf{y}} = \mathbf{u}, \quad \Omega = \mathbb{R}^n$$

Die Bellman-Gleichung lautet in diesem Fall

$$-\frac{\partial V^*}{\partial t} = \max_{\dot{\mathbf{y}}} \left[I(\mathbf{y}, \dot{\mathbf{y}}, t) + \frac{\partial V^*}{\partial \mathbf{y}} \cdot \dot{\mathbf{y}} \right].$$

Existiert das Maximum, ergeben sich als notwendige Bedingungen:

$$\begin{aligned} (1) \quad & \frac{\partial}{\partial \dot{\mathbf{y}}} \left[I(\mathbf{y}, \dot{\mathbf{y}}, t) + \frac{\partial V^*}{\partial \mathbf{y}} \cdot \dot{\mathbf{y}} \right] = \mathbf{0} \\ (2) \quad & \frac{\partial^2}{\partial \dot{\mathbf{y}}^2} \left[I(\mathbf{y}, \dot{\mathbf{y}}, t) + \frac{\partial V^*}{\partial \mathbf{y}} \cdot \dot{\mathbf{y}} \right] \leq \mathbf{0} \end{aligned}$$

Da $\frac{\partial V^*}{\partial \dot{\mathbf{y}}}$ von $\dot{\mathbf{y}}$ unabhängig ist, ergibt sich aus (2) sofort die Legendre-Bedingung. Aus dem gleichen Grund vereinfacht sich die erste Gleichung zu

$$\frac{\partial I}{\partial \dot{\mathbf{y}}} + \frac{\partial V^*}{\partial \dot{\mathbf{y}}} = 0$$

Bildet man hiervon die totale Zeitableitung, folgt weiter:

$$(3) \quad \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{\mathbf{y}}} \right) = - \frac{d}{dt} \left(\frac{\partial V^*}{\partial \dot{\mathbf{y}}} \right) = - \frac{\partial^2 V^*}{\partial t \partial \dot{\mathbf{y}}} - \frac{\partial^2 V^*}{\partial \dot{\mathbf{y}}^2} \cdot \dot{\mathbf{y}}$$

Aus der Bellman-Gleichung folgt aber speziell *im* Optimum $\dot{\mathbf{y}}^*$

$$(4) \quad - \frac{\partial}{\partial \dot{\mathbf{y}}} \left(\frac{\partial V^*}{\partial t} \right) = \frac{\partial I}{\partial \dot{\mathbf{y}}} + \frac{\partial^2 V^*}{\partial \dot{\mathbf{y}}^2} \cdot \dot{\mathbf{y}}$$

Ist V^* zweimal stetig differenzierbar, folgt $\frac{\partial^2 V^*}{\partial t \partial \dot{\mathbf{y}}} = \frac{\partial^2 V^*}{\partial \dot{\mathbf{y}} \partial t}$ und Einsetzen von (4) in (3) führt zur Euler-Lagrange-Gleichung

$$\frac{\partial I}{\partial \dot{\mathbf{y}}} - \frac{d}{dt} \left(\frac{\partial I}{\partial \dot{\mathbf{y}}} \right) = 0.$$

- ii) Auch Restriktionen können in die Bellman-Gleichung eingearbeitet werden. Liegt zum Beispiel eine isoperimetrische Nebenbedingung vor

$$\int_{t_0}^{t_1} g(\mathbf{y}, \dot{\mathbf{y}}, t) dt = k,$$

so lautet die Bellman-Gleichung

$$- \frac{\partial V^*}{\partial t} = \max_{\dot{\mathbf{y}}} \left[I(\mathbf{y}, \dot{\mathbf{y}}, t) + \frac{\partial V^*}{\partial \dot{\mathbf{y}}} \cdot \dot{\mathbf{y}} + \frac{\partial V^*}{\partial k} \cdot g(\mathbf{y}, \dot{\mathbf{y}}, t) \right]$$

Analog zu Teil i), ergibt sich dann die aus der Variationsrechnung bekannte Bedingung. Die partiellen Ableitungen von V^* beschreiben die Sensitivität der Lösung und können daher als Lagrangemultiplikatoren interpretiert werden. $\frac{\partial V^*}{\partial k}$ gibt an, wie der optimale Wert V^* auf eine Änderung der Restriktions-Konstanten k reagiert und stimmt daher mit dem λ aus der Variationsrechnung überein.

- iii) Sei $\mathbf{u}^*(t)$ eine Lösung des Maximierungsproblems auf der rechten Seite der Bellman-Gleichung. Definiert man die Funktion $H(\mathbf{y}, \frac{\partial V^*}{\partial \dot{\mathbf{y}}}, t)$ durch

$$\begin{aligned} H(\mathbf{y}, \frac{\partial V^*}{\partial \dot{\mathbf{y}}}, t) &= I(\mathbf{y}, \mathbf{u}^*, t) + \frac{\partial V^*}{\partial \dot{\mathbf{y}}} \cdot \mathbf{f}(\mathbf{y}, \mathbf{u}^*, t) \\ &= \max_{\mathbf{u}} \left[I(\mathbf{y}, \mathbf{u}, t) + \frac{\partial V^*}{\partial \dot{\mathbf{y}}} \cdot \mathbf{f}(\mathbf{y}, \mathbf{u}, t) \right], \end{aligned}$$

so wird die Bellman-Gleichung zur sog. *Hamilton-Jakobi Gleichung*

$$H(\mathbf{y}, \frac{\partial V^*}{\partial \dot{\mathbf{y}}}, t) + \frac{\partial V^*}{\partial t} = 0.$$

□

Im allgemeinen besitzt die Bellman-Gleichung keine analytische Lösung und man ist auf numerische Lösungen angewiesen, was zu einer Diskretisierung der auftretenden Größen führt. Da in der Praxis zeitdiskrete Probleme häufig in natürlicher Weise auftreten, soll die Vorgehensweise in diesem Fall kurz beschrieben werden. Wir nehmen hierzu an, dass die Zeitvariable t die Werte

$$t_0, t_0+1, \dots, t_1$$

annehmen kann. Aufgrund der diskreten Zeitwerte werden wir die auftretenden Größen entsprechend indizieren, also $\mathbf{x}_t$ und $\mathbf{u}_t$ schreiben, statt $\mathbf{x}(t)$ und $\mathbf{u}(t)$. Die Bewegungsgleichung lautet

$$\mathbf{y}_{t+1} = \mathbf{f}_t(\mathbf{y}_t, \mathbf{u}_t), \quad t = t_0, t_0+1, \dots, t_1-1,$$

wobei $\mathbf{f}_{t_0}, \dots, \mathbf{f}_{t_1-1}$ stetig differenzierbare Funktionen (*Übergangsfunktionen*) sind. Der Anfangs- bzw. Endzustand sei gleich einem vorgegebenen $\mathbf{y}_0$ bzw. $\mathbf{y}_1$, d.h.

$$\mathbf{y}_{t_0} := \mathbf{y}_0 \quad \text{bzw.} \quad \mathbf{y}_{t_1} := \mathbf{y}_1.$$

Die Zielfunktion lautet

$$V(\mathbf{u}) = \sum_{t=t_0}^{t_1-1} I_t(\mathbf{y}_t, \mathbf{u}_t) + F(\mathbf{y}_{t_1}, t_1).$$

Diese soll durch Wahl einer Folge von Steuer- oder Kontrollvektoren

$$\{\mathbf{u}_{t_0}, \mathbf{u}_{t_0+1}, \dots, \mathbf{u}_{t_1-1}\}$$

maximiert werden⁸, wobei die $\mathbf{u}_t$ aus einer Kontrollmenge Ω zu wählen sind:

$$\mathbf{u}_t \in \Omega, \quad \text{für } t = t_0, t_0+1, \dots, t_1-1$$

Die Analogie zum stetigen Fall ist offensichtlich. Wie dort führt man die Funktion

$$V^*(\mathbf{x}, t) = V^*(\mathbf{x}_t)$$

ein, die den optimalen Wert von V liefert, wenn der Anfangswert $\mathbf{x}$ und der Anfangszeitpunkt t ist. Nach dem Optimalitätsprinzip 4.14 gilt wegen $\Delta t = 1$

$$V^*(\mathbf{y}_t) = \max_{\mathbf{u}_t} [I_t(\mathbf{y}_t, \mathbf{u}_t) + V^*(\mathbf{y}_{t+1})],$$

d.h. der optimale Wert von V beim Startwert $\mathbf{y}_t$ ist gleich dem optimalen Wert $V^*(\mathbf{y}_{t+1})$ für den Weg von $\mathbf{y}_{t+1}$ nach $\mathbf{y}_{t_1}$, vermehrt um den maximalen Beitrag den I von $\mathbf{y}_t$ bis $\mathbf{y}_{t+1}$ zu $V^*(\mathbf{y}_{t+1})$ beitragen kann. Wegen $\mathbf{y}_{t+1} = \mathbf{f}_t(\mathbf{y}_t, \mathbf{u}_t)$ lautet damit die Rekursionsformel:

$$V^*(\mathbf{y}_t) = \max_{\mathbf{u}_t} [I_t(\mathbf{y}_t, \mathbf{u}_t) + V^*(\mathbf{f}_t(\mathbf{y}_t, \mathbf{u}_t))]$$

⁸Man nennt die $\mathbf{u}_{t_j}$ oft auch *Politiken*.

Aufgrund der Randbedingung

$$V^*(\mathbf{y}_{t_1}) = F(\mathbf{y}_{t_1})$$

können damit sukzessive die Werte

$$V^*(\mathbf{y}_{t_1-1}), V^*(\mathbf{y}_{t_1-2}), \dots, V^*(\mathbf{y}_0)$$

berechnet werden.

Das folgende Beispiel demonstriert diese Vorgehensweise in einem konkreten Fall.

4.30 Beispiel

Betrachten wir das in 4.5 vorgestellte zeitdiskrete Transportbeispiel. Es gilt $t_0 = 0$, $t_1 = 5$ und $y(t_0) = y_0 = A$, $y(t_1) = y_1 = Z$.

Da an jedem Knotenpunkt maximal zwei Wegalternativen zur Auswahl stehen, kann die Kontrollmenge Ω z.B. gleich $\{-1, 1\}$ gewählt werden, wobei $u = 1$ bedeutet, wähle den Weg nach „oben“, $u = -1$ heißt, wähle den Weg nach „unten“.

Die intermediäre Funktion I liefert gerade die Kosten (negativ, um ein Max-problem zu erhalten) des entsprechenden Wegstücks. Die finale Funktion F ist identisch Null. Ist z.B. $t = 3$, $y_3 = H$ so gilt

$$y_4 = y_{3+1} = f_3(y_3, u_3) = f_3(H, u_3) = \begin{cases} J & \text{falls } u_3 = +1 \\ K & \text{falls } u_3 = -1 \end{cases}$$

und

$$I(y_3, u_3) = I(H, u_3) = \begin{cases} -4 & \text{falls } u_3 = +1 \\ -6 & \text{falls } u_3 = -1 \end{cases}$$

Nun gilt aufgrund der Randbedingung

$$V^*(y_5) = V^*(Z, 5) = F(y_5) = F(Z) = 0.$$

Nach dem Optimalitätsprinzip folgt daher

$$V^*(y_4) = \max_{u_4} [I(y_4, u_4) + V^*(f(y_4, u_4), 5)].$$

Da $f(y_4, u_4) = Z$ gilt, unabhängig von y_4 und u_4 , folgt

$$\begin{aligned} V^*(y_4) &= \max_{u_4} [I(y_4, u_4) + V^*(Z, 5)] \\ &= \max_{u_4} I(y_4, u_4) = \begin{cases} -3 & \text{falls } y_4 = I \\ -1 & \text{falls } y_4 = J \\ -2 & \text{falls } y_4 = K \end{cases} \end{aligned}$$

Damit ist es nun möglich die optimalen Werte auf der Stufe $t = 3$ zu berechnen. Da y auf der Stufe $t = 3$ nur die Werte G oder H annehmen kann, sind die Werte

$$V^*(G, 3) \text{ und } V^*(H, 3)$$

zu berechnen. Die Rekursionsformel liefert wieder:

$$\begin{aligned} V^*(G, 3) &= \max_{u_3} [I(G, u_3) + V^*(f(G, u_3), 4)] \\ &= \max [I(G, 1) + V^*(f(G, 1), 4) ; I(G, -1) + V^*(f(G, -1), 4)] \\ &= \max [-2 + V^*(I, 4) ; -8 + V^*(J, 4)] \\ &= \max [-5 ; -9] = -5 \end{aligned}$$

Analog folgt

$$\begin{aligned} V^*(H, 3) &= \max_{u_3} [I(H, u_3) + V^*(f(H, u_3), 4)] \\ &= \max [I(H, 1) + V^*(J, 4) ; I(H, -1) + V^*(K, 4)] \\ &= \max [-5 ; -8] = -5 \end{aligned}$$

Die Fortsetzung dieses Verfahrens liefert für Stufe $t = 2$

$$V^*(D, 2) = -8 , \quad V^*(E, 2) = -8 , \quad V^*(F, 2) = -10 ,$$

für Stufe $t = 1$

$$V^*(B, 1) = -13 , \quad V^*(C, 1) = -10$$

und damit schließlich den gesuchten Wert

$$\begin{aligned} V^*(A, 0) &= \max_{u_0} [I(A, u_0) + V^*(f(A, u_0), 1)] \\ &= \max [I(A, 1) + V^*(B, 1) ; I(A, -1) + V^*(C, 1)] \\ &= \max [-15 ; -14] = -14 . \end{aligned}$$

Man sieht schon an diesem kleinen Beispiel, dass das Optimierungsproblem dadurch gelöst wird, dass man eine größere Zahl von Optimierungsfragen löst, nämlich die Bestimmung der optimalen Werte

$$V^*(B, 1) , \quad V^*(C, 1) , \quad V^*(D, 2) , \quad V^*(E, 2) , \quad V^*(F, 2) \text{ usw.},$$

die sich für V ergeben, wenn - bei gleichem Endpunkt - die Startpunkte B, C, D usw. sind. Es sollte auch beachtet werden, dass die Lösung des Problems zunächst nur den optimalen Zielfunktionswert $V^*(A, 0)$ liefert und nicht den optimalen Zielpfad. Dieser kann erst nach der Berechnung von $V^*(A, 0)$ bestimmt werden. Hier ergibt sich aus $u_0 = -1$ der Wert $y_1 = C$, daraus wiederum $u_1 = 1, y_2 = E$ usw. $\square$

4.3. Optimale Kontrolltheorie

Wie früher schon erwähnt, bereitet die Variationsrechnung Schwierigkeiten beim Auftreten von Restriktionen, z.B., wenn die Kontrollvariablen auf eine vorgegebene Kontrollmenge eingeschränkt werden. Es ist einer der wesentlichen Vorteile der dynamischen Programmierung und der Methode der optimalen Kontrolle, dass in diesen Theorien solche Restriktionen ohne größere Probleme verarbeitet werden können. Dies äußert sich in unterschiedlichen Vorgehensweisen bei den einzelnen Optimierungsmethoden.

In der Variationsrechnung wird direkt der optimale Zustandspfad $\mathbf{y}^*$ bestimmt, wohingegen in der dynamischen Programmierung zunächst der optimale Zielfunktionswert V^* und erst danach der optimale Zustandspfad $\mathbf{y}^*$ bestimmt wird.

Die optimale Kontrolltheorie dagegen bestimmt zuerst den optimalen Steuervektor $\mathbf{u}^*$, indem man wie in der Variationsrechnung die Steuergröße $\mathbf{u}$ variiert. Die Variation von $\mathbf{u}$ führt jedoch dazu, dass das einfachste Problem der Kontrolltheorie - im Gegensatz zur Variationsrechnung - das mit freiem Endzustand ist. Liegt nämlich ein fester Endzustand vor, so führt eine beliebige Änderung $\Delta \mathbf{u}$ der Kontrollvariablen zu einer Veränderung $\Delta \mathbf{y}$ der Zustandsvariablen, die i.a. mit einem vorgegebenen Endzustand nicht kompatibel ist.

Wir betrachten im folgenden das zu Beginn beschriebene allgemeine Kontrollproblem:

$$\begin{aligned} V(\mathbf{u}) &= \int_{t_0}^{t_1} I(\mathbf{y}, \mathbf{u}, t) dt + F(\mathbf{y}_1, t_1) \rightarrow \max \\ \dot{\mathbf{y}} &= \mathbf{f}(\mathbf{y}, \mathbf{u}, t) \\ \mathbf{y}(t_0) &= \mathbf{y}_0, \quad \mathcal{T}(\mathbf{y}(t_1), t_1) = 0 \\ \mathbf{u}(t) &\in \Omega \end{aligned}$$

Hierbei sind I , F , $\mathbf{f}$ stetig differenzierbar. Die Steuergröße $\mathbf{u}$ gehöre zur Kontrollmenge Ω , d.h. $\mathbf{u}$ ist - in Abhängigkeit von der Zeit - stückweise stetig, mit $\mathbf{u}(t) \in \Omega$, wobei $\Omega \subset \mathbb{R}^k$ eine nichtleere, kompakte Teilmenge sei.

Die optimale Kontrolltheorie stellt im Prinzip eine Verallgemeinerung der Methode der Lagrangemultiplikatoren dar. Man führt eine geeignete Lagrangefunktion L ein und bestimmt einen Sattelpunkt, d.h. einen Punkt, in dem L bzgl. der Zustandsvariablen maximiert und bzgl. der Lagrangemultiplikatoren minimiert wird.

Der Einfachheit halber sei bei den nachfolgenden Überlegungen vorausgesetzt, dass die Kontrollvariablen frei sind, so dass $\Omega = \mathbb{R}^k$ gilt. Wir fassen nun die Bewegungsgleichung $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}, \mathbf{u}, t)$ als Nebenbedingung der Form $\mathbf{f}(\mathbf{y}, \mathbf{u}, t) - \dot{\mathbf{y}} = \mathbf{0}$ auf und führen wie im statischen Fall $\boldsymbol{\lambda}(t) = (\lambda_1(t), \dots, \lambda_n(t))$ als neue Variablen ein (*Costate-Variablen*). Sie stellen das dynamische Analogon zu den Lagrange-Multiplikatoren dar. Die zugehörige Lagrange-

funktion L lautet⁹

$$\begin{aligned} L(\mathbf{u}, \boldsymbol{\lambda}) &:= V(\mathbf{u}) + \int_{t_0}^{t_1} \boldsymbol{\lambda} \cdot [\mathbf{f}(\mathbf{y}, \mathbf{u}, t) - \dot{\mathbf{y}}] dt \\ &= \int_{t_0}^{t_1} [I(\mathbf{y}, \mathbf{u}, t) + \boldsymbol{\lambda} \cdot \mathbf{f}(\mathbf{y}, \mathbf{u}, t) - \boldsymbol{\lambda} \cdot \dot{\mathbf{y}}] dt + F(\mathbf{y}_1, t_1). \end{aligned}$$

Wie im statischen Fall liefert ein Sattelpunkt von L die Lösung. Zur Berechnung von Sattelpunkten leiten wir nun notwendige Bedingungen her.

Sei hierzu $(\mathbf{u}^*, \boldsymbol{\lambda}^*)$ ein Sattelpunkt mit der zugehörigen Zustandstrajektorie $\mathbf{y}^*$. Wie früher sei $\mathbf{h}$ ein beliebiger Störpfad. Da L im Sattelpunkt bzgl. der Variablen $\boldsymbol{\lambda}$ ein lokales Minimum besitzt, muss die erste Variation von L in $(\mathbf{u}^*, \boldsymbol{\lambda}^*)$ bzgl. der Variablen $\boldsymbol{\lambda}$ gleich 0 sein. Wegen

$$\frac{d}{d\varepsilon} L(\mathbf{u}^*, \boldsymbol{\lambda}^* + \varepsilon \mathbf{h}, \mathbf{y}^*) = \int_{t_0}^{t_1} \mathbf{h}(t) \cdot [\mathbf{f}(\mathbf{y}^*, \mathbf{u}^*, t) - \dot{\mathbf{y}}^*] dt$$

folgt daher:

$$\left. \frac{d}{d\varepsilon} L \right|_{\varepsilon=0} = \int_{t_0}^{t_1} \mathbf{h}(t) \cdot [\mathbf{f}(\mathbf{y}^*, \mathbf{u}^*, t) - \dot{\mathbf{y}}^*] dt \stackrel{!}{=} 0$$

Da $\mathbf{h}$ beliebig war, muss nach dem Fundamentallemma der Variationsrechnung $\mathbf{y}^*$ wie im statischen Fall die Nebenbedingung - hier also die Bewegungsgleichung $\dot{\mathbf{y}}^* = \mathbf{f}(\mathbf{y}^*, \mathbf{u}^*, t)$ - erfüllen.

Zur Herleitung der zweiten Sattelpunkt-Bedingung führen wir die Kurzbezeichnung

$$H(\mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, t) := I(\mathbf{y}, \mathbf{u}, t) + \boldsymbol{\lambda} \cdot \mathbf{f}(\mathbf{y}, \mathbf{u}, t)$$

ein. H wird als *Hamilton-Funktion* bezeichnet. Die Lagrange-Funktion lautet dann

$$L(\mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}) = \int_{t_0}^{t_1} [H(\mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, t) - \boldsymbol{\lambda} \dot{\mathbf{y}}] dt + F(\mathbf{y}_1, t_1).$$

Integriert man den zweiten Term unter dem Integral partiell, folgt

$$L(\mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}) = \int_{t_0}^{t_1} [H(\mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, t) + \dot{\boldsymbol{\lambda}} \cdot \mathbf{y}] dt + F(\mathbf{y}_1, t_1) - [\boldsymbol{\lambda}(t_1) \mathbf{y}(t_1) - \boldsymbol{\lambda}(t_0) \mathbf{y}(t_0)]$$

Da L im Sattelpunkt $(\mathbf{u}^*, \boldsymbol{\lambda}^*)$ bzgl. der Variablen $\mathbf{u}$ ein lokales Maximum besitzt, muss die erste Variation von L in $(\mathbf{u}^*, \boldsymbol{\lambda}^*)$ bzgl. der Variablen $\mathbf{u}$ gleich 0 sein. Hierbei besteht allerdings die Schwierigkeit, dass eine Störung von $\mathbf{u}$, nämlich

$$\tilde{\mathbf{u}}(t) = \mathbf{u}^*(t) + \varepsilon \cdot \mathbf{h}(t)$$

⁹Man achte wieder darauf, dass die Nebenbedingung in integrierter Form zu V addiert wird.

über die Bewegungsgleichung zu einer Störung $\tilde{\mathbf{y}} = \mathbf{y}^*(t) + \varepsilon \cdot \mathbf{g}(t)$ der optimalen Zustandsgröße $\mathbf{y}^*(t)$ führt. Daher gilt unter Beachtung von $t_1 = t_1(\varepsilon)$

$$\begin{aligned}
 \frac{d}{d\varepsilon} L(\tilde{\mathbf{y}}(t), \tilde{\mathbf{u}}(t), \boldsymbol{\lambda}^*) &= \frac{d}{d\varepsilon} \left[\int_{t_0}^{t_1(\varepsilon)} [H(\tilde{\mathbf{y}}, \tilde{\mathbf{u}}, \boldsymbol{\lambda}^*, t) + \dot{\boldsymbol{\lambda}}^* \tilde{\mathbf{y}}] dt \right. \\
 &\quad \left. + F(\tilde{\mathbf{y}}(t_1), t_1) - [\boldsymbol{\lambda}^*(t) \tilde{\mathbf{y}}(t)]_{t=t_0}^{t=t_1(\varepsilon)} \right] \\
 &\stackrel{4.14}{=} \int_{t_0}^{t_1(\varepsilon)} \left\{ \left(\frac{\partial H}{\partial \mathbf{y}} \cdot \mathbf{g}(t) + \frac{\partial H}{\partial \mathbf{u}} \cdot \mathbf{h}(t) \right) + \dot{\boldsymbol{\lambda}}^* \cdot \mathbf{g}(t) \right\} dt \\
 &\quad + [H + \dot{\boldsymbol{\lambda}}^*(t) \tilde{\mathbf{y}}(t)]_{t=t_1(\varepsilon)} \cdot \frac{dt_1}{d\varepsilon} + \frac{\partial F}{\partial \mathbf{y}}(t_1) \cdot \frac{d}{d\varepsilon} \tilde{\mathbf{y}}(t_1) \\
 &\quad + \frac{\partial F}{\partial t} \cdot \frac{dt_1}{d\varepsilon} - \boldsymbol{\lambda}^*(t_1) \cdot \frac{d}{d\varepsilon} \tilde{\mathbf{y}}(t_1) - \dot{\boldsymbol{\lambda}}^*(t_1) \cdot \frac{dt_1}{d\varepsilon} \cdot \tilde{\mathbf{y}}(t_1) \\
 &= \int_{t_0}^{t_1(\varepsilon)} \left\{ \left(\frac{\partial H}{\partial \mathbf{y}} + \dot{\boldsymbol{\lambda}}^* \right) \mathbf{g}(t) + \frac{\partial H}{\partial \mathbf{u}} \mathbf{h}(t) \right\} dt \\
 &\quad + \left[\left(H + \frac{\partial F}{\partial t} \right) \frac{dt_1}{d\varepsilon} + \left(\frac{\partial F}{\partial \mathbf{y}} - \boldsymbol{\lambda}^*(t) \right) \frac{d}{d\varepsilon} \tilde{\mathbf{y}}(t) \right]_{t=t_1(\varepsilon)}
 \end{aligned}$$

Für $\varepsilon = 0$ ergibt sich mit

$$\Delta t := \left. \frac{dt_1}{d\varepsilon} \right|_{\varepsilon=0}, \quad \Delta \mathbf{y} := \left. \frac{d}{d\varepsilon} \tilde{\mathbf{y}}(t_1(\varepsilon)) \right|_{\varepsilon=0}$$

und $t^* := t_1(0)$

$$\begin{aligned}
 0 &\stackrel{!}{=} \int_{t_0}^{t^*} \left(\frac{\partial H}{\partial \mathbf{y}} + \dot{\boldsymbol{\lambda}}^* \right) \mathbf{g}(t) dt + \int_{t_0}^{t^*} \frac{\partial H}{\partial \mathbf{u}} \mathbf{h}(t) dt \\
 &\quad + \left[\left(H + \frac{\partial F}{\partial t} \right) \cdot \Delta t + \left(\frac{\partial F}{\partial \mathbf{y}} - \boldsymbol{\lambda}^* \right) \cdot \Delta \mathbf{y} \right]_{t=t^*}
 \end{aligned}$$

Die letzte Summe setzt sich aus verschiedenen beliebigen Störungen zusammen, die Störung von $\mathbf{y}^*$, $\mathbf{u}^*$ und die Störung von t^* . Nach dem Fundamentallemma der Variationsrechnung 4.9 müssen daher alle drei Terme 0 sein (der letzte ist nur vom Endzustand abhängig und

damit vom Verlauf von $\mathbf{g}$ und $\mathbf{h}$ unabhängig).

$$\begin{aligned} (1) \quad & \frac{\partial H}{\partial \mathbf{u}} = 0 \\ (2) \quad & \frac{\partial H}{\partial \mathbf{y}} + \dot{\boldsymbol{\lambda}}^* = 0 \\ (3) \quad & \left[\left(H + \frac{\partial F}{\partial t} \right) \Delta t + \left(\frac{\partial F}{\partial \mathbf{y}} - \boldsymbol{\lambda}^* \right) \cdot \Delta \mathbf{y} \right]_{t=t^*} = 0 \end{aligned}$$

Wie schon im Beweis von Satz 4.15 gesehen, gilt

$$\frac{\partial \mathcal{T}}{\partial t} \cdot \Delta t + \frac{\partial \mathcal{T}}{\partial \mathbf{y}} \cdot \Delta \mathbf{y} = 0$$

Multiplikation von Gleichung (3) mit $\frac{\partial \mathcal{T}}{\partial t}$ und Einsetzen liefert dann

$$\left[- \left(H + \frac{\partial F}{\partial t} \right) \frac{\partial \mathcal{T}}{\partial \mathbf{y}} + \left(\frac{\partial F}{\partial \mathbf{y}} - \boldsymbol{\lambda}^* \right) \frac{\partial \mathcal{T}}{\partial t} \right]_{t=t^*} = 0$$

Da die Bewegungsgleichung $\dot{\mathbf{y}}^* = \mathbf{f}(\mathbf{y}^*, \mathbf{u}^*, t)$ auch in der Form $\dot{\mathbf{y}} = \frac{\partial H}{\partial \boldsymbol{\lambda}}$ geschrieben werden kann, erhält man folgenden Satz, wenn man die Maximierungsbedingung $\frac{\partial H}{\partial \mathbf{u}} = 0$ durch die allgemeinere Forderung $\max_{\mathbf{u}(t) \in \Omega} H(\mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, t)$ ersetzt.

4.31 Satz (Maximumprinzip)

Notwendige Bedingungen für die Lösung $\mathbf{y}^*, \mathbf{u}^*$ des allgemeinen Kontrollproblems sind:

$$\begin{aligned} \mathbf{u}^* &= \operatorname{argmax}_{\mathbf{u}(t) \in \Omega} H(\mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, t) \\ \dot{\mathbf{y}}^* &= \left. \frac{\partial H}{\partial \boldsymbol{\lambda}} \right|^*, \quad \mathbf{y}^*(t_0) = \mathbf{y}_0 \quad (\text{Bewegungsgleichung für } \mathbf{y}) \\ \dot{\boldsymbol{\lambda}}^* &= - \left. \frac{\partial H}{\partial \mathbf{y}} \right|^* \\ \left[\left(\frac{\partial F}{\partial \mathbf{y}} - \boldsymbol{\lambda}^* \right) \frac{\partial \mathcal{T}}{\partial t} - \left(H + \frac{\partial F}{\partial t} \right) \frac{\partial \mathcal{T}}{\partial \mathbf{y}} \right]_{t=t^*} &= 0 \end{aligned}$$

□

4.32 Bemerkung

i) Im allgemeinen muss die Hamiltonfunktion H in der Form

$$H = \lambda_{a_0} \cdot I(\mathbf{y}, \mathbf{u}, t) + \boldsymbol{\lambda}(t) \cdot \mathbf{f}(\mathbf{y}, \mathbf{u}, t)$$

mit $\lambda_{a_0} \geq 0$ angesetzt werden. Erst unter Voraussetzung gewisser Regularitätsbedingungen ergibt sich $\lambda_{a_0} > 0$. In diesem Fall kann das Problem normiert werden, indem man $\lambda_{a_0} = 1$ setzt. Ohne weitere Voraussetzungen kann jedoch $\lambda_{a_0} = 0$ für die gesuchte Lösung gelten.

- ii) Die Bedingungen des letzten Satzes sind im allgemeinen weder hinreichend, noch liefern sie eine eindeutige Lösung oder ein globales Maximum. Die Bedingungen sind jedoch hinreichend, wenn I und $\mathbf{f}$ differenzierbar und konkav sind und $\lambda^*(t) \geq 0$ für alle t ist (Mangasarian, 1966).
- iii) Wie schon in Bemerkung 4.16 gesehen, ist es nützlich, häufig auftretende Formen von $\mathcal{T}$ gesondert zu betrachten.

1. Fall:

$t_1=t^*$ liegt fest, $\mathbf{y}(t_1)$ ist frei. Dann gilt $\mathcal{T}(\mathbf{y}, t) = t - t^*$. Wegen $\frac{\partial \mathcal{T}}{\partial \mathbf{y}} = \mathbf{0}$ und $\frac{\partial \mathcal{T}}{\partial t} = 1$ lautet die Transversalitätsbedingung in diesem Fall

$$\lambda^*(t^*) = \frac{\partial F}{\partial \mathbf{y}}(\mathbf{y}^*, t^*),$$

also $\lambda^*(t^*) = 0$, wenn $F \equiv 0$ ist.

2. Fall:

$\mathcal{T}(\mathbf{y}, t)$ kann nach $\mathbf{y}$ aufgelöst werden, d.h. es gilt

$$\mathcal{T}(\mathbf{y}, t) = \mathbf{y} - \mathbf{g}(t)$$

Wegen $\frac{\partial \mathcal{T}}{\partial \mathbf{y}} = \mathbf{1}$ und $\frac{\partial \mathcal{T}}{\partial t} = -\dot{\mathbf{g}}(t)$ lautet die Transversalitätsbedingung:

$$\left[\left(H + \frac{\partial F}{\partial t} \right) + \left(\frac{\partial F}{\partial \mathbf{y}} - \lambda^* \right) \dot{\mathbf{g}}(t) \right]_{t=t^*} = 0,$$

also $\left[H - \lambda^*(t) \dot{\mathbf{g}}(t) \right]_{t=t^*} = 0$ für $F \equiv 0$.

Ist speziell t_1 frei und $\mathbf{y}(t_1) = \mathbf{y}_1$ vorgegeben, gilt

$$\mathcal{T}(\mathbf{y}, t) = \mathbf{y} - \mathbf{y}_1, \quad \text{d.h. } \mathbf{g}(t) = \mathbf{y}_1 = \text{konstant}$$

Man erhält in diesem Fall wegen $\dot{\mathbf{g}}(t) = \mathbf{0}$

$$\left[H + \frac{\partial F}{\partial t} \right]_{t=t^*} = 0$$

bzw. $\left[H \right]_{t=t^*} = 0$ falls $F \equiv 0$.

- iv) Ist der Endzustand $\mathbf{y}_1$ und der Endzeitpunkt t_1 vorgegeben, gilt das Maximumprinzip weiterhin. Die Transversalitätsbedingung ist in diesem Fall zu ersetzen durch die Bedingung

$$\mathbf{y}(t_1) = \mathbf{y}_1.$$

Die folgende Tabelle zum Maximumprinzip fasst die wichtigsten Fälle der Transversalitätsbedingung

$$\left[\left(\frac{\partial F}{\partial \mathbf{y}} - \boldsymbol{\lambda}^* \right) \cdot \frac{\partial \mathcal{T}}{\partial t} - \left(H + \frac{\partial F}{\partial t} \right) \cdot \frac{\partial \mathcal{T}}{\partial \mathbf{y}} \right]_{t=t^*} = 0$$

zusammen.

Fall	$\mathcal{T}$	Randbedingung
Endzeitpunkt t^* gegeben	$\mathcal{T}(t, \mathbf{y}) = t - t^*$	
1.) $\mathbf{y}(t^*) = \mathbf{y}_1$ gegeben		$\mathbf{y}(t_0) = \mathbf{y}_0, \mathbf{y}(t_1) = \mathbf{y}_1$
2.) $\mathbf{y}(t^*)$ frei		$\mathbf{y}(t_0) = \mathbf{y}_0, \boldsymbol{\lambda}^*(t_1) = \frac{\partial F}{\partial \mathbf{y}}$
Endzeitpunkt t^* frei		
1.) $\mathbf{y}(t^*) = \mathbf{y}_1$ gegeben	$\mathcal{T}(t, \mathbf{y}) = \mathbf{y} - \mathbf{y}_1$	$\mathbf{y}(t_0) = \mathbf{y}_0, \mathbf{y}(t^*) = \mathbf{y}_1, \left(H + \frac{\partial F}{\partial t} \right)_{t=t^*} = 0$
2.) $\mathbf{y}(t^*)$ frei		$\mathbf{y}(t_0) = \mathbf{y}_0, \boldsymbol{\lambda}^*(t^*) = \frac{\partial F}{\partial \mathbf{y}}, \left(H + \frac{\partial F}{\partial t} \right)_{t=t^*} = 0$

v) Man beachte, dass die Forderung $\max_{\mathbf{u}(t) \in \Omega} H$ allgemeiner als $\frac{\partial H}{\partial \mathbf{u}} = \mathbf{0}$ ist, da bei abgeschlossener Kontrollmenge Ω (z.B. $\Omega = [a, b]$) das Maximum am Rand auftreten, d.h. $\frac{\partial H}{\partial \mathbf{u}} \neq \mathbf{0}$ gelten kann.

vi) Für die zeitliche Ableitung der Hamiltonfunktion gilt wegen $\frac{\partial H}{\partial \boldsymbol{\lambda}} = \dot{\mathbf{y}}$:

$$\begin{aligned} \frac{dH}{dt} &= \frac{\partial H}{\partial \mathbf{y}} \cdot \dot{\mathbf{y}} + \frac{\partial H}{\partial \mathbf{u}} \cdot \dot{\mathbf{u}} + \frac{\partial H}{\partial \boldsymbol{\lambda}} \dot{\boldsymbol{\lambda}} + \frac{\partial H}{\partial t} \\ &= \left(\frac{\partial H}{\partial \mathbf{y}} + \dot{\boldsymbol{\lambda}} \right) \cdot \mathbf{f}(\mathbf{y}, \mathbf{u}, t) + \frac{\partial H}{\partial \mathbf{u}} \cdot \dot{\mathbf{u}} + \frac{\partial H}{\partial t} \end{aligned}$$

Da entlang der optimalen Trajektorie die Bewegungsgleichung

$$\dot{\boldsymbol{\lambda}} = - \frac{\partial H}{\partial \mathbf{y}}$$

erfüllt ist, ergibt sich

$$\frac{dH}{dt} = \frac{\partial H}{\partial \mathbf{u}} \dot{\mathbf{u}} + \frac{\partial H}{\partial t}.$$

Weiter gilt $\frac{\partial H}{\partial \mathbf{u}} \dot{\mathbf{u}} = \mathbf{0}$, denn für eine innere Lösung ist $\frac{\partial H}{\partial \mathbf{u}} = \mathbf{0}$ und für eine Randlösung gilt $\dot{\mathbf{u}} = \mathbf{0}$ (entweder gilt $\frac{\partial H}{\partial u_j} = 0$ oder $\dot{u}_j = 0$). Damit folgt

$$\frac{dH}{dt} = \frac{\partial H}{\partial t}.$$

Speziell ergibt sich für autonome Probleme, in denen H von t nicht explizit abhängt,

$$\frac{dH}{dt} = 0.$$

Dies bedeutet, dass im autonomen Fall H entlang der optimalen Trajektorie zeitlich konstant ist. $\square$

4.33 Beispiel

- i) Man bestimme den kürzesten Abstand eines gegebenen Punktes $P = (x_0, y_0)$ zu einer vorgegebenen Kurve mit der Funktionsvorschrift $x \mapsto g(x)$ (siehe auch Beispiel 4.17 i)). Das Problem lautet:

$$\begin{aligned} V(u) &= \int_{x_0}^{x_1} -(1 + u^2)^{\frac{1}{2}} dx \rightarrow \max \\ y' &= u = f(u, y) \\ y(x_0) &= y_0, \quad y(x_1) = g(x_1) \quad (\text{d.h. } \mathcal{T}(y, x) = y(x) - g(x)) \end{aligned}$$

1. Schritt: Bestimmung der Hamiltonfunktion und Berechnung von u^*

$$H = -(1 + u^2)^{\frac{1}{2}} + \lambda u$$

$$\frac{\partial H}{\partial u} = 0 \quad \text{liefert} \quad \lambda^* = \frac{u^*}{\sqrt{1 + u^{*2}}} \iff u^*(x) = \lambda^*(1 - \lambda^{*2})^{-\frac{1}{2}}$$

Wegen

$$\frac{\partial^2 H}{\partial u^2} = -(1 + u^2)^{-\frac{3}{2}} < 0$$

liegt ein absolutes Maximum von H vor. Zur Bestimmung von u^* muss λ^* bestimmt werden.

2. Schritt: Bestimmung der Costate-Variablen $\lambda(x)$

Aus $\lambda^{*'} = -\frac{\partial H}{\partial y} = 0$ folgt, dass λ^* und damit u^* konstant ist. Wegen $y^{*'} = u^*$ ist damit der gesuchte kürzeste Weg y^* eine Gerade, d.h. $y^*(x) = ax + b$ mit $a, b \in \mathbb{R}$ konstant.

3. Schritt: Bestimmung der optimalen Trajektorie $y^*(x)$

Setzt man in die Transversalitätsbedingung ein, folgt wegen $\lambda^* = u^*(1 + u^{*2})^{-\frac{1}{2}}$:

$$\begin{aligned}
 0 &= [H - \lambda^*(x)g'(x)]_{x=x_1} \\
 &= \left[-(1 + u^{*2})^{\frac{1}{2}} + \lambda^* \cdot u^* - \lambda^* g'(x) \right]_{x=x_1} \\
 &= -(1 + u^{*2})^{\frac{1}{2}} + u^{*2}(1 + u^{*2})^{-\frac{1}{2}} - u^*(1 + u^{*2})^{-\frac{1}{2}} g'(x_1) \\
 \iff 0 &= -1 - u^* \cdot g'(x_1) \iff -1 = y^{*'}(x_1)g'(x_1)
 \end{aligned}$$

Dies ist das bekannte Ergebnis, dass die kürzeste Verbindungsstrecke eine Gerade ist, die im Punkt kürzesten Abstands senkrecht auf der Kurve steht. Damit folgt wegen $-\frac{1}{g'(x_1)} = y^{*'}(x_1) = a$ und $y^*(x_0) = y_0$:

$$y^*(x) = -\frac{1}{g'(x_1)} \cdot x + y_0 + \frac{x_0}{g'(x_1)}$$

Da sich y^* und g im (noch unbekannten) Punkt x_1 schneiden, gilt die Gleichung

$$y^*(x_1) = -\frac{1}{g'(x_1)} \cdot x_1 + y_0 + \frac{x_0}{g'(x_1)} = g(x_1) .$$

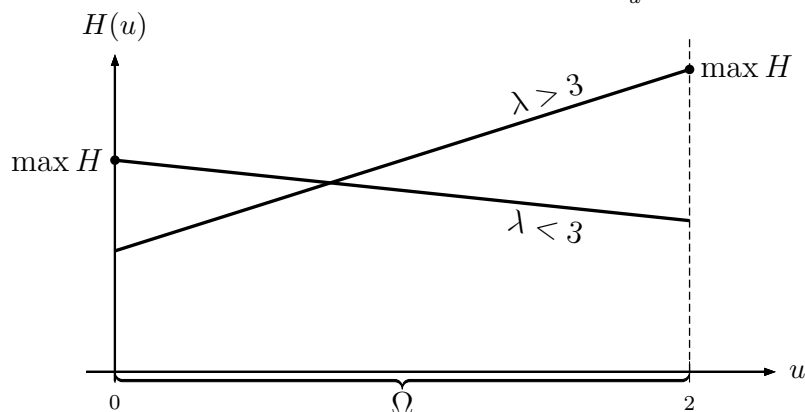
Da g bekannt ist, stellt dies eine Gleichung für x_1 dar.

$$\begin{aligned}
 \text{ii)} \quad V &= \int_0^2 (2y - 3u) dt \rightarrow \max \\
 \text{u.d.N.} \quad \dot{y} &= y + u \\
 y(0) &= 4, \quad y(2) \text{ frei} \\
 u(t) &\in \Omega = [0, 2]
 \end{aligned}$$

1. Schritt: Bestimmung der Hamiltonfunktion und Berechnung von u^*

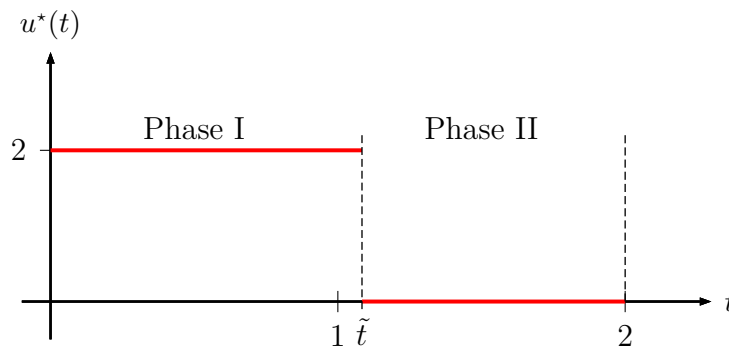
$$H = 2y - 3u + \lambda(y + u) = (2 + \lambda)y + (\lambda - 3)u$$

Damit ist H linear in u , mit der Steigung $\lambda - 3$. Dies bedeutet insbesondere, dass $\frac{\partial H}{\partial u}$ in diesem Beispiel nicht zur Maximierung verwendet werden kann. Besitzt H eine positive Steigung bzgl. u ($\lambda > 3$), wird $\max_u H$ in $u^* = 2$ angenommen. Für eine negative Steigung von H bzgl. u ($\lambda < 3$) wird $\max_u H$ in $u^* = 0$ angenommen.



Die optimale Kontrolle lautet daher:

$$u^*(t) = \begin{cases} 2 & \text{falls } \lambda^*(t) > 3 \\ 0 & \text{falls } \lambda^*(t) < 3 \end{cases}$$



2. Schritt: Bestimmung der Costate-Variablen $\lambda^*(t)$

$$\dot{\lambda}^* = -\frac{\partial H}{\partial y} = -2 - \lambda^* \iff \dot{\lambda}^* + \lambda^* = -2$$

Damit folgt

$$\lambda^*(t) = k \cdot e^{-t} - 2 \quad (k \text{ konstant})$$

Aufgrund der Transversalitätsbedingung

$$0 = \lambda^*(t^*) \stackrel{t^*=2}{=} \lambda^*(2)$$

folgt $k = 2 \cdot e^2$ und damit

$$\lambda^*(t) = 2e^{2-t} - 2.$$

λ^* ist offensichtlich streng monoton fallend mit $\lambda^*(0) > 3$ und $\lambda^*(2) = 0 < 3$. Für $\tilde{t} = 2 - \ln\left(\frac{5}{2}\right) \approx 1.084$ gilt $\lambda^*(\tilde{t}) = 3$. Daraus folgt:

$$u^*(t) = \begin{cases} 2 & \text{falls } t < \tilde{t} \\ 0 & \text{falls } t > \tilde{t} \end{cases}$$

3. Schritt: Bestimmung der optimalen Trajektorie $y^*(t)$

Sie ergibt sich aus der Bewegungsgleichung

$$\dot{y}^* = f(y^*, u^*, t) = y^* + u^* = \begin{cases} y^* + 2 & \text{für } t < \tilde{t} \\ y^* & \text{für } t > \tilde{t} \end{cases}$$

Wegen $y^*(0) = 4$ liefert dies für $t < \tilde{t}$

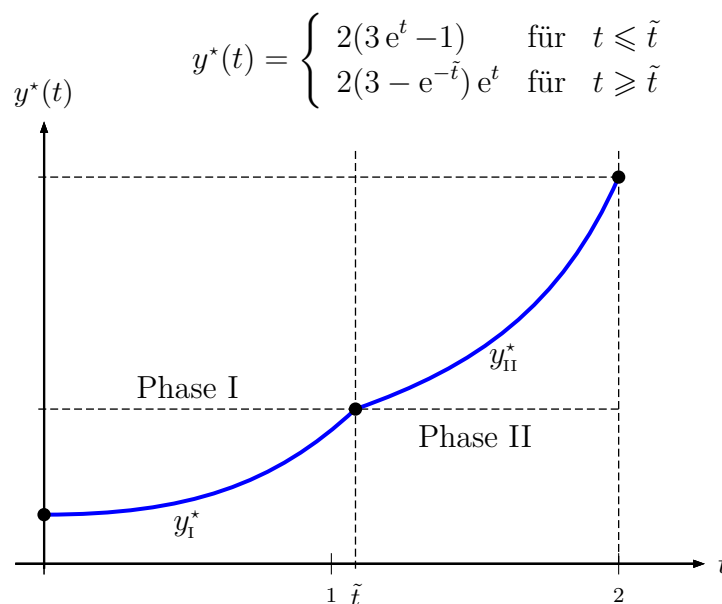
$$y_I^*(t) = 2(3e^t - 1).$$

Im Zeitintervall $[\tilde{t}, 2]$ lautet die Lösung y^*

$$y_{II}^*(t) = c \cdot e^t \quad \text{für } t > \tilde{t}.$$

Die Konstante c kann weder aus der Anfangsbedingung $y(0) = 4$ bestimmt werden, da wir uns im Zeitintervall $[\tilde{t}, 2]$ befinden, noch aus einer Endbedingung zum Zeitpunkt 2, da $y(2)$ frei ist. Da y^* jedoch auf dem gesamten Intervall $[0, 2]$ stetig sein soll, folgt

$$y_I^*(\tilde{t}) = y_{II}^*(\tilde{t}) \iff 2(3e^{\tilde{t}} - 1) = c e^{\tilde{t}} \iff c = 2(3 - e^{-\tilde{t}})$$



In $(\tilde{t}, y^*(\tilde{t}))$ liegt ein Eckpunkt vor, da die links- und rechtsseitige Ableitung von y^* zum Zeitpunkt $\tilde{t}$ verschieden sind. Es gilt nämlich $\dot{y}_-^*(\tilde{t}) = 6e^{\tilde{t}} \neq \dot{y}_+^*(\tilde{t}) = 6e^{\tilde{t}} - 2$.

$$\begin{aligned} \text{iii)} \quad V(u) &= \int_0^1 -u^2 dt \rightarrow \max \\ \text{u.d.N.} \quad \dot{y} &= y + u \\ y(0) &= 1, \quad y(1) = 0 \end{aligned}$$

1. Schritt: Bestimmung der Hamiltonfunktion und Berechnung von u^*

$$H = -u^2 + \lambda(y + u)$$

Da u keiner Restriktion unterliegt, können wir das Maximum von H durch $\frac{\partial H}{\partial u} = 0$ bestimmen. Dies liefert:

$$-2u^* + \lambda^* = 0 \iff u^*(t) = \frac{1}{2}\lambda^*(t)$$

Wegen $\frac{\partial^2 H}{\partial u^2} = -2$ liegt ein Maximum in H vor.

2. Schritt: Bestimmung der Costate-Variablen $\lambda^*(t)$

$$\dot{\lambda}^* = -\frac{\partial H}{\partial y} = -\lambda^* \implies \lambda^*(t) = k e^{-t}$$

Da die Randbedingungen für y gegeben sind, muss zunächst y^* berechnet werden.

3. Schritt: Bestimmung der optimalen Trajektorie $y^*(t)$

Die Bewegungsgleichung $\dot{y}^* = y^* + u^* = y^* + \frac{1}{2}\lambda^* = y^* + \frac{1}{2}k e^{-t}$ liefert

$$y^*(t) = c e^t - \frac{1}{4}k e^{-t}$$

und aus den Anfangsbedingungen $y(0) = 1$, $y(1) = 0$ folgt

$$c = \frac{1}{1 - e^2}, \quad k = 4(c - 1) = \frac{4e^2}{1 - e^2}.$$

Die optimalen Pfade lauten also:

$$y^*(t) = \frac{e^t}{1 - e^2} - \frac{e^{2-t}}{1 - e^2}, \quad \lambda^*(t) = \frac{4e^{2-t}}{1 - e^2}, \quad u^*(t) = \frac{2e^{2-t}}{1 - e^2}$$

Der Lösungsweg ist relativ umständlich, weil der einfachste Fall der optimalen Kontrolltheorie der des freien Endzustandes ist (wie in Teil ii) dieses Beispiels). Sind wie in diesem Fall Anfangs- und Endzustand vorgegeben, kann λ und damit u erst relativ spät berechnet werden.

Man beachte, dass gemäß Bemerkung 4.32 vi) die Hamiltonfunktion entlang des optimalen Pfades konstant ist. Es gilt $H(y^*, u^*, \lambda^*, t) = \frac{4e^2}{(1 - e^2)^2} = \text{konst.}$

- iv) Betrachten wir das letzte Beispiel nochmals, allerdings mit der Bedingung $y(1) \geq 3$. Wie vorher ist also $t_1 = t^* = 1$, jedoch ist der Endzustand „fast“ frei. Wir bestimmen zunächst die Lösung y^* für „völlig freies“ $y(1)$. Wie vorher gilt

$$u^*(t) = \frac{1}{2}\lambda^*(t) \quad \text{und} \quad \lambda^*(t) = k e^{-t}$$

Die Transversalitätsbedingung $\lambda^*(t^*) = \lambda^*(1) = 0$ impliziert $k = 0$, d.h. $\lambda^* \equiv 0$ und damit $u^*(t) \equiv 0$. Aus der Bewegungsgleichung

$$\dot{y}^* = y^* + u^* = y^*$$

folgt daher $y^*(t) = c e^t$ mit $c = 1$ wegen $y(0) = 1$. Wir überprüfen, ob $y^*(1) \geq 3$ gilt. Dies ist nicht erfüllt. Um $y(1) \geq 3$ zu erfüllen, muss daher $y(1) = 3$ gelten. Dieses Problem wurde im vorangehenden Beispiel mit anderen Zahlenwerten gelöst.

$$\begin{aligned}
 \text{v)} \quad V &= \int_0^{t_1} -1 dt \rightarrow \max \\
 \text{u.d.N.} \quad \dot{y} &= y + u \\
 y(0) &= 5, \quad y(t_1) = 11, \quad t_1 \text{ frei} \\
 u(t) &\in \Omega = [-1, 1]
 \end{aligned}$$

Offensichtlich liegt ein Zeitminimierungsproblem vor, wegen

$$V = -t_1 \rightarrow \max \iff t_1 \rightarrow \min$$

1. Schritt: Bestimmung der Hamiltonfunktion und Berechnung von u^*

$$H = -1 + \lambda(y + u)$$

da H in u linear ist, wird das Maximum von H bzgl. u in einem der Randpunkte von Ω angenommen. Dies ist abhängig von der Steigung, die H bzgl. u besitzt. Offensichtlich gilt

$$u^*(t) = \begin{cases} +1 & \text{falls } \lambda > 0 \\ -1 & \text{falls } \lambda < 0 \end{cases}$$

2. Schritt: Bestimmung der Costate-Variablen $\lambda^*(t)$

$$\dot{\lambda}^* = -\frac{\partial H}{\partial y} = -\lambda^* \implies \lambda^* = k e^{-t}$$

Da das Vorzeichen von λ^* gleich dem Vorzeichen von k ist, hat $\lambda^*(t)$ auf dem gesamten Bereich ein einheitliches Vorzeichen. Insbesondere gilt

$$u^* \equiv 1 \quad \text{oder} \quad u^* \equiv -1$$

das Vorzeichen von k kann mit der Transversalitätsbedingung

$$\left[H \right]_{t=t^*} = 0$$

ermittelt werden. Wegen $y^*(t^*) = 11$ folgt aus der Transversalitätsbedingung

$$0 = \left[H \right]_{t=t^*} = -1 + k e^{-t^*} (11 + u^*).$$

da u^* nur die Werte ± 1 annehmen kann und $e^{-t^*} > 0$ gilt, folgt $k = \frac{e^{t^*}}{11 + u^*} > 0$. Damit gilt $\lambda^*(t) > 0$ und $u^*(t) \equiv 1$.

3. Schritt: Bestimmung der optimalen Trajektorie $y^*(t)$

Die Bewegungsgleichung lautet:

$$\dot{y}^* = y^* + u^* \iff \dot{y}^* - y^* = 1$$

Wegen $y(0) = 5$ ergibt sich

$$y^*(t) = 6e^t - 1$$

4. Schritt: Bestimmung von λ^*

$$\begin{aligned} 0 = \left[H \right]_{t=t^*} &= -1 + \lambda^*(y^* + u^*) \\ &= -1 + k e^{-t^*} (6e^{t^*} - 1 + 1) \\ &= -1 + 6k \\ \implies k = \frac{1}{6} &\implies \lambda^*(t) = \frac{1}{6} e^{-t} \end{aligned}$$

5. Schritt: Bestimmung von t^* .

t^* ergibt sich aus der Forderung $y^*(t^*) \stackrel{!}{=} 11$

$$\implies 6e^{t^*} - 1 = 11 \implies t^* = \ln(2) \approx 0.6931$$

- vi) Das folgende Modell stellt die ökonomische Interpretation eines auf Pontrjagin zurückgehenden mathematischen Beispiels dar.

$Y(t)$ bezeichne die Auslandsverschuldung eines Staates, $\dot{Y}_t$ deren Änderungsrate. Aufgrund internationaler Kreditwürdigkeit, ökonomischer Leistungsfähigkeit und sonstiger politischer Gründe sei $\ddot{Y}(t)$, d.h. die Fähigkeit zur Ab- bzw. Zunahme begrenzt, z.B. $|\ddot{Y}| \leq 1$.

Bei gegebener Anfangsverschuldung Y_0 und Änderungsrate $\dot{Y}_0$ Sorge man in kürzest möglicher Zeit für eine vollkommene Entschuldung. $\square$

Literaturverzeichnis

Kapitel 1, 2

- [1] M. Braun. Differentialgleichungen. Springer 1977 (3. Aufl.)
- [2] A. C. Chiang. Fundamental Methods of Mathematical Economics, Mc Graw Hill 1984
- [3] G. Gandolfo. Economic Dynamics. Springer 1977 (3. Aufl.)
- [4] M. Hirsch/ S. Smale . Differential Equations, Dynamical Systems and Linear Algebra. Academic Press 1974
- [5] E.Kamke, Gewöhnliche Differentialgleichungen. Akademische Verlagsgesellschaft. Leipzig 1969 (6. Aufl.)
- [6] D. Luenberger. Introduction to Dynamical Systems. J. Wiley & Sons, 1979
- [7] H. Meschkowski. Differenzengleichungen BI-Wissenschaftsverlag. Mannheim 1959
- [8] H. Rommelfanger. Differenzen- und Differentialgleichungen. BI-Verlag. Mannheim
- [9] K. Schindler. Mathematik für Ökonomen. DUV-Verlag. Wiesbaden 2005

Kapitel 3, 4

- [10] A. C. Chiang. Elements of Dynamic Optimization. Mc Graw Hill 1992
- [11] M. R. Hestenes. Calculus of Variations and Optimal Control Theory. J. Wiley & Sons, 1966
- [12] M. D. Intriligator. Mathematical Optimization and Economic Theory. Prentice-Hall.
- [13] J. LaSalle/ S. Lefshetz. Stability by Lajapunovs Direct Method. Academic Press 1961
- [14] P. N. V. Tu. Dynamical Systems. Springer 1994 (2. Aufl.)
- [15] K. Schindler. Mathematik für Ökonomen. DUV-Verlag. Wiesbaden 2005



Differentialrechnung

A.1 Definition

Eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ heißt *differenzierbar im Punkt t_0* , wenn der Grenzwert

$$a := \lim_{t \rightarrow t_0} \frac{f(t) - f(t_0)}{t - t_0}$$

existiert. $a \in \mathbb{R}$ wird *Differential* oder *Ableitung von f im Punkt t_0* genannt und man schreibt

$$df(t_0), \frac{df}{dt}(t_0), f'(t_0) \text{ oder } \dot{f}(t_0).$$

Die Funktion $f' : \mathbb{R} \rightarrow \mathbb{R}$ mit $t \mapsto f'(t)$ heißt *Differential* oder *Ableitung von f* und gibt das momentane absolute Änderungsverhalten von f im Punkt t an. Alternative Schreibweisen sind df oder $\dot{f}$. $\square$

Wesentlich bei der Ableitungsdefinition ist die Konvergenzuntersuchung, ob der „Differenzenquotient“ $\frac{f(t)-f(t_0)}{t-t_0}$ gegen a konvergiert. Dies geschieht, indem man zeigt, dass

$$\left| \frac{f(t) - f(t_0)}{t - t_0} - a \right|$$

beliebig klein wird, wenn t nur hinreichend nahe bei t_0 liegt.

Definition A.1 soll zunächst auf vektor- bzw. matrisenwertige Abbildungen in einer (!!)-Veränderlichen ausgedehnt werden. Sei hierzu $\mathbf{f} : \mathbb{R} \rightarrow \mathbb{R}^{m \times n}$ eine vorgegebene matrisenwertige Abbildung. Offensichtlich kann auch in diesem Fall der Differenzenquotient

$$\frac{\mathbf{f}(t) - \mathbf{f}(t_0)}{t - t_0} = \frac{1}{t - t_0} \cdot (\mathbf{f}(t) - \mathbf{f}(t_0))$$

gebildet werden. Zur Verallgemeinerung der Ableitungsdefinition müssen wir nur noch entscheiden, ob dieser Differenzenquotient für $t \rightarrow t_0$ konvergiert. Das Problem hierbei ist jedoch, dass $\mathbf{f}(t) - \mathbf{f}(t_0)$ und damit auch $\frac{1}{t - t_0} \cdot (\mathbf{f}(t) - \mathbf{f}(t_0))$ eine $(m \times n)$ -Matrix ist¹.

Wir benötigen daher eine Funktion $\|\cdot\|$ auf der Menge $\mathbb{R}^{m \times n}$ mit den gleichen Eigenschaften wie der Betrag auf den reellen Zahlen. Beispielhaft für solche Normen geben wir die *p-Norm* und die *Maximumsnorm* an.

¹Man beachte, dass $\mathbb{R}^{m \times n} = \mathbb{R}^{m \cdot n}$ gilt, so dass es im Prinzip genügt, Vektorräume der Form $\mathbb{R}^k$ als Bildbereich der auftretenden Abbildungen zu betrachten.

Für eine konstante reelle Zahl $p \geq 1$ definiert man die p -Norm $\|\cdot\|_p : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ durch

$$\|\mathbf{A}\|_p := \left(\sum_{i,j} |a_{ij}|^p \right)^{\frac{1}{p}} \quad \text{für } \mathbf{A} = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}} \in \mathbb{R}^{m \times n}$$

Im Spezialfall $p = \infty$ ergibt sich die sog. ∞ -Norm oder *Maximumsnorm*

$$\|\mathbf{A}\|_\infty = \|\mathbf{A}\|_{\max} := \max_{i,j} |a_{ij}|$$

Da diese Normen äquivalent sind² und daher der Stetigkeits- und Ableitungsbegriff norm-unabhängig sind, werden wir in Zukunft nicht explizit angeben, welche Norm verwendet wird. Um die Analogie zu den reellen Zahlen möglichst groß zu halten (der Betrag ist eine Norm auf $\mathbb{R}$), werden wir in Zukunft auch die Betragsstriche $|\cdot|$ statt der Normstriche $\|\cdot\|$ verwenden.

A.2 Definition

Eine Abbildung $\mathbf{f} : \mathbb{R} \rightarrow \mathbb{R}^{m \times n}$ heißt *differenzierbar im Punkt t_0* , wenn der Grenzwert

$$\mathbf{A} := \lim_{t \rightarrow t_0} \frac{\mathbf{f}(t) - \mathbf{f}(t_0)}{t - t_0}$$

existiert. Die Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ wird das *Differential* oder die *Ableitung von $\mathbf{f}$ im Punkt t_0* genannt. Man schreibt

$$d\mathbf{f}(t_0), \frac{d\mathbf{f}}{dt}(t_0), \mathbf{f}'(t_0) \text{ oder } \dot{\mathbf{f}}(t_0).$$

Die Abbildung $\mathbf{f}' : \mathbb{R} \rightarrow \mathbb{R}^{m \times n}$ mit $t \mapsto \mathbf{f}'(t) = d\mathbf{f}(t)$ heißt *Differential* oder *Ableitung* von $\mathbf{f}$ und man schreibt $d\mathbf{f}$, $\mathbf{f}'$ oder $\dot{\mathbf{f}}$. Interpretiert man $\mathbf{f}$ als Bahn eines (oder mehrerer) Teilchen, gibt $\mathbf{f}'(t_0)$ die momentane Geschwindigkeit, d.h. das momentane Ortsänderungsverhalten des Teilchens, an. $\square$

Am einfachsten ist die Verwendung der Maximumsnorm bei der Grenzwertdefinition. Man sieht dabei nämlich sofort, dass eine Folge von Matrizen $(\mathbf{A}_k)_{k=1}^\infty$ in $\mathbb{R}^{m \times n}$ mit $\mathbf{A}_k = (a_{ij}^{(k)})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$ genau dann gegen die Matrix $\mathbf{G} = (g_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$ konvergiert, wenn jede Komponente der Matrizenfolge $(\mathbf{A}_k)_{k=1}^\infty$ einzeln in $\mathbb{R}$ (!) gegen die entsprechende Komponente von $\mathbf{G}$ konvergiert, d.h.

$$\lim_{k \rightarrow \infty} \mathbf{A}_k = \mathbf{G} \iff \forall i = 1, \dots, m \quad \forall j = 1, \dots, n : \lim_{k \rightarrow \infty} a_{i,j}^{(k)} = g_{i,j}.$$

Dies führt zu folgender Ableitungsregel.

²Dies bedeutet, dass die Konvergenz einer Folge in $\mathbb{R}^{m \times n}$ unabhängig von der Wahl der gewählten Norm ist.

A.3 Satz

Ist $\mathbf{f} : \mathbb{R} \rightarrow \mathbb{R}^{m \times n}$ differenzierbar mit $\mathbf{f}(t) = \begin{pmatrix} f_{11}(t) & \dots & f_{1n}(t) \\ \vdots & \ddots & \vdots \\ f_{m1}(t) & \dots & f_{mn}(t) \end{pmatrix}$, so gilt:

$$\mathbf{f}'(t) = \begin{pmatrix} f'_{11}(t) & \dots & f'_{1n}(t) \\ \vdots & \ddots & \vdots \\ f'_{m1}(t) & \dots & f'_{mn}(t) \end{pmatrix} \quad \square$$

Eine Verallgemeinerung dieser Ableitungsdefinition auf Abbildungen, deren Definitionsbereich mehrdimensional ist, d.h., bei denen mehr als eine Variable vorliegt, ist nicht ohne weiteres möglich. In diesem Fall kann nämlich der Differenzenquotient $\frac{1}{t-t_0} \cdot (\mathbf{f}(t) - \mathbf{f}(t_0))$ nicht mehr gebildet werden, weil $t-t_0$ ein Vektor ist, durch den man nicht „dividieren“ kann. Aus diesem Grund benötigt man eine äquivalente Ableitungsdefinition, die auch auf den Fall mehrerer Veränderlicher übertragen werden kann. Dies leistet der folgende Satz.

A.4 Satz

Eine Abbildung $\mathbf{f} : \mathbb{R} \rightarrow \mathbb{R}^{m \times n}$ ist im Punkt t_0 genau dann differenzierbar, wenn eine $(\ell \times n)$ -Matrix $\mathbf{A}$ existiert mit

$$\lim_{t \rightarrow t_0} \frac{|\mathbf{f}(t) - [\mathbf{f}(t_0) + \mathbf{A} \cdot (t - t_0)]|}{|t - t_0|} = 0$$

In diesem Fall gilt $\mathbf{f}'(t_0) = \mathbf{A}$. □

Der letzte Satz zeigt, wie sich die Ableitungsdefinition auf Funktionen verallgemeinern lässt, deren Definitionsbereich aus Vektoren besteht.

A.5 Definition

Eine Abbildung $\mathbf{f} : \mathbb{R}^k \rightarrow \mathbb{R}^\ell$ heißt im Punkt $\mathbf{x}_0$ (*total*) *differenzierbar*, wenn eine $(\ell \times k)$ -Matrix $\mathbf{A}$ existiert mit³

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{|\mathbf{f}(\mathbf{x}) - [\mathbf{f}(\mathbf{x}_0) + \mathbf{A}(\mathbf{x} - \mathbf{x}_0)]|}{|\mathbf{x} - \mathbf{x}_0|} = 0.$$

$\mathbf{A}$ heißt das *Differential* oder die *Ableitung von $\mathbf{f}$ an der Stelle $\mathbf{x}_0$* . Es sind folgende Schreibweisen für $\mathbf{A}$ üblich:

$$\mathbf{f}'(\mathbf{x}_0), D\mathbf{f}(\mathbf{x}_0), d\mathbf{f}(\mathbf{x}_0) \text{ oder } \frac{d\mathbf{f}}{d\mathbf{x}}(\mathbf{x}_0)$$

Die Abbildung $\mathbf{x} \mapsto d\mathbf{f}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})$ heißt *Differential* bzw. *Ableitung von $\mathbf{f}$* und man schreibt $D\mathbf{f}$ ($d\mathbf{f}$) bzw. $\mathbf{f}'$. □

³Nimmt $\mathbf{f}$ Werte in $\mathbb{R}^{m \times n}$ an, d.h. ist $\ell = m \times n$, ist $\mathbf{A}$ entsprechend eine „dreidimensionale“ $(m \times n) \times k$ -Matrix.

Der große Nachteil von Definition A.5 ist, dass eine explizite Rechenvorschrift zur Bestimmung der Ableitung nicht gegeben wird. Eine direkte Berechnung der Ableitung wie im klassischen Fall als Grenzwert des Differenzenquotienten ist nicht möglich. Definition A.5 ist daher für die direkte Berechnung der Ableitung kaum zu verwenden.

Glücklicherweise lässt sich in den meisten praktischen Fällen die Berechnung der Ableitungsmatrix auf den Fall einer Veränderlichen zurückführen. Hierzu arbeitet man mit den sog. partiellen Ableitungen. Diese entstehen dadurch, dass man eine Abbildung in mehreren Veränderlichen künstlich zu einer Abbildung in einer Veränderlichen macht, indem man nur noch eine Variable verändert und die anderen konstant hält.

A.6 Definition

Sei $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^\ell$, $\mathbf{x} = (x_1, \dots, x_n) \mapsto \mathbf{f}(\mathbf{x})$ eine Abbildung. Hält man alle Variablen mit Ausnahme der i -ten konstant, entsteht eine Abbildung in einer Veränderlichen:

$$x \mapsto f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) .$$

Ist diese eindimensional differenzierbar im Sinne von Definition A.2, so wird die zugehörige Ableitung als *partielle Ableitung* von $\mathbf{f}$ bezeichnet. Man schreibt

$$D_i \mathbf{f} \text{ oder } \frac{\partial \mathbf{f}}{\partial x_i} .$$

□

Mit dem Begriff der partiellen Ableitung lässt sich die gesuchte Ableitungsmatrix von $\mathbf{f}$ nun einfach berechnen.

A.7 Satz

Sei $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^\ell$ eine Abbildung. $\mathbf{f}$ bestehe aus den Komponentenfunktionen $f_1, \dots, f_\ell$, d.h. es ist $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_\ell(\mathbf{x}))$. Dann folgt:

a) Ist $\mathbf{f}$ in $\mathbf{x}_0$ differenzierbar, so gilt:

$$\mathbf{f}'(\mathbf{x}_0) = (D_1 \mathbf{f}, \dots, D_n \mathbf{f})(\mathbf{x}_0) = \begin{pmatrix} D_1 f_1(\mathbf{x}_0) & D_2 f_1(\mathbf{x}_0) & \dots & D_n f_1(\mathbf{x}_0) \\ D_1 f_2(\mathbf{x}_0) & D_2 f_2(\mathbf{x}_0) & \dots & D_n f_2(\mathbf{x}_0) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 f_\ell(\mathbf{x}_0) & D_2 f_\ell(\mathbf{x}_0) & \dots & D_n f_\ell(\mathbf{x}_0) \end{pmatrix}$$

bzw. in Differentialschreibweise

$$d\mathbf{f} = \mathbf{f}' = \frac{\partial \mathbf{f}}{\partial x_1} dx_1 + \dots + \frac{\partial \mathbf{f}}{\partial x_n} dx_n = (D_1 \mathbf{f}) dx_1 + \dots + (D_n \mathbf{f}) dx_n .$$

b) $\mathbf{f}'$ existiert und ist stetig genau dann, wenn alle partiellen Ableitungen $D_j f_i$ existieren und stetig sind. $\mathbf{f}$ heißt in diesem Fall *stetig differenzierbar*. □

Beachtet man, dass durch die Abbildung $\mathbf{x} \mapsto \mathbf{f}(\mathbf{x}_0) + \mathbf{A}(\mathbf{x} - \mathbf{x}_0)$ gerade die Tangente bzw. Tangentialebene an den Graphen von $\mathbf{f}$ beschrieben wird, ergibt sich aus Definition

A.5 der allgemeinen Ableitung die bekannte Eigenschaft, dass in einer Umgebung von $\mathbf{x}_0$ die Tangente bzw. Tangentialebene, d.h. die Abbildung $\mathbf{x} \mapsto \mathbf{f}(\mathbf{x}_0) + \mathbf{f}'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$ die beste affin-lineare Approximation für $\mathbf{f}$ ist. Führt man nämlich in Definition A.5 die Bezeichnung

$$\mathbf{R}(\mathbf{x}) := \frac{\mathbf{f}(\mathbf{x}) - [\mathbf{f}(\mathbf{x}_0) + \mathbf{A}(\mathbf{x} - \mathbf{x}_0)]}{|\mathbf{x} - \mathbf{x}_0|}$$

ein, erhält man die folgende Aussage.

A.8 Satz

Sei $\mathbf{f} : \mathbb{R}^k \rightarrow \mathbb{R}^\ell$ im Punkt $\mathbf{x}_0$ differenzierbar. Dann gilt

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_0) + \mathbf{f}'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + |\mathbf{x} - \mathbf{x}_0| \cdot \mathbf{R}(\mathbf{x}) .$$

$\mathbf{R}$ hat hierbei die Eigenschaft $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \mathbf{R}(\mathbf{x}) = \mathbf{0}$. □

Satz A.8 ist ein Spezialfall des Satzes von Taylor. Ist die vorgegebene Funktion $\mathbf{f}$ sogar mehrfach differenzierbar, sind bessere Approximationen durch Polynome höherer Ordnung möglich. Die zugehörige Aussage (Satz von Taylor) soll für Funktionen in einer Veränderlichen formuliert werden, gilt aber analog für Funktionen in mehreren Veränderlichen.

A.9 Satz (Taylor)

Sei $f :]a, b[\rightarrow \mathbb{R}$ eine $(n+1)$ -mal differenzierbare Funktion und sei $x_0 \in]a, b[$. Dann existiert zu jedem $x \in]a, b[$ ein ξ zwischen x und x_0 , so dass gilt

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n \\ &\quad + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{n+1}. \end{aligned} \quad \square$$

Nachfolgende Reihenentwicklungen (Taylorreihen) einiger wichtiger Funktionen ergeben sich, wenn man im Satz von Taylor $x_0=0$ und $n=\infty$ wählt.

$$\begin{aligned} \sin(x) &= x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \frac{1}{9!}x^9 - \dots = \sum_{n=0}^{\infty} (-1)^n \cdot \frac{x^{2n+1}}{(2n+1)!} \\ \cos(x) &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + \frac{1}{8!}x^8 - \dots = \sum_{n=0}^{\infty} (-1)^n \cdot \frac{x^{2n}}{(2n)!} \\ e^x &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \dots = \sum_{n=0}^{\infty} \frac{x^n}{n!} \\ \ln(1+x) &= x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \dots = \sum_{n=0}^{\infty} (-1)^{n+1} \cdot \frac{x^n}{n} ; \quad -1 < x \leq 1 \\ \ln(1-x) &= -(x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \frac{1}{4}x^4 + \dots) = -\sum_{n=0}^{\infty} \frac{x^n}{n} ; \quad -1 \leq x < 1 \end{aligned}$$

Die folgenden Taylorreihen lassen auch den Ursprung der Bezeichnungen *Sinus* bzw. *Cosinus hyperbolicus* für die Funktionen $\frac{1}{2}(e^x - e^{-x})$ bzw. $\frac{1}{2}(e^x + e^{-x})$ erkennen.

$$\begin{aligned}\ln\left(\frac{1+x}{1-x}\right) &= 2\left(x + \frac{1}{3}x^3 + \frac{1}{5}x^5 + \frac{1}{7}x^7 + \dots\right) = 2\sum_{n=0}^{\infty} \frac{x^{2n+1}}{2n+1}; \quad -1 < x < 1 \\ \sinh(x) &:= \frac{e^x - e^{-x}}{2} = x + \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + \frac{1}{7!}x^7 + \dots = \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!} \\ \cosh(x) &:= \frac{e^x + e^{-x}}{2} = 1 + \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + \frac{1}{6!}x^6 + \dots = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} \\ \arctan(x) &= x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \dots = \sum_{n=0}^{\infty} (-1)^n \cdot \frac{x^{2n+1}}{2n+1} \\ \frac{1}{x-a} &= -\sum_{n=0}^{\infty} \frac{x^n}{a^{n+1}} = -\frac{1}{a} \sum_{n=0}^{\infty} \left(\frac{x}{a}\right)^n; \quad |x| < |a| \text{ (Geometrische Reihe)} \\ \frac{1}{x} &= \sum_{n=0}^{\infty} (1-x)^n; \quad 0 < x < 2\end{aligned}$$

Die Taylordarstellung von Funktionen gestattet es, den Definitionsbereich von Funktionen auf Matrizen oder Operatoren auszuweiten. Für den Differentialoperator D erhält man z.B.⁴

$$\sin(D) = D - \frac{1}{3!}D^3 + \frac{1}{5!}D^5 - \frac{1}{7!}D^7 + \frac{1}{9!}D^9 - \dots$$

Ist $\mathbf{M}$ eine quadratische Matrix, definiert man

$$e^{\mathbf{M}} = 1 + \mathbf{M} + \frac{\mathbf{M}^2}{2!} + \frac{\mathbf{M}^3}{3!} + \frac{\mathbf{M}^4}{4!} + \frac{\mathbf{M}^5}{5!} + \dots$$

Das folgende Beispiel zeigt, wie man konkret die Exponentialmatrix $e^{\mathbf{M}}$ berechnet.

A.10 Beispiel

i) Für eine Blockdiagonalmatrix

$$\mathbf{M} = \text{diag}(\mathbf{B}_1, \dots, \mathbf{B}_k) = \begin{pmatrix} \mathbf{B}_1 & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \mathbf{B}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{O} \\ \mathbf{O} & \dots & \mathbf{O} & \mathbf{B}_k \end{pmatrix}$$

mit quadratischen Matrizen $\mathbf{B}_1, \dots, \mathbf{B}_k$ und passenden Nullmatrizen $\mathbf{O}$ gilt offensichtlich:

$$\mathbf{M}^j = \begin{pmatrix} \mathbf{B}_1^j & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \mathbf{B}_2^j & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{O} \\ \mathbf{O} & \dots & \mathbf{O} & \mathbf{B}_k^j \end{pmatrix}$$

⁴Hierbei ist immer die Frage der Konvergenz zu klären.

Daraus ergibt sich:

$$e^{\mathbf{M}} = e^{\text{diag}(\mathbf{B}_1, \dots, \mathbf{B}_k)} = \text{diag}(e^{\mathbf{B}_1}, \dots, e^{\mathbf{B}_k}) = \begin{pmatrix} e^{\mathbf{B}_1} & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & e^{\mathbf{B}_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{O} \\ \mathbf{O} & \dots & \mathbf{O} & e^{\mathbf{B}_k} \end{pmatrix}$$

Insbesondere erhält man für eine Diagonalmatrix $\mathbf{M} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$ mit $\lambda_i \in \mathbb{C}$:

$$e^{\mathbf{M}} = \text{diag}(e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_k}) = \begin{pmatrix} e^{\lambda_1} & 0 & \dots & 0 \\ 0 & e^{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & e^{\lambda_k} \end{pmatrix}$$

ii) Sind $\mathbf{M}$, $\mathbf{N}$ und $\mathbf{B}$ quadratische $(n \times n)$ -Matrizen mit $\mathbf{M} = \mathbf{B} \cdot \mathbf{N} \cdot \mathbf{B}^{-1}$, so folgt:

$$\mathbf{M}^j = (\mathbf{B} \mathbf{N} \mathbf{B}^{-1})^j = (\mathbf{B} \mathbf{N} \mathbf{B}^{-1}) \cdot (\mathbf{B} \mathbf{N} \mathbf{B}^{-1}) \cdots (\mathbf{B} \mathbf{N} \mathbf{B}^{-1}) = \mathbf{B} \mathbf{N}^j \mathbf{B}^{-1}$$

woraus wiederum folgt:

$$e^{\mathbf{M}} = \sum_{j=0}^{\infty} \frac{(\mathbf{B} \mathbf{N} \mathbf{B}^{-1})^j}{j!} = \sum_{j=0}^{\infty} \mathbf{B} \frac{\mathbf{N}^j}{j!} \mathbf{B}^{-1} = \mathbf{B} \left(\sum_{j=0}^{\infty} \frac{\mathbf{N}^j}{j!} \right) \mathbf{B}^{-1} = \mathbf{B} e^{\mathbf{N}} \mathbf{B}^{-1}$$

Diese Eigenschaft zur Berechnung von $e^{\mathbf{M}}$ ist immer dann nützlich, wenn eine Basis von Eigenvektoren der Matrix $\mathbf{M}$ existiert. In genau diesem Fall kann $\mathbf{M}$ nämlich in der Form $\mathbf{B} \cdot \text{diag}(\lambda_1, \dots, \lambda_n) \cdot \mathbf{B}^{-1}$ dargestellt werden, wobei λ_i (evtl. gleiche) Eigenwerte sind und die Spalten der Matrix $\mathbf{B}$ aus Eigenvektoren (der Stufe 1) von $\mathbf{M}$ bestehen und man erhält

$$e^{\mathbf{M}} = \mathbf{B} \cdot \text{diag}(e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_k}) \cdot \mathbf{B}^{-1}$$

iii) Der unangenehmste Fall bei der Berechnung der Exponentialmatrix $e^{\mathbf{M}}$ liegt vor, wenn keine Basis von Eigenvektoren (der Stufe 1) der Matrix $\mathbf{M}$ existiert, da dann eine Diagonalisierung von $\mathbf{M}$ nicht mehr möglich ist. In dieser Situation greift man auf die sog. *Jordan-Normalform* von Matrizen zurück (siehe Satz D.6 dieses Anhangs), die eine zu $\mathbf{M}$ äquivalente „fast“ diagonale (*bidiagonale*) Jordanmatrix $\mathbf{J}$ liefert. Man bestimmt hierzu eine invertierbare Matrix $\mathbf{B}$ mit

$$\mathbf{M} = \mathbf{B} \cdot \text{diag}(\mathbf{J}_1, \dots, \mathbf{J}_\ell) \cdot \mathbf{B}^{-1}$$

und den Jordanblöcken $\mathbf{J}_1, \dots, \mathbf{J}_\ell$. Nach Teil ii) dieses Beispiels gilt dann

$$e^{\mathbf{M}} = \mathbf{B} \cdot \text{diag}(e^{\mathbf{J}_1}, \dots, e^{\mathbf{J}_\ell}) \cdot \mathbf{B}^{-1},$$

so dass wir also nur noch klären müssen, wie e^{J_k} aussieht.

Definiert man die „Nebendiagonaleinheitsmatrix“ $E^{(k)} \in \mathbb{R}^{n \times n}$, die auf der k -ten Nebendiagonale nur Einsen besitzt, d.h.

$$[E^{(k)}]_{i,j} := \delta_{i,j+k},$$

so lässt sich ein Jordanblock $J \in \mathbb{R}^{n \times n}$

$$J = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \lambda & 1 \\ 0 & \dots & \dots & 0 & \lambda \end{pmatrix} = \begin{pmatrix} \lambda & 0 & 0 & \dots & 0 \\ 0 & \lambda & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \lambda & 0 \\ 0 & \dots & \dots & 0 & \lambda \end{pmatrix} + \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & 1 \\ 0 & \dots & \dots & 0 & 0 \end{pmatrix}$$

darstellen in der Form

$$J = \lambda \cdot \mathbf{1} + E^{(1)}.$$

Mit den Rechenregeln

- 1) $E^{(k)} = \mathcal{O}$ für $|k| > n-1$
- 2) $E^{(k)} \cdot E^{(\ell)} = E^{(k+\ell)}$
- 3) $(E^{(1)})^\ell = E^{(\ell)}$

soll nun die Exponentialmatrix $e^{J \cdot t}$ berechnet werden⁵. Mit dem binomischen Lehrsatz und Rechenregeln 1) und 3) folgt zunächst

$$J^m = (\lambda \cdot \mathbf{1} + E^{(1)})^m = \sum_{j=0}^m \binom{m}{j} \lambda^{m-j} (E^{(1)})^j = \sum_{j=0}^{n-1} \binom{m}{j} \lambda^{m-j} E^{(j)}$$

Damit ergibt sich

$$\begin{aligned} e^{J \cdot t} &= \sum_{m=0}^{\infty} \frac{t^m}{m!} \cdot J^m = \sum_{m=0}^{\infty} \frac{t^m}{m!} \cdot \sum_{j=0}^{n-1} \underbrace{\binom{m}{j}}_{0 \text{ für } m < j} \lambda^{m-j} E^{(j)} \\ &= \sum_{m=j}^{\infty} \frac{t^m}{m!} \cdot \sum_{j=0}^{n-1} \binom{m}{j} \lambda^{m-j} E^{(j)} \\ &= \sum_{j=0}^{n-1} \frac{t^j}{j!} \cdot E^{(j)} \overbrace{\sum_{m=j}^{\infty} \frac{t^{m-j} \lambda^{m-j}}{(m-j)!}}^{= e^{\lambda t}} = e^{\lambda t} \cdot \sum_{j=0}^{n-1} \frac{t^j}{j!} \cdot E^{(j)} \end{aligned}$$

⁵Der Faktor t tritt bei der Lösung von DGLen auf.

Als Matrix hat $e^{\mathbf{J} \cdot t}$ folgendes Aussehen (beachte: $\mathbf{E}^{(0)} = \mathbf{1}$)

$$e^{\mathbf{J} \cdot t} = e^{\lambda t} \cdot \begin{pmatrix} 1 & t & \frac{t^2}{2} & \dots & \frac{t^{n-1}}{(n-1)!} \\ 0 & 1 & t & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \frac{t}{2} \\ \vdots & \ddots & \ddots & 1 & t \\ 0 & \dots & \dots & 0 & 1 \end{pmatrix}$$

□



Banachscher Fixpunktsatz

Mathematische Modelle führen fast zwangsläufig auf Gleichungen bzw. Gleichungssysteme der Form $\mathbf{g}(\mathbf{x}) = \boldsymbol{\alpha}$ ($\boldsymbol{\alpha}$ konstant) mit einer i.d.R. nichtlinearen Funktion $\mathbf{g}$. Diese Gleichungen können äquivalent als Fixpunktproblem formuliert werden, bei denen man für eine vorgegebene Funktion $\mathbf{F}$ ein $\mathbf{x}$ mit $\mathbf{F}(\mathbf{x}) = \mathbf{x}$ sucht. So ist die Gleichung $\mathbf{g}(\mathbf{x}) = \boldsymbol{\alpha}$ äquivalent zu dem Fixpunktproblem

$$\mathbf{F}(\mathbf{x}) := \mathbf{g}(\mathbf{x}) - \boldsymbol{\alpha} + \mathbf{x} = \mathbf{x}$$

Der folgende Fixpunktsatz liefert ein elegantes Verfahren zur Berechnung solcher Fixpunkte. Mit seiner Hilfe kann z.B. auch das Newton-Verfahren zur numerischen Berechnung von Nullstellen abgeleitet werden. Er gilt auch in beliebigen normierten Räumen und nicht nur im $\mathbb{R}^n$

B.1 Satz (Banachscher Fixpunktsatz)

Sei $\mathbb{D} \subset \mathbb{R}^n$ eine abgeschlossene Menge und $\Phi : \mathbb{D} \rightarrow \mathbb{D}$ eine Kontraktion¹, d.h. es existiert eine Konstante K mit $0 < K < 1$, so dass $|\Phi(\mathbf{x}) - \Phi(\mathbf{y})| \leq K \cdot |\mathbf{x} - \mathbf{y}|$ gilt. Dann besitzt die Gleichung $\Phi(\mathbf{x}) = \mathbf{x}$ genau eine Lösung $\mathbf{z}$, und für jeden Startwert $\mathbf{x}_0 \in \mathbb{D}$ konvergiert die iterativ definierte Folge $(\mathbf{x}_k)_{k=0}^{\infty}$ mit $\mathbf{x}_{k+1} := \Phi(\mathbf{x}_k)$ gegen den *Fixpunkt* $\mathbf{z}$. $\square$

Beweis:

Wir zeigen zunächst die Eindeutigkeit. Seien hierzu $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{D}$ zwei Fixpunkte. Dann gilt nach Voraussetzung:

$$|\mathbf{z}_1 - \mathbf{z}_2| = |\Phi(\mathbf{z}_1) - \Phi(\mathbf{z}_2)| \leq K \cdot |\mathbf{z}_1 - \mathbf{z}_2|$$

Wegen $0 < K < 1$ muss daher $|\mathbf{z}_1 - \mathbf{z}_2| = 0$ und damit $\mathbf{z}_1 - \mathbf{z}_2 = \mathbf{0}$ gelten.

Wir zeigen nun, dass bei beliebig vorgegebenem $\mathbf{x}_0 \in \mathbb{D}$ die Folge $(\mathbf{x}_k)_{k=0}^{\infty}$ konvergiert und dass der Grenzwert ein Fixpunkt von Φ ist.

Um die Konvergenz der Folge $(\mathbf{x}_k)_{k=1}^{\infty}$ zu beweisen, genügt es zu zeigen, dass $(\mathbf{x}_k)_{k=0}^{\infty}$ eine Cauchy-Folge ist. Hierzu beachte man zunächst die Ungleichung

$$|\mathbf{x}_{j+1} - \mathbf{x}_j| = |\Phi(\mathbf{x}_j) - \Phi(\mathbf{x}_{j-1})| \leq K \cdot |\mathbf{x}_j - \mathbf{x}_{j-1}|.$$

¹Daraus folgt insbesondere die Stetigkeit von Φ .

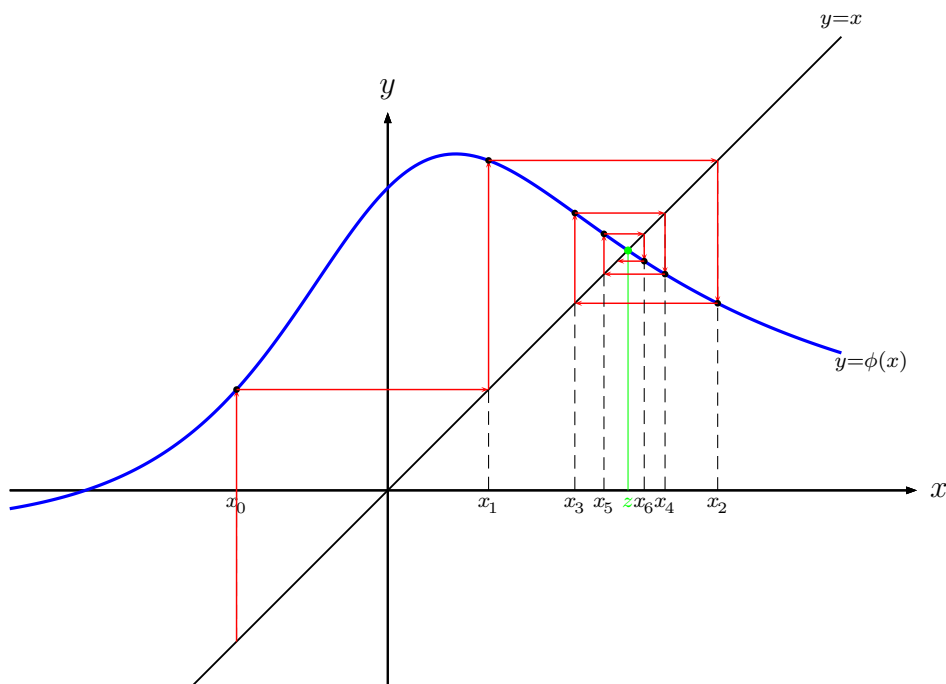


Abbildung B.1.: Iteration zur Fixpunktbestimmung im Fall $n = 1$

Mehrfache Anwendung dieser Ungleichung liefert

$$|\mathbf{x}_j - \mathbf{x}_{j-1}| \leq K^{j-1} \cdot |\mathbf{x}_1 - \mathbf{x}_0|.$$

Sind nun $m, n \in \mathbb{N}$ zwei natürliche Zahlen mit $n > m$, so ergibt sich aus der Dreiecksungleichung und der Summenformel für geometrische Reihen:

$$\begin{aligned} |\mathbf{x}_n - \mathbf{x}_m| &= \left| \sum_{j=m+1}^n (\mathbf{x}_j - \mathbf{x}_{j-1}) \right| \leq \sum_{j=m+1}^n |\mathbf{x}_j - \mathbf{x}_{j-1}| \leq |\mathbf{x}_1 - \mathbf{x}_0| \sum_{j=m+1}^n K^{j-1} \\ &= \frac{K^n - K^m}{K - 1} \cdot |\mathbf{x}_1 - \mathbf{x}_0| \leq \frac{K^n}{K - 1} \cdot |\mathbf{x}_1 - \mathbf{x}_0| \end{aligned}$$

Da der letzte Ausdruck mit $n \rightarrow \infty$ gegen 0 konvergiert, ist $(\mathbf{x}_k)_{k=0}^\infty$ eine Cauchy-Folge und damit konvergent. Wegen der Abgeschlossenheit von $\mathbb{D}$ liegt $\mathbf{z} := \lim_{k \rightarrow \infty} \mathbf{x}_k$ in $\mathbb{D}$. Aus der Stetigkeit von Φ folgt schließlich die Fixpunkteigenschaft von $\mathbf{z}$.

$$\mathbf{z} = \lim_{k \rightarrow \infty} \mathbf{x}_{k+1} = \lim_{k \rightarrow \infty} \Phi(\mathbf{x}_k) = \Phi\left(\lim_{k \rightarrow \infty} \mathbf{x}_k\right) = \Phi(\mathbf{z}).$$

■



Komplexe Zahlen

Der Aufbau des Zahlensystems ist am leichtesten zu verstehen, wenn man ihn im Zusammenhang mit der Lösung von Gleichungen betrachtet. Zum Beispiel besitzt die Gleichung

$$a \cdot x = b \quad (a, b \in \mathbb{Z})$$

nur in Ausnahmefällen eine ganzzahlige Lösung x . Erst durch die Konstruktion der rationalen Zahlen $\mathbb{Q}$ ist es möglich, diese Gleichung für alle $a \neq 0$ zu lösen. Die Grundidee, die sich bei den reellen und komplexen Zahlen in gleicher Weise finden lässt, besteht darin, eine größere Zahlenmenge $\mathbb{Q}$ zu schaffen, in der die ganzen Zahlen $\mathbb{Z}$ „eingebettet“ sind. Diese größere Menge muss also zwei Eigenschaften haben:

1. Sie muss die ganzen Zahlen $\mathbb{Z}$ als Teilmenge enthalten, bzw. eine Teilmenge besitzen, die mit den ganzen Zahlen in eindeutiger Weise identifiziert werden kann.
2. Die größere Menge muss eine algebraische Struktur (Addition, Multiplikation) besitzen, die auf der Teilmenge $\mathbb{Z}$ mit der ursprünglichen algebraischen Struktur übereinstimmt.

Diese Erweiterung der Zahlenmenge wird man als sinnvoll empfinden, wenn die Gleichung

$$a \cdot x = b \quad (a, b \in \mathbb{Z})$$

innerhalb des neuen Zahlenkörpers in mehr Fällen als früher lösbar ist. Akzeptabel wird diese Vorgehensweise vor allem dann, wenn man diese „neuen“ Lösungen auch praktisch verwerten kann. Dies liegt daran, dass die nicht ganzzahligen Lösungen neue Objekte sind, die erst einer Interpretation bedürfen. Diese nicht zu unterschätzende Schwierigkeit wird z.B. bei Kindern deutlich, für die eine Zahl wie 1,734 zunächst unverständlich ist.

Wie sieht der Übergang von $\mathbb{Z}$ zu $\mathbb{Q}$ nun mathematisch aus.

Wir definieren $\mathbb{Q} := \mathbb{Z}^2 = \left\{ \begin{pmatrix} p \\ q \end{pmatrix} \mid p, q \in \mathbb{Z}, q \neq 0 \right\}$ und überprüfen die vorher erwähnten zwei Kriterien. Hier sollte beachtet werden, dass $\begin{pmatrix} p \\ q \end{pmatrix}$ später als „Bruch“ $\frac{p}{q}$ bezeichnet wird.

Punkt 1: $\mathbb{Z}$ ist eine Teilmenge von $\mathbb{Q}$.

Da $\mathbb{Q}$ eine Menge von Zahlenpaaren ist, kann $\mathbb{Z}$ unmöglich eine Teilmenge von $\mathbb{Q}$ sein. Wir erreichen das Ziel jedoch dadurch, dass wir die Zahlenpaare der Form $\begin{pmatrix} p \\ 1 \end{pmatrix}$ mit der ganzen Zahl p identifizieren. Dies entspricht gerade der Identifikation $\frac{p}{1} \hat{=} p$.

Punkt 2: Definition von Addition und Multiplikation auf $\mathbb{Q}$.

Wir definieren für $\begin{pmatrix} p_1 \\ q_1 \end{pmatrix}, \begin{pmatrix} p_2 \\ q_2 \end{pmatrix} \in \mathbb{Q}$:

$$\begin{pmatrix} p_1 \\ q_1 \end{pmatrix} \oplus \begin{pmatrix} p_2 \\ q_2 \end{pmatrix} := \begin{pmatrix} p_1 \cdot q_2 + p_2 \cdot q_1 \\ q_1 \cdot q_2 \end{pmatrix}$$

$$\begin{pmatrix} p_1 \\ q_1 \end{pmatrix} \odot \begin{pmatrix} p_2 \\ q_2 \end{pmatrix} := \begin{pmatrix} p_1 \cdot p_2 \\ q_1 \cdot q_2 \end{pmatrix}$$

Trotz der im Vergleich zu den ganzen Zahlen befremdend wirkenden Addition (Hauptnenner!) lässt sich schnell nachweisen, dass diese beiden Verknüpfungen auf den ganzen Zahlen mit der ursprünglichen Addition bzw. Multiplikation übereinstimmen.

Sind nämlich $\begin{pmatrix} z_1 \\ 1 \end{pmatrix}, \begin{pmatrix} z_2 \\ 1 \end{pmatrix} \in \mathbb{Q}$, so gilt (mit der Identifikation $\begin{pmatrix} z \\ 1 \end{pmatrix} \hat{=} z$)

$$z_1 \oplus z_2 \hat{=} \begin{pmatrix} z_1 \\ 1 \end{pmatrix} \oplus \begin{pmatrix} z_2 \\ 1 \end{pmatrix} \stackrel{\text{Def.}}{=} \begin{pmatrix} z_1 \cdot 1 + z_2 \cdot 1 \\ 1 \cdot 1 \end{pmatrix} = \begin{pmatrix} z_1 + z_2 \\ 1 \end{pmatrix} \hat{=} z_1 + z_2$$

und

$$z_1 \odot z_2 \hat{=} \begin{pmatrix} z_1 \\ 1 \end{pmatrix} \odot \begin{pmatrix} z_2 \\ 1 \end{pmatrix} \stackrel{\text{Def.}}{=} \begin{pmatrix} z_1 \cdot z_2 \\ 1 \cdot 1 \end{pmatrix} = \begin{pmatrix} z_1 \cdot z_2 \\ 1 \end{pmatrix} \hat{=} z_1 \cdot z_2$$

Führt man die Kurzschreibweise $\frac{p}{q}$ anstelle von $\begin{pmatrix} p \\ q \end{pmatrix}$ ein¹ und notiert die Addition und Multiplikation wie vorher, erhält man die gewohnte Menge der rationalen Zahlen mit ihrer gewohnten algebraischen Struktur. Insbesondere besitzt die Gleichung $a \cdot x = b$ in $\mathbb{Q}$ die Lösung

$$x = \begin{pmatrix} b \\ a \end{pmatrix} \hat{=} \frac{b}{a}$$

Die Vorgehensweise beim Aufbau der komplexen Zahlen ist vollkommen analog. Ausgangspunkt ist auch hier die Feststellung, dass die reellen Zahlen unvollkommen sind, da sich bestimmte „elementare“ Gleichungen innerhalb der reellen Zahlen nicht lösen lassen, z.B. die Gleichung

$$x^2 = -1 \text{ .}$$

Die Ursache dieser Nichtlösbarkeit liegt darin, dass das Produkt einer reellen Zahl mit sich selbst immer eine positive reelle Zahl liefert. Anders ausgedrückt heißt dies, dass die Wurzel aus einer negativen Zahl (in $\mathbb{R}$) nicht existiert.

Die Vorgehensweise zur Erweiterung des Zahlenbereiches folgt der Konstruktion der rationalen Zahlen. Wir definieren

$$\mathbb{C} := \mathbb{R}^2 = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \mid a, b \in \mathbb{R} \right\}$$

¹Das ist so nicht ganz richtig, da es bedeuten würde, dass $\begin{pmatrix} 1 \\ 2 \end{pmatrix} \neq \begin{pmatrix} 2 \\ 4 \end{pmatrix}$, d.h. $\frac{1}{2} \neq \frac{2}{4}$ gilt. Dieses Problem der Mehrdeutigkeit bei der Darstellung von Brüchen, wird beseitigt, indem man einen Bruch als Äquivalenzklasse von Zahlenpaaren auffasst. So ist etwa $\frac{1}{2} = \{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ 6 \end{pmatrix}, \begin{pmatrix} 4 \\ 8 \end{pmatrix}, \dots \}$.

und überprüfen nach Wahl einer geeigneten Addition und Multiplikation die früher erwähnten zwei Kriterien.

Punkt 1: $\mathbb{R}$ ist eine Teilmenge von $\mathbb{C}$.

Da $\mathbb{C}$ eine Menge von Zahlenpaaren ist, kann $\mathbb{R}$ keine Teilmenge sein. Wir erreichen das Ziel in diesem Fall dadurch, dass wir die Zahlenpaare der Form $\begin{pmatrix} p \\ 0 \end{pmatrix}$ mit der reellen Zahl p identifizieren. Dies sind gerade die komplexen Zahlen, deren „Imaginärteil“ 0 ist.

Punkt 2: Definition von Addition und Multiplikation auf $\mathbb{C}$.

Für $\begin{pmatrix} p_1 \\ q_1 \end{pmatrix}, \begin{pmatrix} p_2 \\ q_2 \end{pmatrix} \in \mathbb{C}$ sei

$$\begin{aligned} \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} \oplus \begin{pmatrix} p_2 \\ q_2 \end{pmatrix} &:= \begin{pmatrix} p_1 + p_2 \\ q_1 + q_2 \end{pmatrix} \\ \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} \odot \begin{pmatrix} p_2 \\ q_2 \end{pmatrix} &:= \begin{pmatrix} p_1 \cdot p_2 - q_1 \cdot q_2 \\ p_1 \cdot q_2 + p_2 \cdot q_1 \end{pmatrix} \end{aligned}$$

Überraschend - wenn nicht gar unverständlich - wirkt hier die Definition der Multiplikation. Wir müssen überprüfen, dass auf den reellen Zahlen, also für die „Zahlen“ der Form $\begin{pmatrix} p \\ 0 \end{pmatrix}$, die ursprüngliche Definition der Addition und Multiplikation entsteht.

Seien hierzu $\begin{pmatrix} x_1 \\ 0 \end{pmatrix}, \begin{pmatrix} x_2 \\ 0 \end{pmatrix} \in \mathbb{C}$. Dann gilt (mit der Identifikation $\begin{pmatrix} x \\ 0 \end{pmatrix} \hat{=} x$)

$$\begin{aligned} x_1 \oplus x_2 &\hat{=} \begin{pmatrix} x_1 \\ 0 \end{pmatrix} \oplus \begin{pmatrix} x_2 \\ 0 \end{pmatrix} \stackrel{\text{Def.}}{=} \begin{pmatrix} x_1 + x_2 \\ 0 + 0 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ 0 \end{pmatrix} \hat{=} x_1 + x_2 \\ x_1 \odot x_2 &\hat{=} \begin{pmatrix} x_1 \\ 0 \end{pmatrix} \odot \begin{pmatrix} x_2 \\ 0 \end{pmatrix} \stackrel{\text{Def.}}{=} \begin{pmatrix} x_1 \cdot x_2 - 0 \cdot 0 \\ x_1 \cdot 0 + x_2 \cdot 0 \end{pmatrix} = \begin{pmatrix} x_1 \cdot x_2 \\ 0 \end{pmatrix} \hat{=} x_1 \cdot x_2 \end{aligned}$$

Welchen Vorteil bieten die komplexen Zahlen? Betrachten wir hierzu die komplexe Zahl

$$\mathbf{i} := \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Für diese gilt

$$\mathbf{i}^2 = \mathbf{i} \odot \mathbf{i} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \odot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \cdot 1 - 1 \cdot 1 \\ 0 \cdot 1 + 0 \cdot 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \hat{=} -1,$$

so dass $\mathbf{i}$ eine Lösung der quadratischen Gleichung $x^2 = -1$ ist. Wir haben damit das Problem, die Wurzeln negativer Zahlen zu berechnen, beseitigt. Zum Beispiel gilt

$$\sqrt{-36} = \sqrt{36 \cdot (-1)} = \sqrt{36} \cdot \sqrt{-1} = 6\mathbf{i}$$

Der Rest der Arbeit besteht darin, die Notation den vertrauten reellen Zahlen anzupassen. Hierzu notiert man die Addition und Multiplikation wie früher mit $+$ und $\cdot$. Auch die Darstellung der komplexen Zahlen lässt sich suggestiver gestalten. Hierzu beachte man, dass mit $\mathbf{i} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ und $1 \hat{=} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ für eine beliebige komplexe Zahl $\begin{pmatrix} x \\ y \end{pmatrix}$ folgt:

$$\begin{pmatrix} x \\ y \end{pmatrix} = x \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = x \cdot 1 + y \cdot \mathbf{i} = x + \mathbf{i}y$$

x heißt der *Realteil*, y der *Imaginärteil* der komplexen Zahl $\begin{pmatrix} x \\ y \end{pmatrix}$. Diese Darstellung gestattet es, mit komplexen Zahlen wie mit reellen Zahlen umzugehen, weil die zweidimensionale Struktur in dem Symbol i verborgen bleibt. Zum Beispiel kann das Produkt komplexer Zahlen mit der üblichen Distributivregel $a(b+c) = ac+bc$ berechnet werden

$$\begin{aligned}(x_1 + iy_1) \cdot (x_2 + iy_2) &= (x_1 \cdot x_2) + (x_1 \cdot y_2 i) + (y_1 i \cdot x_2) + (y_1 i \cdot y_2 i) \\ &= (x_1 x_2 - y_1 y_2) + (x_1 y_2 + x_2 y_1) i,\end{aligned}$$

weil $y_1 i \cdot y_2 i = y_1 y_2 i^2 = y_1 y_2 (-1) = -y_1 y_2$ gilt.

Die Division komplexer Zahlen ist etwas trickreicher, wenn man den Quotienten in der Form (Realteil + i ·Imaginärteil) darstellen will. Auch hier erweist sich die gerade eingeführte Schreibweise als sehr nützlich, wenn man $(a+ib)(a-ib) = a^2 + b^2$ beachtet.

$$\begin{aligned}\frac{x_1 + iy_1}{x_2 + iy_2} &= \frac{x_1 + iy_1}{x_2 + iy_2} \cdot \frac{x_2 - iy_2}{x_2 - iy_2} \\ &= \frac{(x_1 x_2 + y_1 y_2) + (x_2 y_1 - x_1 y_2) i}{x_2^2 + y_2^2} \\ &= \left(\frac{x_1 x_2 + y_1 y_2}{x_2^2 + y_2^2} \right) + i \left(\frac{x_2 y_1 - x_1 y_2}{x_2^2 + y_2^2} \right)\end{aligned}$$

C.1. Eigenschaften komplexer Zahlen

Wie gerade gesehen, ist es nützlich, mit der komplexen Zahl $x+iy$ auch die komplexe Zahl $x-iy$ zu betrachten. Wir definieren daher

C.1 Definition

Ist $z = x + iy$ eine komplexe Zahl, so heißt $x - iy$ die zu z *konjugiert komplexe Zahl*. Sie wird häufig auch mit $\bar{z}$ bezeichnet. Es gilt also

$$\bar{z} = \overline{x + iy} = x - iy$$

In der komplexen Zahlenebene ist es die an der reellen Achse zu z gespiegelte komplexe Zahl (siehe Abschnitt C.2). $\square$

Offensichtlich gelten folgende leicht einzusehende Aussagen für komplexe Zahlen z, z_1, z_2 .

C.2 Satz

a) $z \in \mathbb{R} \iff z = \bar{z}$

b) $\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$

c) $\overline{z_1 \cdot z_2} = \bar{z}_1 \cdot \bar{z}_2$

$\square$

Der folgende Satz zeigt, dass in polynomialen Gleichungen mit reellen Koeffizienten die komplexen Lösungen immer konjugiert komplex auftreten müssen. Er ist eine unmittelbare Folge von Satz C.2.

C.3 Satz

Sind $a_0, a_1, a_2, \dots, a_m$ reelle Zahlen, so besitzt die Gleichung

$$(*) \quad a_0 + a_1x + a_2x^2 + \dots + a_mx^m = 0$$

genau dann die Lösung z , wenn auch $\bar{z}$ eine Lösung von $(*)$ ist. $\square$

Dass die komplexen Zahlen in der Tat die gewünschten Vollständigkeitseigenschaften besitzen, belegt der folgende Satz.

C.4 Satz (Fundamentalsatz der Algebra)

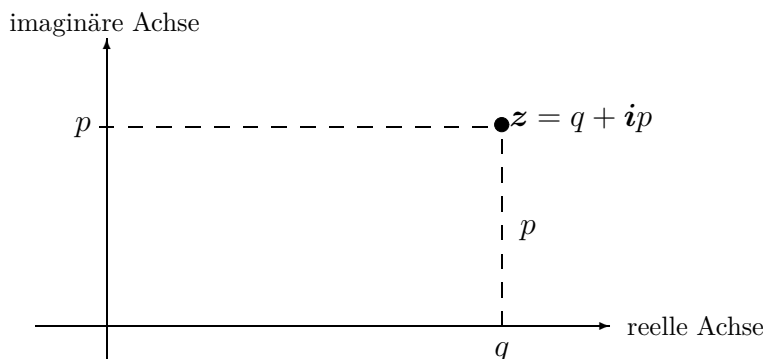
Jedes reelle Polynom $P(x) = a_0 + a_1x + \dots + a_mx^m$ vom Grad n ($a_0, a_1, \dots, a_m \in \mathbb{R}$) besitzt in $\mathbb{C}$ genau m Nullstellen $z_1, \dots, z_m$, so dass gilt:

$$P(x) = c \cdot (x - z_1)(x - z_2) \dots (x - z_m)$$

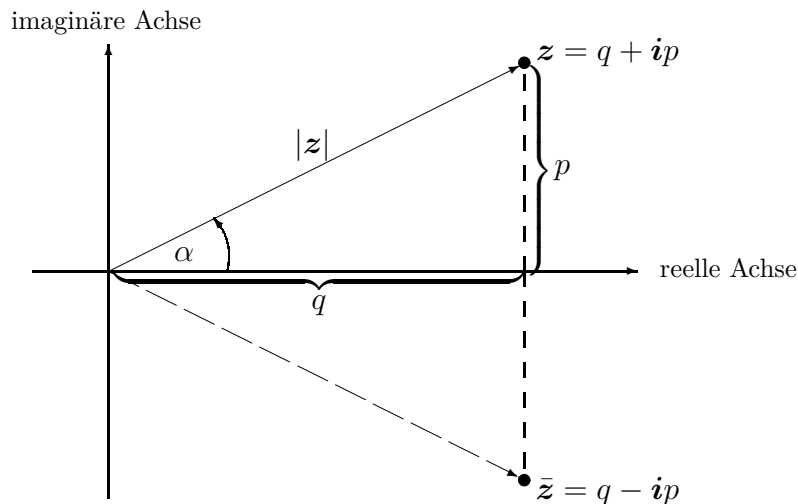
$\square$

C.2. Darstellung komplexer Zahlen

Da komplexe Zahlen geordnete Paare reeller Zahlen sind, können sie mit Hilfe eines zweidimensionalen kartesischen Koordinatensystems dargestellt werden. Wir tragen hierzu den Realteil auf der horizontalen und den Imaginärteil auf der vertikalen Achse ab. Man spricht dann von der komplexen Zahlenebene. Wie früher können die reellen Zahlen, deren Imaginärteil gleich 0 ist, mit der horizontalen (reellen) Achse identifiziert werden.



Diese zweidimensionale Interpretation der komplexen Zahlen lässt eine weitere Darstellungsmöglichkeit erkennen. Zur Beschreibung verwendet man hierbei den Abstand der komplexen Zahl z vom Ursprung und den Winkel, den die Verbindungsstrecke vom Ursprung nach z mit der horizontalen reellen Achse einschließt.



Aus dem Satz von Pythagoras folgt, dass der Abstand von $z = q + ip$ zum Ursprung gleich

$$\sqrt{q^2 + p^2} = \sqrt{(q + ip)(q - ip)} = \sqrt{z\bar{z}} = |z|$$

ist. Weiter gilt:

$$\sin(\alpha) = \frac{p}{|z|} = \frac{\operatorname{Re}(z)}{|z|}, \quad \cos(\alpha) = \frac{q}{|z|} = \frac{\operatorname{Im}(z)}{|z|} \implies \tan(\alpha) = \frac{\sin(\alpha)}{\cos(\alpha)} = \frac{\operatorname{Re}(z)}{\operatorname{Im}(z)}$$

Damit ergibt sich die sog. *Polarkoordinatendarstellung* von z :

$$z = q + ip = |z| \cos(\alpha) + i|z| \sin(\alpha) = |z| (\cos(\alpha) + i \sin(\alpha))$$

Mittels vollständiger Induktion ergibt sich daraus zusammen mit den Additionstheoremen für Sinus und Cosinus folgende Aussage.

C.5 Satz (DeMoivre)

Für eine komplexe Zahl $z = q + ip = |z| \cdot (\cos(\alpha) + i \sin(\alpha))$ und jede natürliche Zahl n gilt:

$$z^n = (q + ip)^n = |z|^n \cdot (\cos(n\alpha) + i \sin(n\alpha)) \quad \square$$

In $\triangle GL$ und DGL ist man oft daran interessiert, mit dem Ausdruck e^z , $z \in \mathbb{C}$ zu arbeiten. Dies ist möglich, wenn man an die Darstellung von e^x für reelle x auf Seite 173 denkt:

$$e^x = \sum_{j=0}^{\infty} \frac{x^j}{j!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots + \frac{x^n}{n!} + \dots$$

C.6 Definition

Für jede komplexe Zahl z definiert man e^z als den Grenzwert

$$e^z = \sum_{j=0}^{\infty} \frac{z^j}{j!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \dots + \frac{z^n}{n!} + \dots \quad \square$$

Man beachte, dass Definition C.6 für rein reelle Zahlen z den vertrauten Wert e^z liefert. Ist $z = it$ rein imaginär ($t \in \mathbb{R}$), ergibt sich wegen $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, ... und der Reihendarstellung von Sinus und Cosinus aus Satz A.9 die sog. *Euler-Gleichung*:

$$\begin{aligned} e^{it} &= 1 + (it) + \frac{(it)^2}{2!} + \frac{(it)^3}{3!} + \frac{(it)^4}{4!} + \dots + \frac{(it)^n}{n!} + \dots \\ &= \left(1 - \frac{t^2}{2!} + \frac{t^4}{4!} + \dots\right) + i \cdot \left(t - \frac{t^3}{3!} + \frac{t^5}{5!} + \dots\right) \\ &= \cos(t) + i \sin(t) \end{aligned}$$

Speziell für $t = \pi$ folgt daraus

$$e^{i\pi} = -1$$

oder in etwas ästhetischerer Form

$$e^{i\pi} + 1 = 0 .$$

Die Polarkoordinatendarstellung von $z = |z| \cdot (\cos(\alpha) + i \sin(\alpha))$ lautet damit in Kurzform

$$z = |z| e^{i\alpha} .$$



Lineare Gleichungssysteme, Eigenwerte, Eigenvektoren

D.1 Satz (Lösung linearer Gleichungssysteme)

Sei $\mathbf{A}$ eine $(m \times n)$ -Matrix und $\mathbf{b} \in \mathbb{R}^m$. Bezeichnet $\mathbb{L} := \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{b}\}$ den Lösungsraum des linearen Gleichungssystems $\mathbf{Ax} = \mathbf{b}$ und $\mathbb{L}_h := \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{0}\}$ den Lösungsraum des homogenen linearen Gleichungssystems $\mathbf{Ax} = \mathbf{0}$, so gelten folgende Aussagen:

- a) $\mathbb{L}_h$ ist ein Untervektorraum des $\mathbb{R}^n$ mit der Dimension $n - \text{Rg}(\mathbf{A})$.
- b) Ist $\mathbf{x}_0$ eine spezielle (inhomogene) Lösung von $\mathbf{Ax} = \mathbf{b}$, so gilt

$$\mathbb{L} = \{\mathbf{x}_0\} + \mathbb{L}_h := \{\mathbf{x}_0 + \mathbf{n} \mid \mathbf{n} \in \mathbb{L}_h\}.$$

- c) Ist $\mathbf{A}$ eine quadratische Matrix, so besitzt das homogene lineare Gleichungssystem $\mathbf{Ax} = \mathbf{0}$ genau dann eine von $\mathbf{0}$ verschiedene Lösung, wenn $\det(\mathbf{A}) = 0$ gilt. $\square$

D.2 Definition

Sei $\mathbf{A}$ eine quadratische $(n \times n)$ -Matrix.

- a) Das Polynom $P(\lambda) := \det(\mathbf{A} - \lambda \mathbf{1})$ heißt das *charakteristische Polynom der Matrix A*. Jede Nullstelle des charakteristischen Polynoms wird als *Eigenwert der Matrix A* bezeichnet.
- b) Ein Vektor $\mathbf{v} \in \mathbb{R}^n$ heißt *Eigenvektor der Stufe k* ($k \in \mathbb{N}$), wenn gilt:

$$(\mathbf{A} - \lambda \mathbf{1})^{k-1} \mathbf{v} \neq \mathbf{0} \text{ und } (\mathbf{A} - \lambda \mathbf{1})^k \mathbf{v} = \mathbf{0}.$$

Ein Eigenvektor der Stufe $k=1$ wird als Eigenvektor bezeichnet, d.h. ohne Angabe der Stufe¹.

¹In der Literatur werden Eigenvektoren der Stufe $k \geq 2$ auch als Hauptvektoren bezeichnet.

c) Ist λ ein Eigenwert der Matrix $\mathbf{A}$, so bezeichnen wir mit

$$E_\lambda := \{\mathbf{v} \in \mathbb{R}^n \mid (\mathbf{A} - \lambda \mathbf{1})\mathbf{v} = \mathbf{0}\}$$

die Menge aller Eigenvektoren der Stufe 1 (inklusive des Nullvektors) und mit

$$H_\lambda := \{\mathbf{v} \in \mathbb{R}^n \mid \exists k \in \mathbb{N} : (\mathbf{A} - \lambda \mathbf{1})^k \mathbf{v} = \mathbf{0}\}$$

die Menge aller Eigenvektoren beliebiger Stufe (inklusive des Nullvektors). $\square$

D.3 Bemerkung

- i) Gemäß Definition D.2 ist insbesondere jeder Eigenvektor ungleich dem Nullvektor.
- ii) Die Dimension von E_λ heißt *geometrische Vielfachheit*. Sie ist i.A. ungleich der *algebraischen Vielfachheit*, mit der λ als Nullstelle des charakteristischen Polynoms auftritt.
- iii) H_λ wird in der Literatur als *Hauptvektorraum* zum Eigenwert λ bezeichnet. $\square$

D.4 Satz

- a) E_λ und H_λ sind (nichtleere) Untervektorräume des $\mathbb{R}^n$ mit $E_\lambda \subset H_\lambda$. Insbesondere ist die geometrische Vielfachheit eines Eigenwertes kleiner gleich der algebraischen Vielfachheit.
- b) Eigenvektoren (beliebiger Stufe) zu verschiedenen Eigenwerten sind linear unabhängig.
- c) Die Dimension des Hauptvektorraumes H_λ ist gleich der algebraischen Vielfachheit. $\square$

Beweis:

Wir beweisen beispielhaft die für „klassische“ Eigenvektoren bekannte Aussage b).

Seien $\mathbf{v}_1 \neq \mathbf{0}$, $\mathbf{v}_2 \neq \mathbf{0}$ Hauptvektoren zu den Eigenwerten $\lambda_1 \neq \lambda_2$. D.h. es existieren natürliche Zahlen k, m mit $k \geq m$ und

$$\begin{aligned} (\mathbf{A} - \lambda_1 \mathbf{1})^{k-1} \mathbf{v}_1 &\neq \mathbf{0}, & (\mathbf{A} - \lambda_1 \mathbf{1})^k \mathbf{v}_1 &= \mathbf{0} \\ (\mathbf{A} - \lambda_2 \mathbf{1})^{m-1} \mathbf{v}_2 &\neq \mathbf{0}, & (\mathbf{A} - \lambda_2 \mathbf{1})^m \mathbf{v}_2 &= \mathbf{0} \end{aligned}$$

Sind nun $\alpha, \beta \in \mathbb{C}$ mit

$$\alpha \mathbf{v}_1 + \beta \mathbf{v}_2 = \mathbf{0},$$

so folgt

$$\begin{aligned}
\mathbf{0} &= (\mathbf{A} - \lambda_1 \mathbf{1})^{k-1} (\mathbf{A} - \lambda_2 \mathbf{1})^m [\alpha \mathbf{v}_1 + \beta \mathbf{v}_2] \\
&= \alpha (\mathbf{A} - \lambda_1 \mathbf{1})^{k-1} \underbrace{(\mathbf{A} - \lambda_2 \mathbf{1})^m}_{(\mathbf{A} - \lambda_1 + \lambda_1 - \lambda_2)^m} \mathbf{v}_1 + \mathbf{0} \\
&= \alpha (\mathbf{A} - \lambda_1 \mathbf{1})^{k-1} \sum_{j=1}^m \binom{m}{j} (\mathbf{A} - \lambda_1 \mathbf{1})^j (\lambda_1 - \lambda_2)^{m-j} \mathbf{v}_1 \\
&\stackrel{(\mathbf{A} - \lambda_1 \mathbf{1})^k \mathbf{v}_1 = \mathbf{0}}{=} \alpha \underbrace{(\lambda_1 - \lambda_2)^m}_{\neq 0} \underbrace{(\mathbf{A} - \lambda_1 \mathbf{1})^{k-1} \mathbf{v}_1}_{\neq 0}
\end{aligned}$$

Daher gilt $\alpha = 0$ und damit $\beta = 0$. ■

D.5 Beispiel
Die Matrix $\mathbf{A} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 1 & -2 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix}$ besitzt das charakteristische Polynom

$$\det(\mathbf{A} - \lambda \mathbf{1}) = \lambda(\lambda+1)^3$$

und damit die Eigenwerte $\lambda_1 = 0$ und $\lambda_2 = -1$, letzteren mit der Vielfachheit 3.

Eigenvektor (der Stufe 1) ist z.B. $(1, 1, 0, 0)^T$ und der zu λ_1 gehörende Eigenraum

$$E_{\lambda_1} = \{\mathbf{x} \in \mathbb{R}^4 \mid (\mathbf{A} - \lambda_1 \mathbf{1})\mathbf{x} = \mathbf{0}\} = LH \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right\}$$

hat die Dimension 1 (=geometrische Vielfachheit).

Der zu λ_2 gehörende Eigenraum

$$E_{\lambda_2} = \{\mathbf{x} \in \mathbb{R}^4 \mid (\mathbf{A} - \lambda_2 \mathbf{1})\mathbf{x} = \mathbf{0}\}$$

hat ebenfalls die Dimension 1 (=geometrische Vielfachheit), da der Rang der Matrix $\mathbf{A} - \lambda_2 \mathbf{1} = \mathbf{A} + \mathbf{1}$ gleich 3 ist. Neben $\mathbf{v}_1 = (1, 0, -1, 0)^T$ gibt es daher keinen weiteren linear unabhängigen Eigenvektor der Stufe 1 zu λ_2 .

Der Eigenwert $\lambda_2 = -1$ hat also die Vielfachheit 3, während der zugehörige Eigenraum E_{λ_2} nur eindimensional ist. Der zugehörige Raum aller Hauptvektoren H_{λ_2} ist dreidimensional. Es gilt nämlich:

$$\mathbf{A} - \lambda_2 \mathbf{1} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 0 \end{pmatrix}, \quad (\mathbf{A} - \lambda_2 \mathbf{1})^2 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

und

$$(\mathbf{A} - \lambda_2 \mathbf{1})^3 = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Daher ist $\mathbf{v}_3 := (0, 1, 0, 0)^T$ ein Hauptvektor der Stufe 3 und $\mathbf{v}_2 := (\mathbf{A} - \lambda_2 \mathbf{1})\mathbf{v}_3 = (0, 1, 1, -1)^T$ ein Hauptvektor der Stufe 2. Mit dem Eigenvektor $\mathbf{v}_1 := (\mathbf{A} - \lambda_2 \mathbf{1})^2 \mathbf{v}_3 = (1, 0, -1, 0)^T$ bilden $\mathbf{v}_2$ und $\mathbf{v}_3$ eine Basis von H_{λ_2} . $\square$

Mit den Eigenvektoren lässt sich jede quadratische Matrix $\mathbf{M}$ in eine äquivalente Diagonal- bzw. Bidiagonalmatrix, die als *Jordanmatrix* bezeichnet wird, transformieren. Ist $\mathbf{M}$ diagonalisierbar, liefert die Jordanmatrix gerade die Diagonalmatrix der Eigenwerte. In diesem Sinne ist die *Jordanform* eine Verallgemeinerung der Diagonalform von Matrizen und als solche zur Berechnung von Exponentialmatrizen besonders nützlich (siehe Beispiel A.10). Es gilt folgender Satz.

D.6 Satz

Ist $\mathbf{M} \in \mathbb{R}^{n \times n}$ eine quadratische Matrix, so existiert eine (evtl. komplexe) invertierbare $(n \times n)$ -Matrix $\mathbf{B}$, die $\mathbf{M}$ in eine äquivalente *Jordanmatrix* $\mathbf{J}$ transformiert, d.h.

$$\mathbf{B}^{-1} \mathbf{M} \mathbf{B} = \text{diag}(\mathbf{J}_1, \dots, \mathbf{J}_\ell) = \begin{pmatrix} \mathbf{J}_1 & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & \mathbf{J}_\ell \end{pmatrix} =: \mathbf{J}$$

Die Matrizen $\mathbf{J}_i$ heißen *Jordanblöcke* und sind quadratische Bidiagonalmatrizen der folgenden Form:

$$\mathbf{J}_i = \begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \lambda_i & 1 \\ 0 & \cdots & \cdots & 0 & \lambda_i \end{pmatrix}$$

Die λ_i sind hierbei die Eigenwerte von $\mathbf{M}$. Zu jedem Eigenwert λ gibt es seiner geometrischen Vielfachheit (=Dimension des Eigenraumes) entsprechend viele Jordanblöcke.

Die Spalten der (nicht eindeutig bestimmten) transformierenden Matrix $\mathbf{B}$ sind Eigenvektoren (evtl. höherer Stufe) von $\mathbf{M}$. $\square$

D.7 Beispiel

i) Die Matrix $\mathbf{M}$ sei definiert durch

$$\mathbf{M} := \begin{pmatrix} 25 & -16 & 30 & -44 & -12 \\ 13 & -7 & 18 & -26 & -6 \\ -18 & 12 & -21 & 36 & 12 \\ -9 & 6 & -12 & 21 & 6 \\ 11 & -8 & 15 & -22 & -3 \end{pmatrix}.$$

Diese besitzt den fünffachen Eigenwert 3. Mit der invertierbaren Matrix

$$\mathbf{B} := \begin{pmatrix} 22 & 1 & -16 & 0 & 2 \\ 13 & 0 & -10 & 1 & 0 \\ -18 & 0 & 12 & 0 & 0 \\ -9 & 0 & 6 & 0 & 1 \\ 11 & 0 & -8 & 0 & 0 \end{pmatrix}$$

erhält man folgende Jordanform der Matrix $\mathbf{M}$:

$$\mathbf{B}^{-1}\mathbf{M}\mathbf{B} = \mathbf{J} = \begin{pmatrix} \boxed{\begin{matrix} 3 & 1 \\ 0 & 3 \end{matrix}} & \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{matrix} & \boxed{\begin{matrix} 3 & 1 \\ 0 & 3 \end{matrix}} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 \end{matrix} & \boxed{3} \end{pmatrix}$$

$\mathbf{J}$ besteht aus drei Jordanblöcken, zwei Blöcke der Größe 2 und ein Block der Größe 1. Die erste, dritte und fünfte Spalte von $\mathbf{B}$ sind Eigenvektoren (der Stufe 1), die zweite und vierte Spalte von $\mathbf{B}$ sind Eigenvektoren der Stufe 2.

ii) Die Matrix

$$\mathbf{M} = \begin{pmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 3 & 1 & -1 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & -1 & -2 & 2 & 2 \end{pmatrix}$$

hat den Eigenwert 3 mit der algebraischen Vielfachheit 2 und den Eigenwert 2 mit der algebraischen Vielfachheit 3.

Der Eigenraum E_3 zum Eigenwert 3 ist zweidimensional, d.h. die geometrische ist gleich der algebraischen Vielfachheit. Eine Basis von E_3 bilden z.B. die Vektoren

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$$

Der Eigenraum E_2 zum Eigenwert 2 ist nur eindimensional, d.h. die geometrische ist ungleich der algebraischen Vielfachheit. Eine Basis von H_3 bilden z.B. die Vektoren

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ -1 \\ 0 \end{pmatrix}.$$

Der Erste ist ein Eigenvektor, der Zweite ein Eigenvektor der Stufe 2 und der Dritte ein Eigenvektor der Stufe 3. Bildet man aus diesen 5 Eigenvektoren die Matrix $\mathbf{B}$ und ihre Inverse, erhält man

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 & 0 \end{pmatrix}, \mathbf{B}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 \end{pmatrix}$$

und es folgt die Jordandarstellung von $\mathbf{M}$

$$\mathbf{B}^{-1}\mathbf{M}\mathbf{B} = \mathbf{J} = \begin{pmatrix} \boxed{3} & 0 & 0 & 0 & 0 \\ 0 & \boxed{3} & 0 & 0 & 0 \\ 0 & 0 & \boxed{2} & 1 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

Die Jordanmatrix besteht in diesem Fall aus drei Blöcken, zwei der Größe 1 und einem der Größe 3. □



Quadratische Formen, Definitheit

Auch im mehrdimensionalen Fall lässt sich die Konvexität mit Hilfe der zweiten Ableitung charakterisieren. Da $f''(\mathbf{x})$ dann jedoch eine Matrix ist, wird ein Positivitätsbegriff für Matrizen benötigt. Dieser muss im 1-dimensionalen Fall mit dem Begriff der Positivität für reelle Zahlen, d.h. für (1×1) -Matrizen, übereinstimmen.

Die Idee für die folgende Definition besteht darin, die Eigenschaft, dass eine reelle Zahl a genau dann positiv ist, wenn das zugehörige quadratische Polynom $x \mapsto a \cdot x^2$ für $x \neq 0$ immer positive Werte liefert, auf Matrizen zu übertragen.

E.1 Definition

Sei $\mathbf{A} = (a_{ij})_{i,j=1}^n$ eine $(n \times n)$ -Matrix. Die zur Matrix $\mathbf{A}$ gehörende *quadratische Form* ist die Funktion $Q_{\mathbf{A}} : \mathbb{R}^n \rightarrow \mathbb{R}$ definiert durch

$$Q_{\mathbf{A}}(\mathbf{x}) = \mathbf{x}^T \cdot \mathbf{A} \cdot \mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} \cdot x_i \cdot x_j$$

- a) • $\mathbf{A}$ heißt *positiv* bzw. *negativ definit* (Schreibweise: $\mathbf{A} > 0$ bzw. $\mathbf{A} < 0$), wenn

$$\forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\} : Q_{\mathbf{A}}(\mathbf{x}) > 0 \quad \text{bzw.} \quad Q_{\mathbf{A}}(\mathbf{x}) < 0 .$$

- $\mathbf{A}$ heißt *positiv* bzw. *negativ semidefinit* ($\mathbf{A} \geq 0$ bzw. $\mathbf{A} \leq 0$), wenn gilt

$$\forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\} : Q_{\mathbf{A}}(\mathbf{x}) \geq 0 \quad \text{bzw.} \quad Q_{\mathbf{A}}(\mathbf{x}) \leq 0 .$$

- $\mathbf{A}$ heißt *indefinit*, wenn $Q_{\mathbf{A}}$ positive und negative Werte annimmt.

- b) Ist $\mathbf{B}$ eine $(k \times n)$ -Matrix, so heißt $\mathbf{A}$ *positiv* bzw. *negativ definit* unter der Nebenbedingung $\mathbf{B} \cdot \mathbf{x} = \mathbf{0}$, wenn gilt

$$\forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\} : \mathbf{B} \cdot \mathbf{x} = \mathbf{0} \implies Q_{\mathbf{A}}(\mathbf{x}) > 0 \quad \text{bzw.} \quad Q_{\mathbf{A}}(\mathbf{x}) < 0 .$$

□

E.2 Beispiel

Betrachten wir die (2×2) -Matrix $\mathbf{A} = \begin{pmatrix} -1 & 1 \\ 1 & -2 \end{pmatrix}$, so ist $Q_{\mathbf{A}} : \mathbb{R}^2 \rightarrow \mathbb{R}$ gegeben durch

$$\begin{aligned} Q_{\mathbf{A}}(\mathbf{x}) &= Q_{\mathbf{A}}(x_1, x_2) = (x_1, x_2) * (-x_1 + x_2, x_1 - 2x_2) \\ &= -x_1^2 + 2x_1x_2 - 2x_2^2 \\ &= -(x_1^2 - 2x_1x_2 + x_2^2) - x_2^2 \\ &= -[(x_1 - x_2)^2 + x_2^2] \end{aligned}$$

Wie man sieht, gilt $Q_{\mathbf{A}}(\mathbf{x}) < 0$ für $\mathbf{x} \neq \mathbf{0}$. $\mathbf{A}$ ist daher negativ definit. $\square$

E.3 Bemerkung

- i) Ist $\mathbf{A} = (a_{ij})$ eine zur Hauptdiagonale *symmetrische* $(n \times n)$ -Matrix, d.h. für die $\mathbf{A} = \mathbf{A}^T$ gilt, lässt sich die Definitheit mittels Determinanten untersuchen.

Bezeichnet man nämlich mit $\mathbf{A}^{ii}$ die Matrix der Elemente von $\mathbf{A}$ deren Zeilenindex und Spaltenindex kleiner gleich i ist (i -ter Hauptabschnitt), so gilt:

- $\mathbf{A}$ ist genau dann positiv definit, wenn alle *Hauptabschnittsdeterminanten* positiv sind, d.h. wenn

$$\det(\mathbf{A}^{rr}) > 0 \quad \text{für } r = 1, \dots, n$$

- $\mathbf{A}$ ist genau dann negativ definit, wenn die *Hauptabschnittsdeterminanten* im Vorzeichen alternieren (beginnend mit < 0), d.h. wenn

$$(-1)^r \det(\mathbf{A}^{rr}) > 0 \quad \text{für } r = 1, \dots, n$$

- ii) Sei $\mathbf{B}$ eine $(k \times n)$ -Matrix. Definiert man die $(k+n) \times (k+n)$ -Matrix $\mathbf{C}$ durch¹

$$\mathbf{C} := \left(\begin{array}{c|c} \mathbf{0} & \mathbf{B} \\ \hline \mathbf{B}^T & \mathbf{A} \end{array} \right),$$

so gilt:

- $\mathbf{A}$ ist genau dann positiv definit unter der Nebenbedingung $\mathbf{B}\mathbf{x} = \mathbf{0}$, wenn die letzten $n-k$ Hauptabschnittsdeterminanten von $\mathbf{C}$ das gleiche Vorzeichen wie $(-1)^k$ besitzen, d.h. wenn

$$(-1)^k \det(\mathbf{C}^{rr}) > 0 \quad \text{für } r = 2k+1, \dots, k+n.$$

- $\mathbf{A}$ ist genau dann negativ definit unter der Nebenbedingung $\mathbf{B}\mathbf{x} = \mathbf{0}$, wenn die letzten $n-k$ Hauptabschnittsdeterminanten von $\mathbf{C}$ im Vorzeichen alternieren, wobei $\text{sgn}(\det(\mathbf{C})) = \text{sgn}(-1)^n$ gilt, d.h. wenn

$$(-1)^{r-k} \det(\mathbf{C}^{rr}) > 0 \quad \text{für } r = 2k+1, \dots, k+n$$

$\square$

¹ $\mathbf{C}$ wird als die zu $\mathbf{A}$ geränderte Matrix bezeichnet.

Für $\mathbf{B} = \mathbf{0}$ ($k=0$) ergibt sich die Charakterisierung der Definitheit ohne Nebenbedingung wie in Teil i).

- iii) Eine $(n \times n)$ -Matrix $\mathbf{A}$ ist offensichtlich genau dann negativ (semi-)definit, wenn $-\mathbf{A}$ positiv (semi-)definit ist.

E.4 Beispiel

Gegeben sei die indefinite² (3×3) -Matrix $\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ und die (1×3) -Matrix $\mathbf{B} = (2, 2, 2)$. Wir wollen $\mathbf{A}$ auf Definitheit unter der Nebenbedingung $\mathbf{B}\mathbf{x} = \mathbf{0}$ untersuchen. Wegen $k = 1$ und $N = 3$, müssen die Determinanten der Matrizen $\mathbf{C}^{rr}$ für $r = 2k+1 = 3$ und $r = k+n = 4$ bestimmt werden. Es gilt

$$\mathbf{C} = \mathbf{C}^{44} = \begin{pmatrix} 0 & 2 & 2 & 2 \\ 2 & 0 & 1 & 1 \\ 2 & 1 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{pmatrix}, \quad \mathbf{C}^{33} = \begin{pmatrix} 0 & 2 & 2 \\ 2 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}.$$

Die Berechnung der Hauptabschnittsdeterminanten ergibt $\det(\mathbf{C}^{33}) = 8$, $\det(\mathbf{C}^{44}) = -12$ und damit:

$$\begin{aligned} (-1)^k \det(\mathbf{C}^{33}) &= -8 < 0, & (-1)^k \det(\mathbf{C}^{44}) &= 12 > 0 \\ (-1)^2 \det(\mathbf{C}^{33}) &= 8 > 0, & (-1)^3 \det(\mathbf{C}^{44}) &= 12 > 0 \end{aligned}$$

Daher ist $\mathbf{A}$ nicht positiv, jedoch negativ definit unter der Nebenbedingung $\mathbf{B}\mathbf{x} = \mathbf{0}$. $\square$

E.5 Bemerkung

Ist die Matrix $\mathbf{A}$ nicht symmetrisch, so kann das Determinantenkriterium aus Bemerkung E.3 dennoch verwendet werden, wenn man beachtet, dass die Matrix $\mathbf{B} := \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)$ symmetrisch ist und dass $Q_{\mathbf{A}} = Q_{\mathbf{B}}$ gilt. $\square$

²Es ist $Q_{\mathbf{A}}(x_1, x_2, x_3) = 2(x_1x_2 + x_1x_3 + x_2x_3)$ und daher $Q_{\mathbf{A}}(1, 1, 0) = 2$ und $Q_{\mathbf{A}}(-1, 1, 0) = -2$.



Statische Optimierung

Ein wesentlicher Nachteil der klassischen Formulierung des Satzes von Lagrange ist, dass die Nebenbedingungen in Gleichungsform vorliegen müssen, so dass Nebenbedingungen in Ungleichungsform nicht direkt verarbeitet werden können. In diesem Sinne stellt der *Satz von Kuhn-Tucker* bzw. das *Theorem von Fritz John* die Verallgemeinerung des Satzes von Lagrange dar.

Für stetig differenzierbare Abbildungen $f : \mathbb{R}^n \rightarrow \mathbb{R}$ und $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ betrachten wir daher folgende Problemstellung:

$$\begin{aligned} f(\mathbf{x}) &\longrightarrow \max \\ \text{u.d.N.} \quad \mathbf{g}(\mathbf{x}) &\leq \mathbf{0}. \end{aligned} \tag{F.1}$$

Wie üblich sei $L(\mathbf{x}, \boldsymbol{\lambda}) := f(\mathbf{x}) - \boldsymbol{\lambda} \cdot \mathbf{g}(\mathbf{x})$ die Lagrange-Funktion des Problems. X bezeichne den *Zulässigkeitsbereich*, d.h. die Menge $X = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$. Dann gilt folgende Aussage:

F.1 Satz (Satz von Kuhn-Tucker)

Ist $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ ein nichtnegativer *Sattelpunkt* der Lagrangefunktion L , d.h. ist $\mathbf{x}^*$ ein Maximum von L bzgl. $\mathbf{x}$ und $\boldsymbol{\lambda}^*$ ein Minimum von L bzgl. $\boldsymbol{\lambda}$, gilt also

$$\forall \mathbf{x} \in X \quad \forall \boldsymbol{\lambda} \geq \mathbf{0} : L(\mathbf{x}, \boldsymbol{\lambda}^*) \leq L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \leq L(\mathbf{x}^*, \boldsymbol{\lambda}), \tag{F.2}$$

so ist $\mathbf{x}^*$ eine Lösung des Maximumproblems (F.1). □

Beweis:

Sei $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ ein nichtnegativer *Sattelpunkt* der Lagrangefunktion L . Wir zeigen

i) $\mathbf{x}^*$ ist zulässig, d.h. es gilt $\mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}$

ii) $\mathbf{x}^*$ maximiert f .

ad i): Aufgrund der Sattelpunktsbedingung $L(\mathbf{x}^*, \boldsymbol{\lambda}) \geq L(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ gilt für alle $\boldsymbol{\lambda} \geq \mathbf{0}$:

$$\mathbf{0} \geq L(\mathbf{x}^*, \boldsymbol{\lambda}^*) - L(\mathbf{x}^*, \boldsymbol{\lambda}) = (\boldsymbol{\lambda} - \boldsymbol{\lambda}^*) \cdot \mathbf{g}(\mathbf{x}^*) \tag{F.3}$$

Wählt man speziell $\lambda = \lambda^* + e_i$ für $i = 1, \dots, m$ folgt $g_i(x^*) \leq 0$ für die i -te Komponente g_i von g , d.h. es gilt $g(x^*) \leq 0$.

Wegen $\lambda^* \geq 0$ folgt $\lambda^* \cdot g(x^*) \leq 0$ und da sich für $\lambda = 0$ außerdem $\lambda^* \cdot g(x^*) \geq 0$ in Ungleichung (F.3) ergibt, muss insbesondere gelten:

$$\lambda^* \cdot g(x^*) = 0$$

ad ii): Ist x zulässig, d.h. gilt $g(x) \leq 0$, so folgt wegen $\lambda^* \geq 0$

$$\begin{aligned} f(x) &\leq f(x) - \lambda^* \cdot g(x) = L(x, \lambda^*) \\ &\stackrel{(F.2)}{\leq} L(x^*, \lambda^*) \\ &= f(x^*) - \underbrace{\lambda^* \cdot g(x^*)}_{=0} = f(x^*) \end{aligned} \quad \blacksquare$$

F.2 Bemerkung

- i) Der Satz von Kuhn-Tucker erscheint plausibel, wenn man beachtet, dass ein Sattelpunkt versucht, die Funktion f zu maximieren unter minimalem Einfluss der Nebenbedingung g .
- ii) Ist f konkav und g konvex, so gilt mit einer kleinen technischen Zusatzbedingung in der Sattelpunkt-Aussage von Kuhn-Tucker sogar die Äquivalenz. $\square$

Da die globalen Kuhn-Tucker-Bedingungen offensichtlich nur schwer nachprüfbar sind, ist es wünschenswert gleichwertige lokale Bedingungen zu haben. Stellt man zusätzlich noch Differenzierbarkeitsforderungen an f und g , erhält man folgende Aussage. Sie ist die Verallgemeinerung des Satzes von Lagrange.

F.3 Satz (Theorem von Fritz John)

Ist x^* eine Lösung von (F.1), so existiert ein λ^* mit folgenden Eigenschaften

- a) $\frac{\partial L}{\partial x}(x^*, \lambda^*) = 0$
- b) $\lambda^* \cdot \frac{\partial L}{\partial \lambda}(x^*, \lambda^*) = 0$ (*komplementärer Schlupf*)
- c) $\lambda^* \geq 0$
- d) $\frac{\partial L}{\partial \lambda}(x^*, \lambda^*) \geq 0$ $\square$

Beweis:

Zum besseren Verständnis des letzten Satzes soll hier eine geometrische Erläuterung des Beweises im Falle zweier Veränderlicher ($n = 2$) gegeben werden. Wir verwenden hierbei zwei bekannte Aussagen aus der mehrdimensionalen Differentialrechnung (siehe z.B. SCHINDLER, Mathematik für Ökonomen):

a) Der Ableitungsvektor einer Funktion f steht *im Definitionsbereich* senkrecht auf den Isoquanten von f und zeigt in Richtung des stärksten Anstiegs der Funktionswerte.

b) Die Richtungsableitung $\frac{\partial f}{\partial \mathbf{v}}$ einer Funktion f in Richtung $\mathbf{v}$ ($\|\mathbf{v}\| = 1$) ist $f' \cdot \mathbf{v}$.

Ist $\mathbf{x}^*$ eine Lösung des Maximumproblems (F.1), so muss insbesondere die Nebenbedingung $g(\mathbf{x}^*) \leq 0$ gelten. Wegen $\frac{\partial L}{\partial \lambda}(\mathbf{x}^*, \lambda^*) = -g(\mathbf{x}^*)$ ist damit Aussage d) automatisch erfüllt und es müssen nur die Aussagen a), b) und c) nachgewiesen werden. Für die Lage des maximalen Punktes $\mathbf{x}^*$ sind nun zwei Möglichkeiten denkbar.

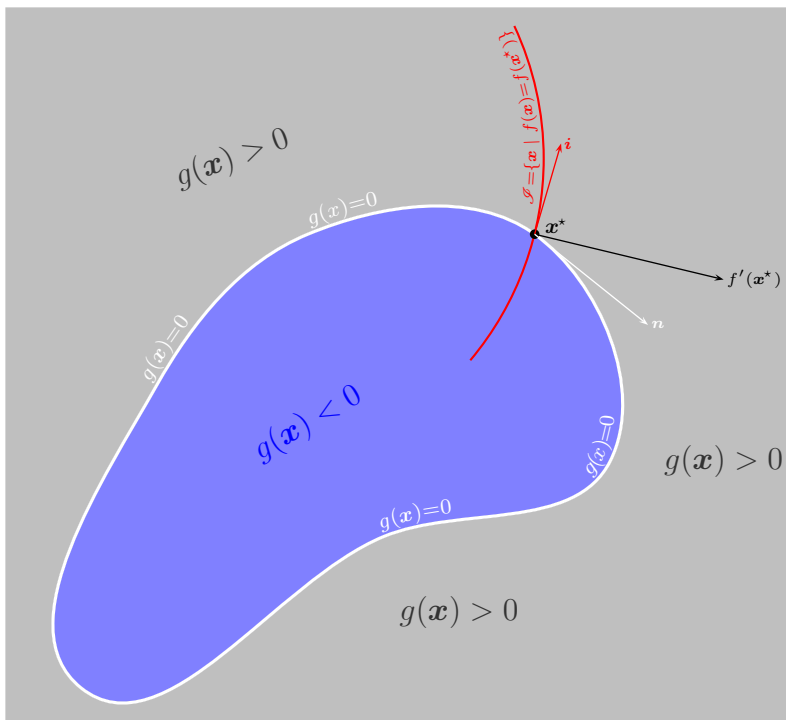
I) $\mathbf{x}^*$ liegt im Innern des Zulässigkeitsbereiches $\{\mathbf{x} \mid g(\mathbf{x}) \leq 0\}$, d.h. es ist $g(\mathbf{x}) < 0$. Da die Menge $\{\mathbf{x} \mid g(\mathbf{x}) < 0\}$ aufgrund der Stetigkeit von g offen ist, liegt ein freies Extremum vor, d.h. es ist $f'(\mathbf{x}^*) = \mathbf{0}$. Wählt man $\lambda^* = 0$ ergibt sich Aussage a) wegen

$$\frac{\partial L}{\partial \mathbf{x}}(\mathbf{x}^*, \lambda^*) = f'(\mathbf{x}^*) - 0 \cdot g'(\mathbf{x}^*) = f'(\mathbf{x}^*) = \mathbf{0}$$

Aussagen b) und c), d.h. $\lambda^* \cdot \frac{\partial L}{\partial \lambda}(\mathbf{x}^*, \lambda^*) = 0$ und $\lambda^* \geq 0$, sind wegen $\lambda^* = 0$ ebenfalls erfüllt.

II) $\mathbf{x}^*$ liegt auf dem Rand des Zulässigkeitsbereiches X , d.h. die Nebenbedingung ist bindend, so dass $g(\mathbf{x}^*) = 0$ gilt und wegen $\frac{\partial L}{\partial \lambda}(\mathbf{x}^*, \lambda^*) = -g(\mathbf{x}^*) = 0$ Aussage b) erfüllt ist. Es bleibt der Nachweis der Aussagen a) und c). Betrachtet man die Isoquante (Niveaulinie) von f durch $\mathbf{x}^*$, d.h. die Menge $\mathcal{J} := \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = f(\mathbf{x}^*)\}$, so können zwei Situationen eintreten.

α) $\mathcal{J}$ schneidet den Rand der Nebenbedingungsmenge $\mathcal{N} := \{\mathbf{x} \in \mathbb{R}^n \mid g(\mathbf{x}) = 0\}$. Bezeichnen $\mathbf{i}$ und $\mathbf{n}$ die Richtungsvektoren, die im Punkt $\mathbf{x}^*$ tangential zu $\mathcal{J}$ und tangential zu $\mathcal{N}$ verlaufen, erhält man nachfolgende Skizze.



Für die Richtungsableitungen $\frac{\partial f}{\partial \mathbf{i}}$ und $\frac{\partial f}{\partial \mathbf{n}}$ im Punkt $\mathbf{x}^*$ gilt:

$$f'(\mathbf{x}^*) \cdot \mathbf{i} = \frac{\partial f}{\partial \mathbf{i}}(\mathbf{x}^*) = 0 \quad (\text{F.4})$$

$$f'(\mathbf{x}^*) \cdot \mathbf{n} = \frac{\partial f}{\partial \mathbf{n}}(\mathbf{x}^*) = 0 \quad (\text{F.5})$$

Gleichung (F.4) gilt, weil f in Richtung $\mathbf{i}$ konstant ist. Gleichung (F.5) gilt, weil f in Richtung der Nebenbedingung ein lokales Maximum besitzt. Wegen der linearen Unabhängigkeit von $\mathbf{i}$ und $\mathbf{n}$ ($\mathcal{I}$ und $\mathcal{N}$ schneiden sich!!), existieren reelle Zahlen $\alpha_1, \alpha_2, \beta_1, \beta_2$ mit:

$$\mathbf{e}_1 = \alpha_1 \mathbf{i} + \beta_1 \mathbf{n}$$

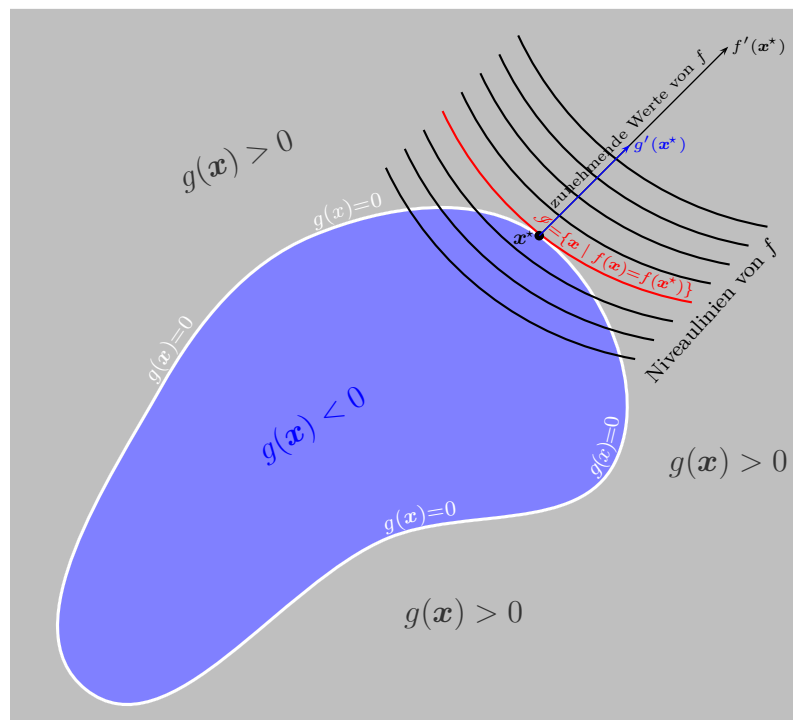
$$\mathbf{e}_2 = \alpha_2 \mathbf{i} + \beta_2 \mathbf{n}$$

Daraus folgt aber $f'(\mathbf{x}^*) = \mathbf{0}$, denn es ist

$$\begin{aligned} \frac{\partial f}{\partial x_1}(\mathbf{x}^*) &= f'(\mathbf{x}^*) \cdot \mathbf{e}_1 = f'(\mathbf{x}^*) \cdot (\alpha_1 \mathbf{i} + \beta_1 \mathbf{n}) \\ &= \alpha_1 f'(\mathbf{x}^*) \cdot \mathbf{i} + \beta_1 f'(\mathbf{x}^*) \cdot \mathbf{n} \\ &= 0. \end{aligned}$$

Analog folgt $\frac{\partial f}{\partial x_2}(\mathbf{x}^*) = 0$ und damit $f'(\mathbf{x}^*) = (\frac{\partial f}{\partial x_1}(\mathbf{x}^*), \frac{\partial f}{\partial x_2}(\mathbf{x}^*)) = \mathbf{0}$. Wir befinden uns daher wieder im Fall I) und können $\lambda^* = 0$ wählen.

β) $\mathcal{I}$ berührt die Menge $\mathcal{N} = \{\mathbf{x} \in \mathbb{R}^n \mid g(\mathbf{x}) = 0\}$.



Da $f'(\mathbf{x}^*)$ und $g'(\mathbf{x}^*)$ jeweils senkrecht auf den Isoquanten $\mathcal{I}$ und $\mathcal{N}$ stehen, müssen beide Vektoren parallel (linear abhängig) sein und es existiert ein $\lambda^* \in \mathbb{R}$ mit $f'(\mathbf{x}^*) = \lambda^* \cdot g'(\mathbf{x}^*)$, d.h. es gilt Aussage a). Da $f'(\mathbf{x}^*)$ in die Richtung des

stärksten Anstiegs von f zeigt, muss dieser Vektor außerdem ins Äußere der Menge X zeigen, da f ansonsten - im Widerspruch zur lokalen Maximalität von $\mathbf{x}^*$ - in X größere Werte annehmen würde. Weil $g'(\mathbf{x}^*)$ ebenfalls in diese Richtung zeigt (die Funktionswerte von g nehmen nach außen zu) zeigen $g'(\mathbf{x}^*)$ und $f'(\mathbf{x}^*)$ in die gleiche Richtung, d.h. es ist $\lambda^* > 0$, womit auch Aussage c) erfüllt ist. ■

F.4 Bemerkung

- i) Der wesentliche Unterschied zur klassischen Lagrangeschen-Multiplikator-Methode besteht in Forderung F.3 c). Da bei einer Ungleichung a priori nicht entschieden werden kann, ob die Restriktion bindend ist ($g(\mathbf{x}^*)=0$) oder nicht ($g(\mathbf{x}^*)<0, \lambda^*=0$), muss die Lagrangebedingung $\frac{\partial L}{\partial \lambda} = 0$ ersetzt werden durch

$$-g(\mathbf{x}^*, \lambda^*) = \frac{\partial L}{\partial \lambda}(\mathbf{x}^*, \lambda^*) \geq 0 \quad \text{und} \quad \lambda^* \cdot g(\mathbf{x}^*, \lambda^*) = \lambda^* \cdot \frac{\partial L}{\partial \lambda}(\mathbf{x}^*, \lambda^*) = 0 .$$

- ii) Aufgrund der Äquivalenz

$$\mathbf{g}=0 \iff \mathbf{g} \leq 0 \wedge -\mathbf{g} \leq 0$$

können in Satz F.3 auch Restriktionen in Gleichungsform behandelt werden. Man erhält dann gerade die klassische Aussage von Lagrange.

- iii) Ist f konkav und $\mathbf{g}$ konvex, so gilt (mit einer kleinen technischen Zusatzbedingung) wie im Sattelpunktsatz von Kuhn-Tucker wiederum die Äquivalenz. □

Index

A

- Ableitung 169
- Abschreibung
 - lineare.....23
- Additionssatz.....68
- Ähnlichkeitssatz.....68
- Algebra
 - Fundamentalsatz der 185
- algebraische Vielfachheit.....64, 190
- allgemeines Kontrollproblem.....114
- Anfangs-
 - wertaufgabe (AWA) 39
 - zeitpunkt.....116
 - zustand 116
- Approximation
 - sukzessive 43
- asymptotische Stabilität.....76
- autonome DGL.....39

B

- Backshiftoperator.....7
- Bedingung
 - von Legendre.....143
- Bellman-Gleichung.....147
- Bernoullische DGL.....49
- Bewegungsgleichung.....116
- Bifurkation.....90
- Black/Scholes DGL.....72
- Bolza
 - Problem von 118
- Brachistochronenproblem.....112

C

- charakteristisches Polynom einer Matrix
189

- Cobweb-Diagramm 12
- Cosinus hyperbolicus 174
- Costate-Variable 154

D

- d'Alembert
 - Reduktionsverfahren von.....51, 53
- Dämpfungssatz.....69
- definit
 - negativ 195
 - positiv 195
- Definitheit 195
- DeMoivre
 - Satz von 186
- Determinante
 - Hauptabschnitts- 196
- Diagramm
 - Cobweb-.....12
- Differential 169
- Differential-
 - gleichung (DGL)
 - autonome.....39
 - Bernoullische 49
 - Black/Scholes 72
 - exakte 46
 - explizite 39
 - gebrochen rationale.....46
 - gewöhnliche 37
 - implizite.....38
 - lineare.....51
 - mit getrennten Variablen 44
 - mit homogenen Funktionen 45
 - mit konstanten Koeffizienten.....61
 - partielle 37

Riccatische	50	erste Variation	123
zeitunabhängige	39	Euler-	
-gleichungssystem		-Gleichung	126
explizites	39	-Langrange-Gleichung	123
homogenes	51	-Poisson-Gleichung	130
implizites	39	Euler-Gleichung	187
inhomogenes	51	Evans	
lineares	51	Preisfestsetzungsmodell von	37
-operator	39	exakte	
-rechnung	169	DGL	46
Differentiationssatz	68	Existenzsatz von Peano	41
Differenz	6	explizite	
Differenzen-		DGL	39
-gleichung (ΔGL)	13	Differenzgleichung	13
explizite	13	explizites	
implizite	13	ΔGL -system	14
-gleichungssystem	13	DGL-system	39
explizites	14	Extremal	125
homogenes	17		
implizites	14	F	
inhomogenes	17	Faltung	68
lineares	17	Faltungssatz	69
-intervall	7	Federpendel	78
-operator	6	Fibonaccifolge	34
differenzierbare Funktion	169	finale Funktion	116
dissipatives System	91	Fixpunkt	179
Divisionssatz	69	Form	
Dynamische Programmierung	146	quadratische	195
E		Forwardoperator	6
Eckenbedingung		Fritz John	
von Weierstraß/Erdmann	131	Theorem von	199
Eigen-		Fundamental-	
vektor einer Matrix	189	lemma der Variationsrechnung	123
wert einer Matrix	189	satz der Algebra	185
einfach zusammenhängende Menge	47	Fundamentalsystem	
Elementarkatastrophe	90	einer linearen	
End-		ΔGL	20
zeitpunkt	116	DGL	55
zustand	116	eines linearen	
		ΔGL -Systems	25

DGL-Systems	55
Funktion	
differenzierbare	169
finale	116
intermediäre	116
Funktional	111
G	
gebrochen rationale DGL	46
geometrische Vielfachheit	64
geränderte Matrix	196
gewöhnliche DGL	37
Gleichung	
von Bellman	147
von Euler	126
von Euler-Poisson	130
von Hamilton-Jakobi	150
Gleichungssystem	
lineares	189
globale asymptotische Stabilität	76
Goldener Schnitt	34
Gradientensystem	90
H	
Hamilton-Jakobi Gleichung	150
Harrod	
Wachstumsmodell von	12
Haupt-	
vektor einer Matrix	189
Hauptabschnittsdeterminante	196
Hauptvektorraum	190
homogene Lösung	
einer linearen Δ GL	20
eines linearen Gleichungssystems .	189
homogenes	
Δ GL-system	17
I	
Imaginärteil	184
implizite	
DGL	38

Differenzengleichung	13
implizites	
Δ GL-system	14
DGL-system	39
indefinit	195
inhomogene Lösung	
einer linearen Δ GL	20
eines linearen Gleichungssystems .	189
inhomogenes	
Δ GL-system	17
Integrationssatz	68
intermediäre Funktion	116
invariante Menge	89
Isoklinen	94
isoperimetrisches Problem	139
J	
Jordanblock	192
Jordanform	192
Jordanmatrix	192
K	
klassisches Variationsproblem	122
komplexe Zahl	181
konjugiert komplexe Zahl	184
Kontroll-	
-menge	116
-problem, allgemeines	114
-theorie, optimale	154
-vektor	116
Kuhn-Tucker	
Satz von	199
L	
Lösung	
homogene	
einer linearen Δ GL	20
eines linearen Gleichungssystems	189
inhomogene	
einer linearen Δ GL	20
eines linearen Gleichungssystems	189

partikuläre	
einer linearen DGL	55
einer linearen Δ GL	20
eines linearen Δ GL-Systems	25
schwache	125
starke	125
stationäre	75
Lösungen	
stationäre	95
Lag-Operator	7
Lagoperator	7
Lagrange	
Problem von	118
Satz von	199
Laplace-Transformation	66
Legendre-Bedingung	143
lineare	
Δ GL	20
Abschreibung	23
DGL	51
lineares	
Δ GL-system	17
DGL-system	51
Gleichungssystem	189
Lipschitz-Bedingung	42
Ljapunov-Funktion	87
logistisches Modell	38
lokale asymptotische Stabilität	76
M	
Malthus-Gesetz	38
Matrix	
charakteristisches Polynom	189
definit unter Nebenbedingung	195
Eigenvektor einer	189
Eigenwert einer	189
geränderte	196
Hauptvektor einer	189
indefinite	195
negativ (semi)definite	195
positiv (semi)definite	195
symmetrische	196
Maximumprinzip	157
Maximumsnorm	170
Mayer	
Problem von	118
Menge	
einfach zusammenhängend	47
invariante	89
Modell	
logistisches	38
Spinnweb	12
von Evans	37
von Harrod	12
von Malthus	38
von Verhulst	38
momentane Wachstumsrate	38
Multiplikationssatz	69
N	
Nebenbedingung	
negativ definit unter einer	195
positiv definit unter einer	195
negativ definit	195
negativ definit unter Nebenbedingung	195
Norm	170
O	
optimale Kontrolltheorie	154
Optimalitätsprinzip	147
Orbit	97
P	
partielle Differentialgleichung	37
partikuläre Lösung	
einer linearen	
Δ GL	20
DGL	55
eines linearen	
Δ GL-Systems	25
Peano	

Existenzsatz von	41	von Taylor	173
Phasen-		schwache Lösung	125
diagramm	92, 96, 99	Schweinezyklus	12
portrait	96	semidefinit	
raum	95, 98	positiv	195
vektor	116	Sinus hyperbolicus	174
Polarkoordinaten	186	Spinnweb-Modell	12
Politik	151	stabil	76
positiv definit	195	Stabilität	76
positiv definit unter einer Nebenbedingung		asymptotische	76
195		global asymptotische	76
positiv semidefinit	195	lokal asymptotische	76
Preisfestsetzungsmodell		Sattelpunkts-	107
von Evans	37	starke Lösung	125
Problem		stationäre Lösung	75
isoperimetrisches	139	stationäre Lösungen	95
von Bolza	118	Steuer-	
von Lagrange	118	bereich	116
von Mayer	118	vektor	116
Q		sukzessive Approximation	43
quadratische Form	195	symmetrische Matrix	196
R		System	
Realteil	184	dissipatives	91
Reduktionsverfahren von d'Alembert .. 51,		T	
53		Taylor	
Riccatische DGL	50	Satz von	173
Richtungsfeld	92	Theorem von Fritz John	199
S		Trajektorie	97
Sattelpunkt	199	Transversalitätsbedingung	116
Sattelpunktstabilität	107	U	
Satz		Übergangsfunktion	151
Banachscher Fixpunkt-	179	V	
vom komplementären Schlupf	138	Variable	
von DeMoivre	186	Costate-	154
von Kuhn-Tucker	199	Variation	
von Lagrange	199	der Konstanten	22
von Peano	41	erste	123
von Picard-Lindelöf	42	zweite	123

Variationsproblem	
klassisches	122
Variationsrechnung	122
Fundamentallemma der	123
Verhulst	
Wachstumsmodell von	38
Verschiebungssatz	68
Vielfachheit	
algebraische	64, 190
geometrische	64

W

Wachstumsmodell	
von Harrod	12
von Verhulst	38
Wachstumsrate	
momentane	38
Weierstraß/Erdmann	
Eckenbedingung von	131

Z

Zahl	
komplexe	181
konjugiert komplexe	184
zeitunabhängige DGL	39
Ziel-	
bereich	120
funktional	116
Zulässigkeitsbereich	199
Zustands-	
vektor	116
zweite Variation	123