AIM - BIMalignment revised

uf

December 31, 2025

Contents

Intro

BIM . . . alignment . . .

Chapter 1

Basics

1.1 Characteristic parameters

Have a look at [**grafitoi**]: In order to analyze the various types of transition curves it is also necessary to determine other parameters that characterize the vehicle displacement along the curve:

Rate of change of non-compensated lateral acceleration parallel to the running plane. Lateral jerk ($\frac{m}{s^3}$):

$$v = \frac{da_q}{dt} = \frac{V}{3,6} \cdot \frac{dh(s)}{ds} \tag{1.1.1}$$

Rate of change of the lateral jerk ($\frac{m}{s^4}$):

$$a = \frac{d^2a_q}{dt^2} = \left(\frac{V}{3,6}\right)^2 \cdot \frac{d^2h(s)}{ds^2} \tag{1.1.2}$$

The third time derivative of the non-compensated lateral acceleration parallel to the running plane ($\frac{m}{s^5}$):

$$\dot{a} = \frac{d^3a_q}{dt^3} = \left(\frac{V}{3,6}\right)^3 \cdot \frac{d^3h(s)}{ds^3} \tag{1.1.3}$$

The angular roll velocity ($\frac{\text{rad}}{s}$):

$$\omega = \frac{d\psi}{dt} = \frac{V}{3,6} \cdot \frac{d\psi}{ds} \tag{1.1.4}$$

The angular roll acceleration is equal to the second time derivative of the cant angle ($\frac{\text{rad}}{s^2}$):

$$a = \frac{d^2\psi}{dt^2} = \left(\frac{V}{3,6}\right)^2 \cdot \frac{d^2\psi}{ds^2} \tag{1.1.5}$$

The angular jerk about roll axis is equal to the third time derivative of the cant angle ($\frac{\text{rad}}{s^3}$):

$$\dot{a} = \frac{d^3\psi}{dt^3} = \left(\frac{V}{3,6}\right)^3 \cdot \frac{d^3\psi}{ds^3} \tag{1.1.6}$$

1.2 Differentialgeometrie / Theoretische Physik

1.2.1 Kurven

In der Mathematik ist eine Kurve (von lateinisch curvus „gebogen, gekrümmt“) ein eindimensionales Objekt. Im Gegensatz etwa zu einer Geraden muss eine Kurve grundsätzlich keinen geraden, sondern kann vielmehr jeden beliebigen Verlauf annehmen.

Eindimensional bedeutet dabei informell, dass man sich auf der Kurve nur in eine Richtung (bzw. in die Gegenrichtung) bewegen kann. Ob die Kurve in der zweidimensionalen Ebene liegt (ebene Kurve), in einem höherdimensionalen Raum (siehe Raumkurve), oder gar in einer Mannigfaltigkeit (beispielsweise in einer LORENTZschen Mannigfaltigkeit) ist in diesem begrifflichen Zusammenhang unerheblich.

Je nach Teilgebiet der Mathematik gibt es unterschiedliche Präzisierungen dieser Beschreibung.

Parameterdarstellungen Eine Kurve kann als das Bild (Wertebereich) eines Weges definiert werden. Ein Weg ist (abweichend von der Umgangssprache) eine stetige Abbildung von einem Intervall in den betrachteten Raum, also z.B. in die euklidische Ebene $\mathbb{R}^2$. Ein Weg, dessen Bild eine gegebene Kurve ist, heißt auch Parameterdarstellung dieser Kurve. Wege werden deshalb manchmal auch als parametrisierte Kurven bezeichnet.

Gelegentlich, insbesondere bei historischen Bezeichnungen, wird zwischen Weg und Kurve nicht unterschieden. So ist die interessante Struktur bei der HILBERT-Kurve der Weg; das Bild dieses Weges ist das Einheitsquadrat, besitzt also keinerlei fraktale Struktur mehr.

Parametertransformation Eine Parametertransformation ϕ ist eine umkehrbar stetige Abbildung (Homöomorphismus), der zwei Wege (d. h. parametrisierte Kurven) c, c' ineinander überführt gemäß $c' = c \circ \phi$.

Für zwei Parameterdarstellungen $c\colon I \to C, c'\colon J \to C$ derselben Kurve C ist ein Parameterwechsel daher durch eine Parametertransformation $\phi\colon J \to I$ gegeben, so dass $c' = c \circ \phi$ – und damit umgekehrt auch $c = c' \circ \phi^{-1}$.

Statt Kurven mit den Bildern von Wegen zu identifizieren, könnte man sie auch (im Sinn der Kategorientheorie) äquivalent auch als die Äquivalenzklassen von Wegen mit gleichem Bild beschreiben, die durch Parametertransformationen (Homöomorphismen) ineinander übergeführt werden können. Diese Gleichwertigkeit macht man sich zunutze, um spezielle Klassen von Kurven zu definieren.

Gerichtete Kurven Durch die Parameterdarstellung erhält die Kurve einen „Richtungssinn“ in der Richtung des wachsenden Parameters.

Eine „gerichtete“ (oder „orientierte“) Kurve ist eine Äquivalenzklasse von Wegen (parametrisierten Kurven), die sich durch streng (strikt) monotone steigende Parametertransformationen ineinander überführen lassen.

In Anpassung des Sprachgebrauchs an den vorliegenden Verwendungszweck wird allgemein definiert:

Die „Spur“ einer (parametrisierten, gerichteten oder allgemeinen) Kurve ist die eindeutige Menge der Bildpunkte (einer beliebigen Parameterdarstellung derselben).

Glatte Kurven In diesem Fall verlangt man zusätzlich k-fache stetige Differenzierbarkeit ($k = 1, 2, \ldots, \infty$) für den Weg bzw. die Parameterdarstellungen einer (gerichteten) Kurve. Die entsprechenden Kurvenklassen werden mit C^k bezeichnet.

Gleichungsdarstellungen Eine Kurve kann auch durch eine oder mehrere Gleichungen in den Koordinaten beschrieben werden. Beispiele dafür sind wieder die Bilder der beiden durch die obigen Parameterdarstellungen gegebenen Kurven:

Die Gleichung $x^2 + y^2 = 1$ beschreibt den Einheitskreis in der Ebene. Die Gleichung $y^2 = x^2(x+1)$ beschreibt die oben in Parameterdarstellung angegebene Kurve mit Doppelpunkt. Ist die Gleichung wie hier durch ein Polynom gegeben, nennt man die Kurve algebraisch.

Funktionsgraphen Funktionsgraphen sind ein Spezialfall beider oben angegebenen Formen: Der Graph einer Funktion $f\colon D \to \mathbb{R}, \quad x \mapsto f(x)$ kann entweder als Parameterdarstellung $D \to \mathbb{R}^2, \quad t \mapsto (t, f(t))$ oder als Gleichung $\Gamma_f = \{(x, y) \in \mathbb{R}^2 \mid y = f(x)\}$ angegeben werden.

Differenzierbare Kurven, Krümmung Sei $[a, b] \subset \mathbb{R}$ ein Intervall und $c\colon [a, b] \to \mathbb{R}^n$ eine reguläre Kurve, d. h. $|c'(x)| \neq 0$ für alle $x \in (a, b)$. Die Länge der Kurve ist $l = \int_a^b |c'(t)|\, dt$

Die Funktion $x \mapsto \int_a^x |c'(t)|\, dt$ ist ein Diffeomorphismus $[a, b] \to [0, l]$, und die Verkettung von c mit dem inversen Diffeomorphismus liefert eine neue Kurve $\tilde{c}\colon [0, l] \to \mathbb{R}^n$ mit $|\tilde{c}'(x)| = 1$ für alle $x \in (0, l)$. Man sagt: $\tilde{c}$ ist nach der Bogenlänge parametrisiert.

$[a, b] \subset \mathbb{R}$ ein Intervall und $c\colon [a, b] \to \mathbb{R}^n$ eine nach der Bogenlänge parametrisierte Kurve. Die Krümmung von c an der Stelle s ist definiert als $\kappa(s) = |c''(s)|$. Für ebene Kurven kann man die Krümmung noch mit einem Vorzeichen versehen: Ist J die Drehung um 90°, dann ist $\kappa(s)$ festgelegt durch $c''(s) = \kappa(s) \cdot Jc'(s)$. Positive Krümmung entspricht Linkskurven, negative Rechtskurven.

Geschlossene Kurven Eine ebene Kurve $c\colon [0, 1] \to \mathbb{R}^2$ heißt geschlossen, wenn $c(0) = c(1)$, und einfach geschlossen, wenn zusätzlich c auf $[0, 1)$ injektiv ist. Der JORDANsche Kurvensatz besagt, dass eine einfach geschlossene Kurve die Ebene in einen beschränkten und einen unbeschränkten Teil zerlegt. Ist c eine geschlossene Kurve mit $c(t) \neq (0, 0)$ für alle $t \in [0, 1]$, kann man der Kurve eine Umlaufzahl zuordnen, die angibt, wie oft die Kurve um den Nullpunkt herumläuft.

Glatten geschlossenen Kurven kann man eine weitere Zahl zuordnen, die Tangentenumlaufzahl, die für eine nach der Bogenläge parametrisierte Kurve $c\colon [0, l] \to \mathbb{R}^2$ durch $\frac{1}{2\pi} \int_0^l \kappa(t)\, dt$ gegeben ist. Der Umlaufsatz von HEINZ HOPF besagt, dass eine einfache geschlossene Kurve Tangentenumlaufzahl 1 oder -1 hat.

Sei allgemein X ein topologischer Raum. Statt von geschlossenen Wegen $c\colon [0, 1] \to X$ mit $c(0) = c(1)$ spricht man auch von Schleifen mit Basispunkt $c(0)$. Weil der Quotientenraum $[0, 1]/\{0, 1\}$ homöomorph zum Einheitskreis S^1 ist, identifiziert man Schleifen mit stetigen Abbildungen $S^1 \to X$. Zwei Schleifen c_1, c_2 mit Basispunkt x heißen homotop, wenn man sie unter Beibehaltung des Basispunkts stetig ineinander deformieren kann, d. h. wenn es eine stetige Abbildung $H\colon [0, 1]^2 \to X$ mit $H(s, 0) = c_1(s)$, $H(s, 1) = c_2(s)$ für alle s und $H(0, t) = H(1, t) = x$ für alle t gilt. Die Äquivalenzklassen homotoper Schleifen bilden eine Gruppe, die Fundamentalgruppe von X. Ist $X = \mathbb{R}^2 - \{0\}$, dann ist die Fundamentalgruppe über die Windungszahl isomorph zu $\mathbb{Z}$.

Raumkurven Sei $[a, b] \subset \mathbb{R}$ ein Intervall und $c\colon [a, b] \to \mathbb{R}^3$ eine nach der Bogenlänge parametrisierte Kurve. Die folgenden Bezeichnungen sind Standard:

$$\begin{aligned} t(s) &= c'(s) \\ n(s) &= \frac{t'(s)}{|t'(s)|} \\ b(s) &= t(s) \times n(s) \end{aligned}$$

(definiert, wann immer $t'(s) \neq 0$). $t(s)$ ist der Tangentialvektor, $n(s)$ der Normalenvektor und $b(s)$ der Binormalenvektor, das Tripel (t, n, b) heißt begleitendes Dreibein. Die Krümmung ist $\kappa(s) = |t'(s)| = |c''(s)|$, die Windung $\tau(s)$ definiert durch $b'(s) = -\tau(s)n(s)$. Es gelten die frenetschen Formeln:

$$\begin{array}{lcccc} t' & = & & \kappa n & \\ n' & = & -\kappa t & + & \tau b \\ b' & = & & -\tau n & \end{array}$$

Der Hauptsatz der lokalen Kurventheorie besagt, dass man eine Kurve aus Krümmung und Windung rekonstruieren kann: Sind glatte Funktionen $\kappa, \tau\colon [0, l] \to \mathbb{R}$ mit $\kappa(s) > 0$ für alle $s \in [0, l]$ (der Wert 0 ist für κ also nicht erlaubt), so gibt es bis auf Bewegungen genau eine entsprechende Kurve.

Die von je zwei der drei Vektoren t, n oder b aufgespannten Ebenen durch den Kurvenpunkt tragen besondere Namen:

- Die Oskulationsebene oder Schmiegebene wird von t und n aufgespannt.
- Die Normalebene wird von n und b aufgespannt.
- Die rektifizierende Ebene oder Streckebene wird von t und b aufgespannt.

Kurven als eigenständige Objekte Kurven ohne umgebenden Raum sind in der Differentialgeometrie relativ uninteressant, weil jede eindimensionale Mannigfaltigkeit diffeomorph zur reellen Geraden $\mathbb{R}$ oder zur Einheitskreislinie S^1 ist. Auch Eigenschaften wie die Krümmung einer Kurve sind intrinsisch nicht feststellbar.

In der algebraischen Geometrie und damit zusammenhängend in der komplexen Analysis versteht man unter „Kurven“ in der Regel eindimensionale komplexe Mannigfaltigkeiten, oft auch als Riemannsche Flächen bezeichnet. Diese Kurven sind eigenständige Studienobjekte, das prominenteste Beispiel sind die elliptischen Kurven.

1.2.2 Algebraische Kurve

Eine algebraische Kurve ist eine eindimensionale algebraische Varietät, kann also durch eine Polynomgleichung beschrieben werden. Ein wichtiger Spezialfall sind die ebenen algebraischen Kurven, also algebraische Kurven, die in der affinen oder projektiven Ebene verlaufen.

Geschichtlich beginnt die Beschäftigung mit algebraischen Kurven schon in der Antike mit der Untersuchung von Geraden und Kegelschnitten. Im 17. Jahrhundert wurden sie im Rahmen der analytischen Geometrie Gegenstand der Analysis und Isaac Newton behandelte systematisch Kubiken. Die Beschäftigung mit ihnen erreichte im 19. Jahrhundert durch die Behandlung im Rahmen der projektiven Geometrie einen Höhepunkt (unter anderem August Ferdinand Möbius, Julius Plücker). Dabei wird der Punkt im Unendlichen systematisch mit berücksichtigt. Die natürliche

Betrachtungsweise ist nach dem Fundamentalsatz der Algebra über den komplexen Zahlen, und die klassische Theorie wurde durch die von Bernhard Riemann entdeckte Verbindung zu Riemannschen Flächen – die im Komplexen Kurven sind – auf eine neue Grundlage gestellt. In der Zahlentheorie (arithmetische Geometrie) werden auch Kurven über anderen Körpern als den reellen und komplexen Zahlen und über Ringen betrachtet.

Algebraische Kurven gehören zu den einfachsten Objekten der algebraischen Geometrie, in der sie mit rein algebraischen Methoden behandelt werden und nicht mit Methoden der Analysis. Höherdimensionale Varietäten der algebraischen Geometrie sind zum Beispiel algebraische Flächen. Man kann algebraische Kurven aber auch im Rahmen der komplexen Analysis untersuchen.

Im Folgenden werden die verwendeten Begriffe am einfachsten Fall ebener algebraischer Kurven erläutert. Man kann algebraische Kurven etwa als Schnittkurve algebraischer Flächen auch in mehr als zwei Dimensionen definieren. Ihre Klassifikation in drei Dimensionen nach Grad d und Geschlecht g war Gegenstand von zwei großen Arbeiten zum Steinerpreis in den 1880er Jahren von Max Noether und Georges Henri Halphen, deren Beweise und Arbeit aber noch unvollständig war. Gegenstand der Klassifikation ist festzustellen, welche Paare $(d,\ g)$ existieren. Algebraische Kurven können immer in den dreidimensionalen projektiven Raum eingebettet werden, so dass die Betrachtung von zwei und drei Raumdimensionen reicht.

Definition und wichtige Eigenschaften Eine ebene algebraische Kurve über einem Körper K wird durch ein nichtkonstantes Polynom in zwei Variablen x und y definiert, dessen Koeffizienten aus K stammen. Dabei werden zwei Polynome miteinander identifiziert, wenn das eine durch Multiplikation mit einer von Null verschiedenen Zahl aus K aus dem anderen hervorgeht. Der Grad des Polynoms wird als Grad der Kurve bezeichnet.

Dieser Definition liegt folgende Motivation zu Grunde: Ist f ein solches Polynom, so kann man die Nullstellenmenge $V(f) = \{(x, y) \in K^2 | f(x, y) = 0\}$ in der Ebene K^2 betrachten. Diese Menge stellt häufig ein Objekt dar, das man auch anschaulich als Kurve bezeichnen würde, so ist beispielsweise $\{(x, y) \in \mathbb{R}^2 | x^2 + y^2 - 1 = 0\}$ ein Kreis. Auch bei der Definition von $V(f)$ spielt ein konstanter Faktor keine Rolle.

Ist der Körper K algebraisch abgeschlossen, so kann man nach dem hilbertschen Nullstellensatz aus der Menge $V(f)$ das Polynom f wiedergewinnen, falls dieses in lauter verschiedene irreduzible Faktoren zerfällt. In diesem Fall muss also nicht streng zwischen dem definierenden Polynom und dessen Nullstellenmenge unterschieden werden.

Ist der Körper K dagegen nicht algebraisch abgeschlossen, so stellt $V(f)$ nicht immer eine Kurve in der Ebene dar. So werden durch $\{(x, y) \in \mathbb{R}^2 | x^2 + y^2 + 1 = 0\}$ und $\{(x, y) \in \mathbb{R}^2 | x^2 + y^2 = 0\}$ im Reellen die leere Menge beziehungsweise ein Punkt definiert, beides keine eindimensionalen Objekte. Erst im Komplexen erzeugen diese Polynome Kurven: ein Kreis und ein sich schneidendes Geradenpaar.

Man sagt daher, eine Kurve habe eine Eigenschaft geometrisch, falls die Menge $V(f)$ diese Eigenschaft über dem algebraischen Abschluss von K besitzt.

Abstrakter kann man eine algebraische Kurve auch als ein eindimensionales separiertes algebraisches Schema über einem Körper definieren. Häufig werden noch weitere Voraussetzungen wie geometrische Reduziertheit oder Irreduzibilität in die Definition mit aufgenommen.

Irreduzibilität Ist das definierende Polynom reduzibel, falls es also in zwei nichttriviale Faktoren zerlegt werden kann, so kann auch die Kurve in zwei unabhängige Komponenten zerlegt werden.

Zum Beispiel ist das Polynom $f(x, y) = xy$ reduzibel, da es in die Faktoren x und y zerlegt werden kann. Die durch f definierte Kurve besteht daher aus zwei Geraden.

Bei einem irreduziblen Polynom kann die Kurve nicht zerlegt werden, welche dann ebenfalls irreduzibel genannt wird.

Singularitäten Im Normalfall lässt sich in jedem Punkt der algebraischen Kurve genau eine Tangente an die Kurve zeichnen. In diesem Fall nennt man den Punkt glatt oder nichtsingulär. Es kann aber auch der Fall auftreten, dass die Kurve in einem oder mehreren Punkten einen Selbstschnitt oder eine Spitze besitzt. Im ersten Fall besitzt die Kurve in diesem Punkt zwei oder mehr Tangenten, im zweiten fallen diese Tangenten zu einer mehrfachen Tangente zusammen.

Beispiele für solche singulären Punkte finden sich bei der NEILschen Parabel mit der Gleichung $y^2 = x^3$, diese hat eine Spitze im Nullpunkt. Einen Doppelpunkt, also einen Punkt, der zwei Mal in verschiedenen Richtungen durchlaufen wird, findet man beim kartesischen Blatt, das durch $x^3 + y^3 - 3xy = 0$ gegeben ist.

Projektive Kurven Häufig ist es von Vorteil, algebraische Kurven nicht im Affinen, sondern in der projektiven Ebene zu betrachten. Diese kann durch sogenannte homogene Koordinaten $[x : y : z]$ beschrieben werden, wobei x, y und z nicht gleichzeitig 0 werden dürfen und zwei Punkte als gleich aufgefasst werden, wenn sie durch Multiplikation mit einer von 0 verschiedenen Zahl auseinander hervorgehen. Für $\lambda \neq 0$ gilt also $[x : y : z] = [\lambda \cdot x : \lambda \cdot y : \lambda \cdot z]$. Um im Projektiven algebraische Kurven zu definieren, benötigt man also Polynome in drei Variablen x, y und z. Würde man hier beliebige Polynome verwenden, so ergäben sich große Probleme auf Grund der Tatsache, dass die Darstellung der Punkte nicht eindeutig ist: So sind die Punkte $[1 : 1 : 1]$ und $[2 : 2 : 2]$ gleich, aber das Polynom $f(x, y, z) = x^2 - y$ verschwindet bei der ersten Darstellung, nicht aber bei der zweiten.

Dieses Problem tritt nicht auf, wenn man sich auf homogene Polynome beschränkt: Zwar können sich auch hier die Werte, die das Polynom bei verschiedenen Darstellungen annimmt, unterscheiden, aber die Eigenschaft, ob das Polynom eine Nullstelle hat, ist von der Wahl der Darstellung des Punktes unabhängig.

Um zu einer affinen Kurve die zugehörige projektive Kurve zu finden, homogenisiert man das definierende Polynom: In jedem Term fügt man eine so große z-Potenz ein, dass sich ein homogenes Polynom ergibt: Aus der Gleichung $x^2 - y + 1 = 0$ wird also $x^2 - yz + z^2 = 0$.

Der umgekehrte Vorgang wird als Dehomogenisieren bezeichnet. Hier setzt man in das homogene Polynom für z (oder eine Variable, falls man nach einer anderen Variablen dehomogenisieren möchte) den Wert 1 ein.

Schnitte zweier Kurven Betrachtet man beispielsweise eine Gerade und eine Parabel, so erwartet man im Allgemeinen zwei gemeinsame Punkte. Durch verschiedene Umstände können auch weniger gemeinsame Punkte auftreten, diese Fälle kann man jedoch alle durch spezielle Voraussetzungen oder Definitionen umgehen:

- Die Gerade und die Parabel können gar keinen Schnittpunkt besitzen, dies umgeht man, indem man voraussetzt, dass der zu Grunde liegende Körper algebraisch abgeschlossen ist.
- Die Gerade kann durch den Scheitel der Parabel senkrecht nach oben verlaufen und somit nur einen Punkt mit ihr gemeinsam haben. Dies tritt nicht auf, wenn man sich in der projektiven Ebene befindet, hier haben Gerade und Parabel in diesem Fall einen weiteren Schnittpunkt im Unendlichen.

- Die Gerade kann eine Tangente an die Parabel sein. Auch in diesem Fall existiert nur ein gemeinsamer Punkt. Mit einer geeigneten Definition von Schnittmultiplizität kann dieser Schnittpunkt jedoch doppelt gezählt werden.

Unter den obigen Voraussetzungen gilt der Satz von BÉZOUT: Die Anzahl der gemeinsamen Punkte zweier projektiver ebener algebraischer Kurven von Grad n und m ohne gemeinsame Komponenten beträgt nm.

Beispiele für algebraische Kurven

Kurven nach Grad geordnet

- Die ebenen algebraischen Kurven von Grad 1 sind genau die Geraden. Die Gleichungen $x = 0$ und $y = 0$ beispielsweise beschreiben die Koordinatenachsen, die Gleichung $x = y$ oder äquivalent $x - y = 0$ die erste Winkelhalbierende.
- Die ebenen algebraischen Kurven von Grad 2 (Quadriken) sind genau die Kegelschnitte, darunter der durch $x^2 + y^2 = 1$ beschriebene Einheitskreis und die Normalparabel mit der Formel $y = x^2$. Die reduziblen Kurven sind dabei die entarteten Kegelschnitte.
- Bei Grad 3 (Kubiken) treten zum ersten Mal irreduzible Kurven mit Singularitäten auf, zum Beispiel die NEIlsche Parabel mit der Gleichung $y^3 = x^2$ und das kartesische Blatt, das durch $x^3+y^3-3xy = 0$ gegeben ist. Die elliptischen Kurven sind ebenfalls wichtige Beispiele ebener algebraischer Kurven von Grad 3.
- Eine Spirische Kurve ist eine algebraische Kurve vom Grad 4 (Quartik). Sonderfälle davon sind die CASSINIsche Kurve, Lemniskate von BERNOULLI und Lemniskate von BOOTH.
- Kurven vom Grad 5 werden als Quintiken bezeichnet, Kurven vom Grad 6 als Sextiken.

Kurven nach Geschlecht geordnet

- Kurven vom Geschlecht 0 sind rationale Kurven.
- Kurven vom Geschlecht 1 sind elliptische Kurven.
- Zu den Kurven vom Geschlecht mindestens 2 gehören hyperelliptische Kurven, die KLEINsche Quartik $x^3y + y^3z + z^3x = 0$ und die FERMAT-Kurve $x^n + y^n - z^n = 0$.

Duale Kurve Eine Kurve kann statt durch ihre Punkte auch durch ihre Tangenten beschrieben werden. Ein in diesem Zusammenhang wichtiges Problem ist die Frage, wie viele Tangenten sich „in der Regel“ von einem nicht auf der Kurve liegenden Punkt aus an eine Kurve n-ter Ordnung legen lassen. Diese Anzahl heißt die Klasse der Kurve. Für eine solche Kurve ohne singuläre Punkte (wie etwa Doppelpunkte oder Spitzen) ist diese Klasse gleich $n(n-1)$. Jeder Doppelpunkt verkleinert die Klasse um 2 und jede Spitze um 3. Das ist eine Hauptaussage der PLÜCKERschen Formeln, die sich außerdem noch mit der Anzahl der Wendepunkte und Doppeltangenten befassen. Hierfür muss der Grundkörper algebraisch abgeschlossen sein.

So ist zum Beispiel eine singularitätenfreie Kurve dritter Ordnung von 6. Klasse, besitzt sie einen Doppelpunkt, ist sie von vierter, und wenn sie eine Spitze hat, von dritter Klasse.

Im homogenen Fall haben Geraden, also auch Tangenten, eine Gleichung der Form $ax+by+cz = 0$, wobei a, b und c nicht alle verschwinden dürfen und mit einer beliebigen von 0 verschiedenen Zahl multipliziert werden dürfen. Damit kann man dieser Geraden den Punkt $[a : b : c]$ zuordnen. Aus der Menge der Tangenten an eine gegebene Kurve erhält man somit eine Punktemenge in der projektiven Ebene. Es stellt sich heraus, dass diese Menge selbst wieder eine algebraische Kurve ist, die sogenannte duale Kurve.

Dual zueinander sind folgende Begriffe:

- Kurvenpunkt und Kurventangente
- Doppelpunkt und Doppeltangente
- Wendepunkt und Spitze
- Ordnung und Klasse

Die duale Kurve der dualen Kurve ist wieder die ursprüngliche Kurve.

...

1.3 Kinematik

1.3.1 Bewegungszustand

Als Bewegungszustand bezeichnet man in der Mechanik die momentane Bewegung eines Körpers. Diese kann in einer Translations- und/oder Rotationsbewegung bestehen.

Hinsichtlich der Translationsbewegung wird der Bewegungszustand eines Körpers gekennzeichnet durch die Geschwindigkeit seines Massenmittelpunkts mit ihren momentanen Werten für Betrag und Richtung, also in vektorieller Form. Wird der Körper schneller oder langsamer, oder ändert er auch nur seine Bewegungsrichtung, ändert sich sein Bewegungszustand. Ein quantitatives Maß für den translatorischen Bewegungszustand ist der Impuls.

Rotiert der Körper um seinen Massenmittelpunkt, gehört auch diese Rotationsbewegung zu seinem Bewegungszustand. Ein Maß für diesen Teil des Bewegungszustandes ist der Drehimpuls.

Die gleichförmige Bewegung ist ein Beispiel für eine Bewegung, bei der der Bewegungszustand unverändert bleibt. Dagegen wird bei einer gleichförmigen Kreisbewegung eines Körpers der Bewegungszustand nicht beibehalten, denn hier ändert sich fortwährend die Richtung der Geschwindigkeit.

Der Trägheitssatz oder das erste NEWTONsche Gesetz der Mechanik besagt, dass jeder Körper, der nicht von äußeren Kräften beeinflusst wird, in seinem Bewegungszustand verharrt. Mit anderen Worten ist das Bestreben eines Körpers, seinen Bewegungszustand beizubehalten, Ausdruck seiner Trägheit. Insbesondere bewegt sich bei einem Körper ohne äußere Kräfte der Massenmittelpunkt mit gleichbleibender Geschwindigkeit geradlinig weiter. Im Fall der Rotation um den Massenmittelpunkt bleibt dann der Drehimpuls nach Betrag und Richtung konstant, jedoch nicht unbedingt die Drehachse und Rotationsgeschwindigkeit.

Im scheinbaren Gegensatz zum Trägheitssatz ist es Alltagserfahrung, dass ein sich bewegender Körper gerade dann langsamer wird, wenn keine Kraft feststellbar ist, die ihn antreibt. Das erklärt sich dadurch, dass bei jeder Bewegung Bremskräfte wie der Luftwiderstand und sonstige Reibungskräfte vorhanden sind. Diese sind für die Abbremsung des Körpers, also die Änderung seines Bewegungszustandes, ursächlich. Allgemein besagt das zweite NEWTONsche Gesetz der Mechanik, wie sich der Bewegungszustand ändert, wenn eine resultierende äußere Kraft auf den Körper wirkt.

1.3.2 Geschwindigkeit

Die Geschwindigkeit ist neben dem Ort und der Beschleunigung einer der grundlegenden Begriffe der Kinematik, eines Teilgebiets der Mechanik. Die Geschwindigkeit beschreibt, wie schnell und in welcher Richtung ein Körper oder ein Phänomen (beispielsweise ein Wellenberg) im Lauf der Zeit seinen Ort verändert. Eine Geschwindigkeit wird durch ihren Betrag und die Bewegungsrichtung angegeben; es handelt sich also um eine vektorielle Größe. Als Formelzeichen ist $\vec{v}$ üblich nach dem lateinischen bzw. englischen Wort für Geschwindigkeit (lateinisch velocitas, englisch velocity).

Oft wird mit dem Wort Geschwindigkeit nur ihr Betrag gemeint (Formelzeichen v), der anschaulich gesprochen das momentane Tempo (englisch speed) der Bewegung wiedergibt, wie es beispielsweise im Auto vom Tachometer angezeigt wird. v gibt an, welche Wegstrecke ein Körper innerhalb einer bestimmten Zeitspanne zurücklegt, wenn die Geschwindigkeit entsprechend lange konstant bleibt; es handelt sich um eine skalare Größe. Die international verwendete Einheit ist Meter pro Sekunde (m/s), gebräuchlich sind auch Kilometer pro Stunde (km/h) und – vor allem in der See- und Luftfahrt – Knoten (kn).

Die höchstmögliche Geschwindigkeit, mit der sich die Wirkung einer bestimmten Ursache räumlich ausbreiten kann, ist die Lichtgeschwindigkeit c. Diese Obergrenze gilt also auch für jedwede Informationsübertragung. Körper, die eine Masse besitzen, können sich nur mit geringeren Geschwindigkeiten als c bewegen.

Eine Geschwindigkeitsangabe ist immer relativ zu einem Bezugssystem zu verstehen. Ruht ein Körper in einem Bezugssystem, so hat er in einem anderen Bezugssystem, welches sich gegenüber dem ersten mit der Geschwindigkeit $\vec{v}$ bewegt, die entgegengesetzt gleich große Geschwindigkeit $-\vec{v}$.

Definition Bewegt sich ein Objekt entlang einer Bahnkurve, wobei es sich zum Zeitpunkt t im Punkt A und zu einem späteren Zeitpunkt $t + \Delta t$ im Punkt B befindet, so ergibt sich seine Geschwindigkeit $\vec{v}(t)$ zum Zeitpunkt t (bzw. im Punkt A) näherungsweise aus der Ortsänderung $\Delta\vec{r}$ und der dafür benötigten Zeitspanne Δt gemäß

$$\vec{v}(t) \approx \frac{\Delta\vec{r}}{\Delta t}.$$

Dabei ist $\Delta\vec{r} = \vec{r}(t + \Delta t) - \vec{r}(t) = \vec{r}_B - \vec{r}_A$ der Verbindungsvektor von Punkt A zu Punkt B. Geometrisch entspricht er der Sehne des Kurvenabschnitts zwischen den beiden Punkten. Außerdem gibt er näherungsweise die Richtung der Geschwindigkeit an. Aus der Näherung erhält man die exakte Definition für die Momentangeschwindigkeit zum Zeitpunkt t (bzw. am Punkt A), wenn man das Zeitintervall Δt gegen null gehen lässt. Dabei rückt (aufgrund der Stetigkeit der Bewegung) der Punkt B beliebig nah an den Punkt A heran, so dass auch $\Delta\vec{r}$ gegen null geht; der Quotient $\frac{\Delta\vec{r}}{\Delta t}$ hingegen strebt einem Grenzwert zu, der gerade der Momentangeschwindigkeit entspricht:

$$\vec{v}(t) = \lim_{\Delta t \to 0} \frac{\vec{r}(t + \Delta t) - \vec{r}(t)}{\Delta t}.$$

Hierfür schreibt man auch

$$\vec{v}(t) = \frac{\mathrm{d}\vec{r}}{\mathrm{d}t} \quad \text{oder} \quad \vec{v}(t) = \dot{\vec{r}},$$

da es sich um eine Zeitableitung handelt.

Da die Sehne $\Delta\vec{r}$ beim Grenzübergang die Richtung der Tangente an die Bahnkurve annimmt, ist dies auch die Richtung der Momentangeschwindigkeit.

Der Betrag der Momentangeschwindigkeit (das „Tempo“ oder die Bahngeschwindigkeit) ist durch den Betrag des Geschwindigkeitsvektors

$$v = |\vec{v}| = \left| \lim_{\Delta t \to 0} \frac{\Delta \vec{r}}{\Delta t} \right| = \left| \frac{\mathrm{d}\vec{r}}{\mathrm{d}t} \right| = \left| \dot{\vec{r}} \right|$$

gegeben, wobei $|\vec{r}| = r$ der Betrag des Ortsvektors $\vec{r}$ ist. Die Bahngeschwindigkeit ist nicht dasselbe wie $|\dot{r}|$, wie man beispielsweise an der Kreisbewegung mit $r = \text{konst.},\ v \neq 0,\ \dot{r} = 0$ sehen kann.

Den Betrag der Momentangeschwindigkeit kann man auch erhalten, wenn man statt der dreidimensionalen Bahnkurve nur die Weglänge (Symbol s) entlang der Bahnkurve berücksichtigt. Man bildet hierfür den Grenzwert des Quotienten aus zurückgelegter Weglänge Δs und benötigter Zeit Δt:

$$v = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \frac{\mathrm{d}s}{\mathrm{d}t} = \dot{s}$$

Anfangsgeschwindigkeit Wenn die Geschwindigkeit eines Körpers oder Massenpunkts zu Beginn eines bestimmten Bewegungsabschnittes interessiert, wird sie auch als Anfangsgeschwindigkeit (Formelzeichen meist v_0) bezeichnet.

Die Anfangsgeschwindigkeit ist eine der Anfangsbedingungen beim Lösen der Bewegungsgleichungen in der klassischen Mechanik, zum Beispiel für numerische Simulationen in der Himmelsmechanik. Sie ist ein wichtiger Parameter z. B. für die Flugbahn beim senkrechten und schrägen Wurf sowie für die Reichweite von Schusswaffen oder Raketen.

Endgeschwindigkeit Die Endgeschwindigkeit (auch: Grenzgeschwindigkeit) ist die Geschwindigkeit, die ein Objekt am Ende seiner Beschleunigung erreicht hat.

Ein Objekt erreicht seine Endgeschwindigkeit, wenn die bremsenden Kräften, beim freien Fall insbesondere der mit der Fallgeschwindigkeit wachsende Luftwiderstand, durch Zu- oder Abnahme der Geschwindigkeit so stark geworden sind, dass sich ein Kräftegleichgewicht aller beteiligten Kräfte ausbildet. Die Beschleunigung bei Erreichen der Endgeschwindigkeit ist daher null.

Der Begriff wird auch in der Technik verwendet. Im Automobilsektor spricht man zum Beispiel von Endgeschwindigkeit oder Maximalgeschwindigkeit, wenn sich das Fahrzeug begrenzt durch Motorleistung und äußere Umstände nicht weiter beschleunigen lässt.

Einfache Sonderfälle

Geradlinig gleichförmige Bewegung Von geradlinig gleichförmiger Bewegung spricht man, wenn die Geschwindigkeit $\vec{v}$ des Objekts immer die gleiche ist (d. h. gleich in Betrag und Richtung), was gleichbedeutend mit der Beschleunigung $\vec{a}(t) = \vec{0}$ ist. In diesem Fall bewegt sich das Objekt auf einer Geraden, entlang derer man üblicherweise das Koordinatensystem ausrichtet, so dass die Geschwindigkeit eine skalare Größe v ist. Dann gilt:

$$v = \frac{s}{t}.$$

Hierbei ist s der in der Zeitspanne t zurückgelegte Weg.

Gleichmäßig beschleunigte Bewegung Bei einer gleichmäßig beschleunigten Bewegung hat die Beschleunigung $\vec{a}$ stets den gleichen Betrag und die gleiche Richtung. Ist die Bewegungsrichtung parallel zu $\vec{a}$, so bewegt sich das Objekt auf einer Geraden. Praktischerweise richtet man das Koordinatensystem in Richtung der Bewegung aus und schreibt Beschleunigung und Geschwindigkeit als Skalar. Dann gilt

$$v(t) = a \cdot t + v_0.$$

Hierbei steht v_0 für die Anfangsgeschwindigkeit.

Kreisbewegung Die Geschwindigkeit einer Kreisbewegung bezeichnet man als Umfangsgeschwindigkeit oder allgemein als Bahngeschwindigkeit:

$$\vec{v} = \vec{\omega} \times \vec{r}$$

Hierbei steht ω für die Winkelgeschwindigkeit und r für den Radius der Kreisbewegung.

Bei einer gleichförmigen Kreisbewegung ist der Betrag der Umfangsgeschwindigkeit konstant und kann als Quotient aus der auf der Kreisbahn zurückgelegten Streckenlänge und der dafür benötigten Zeit T ausgedrückt werden:

$$v = \frac{2 \cdot \pi \cdot r}{T}$$

Beziehungen zu anderen physikalischen Größen

Beziehung zum Ort Bewegt sich ein Massepunkt im Raum (dreidimensionale Bewegung), so kann man aus dem zeitlichen Verlauf des Geschwindigkeitsvektors $\vec{v}$ auf die Verschiebung des Massepunkts schließen, indem man $\vec{v}$ über die Zeit integriert:

$$\vec{r}_2 - \vec{r}_1 = \int_{t_1}^{t_2} \vec{v}(t)\,\mathrm{d}t,$$

wobei $\vec{r}_2 = \vec{r}(t_2)$ und $\vec{r}_1 = \vec{r}(t_1)$. Hieraus erhält man die Position des Massepunktes zum Endzeitpunkt als

$$\vec{r}_2 = \vec{r}_1 + \int_{t_1}^{t_2} \vec{v}(t)\,\mathrm{d}t.$$

Bewegt sich der Massepunkt auf einer Geraden (geradlinige bzw. eindimensionale Bewegung), so richtet man das Koordinatensystem üblicherweise entlang dieser Geraden aus. Die Position des Teilchens wird dann allein durch die Koordinate $x = x(t)$ beschrieben. Die oben stehende Formel vereinfacht sich in diesem Fall zu

$$x_2 - x_1 = \int_{t_1}^{t_2} v(t)\,\mathrm{d}t.$$

Dies ist die kinematische Version des Hauptsatzes der Differential- und Integralrechnung.

Beziehung zur Wegstrecke Die zurückgelegte Strecke s erhält man durch Integration des Geschwindigkeitsbetrags $|\vec{v}|$ über die Zeit:

$$s = \int_{t_1}^{t_2} |\vec{v}(t)|\,\mathrm{d}t.$$

Im einfachsten Fall, nämlich bei konstanter Geschwindigkeit, wird daraus $s = v \cdot (t_2 - t_1)$.

Beziehung zu Beschleunigung und Ruck Die erste Zeitableitung der Geschwindigkeit ist die Beschleunigung: $\vec{a}(t) = \dot{\vec{v}}(t) = \ddot{\vec{r}}(t)$.

Umgekehrt gewinnt man die Geschwindigkeit aus der Beschleunigung durch Integration:

$$\vec{v}(t) = \vec{v}_0 + \int_0^t \vec{a}(\tau)\,\mathrm{d}\tau.$$

Findet die Bewegung auf einer Geraden statt, so richtet man das Koordinatensystem praktischerweise in Richtung der Bewegung aus, und erhält die skalare Gleichung

$$v(t) = v_0 + \int_0^t a(\tau)\,\mathrm{d}\tau.$$

Die zweite Zeitableitung der Geschwindigkeit ergibt den Ruck einer Bewegung: $\vec{j}(t) = \ddot{\vec{v}}(t) = \dot{\vec{a}}(t)$. Umgekehrt gewinnt man die Geschwindigkeit aus dem Ruck durch zweifache Integration.

Beziehung zu Impuls und kinetischer Energie Der Impuls – also anschaulich gesprochen der „Schwung“ – eines Körpers der Masse m berechnet sich nach $\vec{p} = m \cdot \vec{v}$, während die kinetische Energie durch $E_{\mathrm{kin}} = \frac{1}{2}mv^2 = \frac{p^2}{2m}$ gegeben ist. Streng genommen gelten die letzten beiden Gleichungen nur näherungsweise für den sogenannten nichtrelativistischen Fall, also für Geschwindigkeiten, die viel kleiner als die Lichtgeschwindigkeit sind.

Geschwindigkeiten und Bezugssystem Je nach verwendetem Bezugssystem bzw. Koordinatensystem haben sich verschiedene Bezeichnungen eingebürgert:

Im homogenen Schwerefeld wird oft ein kartesisches Koordinatensystem verwendet. Geschwindigkeiten, die parallel zur Fallbeschleunigung $\vec{g}$ gerichtet sind, werden meist als Vertikalgeschwindigkeiten, solche, die orthogonal zu dieser Richtung sind, als Horizontalgeschwindigkeiten bezeichnet.

Bei Polarkoordinaten ist die Radialgeschwindigkeit $\vec{v}_{\mathrm{r}}$ die Komponente des Geschwindigkeitsvektors in Richtung des Ortsvektors, also längs der Verbindungslinie zwischen dem bewegten Objekt und dem Koordinatenursprung. Die Komponente senkrecht dazu heißt Umfangsgeschwindigkeit $\vec{v}_{\perp}$. Somit ergibt sich: $\vec{v} = \vec{v}_{\perp} + \vec{v}_{\mathrm{r}}$. Das Vektorprodukt aus der Winkelgeschwindigkeit und dem Ortsvektor ergibt die Umfangsgeschwindigkeit: $\vec{v}_{\perp} = \vec{\omega} \times \vec{r}$.

Bei Bewegungen auf einer Kreisbahn um den Koordinatenursprung, aber auch nur in diesem Fall, ist die Radialgeschwindigkeit null und die Umfangsgeschwindigkeit gleich der Tangentialgeschwindigkeit, also der Bahngeschwindigkeit längs der Tangente an die Bahnkurve.

Aus der Änderung des Abstands zum Koordinatenursprung (Radius) folgt die Radialgeschwindigkeit: $\vec{v}_{\mathrm{r}} = \dot{r}\,\frac{\vec{r}}{|\vec{r}|}$.

Setzt man voraus, dass es ein allgemein gültiges Bezugssystem gibt, so nennt man die Geschwindigkeiten, die in diesem System gemessen werden, Absolutgeschwindigkeiten. Geschwindigkeiten, die sich auf einen Punkt beziehen, der sich selbst in diesem System bewegt, heißen Relativgeschwindigkeiten.

Das Relativitätsprinzip besagt jedoch, dass es keinen physikalischen Grund gibt, warum man ein bestimmtes Bezugssystem herausgreifen und gegenüber anderen Systemen bevorzugen sollte. Sämtliche physikalischen Gesetze, die in einem Inertialsystem gelten, gelten auch in jedem anderen. Welche Bewegungen man als „absolut“ ansieht, ist also vollkommen willkürlich. Deswegen wird der Begriff der Absolutgeschwindigkeit spätestens seit der speziellen Relativitätstheorie vermieden. Stattdessen sind alle Geschwindigkeiten Relativgeschwindigkeiten. Aus diesem Relativitätsprinzip

folgt, zusammen mit der Invarianz der Lichtgeschwindigkeit, dass Geschwindigkeiten nicht – wie im obigen Beispiel stillschweigend angenommen – einfach addiert werden dürfen. Stattdessen gilt das relativistische Additionstheorem für Geschwindigkeiten. Dies macht sich jedoch erst bei sehr hohen Geschwindigkeiten bemerkbar.

Geschwindigkeit zahlreicher Teilchen Betrachtet man ein System aus vielen Teilchen, so ist es meist nicht mehr sinnvoll oder überhaupt möglich, für jedes einzelne Teilchen eine bestimmte Geschwindigkeit anzugeben. Stattdessen arbeitet man mit der Geschwindigkeitsverteilung, die angibt, wie häufig ein bestimmter Bereich von Geschwindigkeiten in dem Teilchenensemble auftritt. In einem idealen Gas gilt beispielsweise die Maxwell-Boltzmann-Verteilung (siehe nebenstehende Abbildung): Die meisten Teilchen haben eine Geschwindigkeit in der Nähe der wahrscheinlichsten Geschwindigkeit, die durch das Maximum der Maxwell-Boltzmann-Verteilung angezeigt wird. Sehr kleine und sehr große Geschwindigkeiten kommen auch vor, werden aber nur von ganz wenigen Teilchen angenommen. Die Lage des Maximums ist temperaturabhängig. Je heißer das Gas ist, desto höher ist die wahrscheinlichste Geschwindigkeit. Mehr Teilchen erreichen dann hohe Geschwindigkeiten. Dies zeigt, dass die Temperatur ein Maß für die mittlere kinetische Energie der Teilchen ist. Doch sind auch bei niedrigen Temperaturen sehr hohe Geschwindigkeiten nicht vollständig ausgeschlossen. Mit der Geschwindigkeitsverteilung lassen sich viele physikalische Transportphänomene erklären, wie z. B. die Diffusion in Gasen.

Strömungsgeschwindigkeit eines Fluids Die mittlere Strömungsgeschwindigkeit eines Gases oder einer Flüssigkeit v_A ergibt sich aus der Volumenstromstärke $Q = \frac{\mathrm{d}V}{\mathrm{d}t}$ durch den Strömungsquerschnitt A:

$$v_\mathrm{A} = \frac{Q}{A}$$

Allerdings können sich die lokalen Strömungsgeschwindigkeiten sehr stark voneinander unterscheiden. Beispielsweise ist die Geschwindigkeit in der Mitte eines idealen Rohres am größten und fällt durch die Reibung zur Wandung hin bis auf Null ab. Man muss daher die Strömung eines Mediums als Vektorfeld auffassen. Wenn die Geschwindigkeitsvektoren zeitlich konstant sind, spricht man von einer stationären Strömung. Verhalten sich die Geschwindigkeiten im Gegensatz dazu chaotisch, so handelt es sich um eine turbulente Strömung. Bei der Charakterisierung des Strömungsverhaltens hilft die Reynoldszahl, die die Strömungsgeschwindigkeit in Relation zu der Abmessungen des angeströmten Körpers und zur Viskosität des Fluids setzt.

Mathematisch wird das Verhalten der Geschwindigkeiten durch die Navier-Stokes-Gleichungen modelliert, die als Differenzialgleichungen die Geschwindigkeitsvektoren mit inneren und äußeren Kräften in Beziehung setzen. Damit haben sie für die Bewegung eines Fluids eine ähnliche Bedeutung wie die Grundgleichung der Mechanik für Massenpunkte und starre Körper.

Geschwindigkeit von Wellen Die komplexe Bewegung von Wellen macht es nötig, verschiedene Geschwindigkeitsbegriffe zu verwenden. (Insbesondere kann mit dem Wort Ausbreitungsgeschwindigkeit verschiedenes gemeint sein.)

Die Auslenkungsgeschwindigkeit mechanischer Wellen wird als Schnelle bezeichnet. Das bekannteste Beispiel ist die Schwingungsgeschwindigkeit der Luftteilchen in einer Schallwelle.

Die Geschwindigkeit, mit der sich ein Punkt bestimmter Phase vorwärts bewegt, heißt Phasengeschwindigkeit. Es gilt: $v_\mathrm{p} = \frac{\lambda}{T} = \frac{\omega}{k}$. Hierbei sind λ die Wellenlänge, T die Periodendauer, ω die

Kreisfrequenz und k die Kreiswellenzahl. Die Geschwindigkeit, mit der sich die Wellenkämme im Meer fortbewegen, ist ein typisches Beispiel für eine Phasengeschwindigkeit.

Die Geschwindigkeit, mit der sich ein ganzes Wellenpaket bewegt, wird Gruppengeschwindigkeit genannt: $v_\mathrm{g} = \frac{\partial\omega}{\partial\mathbf{k}}$.

Phasen- und Gruppengeschwindigkeit stimmen nur in seltenen Fällen überein (z. B. Ausbreitung von Licht im Vakuum). In der Regel unterscheiden sie sich. Ein anschauliches extremes Beispiel ist die Fortbewegung von Schlangen: Fasst man die Schlange als eine Welle auf, so ist die Geschwindigkeit ihres Vorankommens eine Gruppengeschwindigkeit. Die Phasengeschwindigkeit ist beim Schlängeln jedoch Null, denn die Stellen, an denen sich der Körper der Schlange nach rechts oder links krümmt, sind durch den Untergrund vorgegeben und bewegen sich nicht über den Boden.

In aller Regel ist die Phasengeschwindigkeit einer physikalischen Welle von der Frequenz bzw. der Kreiswellenzahl abhängig. Diesen Effekt bezeichnet man als Dispersion. Er ist unter anderem dafür verantwortlich, dass Licht verschiedener Wellenlänge von einem Prisma unterschiedlich stark gebrochen wird.

Relativitätstheorie Aus den Gesetzen der klassischen Physik folgt für Geschwindigkeiten unter anderem:

- Die Messwerte für Längen und Zeiten sind unabhängig vom Bewegungszustand (und damit der Geschwindigkeit) des Beobachters. Insbesondere stimmen alle Beobachter darin überein, ob zwei Ereignisse gleichzeitig stattfinden oder nicht.
- Bei einem Wechsel des Bezugssystems gilt die Galilei-Transformation. Dies bedeutet, dass Geschwindigkeiten von Bewegungen, die sich überlagern, vektoriell addiert werden dürfen.
- Es gibt keine theoretische Obergrenze für die Geschwindigkeit von Bewegungen.
- Zwar wird es von den Gesetzen der klassischen Physik nicht verlangt, aber es wurde vor Einstein allgemein angenommen, dass es für alle Geschwindigkeiten ein universelles Bezugssystem, den „Äther“, gebe. Wenn dem so wäre, müsste die Ausbreitungsgeschwindigkeit von elektromagnetischen Wellen vom Bewegungszustand des Empfängers abhängen.

Letztere Abhängigkeit ließ sich mit dem Michelson-Morley-Experiment nicht nachweisen. Einstein postulierte, dass das Relativitätsprinzip, das bereits aus der klassischen Mechanik bekannt war, auch auf alle anderen Phänomene der Physik, insbesondere die Ausbreitung des Lichts, angewendet werden müsse und dass die Lichtgeschwindigkeit unabhängig vom Bewegungszustand des Senders sei. Daraus folgerte er, dass die oben genannten Aussagen der klassischen Mechanik modifiziert werden müssen.[3] Im Detail heißt dies:

- Die Messwerte für Längen und Zeiten sind abhängig vom Bewegungszustand (und damit der Geschwindigkeit) des Beobachters (siehe Zeitdilatation und Längenkontraktion). Auch die Gleichzeitigkeit ist relativ.
- Bei einem Wechsel des Bezugssystems gilt die Lorentz-Transformation. Dies bedeutet, dass Geschwindigkeiten von Bewegungen, die sich überlagern, nicht einfach vektoriell addiert werden dürfen.
- Bewegungen von Körpern können nur mit Geschwindigkeiten erfolgen, die geringer als die Lichtgeschwindigkeit sind. Auch Informationen können nicht schneller als das Licht übertragen werden.

- Es gibt keinen „Äther“.

Die Effekte, die sich aus der speziellen Relativitätstheorie ergeben, machen sich jedoch erst bei sehr hohen Geschwindigkeiten bemerkbar. Der Lorentz-Faktor, der für Zeitdilatation und Längenkontraktion maßgeblich ist, ergibt erst für Geschwindigkeiten von $v > 4{,}2 \cdot 10^7 \frac{\mathrm{m}}{\mathrm{s}}$ eine Abweichung von mehr als einem Prozent. Folglich stellt die klassische Mechanik selbst für die schnellsten bisher gebauten Raumfahrzeuge eine äußerst präzise Näherung dar.

1.3.3 Beschleunigung

Die Beschleunigung ist eine physikalische Größe, die die Änderung des Bewegungszustands eines Körpers angibt. Je nach Richtung der Beschleunigung wird ein beschleunigter Körper schneller oder langsamer oder es ändert sich seine Bewegungsrichtung. Die Beschleunigung ist eine zentrale Größe in der Kinematik.

Die Beschleunigung ist die momentane zeitliche Änderungsrate der Geschwindigkeit, also $\vec{a} = \frac{\mathrm{d}\vec{v}}{\mathrm{d}t}$. Sie ist damit eine vektorielle Größe.

Für Insassen von Fahrzeugen sind Beschleunigungen durch die damit verbundenen Trägheitskräfte erfahrbar.

Einführung Nach dem ersten Newtonschen Gesetz bewegen sich alle Körper in Inertialsystemen mit konstanter Geschwindigkeit auf geradlinigen Bahnen, wenn keine Kräfte auf sie wirken. Man sagt: Ihr Bewegungszustand ist konstant. Falls doch eine Kraft auf einen Körper einwirkt, ändert sich sein Bewegungszustand.

In der Umgangssprache bezeichnet Beschleunigung oft nur eine Steigerung des „Tempos“, also des Betrags der Geschwindigkeit. Im physikalischen Sinn ist aber jede Änderung einer Bewegung eine Beschleunigung, z. B. auch eine Abnahme des Geschwindigkeitsbetrages – wie ein Bremsvorgang – oder eine reine Richtungsänderung bei gleichbleibendem Geschwindigkeitsbetrag – wie bei einer Kurvenfahrt mit einem Auto.

Zunächst betrachten wir nur Bewegungen entlang einer Geraden, also eindimensionale Bewegungen. Zu zwei Zeitpunkten hat der Körper die Geschwindigkeiten v_1 und v_2. Seine Geschwindigkeit hat sich also in der Zeitspanne dazwischen $\Delta t = t_2 - t_1$ geändert. Die Geschwindigkeitsänderung beträgt $\Delta v = v_2 - v_1$. Man definiert nun die mittlere Beschleunigung als die mittlere Änderungsrate der Geschwindigkeit. Die Beschleunigung a gibt also an, wie schnell diese Geschwindigkeitsänderung erfolgt. Es gilt somit:

$$\overline{a} = \frac{\Delta v}{\Delta t}$$

Wenn die Beschleunigung dasselbe Vorzeichen hat wie die Geschwindigkeit, dann nimmt der Betrag der Geschwindigkeit zu. Wenn sich beide Vorzeichen unterscheiden, nimmt der Betrag der Geschwindigkeit ab (die Richtung der Geschwindigkeit kann sich auch umkehren). Ähnlich wie bei der Durchschnittsgeschwindigkeit lässt sich mit obiger Gleichung nur die durchschnittliche Beschleunigung berechnen. Nur wenn die Geschwindigkeit sich linear mit der Zeit ändert, also im Falle einer konstanten Beschleunigung, entspricht dies auch zu jedem Zeitpunkt der momentanen Beschleunigung. Um auch in anderen Fällen zur momentanen Beschleunigung zu gelangen, muss man den Grenzwert für sehr kleine Zeitintervalle bilden und gelangt so zur zeitlichen Ableitung der Geschwindigkeit:

$$a(t) = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{\mathrm{d}v}{\mathrm{d}t} = \dot{v}(t)$$

Gleichmäßig beschleunigte Bewegung Von einer gleichmäßig beschleunigten Bewegung spricht man, wenn die Beschleunigung konstant ist. Dann gilt für die Geschwindigkeit

$$v(t) = at + v_0$$

und für die zurückgelegte Strecke

$$s(t) = \frac{1}{2}at^2 + v_0 t + s_0$$

mit dem Startpunkt s_0 und der Anfangsgeschwindigkeit v_0.

Allgemeine Definition Im Allgemeinen erfolgt die Bewegung nicht zwangsläufig geradlinig, sondern im zwei- oder dreidimensionalen Fall. Bei einer konstanten Beschleunigung, muss die Differenz der Geschwindigkeiten $\Delta\vec{v} = \vec{v}(t_2) - \vec{v}(t_1)$ vektoriell bestimmt werden, wie in der Abbildung veranschaulicht. Wenn sich die Beschleunigung während der betrachteten Zeitspanne ändert, erhält man mit obiger Rechnung die mittlere Beschleunigung, auch Durchschnittsbeschleunigung genannt.

$$a = \frac{\Delta\vec{v}}{\Delta t}.$$

Um die Beschleunigung für einen bestimmten Zeitpunkt statt für ein Zeitintervall zu berechnen, muss man – wie oben beschrieben – vom Differenzenquotienten zum Differentialquotienten übergehen. Die Beschleunigung ist dann die erste Ableitung der Geschwindigkeit nach der Zeit:

$$\vec{a}(t) = \frac{\mathrm{d}\vec{v}(t)}{\mathrm{d}t} = \dot{\vec{v}}(t).$$

Da die Geschwindigkeit die Ableitung des Ortes nach der Zeit ist, kann man die Beschleunigung auch als zweite Ableitung des Ortsvektors $\vec{r}$ nach der Zeit darstellen:

$$\vec{a}(t) = \frac{\mathrm{d}^2\vec{r}(t)}{\mathrm{d}t^2} = \ddot{\vec{r}}(t).$$

Wenn die Vektoren der Geschwindigkeit und der Beschleunigung in die gleiche Richtung zeigen, bedeutet die Beschleunigung nur eine Zunahme des Geschwindigkeitsbetrags. Entsprechend nimmt der Geschwindigkeitsbetrag ab, wenn die beiden Vektoren antiparallel sind. In beiden Fällen ändert sich aber die Richtung des Geschwindigkeitsvektors nicht. Es handelt sich also um eine geradlinig beschleunigte Bewegung.

Sofern jedoch die Beschleunigung in einem gewissen Winkel zur Bewegungsrichtung steht, ändert sich auch die Richtung der Geschwindigkeit. Die Bewegung beschreibt also eine gekrümmte Bahn. Wenn Beschleunigung und Geschwindigkeit orthogonal zueinander stehen, besitzt die Beschleunigung überhaupt keine Komponente in Richtung der Geschwindigkeit mehr. In diesem Fall ändert sich nur deren Richtung, aber nicht ihr Betrag. Die Bahnkurve ist dann – zumindest momentan – eine Kreisbahn.

Gekrümmte Wege

Spezialfall: Kreisbewegung Bei der gleichförmigen Kreisbewegung ist der Beschleunigungsvektor in jedem Moment orthogonal zur Bewegungsrichtung. Man spricht von der Zentripetalbeschleunigung a_Z. Sie ergibt sich aus dem Abstand r des Massepunktes zur Drehachse und seiner Tangentialgeschwindigkeit v oder der Winkelgeschwindigkeit ω:

$$a_Z = \frac{v^2}{r} = \omega^2 r.$$

Ein typisches Anwendungsbeispiel ist hierbei die Flugbahn von Satelliten in einem niedrigen, kreisförmigen Orbit, wo die Fallbeschleunigung, die stets zum Erdmittelpunkt gerichtet ist, als Zentripetalbeschleunigung fungiert.

Bezüglich eines mitrotierenden (und daher beschleunigten) Bezugssystems wird ein Objekt vom Mittelpunkt weg nach außen beschleunigt, dann wird die Bezeichnung Zentrifugalbeschleunigung verwendet. Eine Zentrifuge nutzt diesen Effekt, um Dinge einer konstanten Beschleunigung auszusetzen. Der Krümmungsradius entspricht dabei, da es sich um eine Kreisbewegung handelt, dem Abstand r des Zentrifugiergutes zur Drehachse. Der Betrag der Zentrifugalbeschleunigung berechnet sich nach derselben Formel wie die Zentripetalbeschleunigung.

Allgemeiner Fall Die Beschleunigung eines Körpers, der sich entlang eines Weges (einer Raumkurve) bewegt, lässt sich mit den Frenetschen Formeln berechnen. Dies ermöglicht eine additive Zerlegung der Beschleunigung in eine Beschleunigung in Bewegungsrichtung (Tangentialbeschleunigung) und eine Beschleunigung senkrecht zur Bewegungsrichtung (Normalbeschleunigung oder Radialbeschleunigung).

Der Vektor der Geschwindigkeit $\vec{v}$ kann als Produkt aus seinem Betrag v und dem Tangenteneinheitsvektor $\hat{t}$ dargestellt werden:

$$\vec{v} = v\,\hat{t}$$

Der Tangenteneinheitsvektor ist ein Vektor der Länge 1, der an jedem Punkt des Weges die Richtung der Bewegung anzeigt. Die Ableitung dieses Ausdrucks mithilfe der Produktregel liefert die Beschleunigung:

$$\vec{a} = \frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = \left(\frac{\mathrm{d}v}{\mathrm{d}t}\right)\hat{t} + v\left(\frac{\mathrm{d}\hat{t}}{\mathrm{d}t}\right)$$

Die zeitliche Ableitung des Tangenteneinheitsvektors kann über die Bogenlänge s berechnet werden:

$$\frac{\mathrm{d}\hat{t}}{\mathrm{d}t} = \underbrace{\frac{\mathrm{d}\hat{t}}{\mathrm{d}s}}_{\hat{n}/\rho}\underbrace{\frac{\mathrm{d}s}{\mathrm{d}t}}_{v} = \frac{v}{\rho}\hat{n}$$

Dabei führt man den Krümmungsradius ρ und den Normaleneinheitsvektor $\hat{n}$ ein. Der Krümmungsradius ist ein Maß für die Stärke der Krümmung und der Normaleneinheitsvektor zeigt senkrecht zur Bahnkurve in Richtung des Krümmungsmittelpunkts. Man definiert die Tangentialbeschleunigung a_t und Radialbeschleunigung a_n so:

$$a_t = \dot{v}$$
$$a_n = \frac{v^2}{\rho}$$

Die Beschleunigung lässt sich damit in zwei Komponenten zerlegen:

$$\vec{a} = a_t\hat{t} + a_n\hat{n}$$

Ist die Tangentialbeschleunigung null, so ändert der Körper nur seine Bewegungsrichtung. Der Betrag der Geschwindigkeit bleibt dabei erhalten. Um den Betrag der Geschwindigkeit zu ändern, muss also eine Kraft wirken, die eine Komponente in Richtung des Tangentialvektors besitzt.

Ruck Die zeitliche Ableitung der Beschleunigung (also die dritte Ableitung des Ortsvektors nach der Zeit) wird Ruck $\vec{\jmath}$ genannt:

$$\vec{\jmath}(t) = \dot{\vec{a}}(t) = \frac{\mathrm{d}^3\vec{r}(t)}{\mathrm{d}t^3}$$

Zusammenhang zwischen Beschleunigung und Kraft Der Zusammenhang zwischen Beschleunigungen und Kräften wird durch die Newtonschen Gesetze beschrieben:

- In einem Inertialsystem erfahren kräftefreie Körper keine Beschleunigung.
- Falls Kräfte angreifen, ist die Beschleunigung proportional zum Betrag der resultierenden Kraft und erfolgt in deren Richtung: $\vec{F} = m\vec{a}$.

Wenn die resultierende Kraft proportional zur Masse eines Körpers ist – wie das beispielsweise für die Gewichtskraft der Fall ist – ist die Beschleunigung von der Masse des Körpers unabhängig. Das ist der Grund, warum die Fallbeschleunigung beim freien Fall unabhängig von der Masse ist: Alle Körper fallen unabhängig von ihrer Masse gleich schnell, auf der Erde mit rund 9,81 $\mathrm{m/s^2}$.

In der speziellen Relativitätstheorie gilt die Newton'sche Beziehung nicht exakt; die Beschleunigung ist nicht genau parallel zur Kraft (siehe Beschleunigung (spezielle Relativitätstheorie)).

Trägheitskräfte Soll die Bewegung in einem beschleunigten Bezugssystem beschrieben werden, so sind zusätzlich Trägheitskräfte zu berücksichtigen. Damit ist folgendes gemeint:

Ein Körper, der in einem Inertialsystem ruht, erfährt in einem Bezugssystem, das gegenüber dem Inertialsystem mit $\vec{a}$ beschleunigt, eine Beschleunigung von $\vec{a}^* = -\vec{a}$. Ein mitbewegter Beobachter macht dafür eine Kraft $\vec{F}^* = m\vec{a}^*$ verantwortlich, für die es in seinem Bezugssystem keine erkennbare Ursache gibt. Dies ist die Trägheitskraft. Beispiel: Ein Ball, der auf dem Boden einer U-Bahn liegt, rollt plötzlich nach hinten, wenn die Bahn anfährt. Ein naiver Fahrgast könnte vermuten, dass der Ball von einer mysteriösen Kraft beschleunigt wird. Ein am Bahnsteig stehender Beobachter würde hingegen sagen, dass die U-Bahn beschleunigt und der Ball aufgrund seiner Trägheit zunächst zurückbleibt.

Beschleunigung in der speziellen Relativitätstheorie Ebenso wie in der klassischen Mechanik können Beschleunigungen auch in der speziellen Relativitätstheorie (SRT) als Ableitung der Geschwindigkeit nach der Zeit dargestellt werden. Da der Zeitbegriff aufgrund der Lorentz-Transformation und Zeitdilatation in der SRT jedoch komplexer ausfällt, führt dies auch zu komplexeren Formulierungen der Beschleunigung und ihres Zusammenhangs mit der Kraft. Insbesondere ergibt sich, dass kein massebehafteter Körper auf Lichtgeschwindigkeit beschleunigt werden kann.

Äquivalenzprinzip und allgemeine Relativitätstheorie Das Äquivalenzprinzip besagt, dass in einem frei fallenden Bezugssystem lokal keine Gravitationsfelder existieren. Es geht auf die Überlegungen von Galileo Galilei und Isaac Newton zurück, die erkannt haben, dass alle Körper unabhängig von ihrer Masse von der Gravitation gleich beschleunigt werden. Ein Beobachter in

einem (kleinen) Labor kann nicht feststellen, ob sich sein Labor in der Schwerelosigkeit oder im freien Fall befindet. Er kann innerhalb seines Labors auch nicht feststellen, ob sein Labor gleichförmig beschleunigt bewegt wird oder ob es sich in einem äußeren homogenen Gravitationsfeld befindet.

Mit der allgemeinen Relativitätstheorie lässt sich ein Gravitationsfeld durch die Metrik der Raumzeit, also die Maßvorschrift in einem vierdimensionalen Raum aus Orts- und Zeitkoordinaten ausdrücken. Ein Inertialsystem hat eine flache Metrik. Nichtbeschleunigte Beobachter bewegen sich immer auf dem kürzesten Weg (einer Geodäte) durch die Raumzeit. In einem flachen Raum, also einem Inertialsystem, ist dies eine gerade Weltlinie. Gravitation bewirkt eine Raumkrümmung. Das bedeutet, dass die Metrik des Raumes nicht mehr flach ist. Dies führt dazu, dass die Bewegung, die in der vierdimensionalen Raumzeit einer Geodäte folgt, im dreidimensionalen Anschauungsraum vom außenstehenden Beobachter meist als beschleunigte Bewegung längs einer gekrümmten Kurve wahrgenommen wird.

...

1.3.4 Ruck

Ruck ist ein Begriff aus der Kinematik. Er ist die momentane zeitliche Änderungsrate der Beschleunigung eines Körpers.[1] Die SI-Einheit des Rucks ist $\mathrm{m/s^3}$. Als Formelzeichen wird üblicherweise j gewählt in Anlehnung an die englischen Bezeichnungen jerk oder jolt. In deutscher Literatur ist auch r oder h im Gebrauch.

Definition Formal ist der Ruck die Ableitung der Beschleunigung nach der Zeit, also die zweite zeitliche Ableitung der Geschwindigkeit und die dritte zeitliche Ableitung des Wegs:

$$j(t) = \dot{a}(t) = \ddot{v}(t) = \dddot{x}(t)$$

wobei t die Zeit, a die Beschleunigung, v die Geschwindigkeit und x der Ort sind.

Wird von einem körperfesten Koordinatensystem ausgegangen, so kann der Ruck für jede Koordinatenrichtung getrennt bestimmt werden, z. B. als Längsruck oder Querruck, oder allgemein vektoriell als Ableitung der Beschleunigung bezüglich dieses Bezugssystems. Insbesondere stellt diese Definition sicher, dass eine gleichförmige Kreisbewegung ruckfrei ist, was dem allgemeinen Sprachgebrauch sowie der Anwendung in der Technik entspricht. Bei Stoßvorgängen ist der Ruck nicht definiert.

Obwohl die physikalische Größe ‚Ruck' bei jeder Beschleunigungsänderung definiert ist, wird der Begriff umgangssprachlich in der Regel nur bei kurzen „ruckartigen" Beschleunigungsänderungen verwendet (siehe Weblinks). Diese treten z. B. beim Anfahren mit einem nicht vorgespannten Abschleppseil auf. „Ruckartig" bedeutet hier, dass der Gradient des kinematischen Rucks einen hohen Betrag hat.

Praxis / Anwendungen ...

Landfahrzeuge Bei Fahrzeugen ist der Grund für Rucke häufig ein Lastwechsel (z. B. beim Teillastruckeln). Unterschieden werden:

- der Längsruck, die zeitliche Änderung der Längsbeschleunigung
- der Querruck, die zeitliche Änderung der Querbeschleunigung.

Anschaulich bedeutet dies, dass der Längsruck bei einem Fahrzeug durch plötzliches Anfahren oder Bremsen verursacht wird, der Querruck dagegen durch plötzliche Änderung des Lenkradwinkels bei einem fahrenden Automobil.

Bei elektronischen Lenksystemen können durch die Zusatzfunktionen auch Querrucke ohne Betätigung des Lenkrads auftreten. Diese müssen aus Sicherheitsgründen auf 5 m/s3 begrenzt sein (ECE R79).

Die Bezeichnungen längs und quer deuten schon an, dass diese Beschleunigungen Komponenten in einem fahrzeugfesten Bezugssystem sind. Ändern sich die Komponenten nicht, so ist der Ruck Null. Bei stationärer Kreisfahrt zeigt der Beschleunigungsvektor immer zum Kreismittelpunkt (Zentripetalkraft), von außen betrachtet ändert er sich also; im fahrzeugfesten Koordinatensystem dagegen bleibt derselbe Beschleunigungsvektor konstant.

Längsruck Je schneller eine Bremsung eingeleitet oder beendet wird, desto höher ist der Ruck. Eine abrupt eingeleitete Bremsung (Notbremsung) ist mit einem hohen Ruck verbunden. Wenn sich der Insasse nicht schnell genug darauf eingestellt hat und sich nicht abstützt, wird er bei Vorwärtsfahrt nach vorne geworfen (im Auto vom Gurt abgefangen), bei Rückwärtsfahrt in den Sitz gedrückt. Da die Betätigung der Bremse selbst bei einer Notbremsung noch eine gewisse Zeit beansprucht, bleibt der Ruck ein endlicher Wert.

Bleibt die Bremse bis zum Stillstand mit ihrer maximalen Kraft wirksam, so tritt am Ende des Bremsweges ein theoretisch unendlich hoher Ruck (Schlussruck) auf, weil die Verzögerung (= negative Beschleunigung) plötzlich, also in der Zeitdauer null, endet. Dadurch wird der Insasse durch seine eigene Muskelkraft (Abstützkraft) oder, wenn er sich völlig passiv verhalten hat, durch die vom Gurt ausgeübte Kraft in den Sessel geschleudert und von der Federkraft des Sessels dann zurückgeschleudert. Für diese Bewegungen vergeht allerdings Zeit. Dadurch wird der Schlussruck endlich, also gemildert. Außerdem entspannen sich elastische Elemente am Fahrzeug (Reifen, Fahrzeug-Federung, Eisenbahn-Puffer u. a.), was ebenfalls eine kurze Zeit dauert. Das Fahrzeug fährt dabei scheinbar ein kleines Stück zurück.

Im Normalbetrieb löst der routinierte Fahrer die Bremse langsam vor Erreichen des Stillstandes und dehnt damit die Abnahme der Verzögerung zeitlich aus, so dass der Schlussruck auf ein Minimum herabgesetzt wird.

Fahrzeuge mit Elektroantrieb entwickeln bei einfachen (stufigen) Steuerungskonzepten des Motorstromes einen starken Längsruck bei jeder Beschleunigungsänderung. Der Fahrkomfort beim Anfahren, Beschleunigen und rekuperativen Bremsen wird durch sanft reagierende Fahrdynamik verbessert, jedoch kann ein sogenannter Warnruck bei autonomen Fahrzeugen genutzt werden, die Aufmerksamkeit zur Überwachung herzustellen.[6]

Querruck Der Querruck k als Spezialfall des Rucks ist die Änderung der Zentripetalbeschleunigung a_r in Abhängigkeit von der Zeit t:

$$k = \frac{\mathrm{d}a_r(t)}{\mathrm{d}t}$$

Die Zentripetalbeschleunigung eines Fahrzeugs ist abhängig von seiner Geschwindigkeit v sowie der Krümmung $\kappa = \frac{1}{r}$ der Bahn, wobei r der Radius des Krümmungskreises ist:

$$a_r = \frac{v^2}{r} = v^2\,\kappa$$

$$\Rightarrow k = v^2\,\dot{\kappa} \quad \text{für} \quad v = \text{konst.}$$

Die Krümmung ist bei den verwendeten Trassierungselementen als Funktion der Wegstrecke s gegeben: $\kappa = \kappa(s)$

Mit $\dot{\kappa} = \frac{\mathrm{d}\kappa}{\mathrm{d}t} = \frac{\mathrm{d}s}{\mathrm{d}t} \cdot \frac{\mathrm{d}\kappa(s)}{\mathrm{d}s} = v \cdot \frac{\mathrm{d}\kappa(s)}{\mathrm{d}s}$ ergibt sich für den Querruck somit:

$$k = v^3\,\frac{\mathrm{d}\kappa(s)}{\mathrm{d}s}.$$

Ein Querruck tritt also beispielsweise auf, wenn sich der Radius einer Kreisbewegung ändert. Wenn in einer Trasse, z. B. einem Bahngleis, ein Kreisbogen unmittelbar auf eine Gerade folgt, so ändert sich an dieser Stelle die Zentripetalbeschleunigung bei schienengebundenen Fahrzeugen sprungartig. Das heißt, die Zeit für diese Änderung ist fast null, und der Querruck wird extrem groß. Verwendet man als Verbindungselement zwischen Gerade und Kreisbogen eine Klothoide, so ändert sich die Zentripetalbeschleunigung linear während der Zeit, die zum Durchfahren der Klothoide benötigt wird. Daher wird der Querruck entsprechend geringer.

In Abschnitten, in denen das Fahrzeug sich auf einer Geraden oder mit konstanter Geschwindigkeit auf einer Kreisbahn bewegt, ändert sich die Zentripetalbeschleunigung nicht. Der Querruck ist somit null.

Trassenbau Bei der Planung von Trassen ist je nach der Bemessungsgeschwindigkeit und dem Fahrkomfort, den man für eine Strecke erreichen will, darauf zu achten, dass der Querruck einen Grenzwert von 0,4 bis 0,6 $\mathrm{m/s^3}$ nicht übersteigt. Bei Schienenfahrzeugen wird durch die Wahl der Trassierungselemente eine möglichst ruckarme Fahrt beim Übergang in Kurven sichergestellt. Auch bei Achterbahnen wird durch entsprechende Übergänge die Belastung auf den menschlichen Körper reduziert. Im Extremfall, etwa bei Hochgeschwindigkeitszügen, kann durch Verwendung anderer Übergangsbögen als der Klothoide erreicht werden, dass der Querruck am Anfang des Übergangsbogens nicht sprunghaft, sondern allmählich einsetzt.

Der Querruck bei Lenkmanövern von Straßenfahrzeugen ist wegen der erforderlichen Lenkraddrehung generell begrenzt. Der sanfte Verlauf des Querruckes beim autonomen Fahren ist Forschungsgegenstand, um die Vorhersehbarkeit und den Komfort eines Lenkmanövers zu verbessern, der Ruck würde aufgrund rein mathematischer Algorithmen ansonsten plötzlich und überraschend einsetzen.

Ruckänderung Die Ruckänderung s (engl. jounce, snap), manchmal Knall genannt, ist ein Begriff aus der analytischen Modellierung der Fahrdynamik von Schienenfahrzeugen und die erste Ableitung des Rucks nach der Zeit:

$$s(t) = \frac{\mathrm{d}j}{\mathrm{d}t}$$

wobei t die Zeit und j der Ruck ist. Die SI-Einheit der Ruckänderung ist dementsprechend $\frac{\mathrm{m}}{\mathrm{s}^4}$.

Die Ruckänderung spielt in diesen Modellen vor allem eine theoretische Rolle, in dem zumindest bei einem stückweise stetigen Differenzieren oder Integrieren die Ruckänderung jeweils als gleich Null vorausgesetzt wird und auf diese Weise eine Lösung der zugehörigen Gleichungssysteme möglich wird.

...

1.4 Optimierungstheorie

1.4.1 Grundlagen der Optimierungstheorie

Optimierungsaufgaben treten in den Wirtschaftswissenschaften (Operations Research), in der Technik und in den Naturwissenschaften in vielfältiger Art und Weise auf.

Glossary

$\mathbb{R}$, $\mathbb{N}$, $\mathbb{C}$	wie üblich
k, k, ℓ	Iterations-Indizes
f, g, h	Zielfkt., Fkt.en der Ungleichungsrestr., Fkt.en der Gleichheitsrestr.
m, p	Anzahl Un- resp. Gleichheitsrestr.
$\mathcal{I}$, $\mathcal{J}$; $\mathcal{A}$	Indexmengen; active set
$\mathcal{L}$; μ, ν	Lagrange-Funktion; -multiplikatoren

Table 1.1: default

...

Aufgabenstellung

Für gegebene (und hinreichend oft differenzierbare) Funktionen

$$\begin{aligned} f &: \mathbb{R}^n \to \mathbb{R} \\ g = (g_1, \ldots, g_m)^T &: \mathbb{R}^n \to \mathbb{R}^m \\ h = (h_1, \ldots, h_p)^T &: \mathbb{R}^n \to \mathbb{R}^p \end{aligned}$$

betrachten wir das restringierte Standard-Optimierungsproblem:

Problem 1.4.1.1. (Standard-Optimierungsproblem) *Finde $x \in \mathbb{R}^n$, so dass $f(x)$ minimal wird unter den Nebenbedingungen*

$$\begin{aligned} g_i(x) &\leq 0, \quad i = 1, \ldots, m \\ h_j(x) &= 0, \quad j = 1, \ldots, p \end{aligned}$$

In Kurzform schreiben wir auch:

$$\begin{aligned} f(x) &\to \min \\ g(x) &\leq 0 \\ h(x) &= 0 \end{aligned}$$

□

Mit

$$\Sigma := \{x \in \mathbb{R}^n \ : \ g_i(x) \leq 0, \ i = 1, \ldots, m, \ h_j(x) = 0, \ j = 1, \ldots, p\} \tag{1.4.1}$$

bezeichnen wir die *zulässige Menge* (oder den *zulässigen Bereich*) des Standard-Optimierungsproblems. Die Menge Σ in (1.4.1) ist eine abgeschlossene Menge, falls die Funktionen g_i und h_j stetig sind.

Ein Punkt $x \in \mathbb{R}^n$ heißt

- *zulässig*, falls $x \in \Sigma$ gilt, und
- *unzulässig*, falls $x \notin \Sigma$ gilt.

Für zulässige Punkte $x \in \Sigma$ bezeichnen wir mit

$$\mathcal{A}(x) := \{i \in \{1, \ldots, m\} \ : \ g_i(x) = 0\}$$

die *Indexmenge der (in x) aktiven Ungleichungsrestriktionen* und nennen die Restriktion $g_i(x) \leq 0$

- *aktiv in x*, wenn $g_i(x) = 0$ gilt, und
- *inaktiv in x*, wenn $g_i(x) < 0$ gilt.

Das Standard-Optimierungsproblem enthält die folgenden Problemklassen als Spezialfälle:

- *Unrestringiertes Optimierungsproblem:*

  $$\min_{x \in \mathbb{R}^n} f(x)$$

  Hierin treten keine Ungleichungs- und Gleichungsrestriktionen auf.

- *Konvexes Optimierungsproblem:*

  $$\min_{\substack{g(x) \leq 0 \\ Ax = b}} f(x)$$

  Spezialfall, wobei f und g konvexe Funktionen, $A \in \mathbb{R}^{m \times n}$ eine Matrix und $b \in \mathbb{R}^m$ ein Vektor sind.

- *Lineares Optimierungsproblem (in primaler Normalform):*

  $$\min_{\substack{Ax = b \\ x \geq 0}} c^T x$$

  Spezialfall mit $g(x) = -x$, $h(x) = Ax - b$, wobei $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ Vektoren und $A \in \mathbb{R}^{m \times n}$ eine Matrix sind. Das lineare Optimierungsproblem ist ein konvexes Optimierungsproblem.

- *Linear-quadratisches Optimierungsproblem:*

  $$\min_{\substack{Ax = b \\ x \geq 0}} \frac{1}{2} x^T Q x + c^T x$$

  Spezialfall mit $g(x) = -x$, $h(x) = Ax - b$, wobei $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ Vektoren und $Q \in \mathbb{R}^{n \times n}$ und $A \in \mathbb{R}^{m \times n}$ Matrizen sind. Falls Q symmetrisch und positiv semi-definit ist, liegt ein konvexes Optimierungsproblem vor.

Notwendige Bedingungen für restringierte Optimierungsprobleme

Die folgenden KARUSH-KUHN-TUCKER (KKT) Bedingungen sind die Basis für viele theoretische Untersuchungen und numerische Algorithmen.

Die *Lagrange-Funktion* für das Standard-Optimierungsproblem ist definiert als

$$\begin{aligned}\mathcal{L}(x,\mu,\nu) &:= f(x)+\mu^T g(x)+\nu^T h(x)\\ &= f(x)+\sum_{i=1}^{m}\mu_i g_i(x)+\sum_{j=1}^{p}\nu_j h_j(x)\end{aligned}$$

wobei $\mu=(\mu_1,\ldots,\mu_m)^T\in\mathbb{R}^m$ und $\nu=(\nu_1,\ldots,\nu_p)^T\in\mathbb{R}^p$ als *Lagrange-Multiplikatoren* bezeichnet werden. Es gilt:

Theorem 1.4.1.2. (Notwendige Bedingungen erster Ordnung, KKT-Bedingungen)
Voraussetzungen:

- *$\hat{x}$ ist ein lokales Minimum des Standard-Optimierungsproblems.*
- *Die Funktionen f, g_i, $i=1,\ldots,m$, und h_j, $j=1,\ldots,p$ sind stetig differenzierbar.*
- *Es gilt die Linear Independence Constraint Qualification (LICQ) in $\hat{x}$, d.h. die Vektoren*

  $$\nabla g_i(\hat{x}), i\in\mathcal{A}(\hat{x}) \text{ und } \nabla h_j(\hat{x}), j=1,\ldots,p$$

  sind linear unabhängig.

Dann existieren eindeutig bestimmte Multiplikatoren $\mu=(\mu_1,\ldots,\mu_m)^T\in\mathbb{R}^m$ und $\nu=(\nu_1,\ldots,\nu_p)^T\in\mathbb{R}^p$, so dass die folgenden Bedingungen gelten:

(a) ***Stationarität der Lagrange-Funktion:***

$$\nabla_x\mathcal{L}(\hat{x},\mu,\nu)=0 \tag{1.4.2}$$

bzw.

$$\nabla f(\hat{x})+\sum_{i=1}^{m}\mu_i\cdot\nabla g_i(\hat{x})+\sum_{j=1}^{p}\nu_j\cdot\nabla h_j(\hat{x})=0. \tag{1.4.3}$$

(c) ***Komplementaritätsbedingungen:*** *Für $i=1,\ldots,m$ gilt:*

$$\mu_i\cdot\geq 0 \quad \textit{und} \quad \mu_i g_i(\hat{x})=0 \tag{1.4.4}$$

(d) ***Zulässigkeit:***

$$g(\hat{x})\leq 0 \quad \textit{und} \quad h(\hat{x})=0 \tag{1.4.5}$$

□

Jeden Punkt $(\hat{x},\mu,\nu)$, der die Bedingungen (1.4.2)–(1.4.5) erfüllt, nennen wir *KKT-Punkt* oder *stationären Punkt*. Beachte, dass KKT-Punkte lediglich Kandidaten für optimale Lösungen liefern.

Aus den KKT-Bedingungen ergeben sich folgende Spezialfälle:

- Für das *unrestringierte Optimierungsproblem*

$$\min_{x \in \mathbb{R}^n} f(x)$$

erhält man die notwendige Bedingung

$$\nabla f(\hat{x}) = 0$$

- Für das *lineare Optimierungsproblem (in primaler Normalform)*

$$\min_{\substack{x \geq 0 \\ Ax = b}} c^T x$$

erhält man mit der Lagrange-Funktion $\mathcal{L}(x, \mu, \nu) = c^T x + \mu^T(-x) + \nu^T(b - Ax)$ die notwendigen Bedingungen

$$c - \mu - A^T \nu = 0, \quad \mu \geq 0, \quad \mu^T(-\hat{x}) = 0$$

bzw.

$$A^T \nu \leq c, \qquad \hat{x}^T(c - A^T \nu) = 0$$

- Für das *linear-quadratische Optimierungsproblem*

$$\min_{\substack{x \geq 0 \\ Ax = b}} \frac{1}{2} x^T Q x + c^T x$$

erhält man mit der Lagrange-Funktion $\mathcal{L}(x, \mu, \nu) = \frac{1}{2} x^T Q x + c^T x + \mu^T(-x) + \nu^T(b - Ax)$ die notwendigen Bedingungen

$$Q\hat{x} + c - \mu - A^T \nu = 0, \quad \mu \geq 0, \quad \mu^T(-\hat{x}) = 0$$

- Für das *gleichungsrestringierte Optimierungsproblem*

$$\min_{h(x) = 0} f(x)$$

erhält man mit der Lagrange-Funktion $\mathcal{L}(x, \nu) = f(x) + \nu^T h(x)$ die notwendigen Bedingungen

$$\begin{pmatrix} \nabla_x \mathcal{L}(\hat{x}, \nu) \\ h(\hat{x}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Dies ist i.A. ein nichtlineares Gleichungssystem in den Unbekannten x und ν und kann mit dem NEWTON-Verfahren gelöst werden (siehe LAGRANGE-NEWTON-Verfahren).

Für eine notwendige Bedingung zweiter Ordnung in einem KKT-Punkt $(\hat{x}, \hat{\mu}, \hat{\nu})$ benötigen wir den *kritischen Kegel*

$$\begin{aligned} T_K(x) := \{d \in \mathbb{R}^n \quad : \quad & \nabla g_i(x)^T d \leq 0,\ i \in \mathcal{A}(x),\ \hat{\mu}_i = 0, \\ & \nabla g_i(x)^T d = 0,\ i \in \mathcal{A}(x),\ \hat{\mu}_i > 0, \\ & \nabla h_j(x)^T d = 0,\ j = 1, \ldots, p\} \end{aligned}$$

Der kritische Kegel enthält tangentiale Richtungen d an den zulässigen Bereich für die die Richtungsableitung $\nabla f(\hat{x})^T d$ gleich Null ist. Für diese kritischen Richtungen müssen Bedingungen zweiter Ordnung herangezogen werden.

Es gilt:

Theorem 1.4.1.3. (Notwendige Bedingungen zweiter Ordnung)
Voraussetzungen:

- *f, g_i, $i = 1, \ldots, m$, und h_j, $j = 1, \ldots, p$ sind zweimal stetig differenzierbar.*
- *$(\hat{x}, \mu, \nu)$ ist ein KKT-Punkt.*
- *$\hat{x}$ ist ein lokales Minimum des Standard-Optimierungsproblems.*
- *Es gilt die LICQ in $\hat{x}$.*

Dann gilt

$$d^T \nabla^2_{xx} \mathcal{L}(\hat{x}, \mu, \nu) d \geq 0 \qquad \forall d \in T_K(\hat{x})$$

(Die Hessematrix der Lagrange-Funktion ist positiv semi-definit auf dem kritischen Kegel). □

Treten keine Restriktionen $g(x) \leq 0$ und $h(x) = 0$ auf (unbeschränkter Fall), so ist der kritische Kegel durch $T_K(\hat{x}) = \mathbb{R}^n$ gegeben und die notwendige Bedingung zweiter Ordnung reduziert sich auf die Bedingung

$$H_f(\hat{x}) \text{ ist positiv semidefinit,}$$

wobei $H_f(\hat{x})$ die Hessematrix der Zielfunktion f in $\hat{x}$ bezeichnet.

Hinreichende Bedingungen für restringierte Optimierungsprobleme

Die hinreichenden Bedingungen zweiter Ordnung sind sehr nah an den notwendigen Bedingungen zweiter Ordnung.

Theorem 1.4.1.4. (Hinreichende Bedingung zweiter Ordnung)
Voraussetzungen:

- *f, g_i, $i = 1, \ldots, m$ und h_j, $j = 1, \ldots, p$ sind zweimal stetig differenzierbar.*
- *$(\hat{x}, \mu, \nu)$ ist KKT-Punkt des Standard-Optimierungsproblems.*
- *Es gilt*

$$d^T \nabla^2_{xx} \mathcal{L}(\hat{x}, \mu, \nu) d > 0 \qquad \forall d \in T_K(\hat{x}), d \neq 0 \tag{1.4.6}$$

(Die Hessematrix der Lagrange-Funktion ist positiv-definit auf dem kritischen Kegel).

Dann existiert eine Umgebung U von $\hat{x}$ und ein $\alpha > 0$ mit

$$f(x) \geq f(\hat{x}) + \alpha \cdot \|x - \hat{x}\|^2 \qquad \forall x \in \Sigma \cap U$$

(insbesondere ist $\hat{x}$ also lokales Minimum und f wächst lokal mindestens quadratisch). □

Treten keine Restriktionen $g(x) \leq 0$ und $h(x) = 0$ auf (unbeschränkter Fall), so ist der kritische Kegel durch $T_K(\hat{x}) = \mathbb{R}^n$ gegeben und die hinreichende Bedingung zweiter Ordnung reduziert sich auf die Bedingung (1.4.6)

$$H_f(\hat{x}) \text{ ist positiv definit,}$$

wobei $\hat{x}$ ein stationärer Punkt von f sei.

Verfahren für unrestringierte Optimierungsprobleme

Für die Minimierung der Funktion $f : \mathbb{R}^n \to \mathbb{R}$ haben wir iterative Verfahren kennengelernt, die ausgehend von einer Startschätzung $x^{[0]}$ Näherungslösungen

$$x^{[k+1]} := x^{[k]} + \alpha_k \cdot d^{[k]} \qquad \text{für } k = 0, 1, 2, \ldots$$

berechnen. Hierin ist $d^{[k]}$ eine Suchrichtung und $\alpha_k > 0$ eine Schrittweite.

Zur Bestimmung der Suchrichtung gibt es u.a. die folgenden Ansätze:

- Beim *Gradientenverfahren* wird stets die Richtung des steilsten Abstiegs gewählt, also

  $$d^{[k]} := -\nabla f(x^{[k]})$$

  Diese Richtung ist außer in einem stationären Punkt stets eine Abstiegsrichtung.

- Beim *Newton-Verfahren* wird die Richtung

  $$d^{[k]} := -H_f(x^{[k]})^{-1} \nabla f(x^{[k]})$$

  gewählt. Ist die Hessematrix $H_f(x^{[k]})$ positiv definit, so ist die Newton-Richtung außer in einem stationären Punkt eine Abstiegsrichtung.

- Beim *Quasi-Newton-Verfahren* wird die Richtung

  $$d^{[k]} := -H_k^{-1} \nabla f(x^{[k]})$$

  gewählt, wobei die Matrix H_k stets symmetrisch und positiv definit gewählt wird und durch eine sogenannte Update-Formel in jedem Iterationsschritt aufdatiert wird. Diese Quasi-Newton-Richtung ist außer in einem stationären Punkt stets eine Abstiegsrichtung.

- *Trust-Region-Verfahren:* Die Suchrichtung ist durch Lösen des Trust-Region-Hilfsproblems

  $$\min_{\|d\| \le \Delta_k} \frac{1}{2} d^T H_k d + \nabla f(x^{[i]})^T d$$

  gegeben, wobei der Trust-Region-Radius $\Delta_k > 0$ und die Matrix H_k in jedem Schritt angepasst werden. Als Schrittweite wird beim Trust-Region-Verfahren stets $\alpha_k = 1$ gewählt, da die Schrittlänge durch den Trust-Region-Radius Δ_k gesteuert wird.

Beim Trust-Region-Verfahren arbeitet man mit der Schrittweite $\alpha_k = 1$ und verzichtet auf die nachfolgende Liniensuche, da sie durch die Steuerung des Trust-Region-Radius Δ_k überflüssig ist. Für alle anderen Verfahren kann die Schrittweite $\alpha_k > 0$ mithilfe einer eindimensionalen Liniensuche für die Funktion

$$\varphi(\alpha) := f(x^{[k]} + \alpha \cdot d^{[k]})$$

berechnet werden. Voraussetzung ist aber, dass $d^{[k]}$ eine Abstiegsrichtung ist. Üblicherweise verwendet man dann das Armijo-Verfahren oder darauf aufbauende Verfahren, z.B. die Wolfe-Powell-Regeln.

Algorithm 1.4.1.5. (Armijo-Regel)

(i) *Wähle $\beta \in (0,1)$, $\sigma \in (0,1)$ und setze $\alpha := 1$.*

(ii) *Falls die Bedingung*

$$\varphi(\alpha) \leq \varphi(0) + \sigma \cdot \alpha \cdot \varphi'(0)$$

erfüllt ist, setze $\alpha_k := \alpha$ und beende das Verfahren. Andernfalls gehe zu (iii).

(iii) *Setze $\alpha := \beta \cdot \alpha$ und gehe zu (ii).*

□

Überblick

Zur numerischen Lösung des Standard-Optimierungsproblems gibt es im Wesentlichen die folgenden Herangehensweisen:

(a) *Penalty- und Multiplikator-Verfahren:* Diese Verfahren basieren auf der Ankopplung der Nebenbedingungen an die Zielfunktion mithilfe eines gewichteten Strafterms (Penalty-Term), der unzulässige Punkte bestraft. Dadurch werden die Nebenbedingungen eliminiert und man kann Verfahren der unrestringierten Optimierung (Gradientenverfahren, Newtonverfahren, Quasi-Newton-Verfahren) anwenden. Zielfunktion und Strafterm müssen jedoch geeignet gewichtet werden, damit man letztendlich eine zulässige Lösung bekommt.

(b) *Sequentielle quadratische Programmierung (SQP):* SQP-Verfahren basieren auf der lokalen Approximation des Standard-Optimierungsproblems durch ein linear-quadratisches Optimierungsproblem, dessen Lösung die Suchrichtung in einem iterativen Verfahren liefert. SQP-Verfahren erweitern das Lagrange-Newton-Verfahren auf Probleme mit Ungleichungsrestriktionen.

(c) *Innere-Punkte-Verfahren (IP):* Innere-Punkte-Verfahren verwenden sogenannte Barriere-Funktionen, um Ungleichungsnebenbedingungen zu eliminieren und diese ähnlich wie bei Penalty-Verfahren mithilfe eines gewichteten Stratfterms an die Zielfunktion zu koppeln. Im Gegensatz zu Penalty- und Multiplikator-Verfahren wird dabei nicht nur das Verlassen des zulässigen Bereichs bestraft, sondern es wird bereits die Annäherung an den Rand des zulässigen Bereichs bestraft (den Rand des zulässigen Bereichs kann man sich als unüberwindliche Barriere vorstellen).

(d) *Verfahren für Komplementaritätsprobleme:* Diese Verfahren, zu denen semiglatte Newtonverfahren oder Variationsmethoden gehören, versuchen, die KKT- Bedingungen direkt zu lösen. Dazu werden die Komplementaritätsprobleme entweder als Variationsungleichung umgeschrieben und mit geeigneten Verfahren gelöst, oder sie werden mithilfe von speziellen Funktionen als Gleichung reformuliert. Letzteres führt auf ein nichtdifferenzierbares Gleichungssystem, auf das Varianten des Newtonverfahrens angewendet werden können.

Jede dieser Verfahrensklassen verwendet zusätzlich Strategien, um Konvergenz von beliebigen Startschätzungen zu erreichen (Globalisierungsstrategien). Die üblichen Strategien sind

- eindimensionale Liniensuche (z.B. Armijo-Verfahren)
- Trust-Region-Verfahren (Approximation auf einem Vertrauensbereich)
- Filterverfahren (versuchen, nicht-dominierte Iterierte zu erzeugen, wobei Zielfunktionswert und Verletzung der Restriktionen als zwei Kriterien mitgeführt werden)

1.4.2 Penalty- und Multiplikator-Verfahren

Penalty- und Multiplikatorverfahren sind beliebte Verfahren, die auf der Ankopplung der Nebenbedingungen an die Zielfunktion mithilfe eines gewichteten Strafterms (Penalty-Term) basieren. Der Strafterm bestraft unzulässige Punkte. Der Vorteil der Verfahren ist, dass durch die Ankopplung der Nebenbedingungen an die Zielfunktion Nebenbedingungen eliminiert werden, so dass Verfahren der unrestringierten Optimierung angewendet werden können.

Das Konzept der Penalty-Verfahren für die allgemeine Aufgabenstellung

$$(\mathrm{P}) \qquad \min_{x\in\Sigma} f(x)$$

funktioniert wie folgt. Man benötigt eine Funktion $r : \mathbb{R}^n \to [0,\infty)$ mit der Eigenschaft

$$r(x)\begin{cases} = 0, & \text{falls } x \in \Sigma \\ > 0, & \text{falls } x \notin \Sigma. \end{cases}$$

Dann minimiert man für eine geeignet gewählte Folge von Gewichtungsparametern $\{\eta_k\}_{k\in\mathbb{N}}$ mit $\eta_k > 0$ die unrestringierte Penalty-Funktion

$$P(x;\eta_k) := f(x) + \eta_k \cdot r(x) \tag{1.4.7}$$

Für jedes $\eta_k > 0$ erhält man eine Lösung $x^{[k]} := x(\eta_k)$ und es stellt sich die Frage, wie die Gewichtungsparameter η_k, $k \in \mathbb{N}$, gewählt werden müssen, damit die Folge $\{x^{[k]}\}_{k\in\mathbb{N}}$ gegen ein Minimum von (P) konvergiert. Die Funktion r kann auf verschiedene Arten definiert werden, wobei differenzierbare Varianten ideal sind, um die uns bekannten Verfahren der unrestringierten Verfahren anwenden zu können. Wird r als stetige, aber nicht stetig differenzierbare Funktion gewählt, so gestaltet sich die Lösung des unrestringierten Penalty-Problems schwieriger.

Penalty-Verfahren

Dazu betrachten wir zunächst das folgende gleichungsrestringierte Optimierungsproblem mit stetigen (!) Funktionen $f : \mathbb{R}^n \to R$ und $h_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1,\dots,p$.

Problem 1.4.2.1. (Restringiertes Optimierungsproblem)

$$\min_{x\in\Sigma} f(x)$$

mit

$$\Sigma = \{x \in \mathbb{R}^n \mid h_j(x) = 0,\ j = 1,\dots,p\}.$$

□

Die Idee des Penalty-Verfahrens besteht darin, die Lösung $\hat{x}$ des Ausgangsproblems iterativ durch die Lösungen von unrestringierten Hilfsproblemen zu approximieren. Diese Hilfsprobleme bestehen in der Minimierung der Penalty-Funktion

$$P(x;\eta) := f(x) + \frac{\eta}{2}\sum_{j=1}^{p}(h_j(x))^2$$

für geeignete Werte von $\eta > 0$. Durch die Ankopplung der Nebenbedingungen wird ein Verlassen des zulässigen Bereichs Σ 'bestraft'. Die Konstante η stellt einen Gewichtungsfaktor dar, mit dessen Hilfe die 'Stärke der Bestrafung' gesteuert werden kann. Das Penalty-Verfahren ist durch folgenden Algorithmus gegeben.

Algorithm 1.4.2.2. (Penalty-Verfahren)

(i) *Wähle $\eta_0 > 0$ und setze $k = 0$.*

(ii) *Bestimme $x^{[k]}$ als Lösung von*

$$\min_{x \in \mathbb{R}^n} P(x; \eta_k)$$

(iii) *Ist $h(x^{[k]}) \approx 0$,* `STOP`.

(iv) *Bestimme $\eta_{k+1} > \eta_k$, setze $k := k + 1$ und gehe zu (ii).*

□

Da P i.A. nicht differenzierbar ist, werden für das Hilfsproblem in Schritt (ii) Verfahren der unrestringierten, nichtdifferenzierbaren Optimierung benötigt. Es stellt sich natürlich die Frage, ob das Verfahren tatsächlich gegen eine Lösung des Ausgangsproblems konvergiert.

Theorem 1.4.2.3. (Konvergenzsatz für das Penalty-Verfahren)
Seien f und h_j, $j = 1, \ldots, p$ stetig und $\{\eta_k\}$ streng monoton wachsend mit $\eta_k \to \infty$. Die zulässige Menge $\Sigma := \{x \in \mathbb{R}^n \mid h_j(x) = 0,\ j = 1, \ldots, p\}$ sei nichtleer, und $\{x^{[k]}\}$ sei eine durch Algorithmus (1.4.2.2) erzeugte Folge (die Existenz der Folge sei vorausgesetzt). Dann gelten die folgenden Aussagen:

(a) *Die Folge der Zielfunktionswerte der Penalty-Funktion $\{P(x^{[k]}; \eta_k)\}_{k \in \mathbb{N}}$ ist monoton wachsend.*

(b) *Die Folge der Verletzung der Nebenbedingungen $\{\|h(x^{[k]})\|\}_{k \in \mathbb{N}}$ ist monoton fallend.*

(c) *Die Folge der Zielfunktionswerte $\{f(x^{[k]})\}_{k \in \mathbb{N}}$ ist monoton wachsend.*

(d) *Es gilt $\lim_{k \to \infty} h(x^{[k]}) = 0$.*

(e) *Jeder Häufungspunkt der Folge $\{x^{[k]}\}_{k \in \mathbb{N}}$ ist eine Lösung des Ausgangsproblems.*

□

Remark 1.4.2.4. *Da nur die Stetigkeit der auftretenden Funktionen benötigt wird, ist das Verfahren auch auf Problemstellungen mit Ungleichungsnebenbedingungen*

$$g_i(x) \leq 0, \qquad i = 1, \ldots, m,$$

anwendbar. Denn diese Nebenbedingungen können äquivalent als stetige Nebenbedingungen

$$\max\{0,\ g_i(x)\} = 0, \qquad i = 1, \ldots, m,$$

geschrieben werden. Die Penaltyfunktion lautet dann

$$P(x; \eta) = f(x) + \frac{\eta}{2} \sum_{i=1}^{m} (\max\{0,\ g_i(x)\})^2 + \frac{\eta}{2} \sum_{j=1}^{p} (h_j(x))^2.$$

□

Ein wesentlicher Nachteil des Penalty-Verfahrens ist die Tatsache, dass die Gewichtungsfaktoren η_k gegen ∞ streben müssen, um Konvergenz zu erhalten. Dies führt dazu, dass die Teilprobleme in (ii) des Algorithmus für großes η_k sehr schlecht konditioniert sind[1] und numerisch nur sehr schwer zu lösen sind.

Schätzung der Lagrange-Multiplikatoren

Wir untersuchen, wie aus der Folge $\{x^{[k]}\}_{k\in\mathbb{N}}$ eine Folge $\{\nu^{[k]}\}_{k\in\mathbb{N}}$ von Näherungen der Lagrange-Multiplikatoren konstruiert werden kann, so dass $x^{[k]}$ und $\nu^{[k]}$ gegen einen KKT-Punkt $\hat{x}$ und $\hat{\nu}$ des Ausgangsproblems konvergieren.

Hierzu benötigen wir die stetige Differenzierbarkeit der Funktionen f und h_j, $j = 1, \ldots, p$. Ein KKT-Punkt $(\hat{x}, \hat{\nu})$ des Ausgangsproblems erfüllt

$$0 = \nabla f(\hat{x}) + \sum_{j=1}^{p} \hat{\nu}_j \nabla h_j(\hat{x}).$$

Da $x^{[k]}$ Minimalstelle der Penalty-Funktion mit Gewichtungsparameter η_k ist, gilt notwendig

$$0 = \nabla_x P(x^{[k]}; \eta) = \nabla f(x^{[k]}) + \eta_k \cdot \sum_{j=1}^{p} h_j(x^{[k]}) \nabla h_j(x^{[k]}).$$

Vergleicht man die beiden Ausdrücke, so liegt es nahe,

$$\nu^{[k]} = \eta_k h_j(x^{[k]}) \tag{1.4.8}$$

als Approximation der Lagrange-Multiplikatoren $\hat{\nu}_j$ zu verwenden.

Es gilt:

Theorem 1.4.2.5. *Seien f und h_j, $j = 1, \ldots, p$, stetig differenzierbar und $\{x^{[k]}\}_{k\in\mathbb{N}}$ eine durch das Penalty-Verfahren erzeugte Folge mit $x^{[k]} \to \hat{x}$ für $k \to \infty$. Die Gradienten $\nabla h_j(\hat{x})$, $j = 1, \ldots, p$, seien linear unabhängig, und $\{\nu^{[k]}\}_{k\in\mathbb{N}}$ sei durch (1.4.8) gegeben. Dann gelten:*

(a) Die Folge $\{\nu^{[k]}\}_{k\in\mathbb{N}}$ konvergiert gegen einen Vektor $\hat{\nu}$.

(b) $(\hat{x}, \hat{\nu})$ ist ein KKT-Punkt des Ausgangsproblems.

□

Multiplikator-Penalty-Verfahren

Multiplikator-Penalty-Verfahren ähneln den Penalty-Verfahren. Allerdings arbeiten sie mit einer exakten und differenzierbaren Penalty-Funktion — der erweiterten Lagrange-Funktion.

Wir betrachten wieder das gleichungsrestringierte Problem 1.4.2.1, d.h.

$$\min_{h(x)=0} f(x) \tag{1.4.9}$$

[1] Einige Eigenwerte der Hessematrix $\nabla^2_{xx} P(x^{[k]}; \eta_k)$ streben gegen ∞ und somit strebt die Spektral-Konditionszahl der Hessematrix gegen unendlich.

Darin seien die Funktionen $f : \mathbb{R}^n \to \mathbb{R}$ und $h = (h_1, \ldots, h_p)^T : \mathbb{R}^n \to \mathbb{R}^p$ zweimal stetig differenzierbar.

Sei $\hat{x}$ lokales Minimum des Problems. Dann ist $\hat{x}$ für $\eta > 0$ auch ein lokales Minimum von

$$\min_{h(x)=0} f(x) + \frac{\eta}{2}\|h(x)\|^2$$

Die Lagrange-Funktion für dieses Problem lautet

$$\mathcal{L}_a(x, \nu; \eta) := f(x) + \frac{\eta}{2}\|h(x)\|^2 + \nu^T h(x)$$

und heißt erweiterte *Lagrange-Funktion (augmented Lagrangian)* oder *Multiplikator-Penalty-Funktion.*

Es zeigt sich, dass der Gewichtungsparameter η für $\mathcal{L}_a$ nicht gegen ∞ streben muss, um ein lokales Minimum des Ausgangsproblems zu erreichen.

Theorem 1.4.2.6. *Sei $(\hat{x}, \hat{\nu})$ KKT-Punkt von (1.4.9). Desweiteren sei die hinreichende Bedingung zweiter Ordnung (1.4.6) erfüllt. Dann existiert ein endliches $\bar{\eta} > 0$, so dass $\hat{x}$ für jedes $\eta \geq \bar{\eta}$ ein striktes lokales Minimum von $\mathcal{L}_a(\cdot, \hat{\nu}; \eta)$ ist.* □

Auf Grund dieses Hilfssatzes kann man versuchen, das Ausgangsproblem (1.4.9) indirekt zu lösen, indem die erweiterte Lagrange-Funktion minimiert wird:

$$\min_{x \in \mathbb{R}^n} \mathcal{L}_a(x, \hat{\nu}; \eta)$$

Der Penalty-Parameter η muß jetzt, anders als bei den Penalty-Verfahren, nicht mehr gegen ∞ streben. Darüber hinaus ist $\mathcal{L}_a$ differenzierbar, so dass bekannte Verfahren aus der unrestringierten Optimierung eingesetzt werden können.

Problem: Der optimale Lagrange-Multiplikator $\hat{\nu}$ ist unbekannt.

Wir versuchen nun, $\hat{\nu}$ geeignet zu approximieren. Sei η hinreichend groß und $x^{[k+1]}$ stationärer Punkt des Problems

$$\min_{x \in \mathbb{R}^n} \mathcal{L}_a(x, \nu^{[k]}; \eta)$$

Dann gilt notwendig

$$0 = \nabla_x \mathcal{L}_a(x^{[k+1]}, \nu^{[k]}; \eta) = \nabla f(x^{[k+1]}) + \sum_{j=1}^{p} (\nu_j^{[k]} + \eta \cdot h_j(x^{[k+1]}) \nabla h_j(x^{[k+1]}).$$

Andererseits gilt in einem KKT-Punkt $(\hat{x}, \hat{\nu})$ von (1.4.9) notwendig

$$0 = \nabla_x \mathcal{L}(\hat{x}, \hat{\nu}) = \nabla f(\hat{x}) + \sum \hat{\nu}_j \nabla h_j(\hat{x}).$$

Ein Vergleich beider Ausdrücke liefert die naheliegende Aufdatierungsvorschrift

$$\nu^{[k+1]} := \nu^{[k]} + \eta \cdot h(x^{[k+1]}).$$

Insgesamt entsteht das Multiplier-Penalty-Verfahren:

Algorithm 1.4.2.7. (Multiplikator-Penalty-Verfahren)

(i) *Wähle* $x[0] \in \mathbb{R}^n$, $\nu^{[0]} \in \mathbb{R}^p$, $\eta_0 > 0$, $\sigma \in (0,1)$ *und setze* $k = 0$.

(ii) *Ist* $(x^{[k]}, \nu^{[k]})$ *KKT-Punkt von (1.4.9),* `STOP`.

(iii) *Bestimme* $x^{[k+1]}$ *als Lösung von*

$$\min_{x \in \mathbb{R}^n} \mathcal{L}_a(x, \nu^{[k]}; \eta_k)$$

(iv) *Setze* $\nu^{[k+1]} := \nu^{[k]} + \eta_k h(x^{[k+1]})$.

(v) *Ist* $\|h(x^{[k+1]})\| \geq \sigma \|h(x^{[k]})\|$, *so setze* $\eta_{k+1} := 10\eta_k$, *andernfalls setze* $\eta_{k+1} := \eta_k$.

(vi) *Setze* $k := k + 1$ *und gehe zu (ii).*

□

Anwendung auf Ungleichungen

Das Standard-Optimierungsproblem

$$\min_{\substack{g(x) \leq 0, \\ h(x) = 0}} f(x) \tag{1.4.10}$$

ist durch Einführung von Schlupfvariablen $s = (s_1, \ldots, s_m)^T \in \mathbb{R}^m$ äquivalent mit dem gleichungsrestringierten Problem

$$\min_{\substack{(x,s) \in \mathbb{R}^{n \times m} \\ g_i(x) + s_i^2 = 0,\ i=1,\ldots,m \\ h_j(x) = 0,\ j=1,\ldots,p}} f(x)$$

Die erweiterte Lagrange-Funktion hierfür lautet

$$\bar{\mathcal{L}}_a(x, s, \mu, \nu; \eta) = f(x) + \frac{\eta}{2}\|h(x)\|^2 + \nu^T h(x) + \sum_{i=1}^{m} \left(\mu_i \left(g_i(x) + s_i^2 \right) + \frac{\eta}{2} \left(g_i(x) + s_i^2 \right)^2 \right)$$

Für festes x kann die Minimierung bzgl. s explizit ausgeführt werden und man erhält

$$\hat{s}_i = \left(\max \left\{ 0, -\left(\frac{\mu_i}{\eta} + g_i(x) \right) \right\} \right)^{\frac{1}{2}}, i = 1, \ldots, m.$$

Einsetzen in die erweiterte Lagrange-Funktion liefert

$$\begin{aligned} \mathcal{L}_a(x, \mu, \nu; \eta) &= f(x) \quad + \nu^T h(x) + \frac{\eta}{2}\|h(x)\|^2 \\ &\qquad + \frac{1}{2\eta} \sum_{i=1}^{m} \left((\max\{0, \mu_i + \eta g_i(x)\})^2 - \mu_i^2 \right) \\ &= f(x) \quad + \sum_{j=1}^{p} \left(\nu_j h_j(x) + \frac{\eta}{2} h_j(x)^2 \right) \\ &\qquad + \sum_{i=1}^{m} \begin{cases} \mu_i g_i(x) + \dfrac{\eta}{2} g_i(x)^2, & \text{falls } \mu_i + \eta g_i(x) \geq 0 \\ -\dfrac{\mu_i^2}{2\eta} & \text{sonst.} \end{cases} \end{aligned}$$

Beachte, dass diese Funktion nur noch stetig differenzierbar ist.

Für die Multiplikatoren ergeben sich die Aufdatierungsformeln

$$\begin{aligned}\mu_i^{[k+1]} &:= \max\left\{0, \mu^{[k]} + \eta \cdot g_i(x^{[k+1]})\right\},\ i = 1, \ldots, m \\ \nu^{[k+1]} &:= \nu^{[k]} + \eta \cdot h(x^{[k+1]}),\end{aligned}$$

1.4.3 SQP-Verfahren

Die sequentielle quadratische Programmierung (SQP) wird vielfach in der Literatur behandelt. Es existieren diverse Implementierungen. Zunächst diskutieren wir das lokale SQP-Verfahren (mit Schrittweite 1) und erweitern das lokale Verfahren dann durch eine Globalisierungsstrategie, die auf dem Armijo-Verfahren basiert.

Das lokale SQP-Verfahren

Zur Motivation des SQP-Verfahrens erinnern wir uns an das Lagrange-Newton-Verfahren. Das Lagrange-Newton-Verfahren eignet sich zur Lösung des gleichungsrestringierten Optimierungsproblems

$$\min_{h(x)=0} f(x)$$

wobei $f : \mathbb{R}^n \to \mathbb{R}$ und $h : \mathbb{R}^n \to \mathbb{R}^p$ zweimal stetig differenzierbare Funktionen seien und $\mathcal{L}(x, \nu) = f(x) + \nu^T h(x)$ die Lagrange-Funktion bezeichnet. Das Lagrange-Newton-Verfahren entsteht durch Anwendung des Newton-Verfahrens auf die KKT-Bedingungen

$$\nabla_x \mathcal{L}(x, \nu) = 0 \text{ und } h(x) = 0$$

und lautet wie folgt:

Algorithm 1.4.3.1. (Lagrange-Newton-Verfahren)

(i) Wähle Startschätzungen $x^{[0]} \in \mathbb{R}^n$ und $\nu^{[0]} \in \mathbb{R}^p$, $\epsilon > 0$ und setze $k = 0$.

(ii) Falls $\max\left\{\|\nabla_x \mathcal{L}(x^{[k]}, \nu^{[k]})\|, \|h(x^{[k]})\|\right\} \leq \epsilon$, `STOP`.

(iii) Löse das lineare Gleichungssystem

$$\begin{pmatrix} \nabla^2_{xx} \mathcal{L}(x^{[k]}, \nu^{[k]}) & h'(x^{[k]})^T \\ 0 & h'(x^{[k]}) \end{pmatrix} \begin{pmatrix} d \\ v \end{pmatrix} = - \begin{pmatrix} \nabla_x \mathcal{L}(x^{[k]}, \nu^{[k]}) \\ h(x^{[k]}). \end{pmatrix} \tag{1.4.11}$$

und setze

$$x^{[k+1]} := x^{[k]} + d, \qquad \nu^{[k+1]} := \nu^{[k]} + v \tag{1.4.12}$$

(iv) Setze $k := k + 1$und gehe zu (ii).

□

Wir starten zunächst mit einer Beobachtung. Das lineare Gleichungssystem (1.4.11) in (iii) des Lagrange-Newton-Verfahrens entsteht auch auf andere Art. Wir erinnern uns an das Newton-Verfahren für unrestringierte Optimierungsprobleme. Dort hatten wir das Newtonverfahren auf zwei Arten motiviert:

1. Anwendung des Newtonverfahrens auf die notwendigen Bedingung $\nabla f = 0$ (indirekter Ansatz);

2. lokale Approximation der Zielfunktion durch eine quadratische Funktion (direkter Ansatz).

Beide Ansätze lieferten das gleiche Verfahren.

Zur Abkürzung setzen wir im Folgenden

$$Q_k := \nabla^2_{xx} \mathcal{L}(x^{[k]}, \nu^{[k]}).$$

Wir betrachten nun wieder das gleichungsrestringierte Optimierungsproblem und approximieren es lokal im Punkt $(x^{[k]}, \nu^{[k]})$ durch das quadratische Optimierungsproblem

$$\min_{\substack{d \in \mathbb{R}^n \\ h(x^{[k]}) + h'(x^{[k]})d = 0}} \frac{1}{2} d^T Q_k d + \nabla f(x^{[k]})^T d$$

Die Lagrange-Funktion für das quadratische Optimierungsproblem ist gegeben durch

$$\mathcal{L}_{QP}(d, \eta) := \frac{1}{2} d^T Q_k d + \nabla f(x^{[k]})^T d + \eta^T h(x^{[k]}) + h'(x^{[k]}) d$$

Auswertung der KKT-Bedingungen führt auf das lineare Gleichungssystem

$$\begin{aligned} Q_k d + \nabla f(x^{[k]}) + h'(x^{[k]})^T \eta &= 0 \\ h(x^{[k]}) + h'(x^{[k]}) d &= 0 \end{aligned}$$

bzw.

$$\begin{pmatrix} Q_k & h'(x^{[k]})^T \\ h'(x^{[k]}) & 0 \end{pmatrix} \begin{pmatrix} d \\ \eta \end{pmatrix} = - \begin{pmatrix} \nabla f(x^{[k]}) \\ h(x^{[k]}) \end{pmatrix} \tag{1.4.13}$$

Subtraktion von $h'(x^{[k]})^T \nu^{[k]}$ auf beiden Seiten der ersten Gleichung in (1.4.13) liefert das lineare Gleichungssystem

$$\begin{pmatrix} Q_k & h'(x^{[k]})^T \\ h'(x^{[k]}) & 0 \end{pmatrix} \begin{pmatrix} d \\ \eta - \nu^{[k]} \end{pmatrix} = - \begin{pmatrix} \nabla_x \mathcal{L}(x^{[k]}, \nu^{[k]}) \\ h(x^{[k]}) \end{pmatrix} \tag{1.4.14}$$

Ein Vergleich von (1.4.14) mit (1.4.11) zeigt, dass diese zwei Gleichungssysteme identisch sind, wenn man noch $v := \eta - \nu^{[k]}$ definiert. Die neuen Iterierten in (3.2) lassen sich damit wie folgt berechnen:

$$x^{[k+1]} = x^{[k]} + d, \quad \nu^{[k+1]} = \nu^{[k]} + v = \eta.$$

Summary 1.4.3.2. *Für gleichungsrestringierte Optimierungsprobleme ist das Lagrange-Newton-Verfahren identisch mit dem oben hergeleiteten sukzessiven quadratischen Optimierungsverfahren, wenn der Multiplikator η des quadratischen Hilfsproblems als neue Approximation für den Multiplikator ν des Ausgangsproblems verwendet wird.* □

Diese Beobachtung motiviert die folgende Erweiterung des quadratischen Hilfsproblems für Standard-Optimierungsprobleme mit Gleichungs- und Ungleichungsrestriktionen:

Problem 1.4.3.3. (QP Problem $QP(x^{[k]}, \mu^{[k]}, \nu^{[k]})$)

$$\begin{aligned} \frac{1}{2} d^T Q_k d + \nabla f(x^{[k]})^T d &= \min_{d \in \mathbb{R}^n} ! \\ g(x^{[k]}) + g'(x^{[k]}) d &\leq 0 \\ h(x^{[k]}) + h'(x^{[k]}) d &= 0 \end{aligned}$$

□

Sukzessive quadratische Approximation liefert das lokale SQP Verfahren:

Algorithm 1.4.3.4. (Lokales SQP Verfahren)

(i) Wähle Startwerte $(x^{[0]}, \mu^{[0]}, \nu^{[0]}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ *und setze* $k = 0$.

(ii) Falls $(x^{[k]}, \mu^{[k]}, \nu^{[k]})$ *ein KKT-Punkt des Standard-Optimierungsproblems ist,* `STOP`.

(iii) Berechne einen KKT-Punkt $(d^{[k]}, \mu^{[k+1]}, \nu^{[k+1]}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ *des quadratischen Optimierungsproblems* $QP(x^{[k]}, \mu^{[k]}, \nu^{[k]})$.

(iv) Setze $x^{[k+1]} := x^{[k]} + d^{[k]}$, $k := k + 1$ *und gehe zu (ii).*

□

Remark 1.4.3.5.

- *Es ist nicht notwendig, die Indexmenge* $\mathcal{A}(\hat{x})$ *der aktiven Ungleichungsnebenbedingungen im Voraus zu kennen.*
- *Die Iterierten* $x^{[k]}$ *sind in der Regel nicht zulässig, d.h. es gilt i.A.* $x^{[k]} \notin \Sigma$.

□

Die lokale Konvergenz des SQP Verfahrens wird im folgenden Satz formuliert.

Theorem 1.4.3.6. (Lokale Konvergenz des SQP-Verfahrens)
Voraussetzungen:

(i) $\hat{x}$ *ist lokales Minimum des Standard-Optimierungsproblems und* $\hat{\mu}$ *und* $\hat{\nu}$ *bezeichnen die Lagrange-Multiplikatoren.*

(ii) Die Funktionen f, g_i, $i = 1, \ldots, m$, *und* h_j, $j = 1, \ldots, p$, *sind zweimal stetig differenzierbar mit Lipschitz-stetigen zweiten Ableitungen.*

(iii) Es gilt die 'Linear Independence Constraint Qualification' (LICQ) in $\hat{x}$.

(iv) Die strikte Komplementaritätsbedingung $\hat{\mu}_i - g_i(\hat{x}) \geq 0$ *ist für alle* $i \in \mathcal{A}(\hat{x})$ *erfüllt.*

(v) Es gilt die hinreichende Bedingung zweiter Ordnung:

$$d^T \nabla^2_{xx} \mathcal{L}(\hat{x}, \hat{\mu}, \hat{\nu}) d > 0$$

für alle $0 \neq d \in \mathbb{R}^n$ *mit*

$$\nabla g_i(\hat{x})^T d = 0, \quad i \in \mathcal{A}(\hat{x}), \quad \nabla h_j(\hat{x})^T d = 0, \quad j = 1, \ldots, p.$$

Dann existieren Umgebungen U von $(\hat{x}, \hat{\mu}, \hat{\nu})$ und V von $(0, \hat{\mu}, \hat{\nu})$, so dass alle quadratischen Optimierungsprobleme $QP(x^{[k]}, \mu^{[k]}, \nu^{[k]})$ für beliebige Startwerte

$$(x^{[0]}, \mu^{[0]}, \nu^{[0]}) \in U$$

in V eine eindeutige lokale Lösung $d^{[k]}$ mit eindeutigen Multiplikatoren $\mu^{[k+1]}$ und $\nu^{[k+1]}$ besitzen. Desweiteren konvergiert die Folge $\left\{(x^{[k]}, \mu^{[k]}, \nu^{[k]})\right\}_{k\in\mathbb{N}}$ quadratisch gegen $(\hat{x}, \hat{\mu}, \hat{\nu})$. □

Remark 1.4.3.7. (Approximation der Hessematrix) *Die Verwendung der exakten Hessematrix $Q_k = \nabla^2_{xx}\mathcal{L}(x^{[k]}, \mu^{[k]}, \nu^{[k]})$ im QP-Problem hat zwei Nachteile:*

- *In vielen Anwendungen ist die Hessematrix nicht explizit bekannt. Die numerische Approximation durch finite Differenzen ist sehr aufwendig und ungenau.*
- *Die Hessematrix kann indefinit sein. Dies erschwert die Lösung der QP-Hilfsprobleme erheblich. Es ist daher wünschenswert, die Hessematrix durch eine positiv definite Matrix zu ersetzen (siehe dazu die Idee der Quasi-*Newton*-Verfahren).*

In der Praxis wird die Hessematrix der Lagrange-Funktion Q_k in Iteration k durch eine geeignete Matrix H_k ersetzt. Powell *[Pow78] schlug vor, die modifizierte BFGS-Update-Formel mit*

$$H_{k+1} = H_k + \frac{q^{[k]}(q^{[k]})^T}{(q^{[k]})^T d^{[k]}} - \frac{H_k d^{[k]}(d^{[k]})^T H_k}{(d^{[k]})^T H_k d^{[k]}}, \tag{1.4.15}$$

mit

$$\begin{aligned}
d^{[k]} &= x^{[k+1]} - x^{[k]},\\
q^{[k]} &= \theta_k y^{[k]} + (1-\theta_k) H_k d^{[k]},\\
y^{[k]} &= \nabla_x \mathcal{L}(x^{[k+1]}, \mu^{[k]}, \nu^{[k]}) - \nabla_x \mathcal{L}(x^{[k]}, \mu^{[k]}, \nu^{[k]}),\\
\theta_k &= \begin{cases} 1, & \text{falls } (d^{[k]})^T y^{[k]} \geq 0.2 \cdot (d^{[k]})^T H_k d^{[k]},\\ \dfrac{0.8 \cdot (d^{[k]})^T H_k d^{[k]}}{(d^{[k]})^T H_k d^{[k]} - (d^{[k]})^T y^{[k]}}, & \text{sonst} \end{cases}
\end{aligned}$$

*zu verwenden. Diese Update-Formel garantiert, dass H_{k+1} symmetrisch und positiv definit bleibt, wenn H_k symmetrisch und positiv definit war. Für $\theta_k = 1$ entsteht die BFGS-Formel, welche schon bei den Quasi-*Newton*-Verfahren verwendet wurde (allerdings musste dort durch Wahl einer geeigneten Schrittweiten-Strategie noch die Bedingung $(d^{[k]})^T y^{[k]} > 0$ garantiert werden).*
Wird die modifizierte BFGS-Update-Formel im SQP-Verfahren verwendet, kann immerhin noch superlineare Konvergenz nachgewiesen werden. □

Globalisierung des SQP-Verfahrens

Das Konvergenzresultat zeigt, dass das SQP-Verfahren für alle Startwerte, die in einer Umgebung eines lokalen Minimums liegen, konvergent ist. In der Praxis ist diese Umgebung jedoch unbekannt und kann sehr klein sein. Daher ist es notwendig, das SQP-Verfahren zu globalisieren, so dass es (unter geeigneten Bedingungen) für beliebige Startwerte konvergiert. Wie im unrestringierten Fall wird dies durch Einführung einer Schrittweite $\alpha_k > 0$ erreicht. Die neue Iterierte ist gegeben durch

$$x^{[k+1]} = x^{[k]} + \alpha_k \cdot d^{[k]},$$

wobei $d^{[k]}$ wie zuvor ein quadratisches Hilfsproblem löst. Zur Bestimmung der Schrittweite α_k wird wieder eine eindimensionale Liniensuche in Richtung $d^{[k]}$ durchgeführt. Im Unterschied zur unrestringierten Optimierung tritt jetzt allerdings das folgende Problem auf:

$$\text{Wann ist } x^{[k+1]} \text{ 'besser' als } x^{[k]} \text{ ?}$$

Im unrestringierten Fall konnte diese Frage leicht durch einen Vergleich der Zielfunktionswerte beantwortet werden: $x^{[k+1]}$ ist besser als $x^{[k]}$, wenn $f(x^{[k+1]}) < f(x^{[k]})gilt$.

Im restringierten Fall ist dies nicht mehr so einfach, da die Iterierten $x^{[k]}$ des SQP-Verfahrens i.a. unzulässig sind. Eine Verbesserung kann also sowohl an Hand der Zielfunktionswerte als auch an Hand der Verletzungen der Nebenbedingungen gemessen werden. Dies sind i.A. zwei miteinander konkurrierende Kriterien, da man einen besseren Zielfunktionswert leicht auf Kosten der Zulässigkeit erreichen kann und umgekehrt.

Ein Ansatz, um dieses Dilemma aufzulösen, besteht in der Verwendung von sogenannten Bewertungsfunktionen (engl. merit functions), die im einfachsten Fall Zielfunktion und Verletzung der Nebenbedingungen gewichtet in einer skalar-wertigen Funktion vereinen (siehe Idee der Penalty-Funktion).

Mithilfe der Bewertungsfunktion ist es möglich zu entscheiden, ob die neue Iterierte $x^{[k+1]}$ 'besser' ist als die alte Iterierte $x^{[k]}$. Dabei ist die neue Iterierte besser als die alte, falls entweder ein hinreichender Abstieg in der Zielfunktion f oder eine weniger starke Verletzung der Nebenbedingungen erreicht wird, wobei sich das jeweils andere Kriterium nicht substantiell verschlechtern darf.

Eine allgemeine Klasse von Bewertungsfunktionen wird durch

$$P_r(x;\eta) := f(x) + \eta \cdot r(x) \tag{1.4.16}$$

definiert (vgl. (1.4.7)), wobei $\eta > 0$ einen Gewichtungsparameter und $r : \mathbb{R}^n \to [0,\infty)$ eine stetige Funktion mit der Eigenschaft

$$r(x) \begin{cases} = 0, & \text{falls } x \in \Sigma, \\ > 0, & \text{falls } x \notin \Sigma \end{cases}$$

bezeichnen.

Example 1.4.3.8. (Bewertungsfunktion) *Eine typische Bewertungsfunktion für das Standard-Optimierungsproblem, die auf der* 1*-Norm basiert, ist die* ℓ_1*-Bewertungsfunktion:*

$$\ell_1(x;\eta) := f(x) + \eta \sum \max\{0, g_i(x)\} + \sum |h_j(x)|, \quad \eta > 0.$$

Beachte, dass unzulässige Punkte $x \notin \Sigma$ *durch die Terme*

$$\sum \max\{0, g_i(x)\} + \sum |h_j(x)| > 0$$

bestraft werden. Desweiteren ist ℓ_1 *Lipschitz-stetig (falls* f*,* g_i *und* h_j *differenzierbar sind), aber nicht differenzierbar.*
Allgemeinere Bewertungsfunktionen basieren auf der q*-Norm:*

$$\ell_q(x;\eta) := f(x) + \eta \left(\sum_{i=1}^{m} (\max\{0, g_i(x)\})^q + \sum_{j=1}^{p} |h_j(x)|^q \right)^{\frac{1}{q}}$$

und

$$\ell_\infty(x;\eta) := f(x) + \eta \cdot \max\{0, g_1(x), \ldots, g_m(x), |h_1(x)|, \ldots, |h_p(x)|\}$$

□

Von besonderem Interesse sind die sogenannten exakten Bewertungsfunktionen, da für diese Bewertungsfunktionen lokale Minima des restringierten Ausgangsproblems auch lokale Minima der unrestringierten Bewertungsfunktion sind und der Gewichtungsparameter η dabei endlich gewählt werden kann.

Definition 1.4.3.9. (Exakte Bewertungsfunktion) *Die Bewertungsfunktion $P_r(x;\eta)$ in (1.4.16) heißt exakt in einem lokalen Minimum $\hat{x}$ des Standard-Optimierungsproblems, falls es einen endlichen (!) Parameter $\hat{\eta} > 0$ gibt, so dass $\hat{x}$ ein lokales Minimum von $P_r(\cdot;\eta)$ für alle $\eta \geq \hat{\eta}$ ist.* □

Es wäre wünschenswert, eine differenzierbare exakte Bewertungsfunktion zu haben. Dummerweise kann gezeigt werden, dass Bewertungsfunktionen der Form $P_r(x;\eta)$ aus (1.4.16) in einem lokalen Minimum $\hat{x}$ stets nicht differenzierbar sind, falls sie exakt sind und $\nabla f(\hat{x}) \neq 0$ gilt (letzteres ist der Normalfall in der restringierten Optimierung).

Der folgende Satz sagt aus, dass die Bewertungsfunktionen ℓ_q für $1 \leq q \leq \infty$ exakt sind, wenn eine Regularitätsbedingung gilt.

Theorem 1.4.3.10. *Sei $\hat{x} \in \Sigma$ ein isoliertes lokales Minimum des Standard-Optimierungsproblems, welches die LICQ erfüllt. Dann ist ℓ_q exakt für $1 \leq q \leq \infty$.* □

Im Folgenden beschränken wir uns auf die ℓ_1-Bewertungsfunktion. Es liegt nun nahe, das restringierte Standard-Optimierungsproblem für hinreichend großes $\eta > 0$ durch das unrestringierte Minimierungsproblem

$$\min_{x \in \mathbb{R}^n} \ell_1(x;\eta)$$

zu ersetzen. Diese Idee wird im SQP-Verfahren ausgenutzt, um eine Schrittweite α mittels eindimensionaler Liniensuche (siehe Armijo-Verfahren) für die Funktion

$$\varphi(\alpha) := \ell_1(x^{[k]} + \alpha \cdot d^{[k]};\eta)$$

durchzuführen. Wie oben erwähnt, ist die exakte ℓ_1-Bewertungsfunktion nicht differenzierbar. Allerdings ist sie immerhin noch richtungsdifferenzierbar, d.h. der Grenzwert

$$\ell_1'(x;d;\eta) := \lim_{\alpha \downarrow 0} \frac{\ell_1(x + \alpha \cdot d;\eta) - \ell_1(x;\eta)}{\alpha}$$

existiert für alle $x \in \mathbb{R}^n$ und alle Richtungen $d \in \mathbb{R}^n$, und man kann zeigen, dass ein KKT-Punkt $(d^{[k]}, \mu^{[k+1]}, \nu^{[k+1]})$ mit $d^{[k]} \neq 0$ des QP-Problems $QP(x^{[k]}, \mu^{[k]}, \nu^{[k]})$ die Abschätzung

$$\ell_1'(x^{[k]}; d^{[k]}; \eta) \leq -(d^{[k]})^T Q_k d^{[k]} < 0$$

erfüllt, falls

- Q_k symmetrisch und positiv definit ist (was bei Verwendung des modifizierten BFGS-Updates H_k erfüllt ist) und

- der Gewichtungsparameter η die Bedingung

$$\eta \geq \max\left\{\mu_1^{[k+1]}, \ldots, \mu_m^{[k+1]}, |\nu_1^{[k+1]}|, \ldots, |\nu_p^{[k+1]}|\right\} \tag{1.4.17}$$

  erfüllt. Hierin bezeichnen $\mu_i^{[k+1]}$, $i = 1, \ldots, m$, und $\nu_j^{[k+1]}$, $j = 1, \ldots, p$, die Lagrange-Multiplikatoren des QP-Problems.

Summary 1.4.3.11. *Eine Liniensuche mit dem* Armijo*-Verfahren kann durchgeführt werden, wenn die Hessematrix Q_k positiv definit ist, oder alternativ der modifizierte BFGS-Update H_k im QP verwendet wird, und wenn der Gewichtungsparameter hinreichend groß gewählt wird, was man durch iteratives Anpassen z.B. gemäß der Formel*

$$\eta_{k+1} := \max\left\{\eta_k, \max\{\mu^{[k+1]}, \ldots, \mu^{[k+1]}, |\nu^{[k+1]}|, \ldots, |\nu^{[k+1]}|\} + \epsilon\right\}, \tag{1.4.18}$$

erreichen kann ($\epsilon \geq 0$ ist ein Parameter). □

Insgesamt erhalten wir das globalisierte SQP-Verfahren:

Algorithm 1.4.3.12. (Globalisiertes SQP-Verfahren)

(i) Wähle Startwerte $(x^{[0]}, \mu^{[0]}, \nu^{[0]}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$, $H_0 \in \mathbb{R}^{n \times n}$ symmetrisch und positiv definit, $\beta \in (0,1)$, $\sigma \in (0,1)$ und setze $k = 0$.

(ii) Falls $(x^{[k]}, \mu^{[k]}, \nu^{[k]})$ ein KKT-Punkt des Standard-Optimierungsproblems ist, `STOP`.

(iii) QP-Hilfsproblem: *Berechne einen KKT-Punkt $(d^{[k]}, \mu^{[k+1]}, \nu^{[k+1]}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ des quadratischen Hilfsproblems $QP(x^{[k]}, \mu^{[k]}, \nu^{[k]})$, wobei die Hessematrix Q_k durch die modifizierte BFGS-Update-Matrix H_k ersetzt ist.*

(iv) Wähle $\eta^{[k]}$ hinreichend groß, z.B. gemäß (1.4.18).

(v) Armijo-Regel: *Bestimme eine Schrittweite $\alpha_k = \max\{\beta^j : j = 0, 1, 2, \ldots\}$ mit*

$$\ell_1(x^{[k]} + \alpha_k \cdot d^{[k]}; \eta^{[k]}) \leq \ell_1(x^{[k]}; \eta^{[k]}) + \sigma \cdot \alpha_k \cdot \ell_1'(x^{[k]}; d^{[k]}; \eta^{[k]}).$$

(vi) Modifizierter BFGS-Update: *Berechne H_{k+1} gemäß der Update-Formel (1.4.15).*

(vii) Setze $x^{[k+1]} := x^{[k]} + \alpha_k \cdot d^{[k]}$, $k := k + 1$ und gehe zu (ii).

□

Remark 1.4.3.13. • ...

- *Es gibt auch differenzierbare exakte Bewertungsfunktionen, diese sind allerdings nicht von der Gestalt in (1.4.16). Eine häufig benutzte differenzierbare exakte Bewertungsfunktion für das*

Standard-Optimierungsproblem ist die erweiterte Lagrange-Funktion

$$\begin{aligned}\mathcal{L}_a(x,\mu,\nu;\eta) &= f(x) \quad +\nu^T h(x) + \frac{\eta}{2}\cdot\|h(x)\|^2 \\ &\qquad +\frac{1}{2\eta}\cdot\sum_{i=1}^{m}\left((\max\{0,\mu_i+\eta\cdot g_i(x)\})^2-\mu_i^2\right) \\ &= f(x) \quad +\sum_{j=1}^{p}\left(\nu_j\cdot h_j(x)+2h_j(x)^2\right) \\ &\qquad +\sum_{i=1}^{m}\begin{cases}\mu_i\cdot g_i(x)+\frac{\eta}{2}\cdot g_i(x)^2, & \textit{falls } \mu_i+\eta\cdot g_i(x)\geq 0,\\ -\frac{\mu_i^2}{2\eta}, & \textit{sonst.}\end{cases}\end{aligned}$$

Ein SQP-Verfahren unter Verwendung der erweiterten Lagrange-Funktion wird in Schittkowsi [Sch81, Sch83] diskutiert.

- *In praktischen Anwendungen wird anstatt eines einzelnen Gewichtungsparameters η jeder Summand der Strafterme in der Bewertungsfunktion individuell gewichtet, etwa durch η_i, $i=1,\ldots,m$ und $\hat{\eta}_j$, $j=1,\ldots,p$.* Powell *[Pow78] schlug folgende Update-Formel vor:*

$$\begin{aligned}\eta_i^{[k+1]} &:= \max_{i=1,\ldots,m}\left\{|\mu_i^{[k+1]}|,\frac{1}{2}\left(\eta^{[k]}+|\mu_i^{[k+1]}|\right)\right\},\\ \hat{\eta}_j^{[k+1]} &:= \max_{j=1,\ldots,p}\left\{|\nu_j^{[k+1]}|,\frac{1}{2}\left(\hat{\eta}_j^{[k]}+|\nu_j^{[k+1]}|\right)\right\}\end{aligned}$$

Diese Formel hat sich in der Praxis bewährt.

- ...

□

Inkonsistentes QP Problem

...

Powell schlug vor, die Nebenbedingungen des QP-Problems zu relaxieren, so dass das relaxierte QP-Problem zulässig ist. Das ursprüngliche QP-Problem wird ersetzt durch ein relaxiertes QP-Problem.

Problem 1.4.3.14. (Relaxiertes QP-Problem $RQP(x^{[k]},\mu^{[k]},\nu^{[k]})$)

$$\begin{aligned}\frac{1}{2}d^T H_k d+\nabla f(x^{[k]})^T d+\frac{\eta}{2}\delta^2 &= \min_{\substack{d\in\mathbb{R}^n\\ \delta\in[0,1]}}\\ g_i(x^{[k]})(1-\sigma_i\delta)+\nabla g_i(x^{[k]})^T d &\leq 0, \quad i=1,\ldots,m,\\ h_j(x^{[k]})(1-\delta)+\nabla h_j(x^{[k]})^T d &= 0, \quad j=1,\ldots,p.\end{aligned}$$

□

Hierin ist

$$\sigma_i = \begin{cases} 0, & \text{falls } g_i(x^{[k]}) < 0, \\ 1, & \text{sonst,} \end{cases} \qquad i = 1, \ldots, m.$$

Der Punkt $d = 0$ und $\delta = 1$ ist stets zulässig für $RQP(x^{[k]}, \mu^{[k]}, \nu^{[k]})$. Erfüllt die optimale Lösung (d, δ) von $RQP(x^{[k]}, \mu^{[k]}, \nu^{[k]})$ die Beziehung $\delta = 0$, dann ist d auch optimal für das ursprüngliche QP-Problem $QP(x^{[k]}, \mu^{[k]}, \nu^{[k]})$. Um tatsächlich $\delta = 0$ zu erreichen, muss der Gewichtungsparameter η, der auch in der Bewertungsfunktion auftritt, hinreichend groß sein.

Quadratische Optimierung

Wir widmen uns nun einem Verfahren zur Lösung des wohl einfachsten nichtlinearen Optimierungsproblems – dem quadratischen Optimierungsproblem. Zunächst beschränken wir uns auf den Fall mit Gleichungsrestriktionen.

Problem 1.4.3.15. (Quadratisches Optimierungsproblem mit Gleichungsbeschränkungen) *Für eine symmetrische Matrix $W \in \mathbb{R}^{n \times n}$, eine Matrix $B \in \mathbb{R}^{p \times n}$ und Vektoren $c \in \mathbb{R}^n$ und $v \in \mathbb{R}^p$ minimiere*

$$f(x) := \frac{1}{2} x^T W x + c^T x$$

unter der Nebenbedingung

$$h(x) := Bx - v = 0.$$

□

Die KKT-Bedingungen in einem Minimum $\hat{x}$ mit Lagrange-Multiplikator $\nu \in \mathbb{R}^p$ lauten

$$\begin{pmatrix} W & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \hat{x} \\ \nu \end{pmatrix} = \begin{pmatrix} -c \\ v \end{pmatrix} \tag{1.4.19}$$

Beachte, dass hier wieder die sogenannte KKT-Matrix auftritt. Ist W positiv definit auf dem Kern von B und $Rang(B) = p$, so ist die Matrix invertierbar. Ist W positiv semi-definit, so ist f konvex und jede Lösung des Gleichungssystems ist zugleich globale Lösung des quadratischen Optimierungsproblems.

Im Hinblick auf ein später zu diskutierendes iteratives Verfahren formen wir (1.4.19) um und setzen dazu $\hat{x} = x^{[k]} + d$, wobei $x^{[k]}$ ein beliebiger zulässiger Punkt mit $h(x^{[k]}) = 0$ sei.

Dann ist (1.4.19) äquivalent mit

$$\begin{pmatrix} W & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} x^{[k]} + d \\ \nu \end{pmatrix} = \begin{pmatrix} -c \\ v \end{pmatrix},$$

bzw. mit

$$\begin{pmatrix} W & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} d \\ \nu \end{pmatrix} = \begin{pmatrix} -c \\ v \end{pmatrix} - \begin{pmatrix} W & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \hat{x} \\ 0 \end{pmatrix} = \begin{pmatrix} -\nabla f(x^{[k]}) \\ 0 \end{pmatrix}.$$

Diese Betrachtungen zeigen

Theorem 1.4.3.16. *Ist $x^{[k]} \in \mathbb{R}^n$ zulässig, so erfüllen $x = x^{[k]} + d$ und $\nu \in \mathbb{R}^p$ die KKT-Bedingungen des gleichungsbeschränkten quadratischen Optimierungsproblems, wenn (d, ν) das lineare Gleichungssystem*

$$\begin{pmatrix} W & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} d \\ \nu \end{pmatrix} = \begin{pmatrix} -\nabla f(x^{[k]}) \\ 0 \end{pmatrix}. \tag{1.4.20}$$

löst. □

Wir lassen nun auch Ungleichungen zu und betrachten allgemeine quadratische Optimierungsprobleme:

Problem 1.4.3.17. (Quadratisches Optimierungsproblem) *Für eine symmetrische Matrix* $W \in \mathbb{R}^{n\times n}$*, Vektoren* $c \in \mathbb{R}^n$*,* $a_i \in \mathbb{R}^n$*,* $i \in \mathcal{I}$*,* $b_j \in \mathbb{R}^n$*,* $j \in \mathcal{J}$*, und Zahlen* u_i*,* $i \in \mathcal{I}$*,* $v_j \in \mathbb{R}$*,* $j \in \mathcal{J}$*, minimiere*

$$f(x) := \frac{1}{2}x^T W x + c^T x$$

unter den Nebenbedingungen

$$\begin{aligned} g_i(x) := a_i^T x - u_i &\leq 0, \quad i \in \mathcal{I} \\ h_j(x) := b_j^T x - v_j &= 0, \quad j \in \mathcal{J}. \end{aligned}$$

□

Mit

$$A := \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix}, \quad B := \begin{pmatrix} b_1^T \\ \vdots \\ b_p^T \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix}$$

lautet das Problem in Matrixschreibweise

$$\begin{aligned} \frac{1}{2}x^T W x + c^T x &= \min_{x\in\mathbb{R}^n}! \\ Ax &\leq u \\ Bx &= v \end{aligned}$$

Auswertung der KKT-Bedingungen in einem lokalen Minimum $\hat{x}$ liefert

$$\begin{aligned} \mu_i &\geq 0, \quad i \in \mathcal{I}, \\ \mu_i \cdot g_i(\hat{x}) &= 0, \quad i \in \mathcal{I}, \\ W\hat{x} + c + \sum_{i=1}^{m} \mu_i \cdot a_i + \sum_{j=1}^{p} \nu_j \cdot b_j &= 0, \\ a_i^T \hat{x} - u_i &\leq 0, \quad i \in \mathcal{I}, \\ b_j^T \hat{x} - v_j &= 0, \quad j \in \mathcal{J}. \end{aligned}$$

Dieses System von Gleichungen und Ungleichungen lässt sich nicht so einfach lösen wie im gleichungsbeschränkten Fall. Das Problem ist darin begründet, dass die Indexmenge $\mathcal{A}(\hat{x})$ der aktiven Ungleichungsbeschränkungen unbekannt ist. Wäre sie bekannt, so könnten die inaktiven Ungleichungsbeschränkungen weggelassen werden, da sie keinen Einfluss auf das Optimum haben und man erhielte das äquivalente Problem

$$\begin{aligned} \frac{1}{2}x^T W x + c^T x &= \text{min!} \\ g_i(x) = a_i^T x - u_i &= 0, \quad i \in \mathcal{A}(\hat{x}), \\ h_j(x) = b_j^T x - v_j &= 0, \quad j \in \mathcal{J} \end{aligned} \tag{1.4.21}$$

Dieses ist ein quadratisches Optimierungsproblem mit Gleichungsbeschränkungen, dessen Lösung durch Satz 1.4.3.16 charakterisiert ist.

Die Idee der *Strategie der aktiven Mengen* zur Lösung des allgemeinen quadratischen Optimierungsproblems besteht nun darin, die unbekannte Indexmenge $\mathcal{A}(\hat{x})$ in (1.4.21) durch eine Schätzung $\mathcal{A}_s \subseteq \mathcal{I}$ zu ersetzen und diese iterativ anzupassen.

Sei $x^{[k]}$ der aktuelle Iterationspunkt, $x^{[k]}$ sei zulässig und $\mathcal{A}_s^k \subseteq \mathcal{I}$ sei die aktuelle Schätzung der aktiven Menge. Löse dann das Hilfsproblem

Problem 1.4.3.18. (Hilfsproblem)

$$\begin{aligned} f(x^{[k]}+d) &= \min_{d\in\mathbb{R}^n}! \\ a_i^T d &= 0, \quad i \in \mathcal{A}_s^k \\ b_j^T d &= 0, \quad j \in \mathcal{J}. \end{aligned}$$

□

Die Strategie zur Anpassung der Indexmenge $\mathcal{A}_s^k$ hängt nun ab von der Lösung d und den zugehörigen Lagrange-Multiplikatoren μ_i, $i \in \mathcal{A}_s^k$, und ν_j, $j \in \mathcal{J}$, des Hilfsproblems. Folgende Fälle können eintreten:

(a) Besitzt das Hilfsproblem die Lösung $d = \Theta$, so liefern die KKT-Bedingungen für das Hilfssproblem

$$\nabla f(x^{[k]}) + \sum_{i\in\mathcal{A}_s^k} \mu_i \cdot a_i + \sum_{j\in\mathcal{J}} \nu_j \cdot b_j = 0.$$

(i) Sind alle $\mu_i \geq 0$, $i \in \mathcal{A}_s^k$, so wird $x^{[k]}$ als Lösung akzeptiert, da die KKT-Bedingungen für das Ausgangsproblem erfüllt sind, wenn man noch $\mu_i = 0$, $i \in \mathcal{I} \setminus \mathcal{A}_s^k$ setzt.

(ii) *Deaktivierungsschritt:*
Gibt es einen Index $i \in \mathcal{A}_s^k$ mit $\mu_i < 0$, so erfüllt $x^{[k]}$ die KKT-Bedingungen des Ausgangsproblems nicht, ist also nicht optimal. Andererseits ist $x^{[k]}$ Minimum von f unter den Nebenbedingungen $\mathcal{A}_s^k \cup \mathcal{J}$. Daher muss der zulässige Bereich vergrößert werden, d.h. die Indexmenge $\mathcal{A}_s^k$ wird verkleinert. Bestimme dazu denjenigen Index $q \in \mathcal{A}_s^k$ mit

$$\mu_q = \min_{i\in\mathcal{A}_s^k} \mu_i < 0$$

und setze

$$\mathcal{A}_s^{k+1} := \mathcal{A}_s^k \setminus \{q\}.$$

(b) Besitzt das Hilfsproblem eine Lösung $d \neq \Theta$, so ist d eine Abstiegsrichtung von f im Punkt $x^{[k]}$, die die Nebenbedingungen $\mathcal{A}_s^k \cup \mathcal{J}$ erfüllt, d.h. es gilt

$$a_i^T d = 0,\ i \in \mathcal{A}_s^k, \quad b_j^T d = 0,\ j \in \mathcal{J}. \tag{1.4.22}$$

(i) Ist $x^{[k]} + d$ zulässig für das Ausgangsproblem, d.h. gilt

$$a_i^T(x^{[k]} + d) \leq u_i, \quad i \in \mathcal{I} \setminus \mathcal{A}_s^k,$$

so setze

$$x^{[k+1]} := x^{[k]} + d, \quad \mathcal{A}_s^{k+1} := \mathcal{A}_s^k.$$

(ii) *Aktivierungsschritt:*
Ist $x^{[k]}+d$ unzulässig für das Ausgangsproblem, so bestimme eine möglichst große Schrittweite $\alpha_k \geq 0$, so dass $x^{[k]} + \alpha_k \cdot d$ zulässig bleibt:

$$a_i^T(x^{[k]} + \alpha_k \cdot d) \leq u_i \quad \forall i \in \mathcal{I} \setminus \mathcal{A}_s^k$$

bzw.

$$\alpha_k \cdot a_i^T d \leq u_i - a_i^T x^{[k]} \quad \forall i \in \mathcal{I} \setminus \mathcal{A}_s^k \tag{1.4.23}$$

Beachte, dass $x^{[k]} + \alpha \cdot d$ für alle α zulässig bleibt für die Nebenbedingungen $\mathcal{A}_s^k \cup \mathcal{J}$, da $x^{[k]}$ zulässig ist und (1.4.22) gilt.
Da $x^{[k]}$ zulässig ist und $x^{[k]} + \alpha \cdot d$ mit $\alpha = 1$ unzulässig ist, muß es in (1.4.23) einen Index $i \in \mathcal{I} \setminus \mathcal{A}_s^k$ mit $a_i^T d > 0$ geben. Bestimme also

$$\alpha_k := \min \left\{ \frac{u_i - a_i^T x^{[k]}}{a_i^T d} \mid i \in \mathcal{I} \setminus \mathcal{A}_s^k,\ a_i^T d > 0 \right\}.$$

Sei $r \in \mathcal{I} \setminus \mathcal{A}_s^k$ ein (nicht notwendig eindeutiger) Index, für den dieses Minimum angenommen wird. Der Fall $\alpha = 0$ kann auftreten, wenn mehrere Nebenbedingungen gleichzeitig aktiv werden (Entartung).
Setze

$$x^{[k+1]} := x^{[k]} + \alpha \cdot d, \quad \mathcal{A}_s^{k+1} := \mathcal{A}_s^k \cup \{r\}.$$

Zusammenfassend erhalten wir den folgenden Algorithmus.

Algorithm 1.4.3.19. (Strategie der aktiven Menge)

(i) Sei $x^{[0]}$ zulässig für das quadratische Optimierungsproblem. Setze $k := 0$ und

$$\mathcal{A}_s^0 := \left\{ i \in \mathcal{I} \ :\ a_i^T x^{[0]} = u_i \right\}.$$

(ii) Bestimme eine Lösung $(d, \mu_{\mathcal{A}_s^k}, \nu)$ des Hilfsproblems durch Lösen des Gleichungssystems

$$\begin{pmatrix} W & A_{\mathcal{A}_s^k}^T & B^T \\ A_{\mathcal{A}_s^k} & 0 & 0 \\ B & 0 & 0 \end{pmatrix} \begin{pmatrix} d \\ \mu_{\mathcal{A}_s^k} \\ \nu \end{pmatrix} = \begin{pmatrix} -\nabla f(x^{[k]}) \\ 0 \\ 0 \end{pmatrix}$$

(vgl. (1.4.20)). Hierin sind

$$A_{\mathcal{A}_s^k} := (a_i^T)_{i \in \mathcal{A}_s^k}, \quad \mu_{\mathcal{A}_s^k} := (\mu_i)_{i \in \mathcal{A}_s^k}.$$

(iii) Ist $d = \Theta$ und $\mu_{\mathcal{A}_s^k} \geq 0$, so setze $\mu_i = 0$ für $i \in \mathcal{I} \setminus \mathcal{A}_s^k$ und `STOP`.

(iv) Ist $d = \Theta$ und $\mu_q := \min\{\mu_i \ :\ i \in \mathcal{A}_s^k\} < 0$, so setze

$$\mathcal{A}_s^{k+1} := \mathcal{A}_s^k \setminus \{q\}.$$

Setze $k := k + 1$ und gehe zu (ii).

(v) Gilt $a_i^T(x^{[k]}+d) \le u_i$, $i \in \mathcal{I} \setminus \mathcal{A}_s^k$, *so setze*

$$x^{[k+1]} := x^{[k]} + d, \quad \mathcal{A}_s^{k+1} := \mathcal{A}_s^k.$$

Setze $k := k+1$ *und gehe zu (ii).*

(vi) Bestimme $r \in \mathcal{I} \setminus \mathcal{A}_s^k$ *mit*

$$\alpha_k := \frac{u_r - a_r^T x^{[k]}}{a_r^T d} = \min_{\substack{i \in \mathcal{I}\setminus\mathcal{A}_s^k \\ a_i^T d > 0}} \left\{ \frac{u_i - a_i^T x^{[k]}}{a_i^T d} \right\}$$

und setze

$$x^{[k+1]} := x^{[k]} + \alpha_k \cdot d, \quad \mathcal{A}_s^{k+1} := \mathcal{A}_s^k \cup \{r\}.$$

Setze $k := k+1$ *und gehe zu (ii).*

□

1.4.4 Aktive-Mengen-Strategie (2)

Herleitung des Verfahrens Wir führen aktive und inaktive Mengen ein. Diese Unterscheidung betrifft die Frage, ob eine Ungleichungsnebenbedingung in einem Punkt mit Gleichheit erfüllt wird. Für die Optimallösung $\bar{u}$ von (ROS) ohne gemischte Beschränkung definieren wir

$$\begin{aligned} \Omega_+ &= \{x \in \Omega \,:\, \bar{u}(x) = u_b(x)\}, \\ \Omega_- &= \{x \in \Omega \,:\, \bar{u}(x) = u_a(x)\}, \text{ und} \\ \Omega_B &= \{x \in \Omega \,:\, u_a(x) < \bar{u}(x) < u_b(x)\}. \end{aligned}$$

Diese und weitere Definitionen sind immer nur bis auf Mengen vom Maß Null eindeutig und auch in diesem Sinne zu verstehen.

Wir gehen bei der Strategie des Algorithmus von Iterierten $(u^{[k-1]}, \lambda^{[k-1]})$ aus und werten den Ausdruck $u^{[k-1]} + \lambda^{[k-1]}$ aus. Ist dieser für einen Punkt positiv, so nehmen wir diesen in $\Omega_+^{[k]}$ auf. Ist er negativ, so nehmen wir den Punkt in $\Omega_-^{[k]}$ auf. Dementsprechend setzen wir die neuen aktiven Mengen gemäß

$$\begin{aligned} \Omega_+^{[k]} &= \left\{x \in \Omega \,:\, \left(u^{[k-1]} + \lambda^{[k-1]}\right)(x) > u_b(x)\right\}, \\ \Omega_-^{[k]} &= \left\{x \in \Omega \,:\, \left(u^{[k-1]} + \lambda^{[k-1]}\right)(x) < u_a(x)\right\}, \end{aligned}$$

fest. Die übrigen Punkte ergeben

$$\Omega_B^{[k]} = \Omega \setminus \left(\Omega_-^{[k]} \cup \Omega_+^{[k]}\right).$$

Wir erinnern, daß $\bar{\lambda} \ge 0$ auf Ω_- , $\bar{\lambda} \le 0$ auf Ω_+ und $\bar{\lambda} = 0$ auf Ω_B gilt. Die Änderungen der aktiven Mengen stellen die Schlüsselstelle des Algorithmus dar.

Zur Formulierung des konkreten Algorithmus wollen wir diese Ausdrücke noch etwas genauer analysieren. An erster Stelle steht dabei die Feststellung, dass die Iterierten $(u^{[k-1]}, \lambda^{[k-1]})$ die komplementäre Schlupfbedingung erfüllen.

Die aktiven und inaktiven Mengen stellen eine disjunkte Zerlegung des Gebietes Ω dar

$$\Omega = \Omega_+^{[k]} \dot\cup \Omega_B^{[k]} \dot\cup \Omega_-^{[k]} \quad \forall k.$$

Die Berechnung der Ausdrücke $(u^{[k-1]} + \lambda^{[k-1]})$ über das Optimalitätssystem beinhaltet insbesondere die Forderungen

$$\begin{aligned} x \in \Omega_-^{[k-1]} &\Rightarrow u^{[k-1]}(x) = u_a(x), \\ x \in \Omega_B^{[k-1]} &\Rightarrow \lambda^{[k-1]}(x) = 0, \\ x \in \Omega_+^{[k-1]} &\Rightarrow u^{[k-1]}(x) = u_b(x). \end{aligned}$$

Betrachten wir die Bestimmung der aktiven Mengen unter diesen Gesichtspunkten

$$\begin{aligned} \Omega_+^{[k]} &= \left\{x \in \Omega \,:\, \left(u^{[k-1]} + \lambda^{[k-1]}\right)(x) > u_b(x)\right\} \\ &= \left\{x \in \Omega_+^{[k-1]} \dot\cup \Omega_B^{[k-1]} \dot\cup \Omega_-^{[k-1]} \,:\, \left(u^{[k-1]} + \lambda^{[k-1]}\right)(x) > u_b(x)\right\} \\ &= \left\{x \in \Omega_+^{[k-1]} \,:\, \left(u^{[k-1]} + \lambda^{[k-1]}\right)(x) > u_b(x)\right\} \\ &\cup \left\{x \in \Omega_B^{[k-1]} \,:\, \left(u^{[k-1]} + \lambda^{[k-1]}\right)(x) > u_b(x)\right\} \\ &\cup \left\{x \in \Omega_-^{[k-1]} \,:\, \left(u^{[k-1]} + \lambda^{[k-1]}\right)(x) > u_b(x)\right\}. \end{aligned}$$

Unter der Annahme, daß von $\Omega_-^{[k-1]}$ keine Zugänge für $\Omega_+^{[k]}$ zu erwarten sind, ergibt sich

$$\begin{aligned} \Omega_+^{[k]} &= \left\{x \in \Omega_+^{[k-1]} \,:\, \lambda^{[k-1]}(x) > 0\right\} \cup \left\{x \in \Omega_B^{[k-1]} \,:\, u^{[k-1]}(x) > u_b(x)\right\} \\ &= \left(\Omega_+^{[k-1]} \setminus \left\{x \in \Omega_+^{[k-1]} \,:\, \lambda^{[k-1]}(x) \le 0\right\}\right) \cup \left\{x \in \Omega_B^{[k-1]} \,:\, u^{[k-1]}(x) > u_b(x)\right\}. \end{aligned}$$

Analog gilt:

$$\Omega_-^{[k]} = \Omega_-^{[k-1]} \setminus \left\{x \in \Omega_-^{[k-1]} \,:\, \lambda^{[k-1]}(x) \ge 0\right\} \cup \left\{x \in \Omega_B^{[k-1]} \,:\, u^{[k-1]}(x) < u_a(x)\right\}$$

und damit

$$\begin{aligned} \Omega_B^{[k]} &= \Omega \setminus \left(\Omega_+^{[k]} \cup \Omega_-^{[k]}\right) \\ &= \Omega \setminus \left(\Omega_+^{[k-1]} \cup \Omega_-^{[k-1]}\right) \\ &\quad \setminus \left(\left\{x \in \Omega_B^{[k-1]} \,:\, u^{[k-1]}(x) > u_b(x)\right\} \cup \left\{x \in \Omega_B^{[k-1]} \,:\, u^{[k-1]}(x) < u_a(x)\right\}\right) \\ &\quad \cup \left(\left\{x \in \Omega_+^{[k-1]} \,:\, \lambda^{[k-1]}(x) \le 0\right\} \cup \left\{x \in \Omega_-^{[k-1]} \,:\, \lambda^{[k-1]}(x) \ge 0\right\}\right). \end{aligned}$$

Wir bemerken, dass zur Neubestimmung der aktiven und inaktiven Mengen nicht mehr der Ausdruck $(u^{[k-1]} + \lambda^{[k-1]})$, sondern nur noch $u^{[k-1]}$ und $\lambda^{[k-1]}$ getrennt ausgewertet werden müssen. Dieses Update-Schema können wir aber auch für die gemischte Beschränkung nutzen.

Dazu sei Ω_A die Menge der Punkte aus Ω, in denen die gemischte Nebenbedingung aktiv ist und Ω_M der Rest.

$$\Omega_A^{[k]} = \Omega_A^{[k-1]} \setminus \left\{ x \in \Omega_A^{[k-1]} \ : \ \mu^{[k-1]}(x) \leq 0 \right\} \cup \left\{ x \in \Omega_M^{[k-1]} \ : \ \left(K u^{[k-1]} - c \right)(x) > 0 \right\}$$

und

$$\Omega_M^{[k]} = \Omega_M^{[k-1]} \setminus \left\{ x \in \Omega_M^{[k-1]} \ : \ \left(K u^{[k-1]} - c \right)(x) > 0 \right\} \cup \left\{ x \in \Omega_A^{[k-1]} : \mu^{[k-1]}(x) \leq 0 \right\}$$

Um diese aktiven und inaktiven Mengen anwendbar zu machen, definieren wir über die zugehörigen charakteristischen Funktionen χ_-, χ_B, χ_+ und χ_M, χ_A die Operatoren

$$X_* : v \to \chi_* \cdot v$$

Mit diesen haben die folgenden Ausdrücke den Wert Null:

$$X_B \lambda(x) = \chi_B(x) \cdot \lambda(x) = \begin{cases} 0 \ : \ x \notin \Omega_B \iff \chi_B(x) = 0 \\ 0 \ : \ x \in \Omega_B \iff \lambda(x) = 0 \end{cases}$$

$$\begin{aligned} X_-(u - u_a)(x) = \chi_-(x) \cdot (u - u_a)(x) &= \begin{cases} 0 \ : \ x \notin \Omega_- \iff \chi_-(x) = 0 \\ 0 \ : \ x \in \Omega_- \iff (u - u_a)(x) = 0 \end{cases} \\ X_+(u - u_b)(x) = \chi_+(x) \cdot (u - u_b)(x) &= \begin{cases} 0 \ : \ x \notin \Omega_+ \iff \chi_+(x) = 0 \\ 0 \ : \ x \in \Omega_+ \iff (u - u_b)(x) = 0 \end{cases} \end{aligned}$$

und

$$\begin{aligned} X_M \mu(x) = \chi_M(x) \cdot \mu(x) &= \begin{cases} 0 \ : \ x \notin \Omega_M \iff \chi_M(x) = 0 \\ 0 \ : \ x \in \Omega_M \iff \mu(x) = 0 \end{cases} \\ X_A(Ku - c)(x) = \chi_A(x) \cdot (Ku - c)(x) &= \begin{cases} 0 \ : \ x \notin \Omega_A \iff \chi_A(x) = 0 \\ 0 \ : \ x \in \Omega_A \iff (Ku - c)(x) = 0 \end{cases} \end{aligned}$$

Damit formen wir jetzt das System (ROS) in ein Gleichungssystem um

$$\begin{aligned} \Theta &= Qu + K^* X_A \mu + (X_- + X_+)\lambda + q, \\ \Theta &= X_A(Ku - c) + X_M \mu, \\ \Theta &= X_-(u - u_a) + X_+(u - u_b) + X_B \lambda, \end{aligned}$$

in anderer Form

$$\underbrace{\begin{pmatrix} Q & K^* X_A & X_- + X_+ \\ X_A K & X_M & 0 \\ X_- + X_+ & 0 & X_B \end{pmatrix}}_{B} \underbrace{\begin{pmatrix} u \\ \mu \\ \lambda \end{pmatrix}}_{x} + \underbrace{\begin{pmatrix} q \\ -X_A c \\ -(X_- u_a + X_+ u_b) \end{pmatrix}}_{b} = \Theta.$$

Aktive-Mengen-Methode

1. Wähle als Startwerte $u^{[0]} = \mu^{[0]} = \lambda^{[0]} = \Theta$, und für die Zerlegungsmengen

$$\Omega_B^{[0]} = \Omega_M^{[0]} = \Omega, \quad \Omega_-^{[0]} = \Omega_+^{[0]} = \Omega_A^{[0]} = \Theta,$$
$$\Omega_B^{[-1]} = \Omega_M^{[-1]} = \Theta, \quad \Omega_-^{[-1]} = \Omega_+^{[-1]} = \Omega_A^{[-1]} = \Omega,$$

   und setze $k = 0$.

2. Wenn $\Omega_B^{[k-1]} = \Omega_B^{[k]}$ und $\Omega_M^{[k-1]} = \Omega_M^{[k]}$: `STOP`.

3. - Bestimme die Einträge in $B^{[k]}$ und $b^{[k]}$.
   - Bestimme Lösung $(u^{[k+1]}, \mu^{[k+1]}, \lambda^{[k+1]})$ von $B^{[k]}(u, \mu, \lambda) + b^{[k]} = \Theta$.
   - Berechne das Zerlegungs-Update für die Boxbeschränkung(en)

$$\begin{aligned}
d_- &= \left\{ x \in \Omega_B^{[k]} \ : \ u^{[k+1]}(x) \leq u_a(x) \right\}, \\
d_B^- &= \left\{ x \in \Omega_-^{[k]} \ : \ \lambda^{[k+1]}(x) \geq 0 \right\}, \\
d_B^+ &= \left\{ x \in \Omega_+^{[k]} \ : \ \lambda^{[k+1]}(x) \leq 0 \right\}, \\
d_+ &= \left\{ x \in \Omega_B^{[k]} \ : \ u^{[k+1]}(x) \geq u_b(x) \right\},
\end{aligned}$$

$$\begin{aligned}
\Omega_-^{[k+1]} &= \Omega_-^{[k]} \setminus d_B^- \cup d_-, \\
\Omega_B^{[k+1]} &= \Omega_B^{[k]} \setminus (d_- \cup d_+) \cup (d_B^- \cup d_B^+), \\
\Omega_+^{[k+1]} &= \Omega_+^{[k]} \setminus d_B^+ \cup d_+,
\end{aligned}$$

   und für die Mixbeschränkung(en)

$$\begin{aligned}
d_I &= \left\{ x \in \Omega_A^{[k]} \ : \ \mu^{[k+1]}(x) \leq 0 \right\}, \\
d_A &= \left\{ x \in \Omega_M^{[k]} \ : \ \left(K u^{[k+1]} - c \right)(x) \geq 0 \right\},
\end{aligned}$$

$$\begin{aligned}
\Omega_M^{[k+1]} &= \Omega_M^{[k]} \setminus d_A \cup d_I, \\
\Omega_A^{[k+1]} &= \Omega_A^{[k]} \setminus d_I \cup d_A.
\end{aligned}$$

4. Setze $k = k + 1$ und gehe zu Schritt 2. □

1.4.5 Active-Set-Methoden (3)

Active-Set-Methoden sind eine Klasse iterativer Algorithmen zur Lösung von quadratischen Optimierungsproblemen.[2]

[2] https://de.wikipedia.org/wiki/Active-Set-Methoden

Mathematische Problemstellung

Jedes quadratische Programm kann in eine standardisierte Form überführt werden:

$$\begin{array}{rll} \min_{x\in\mathbb{R}^n} & \frac{1}{2}x^T Hx + c^T x & = f(x) \\ \text{s.t.} & a_i^T x \geq u_i & \forall i \in \mathcal{I} \\ & b_j^T x = v_j & \forall j \in \mathcal{J} \end{array}$$

wobei n die Anzahl der Entscheidungsvariablen ist. In der Ziel(=Kosten-)funktion $f(x)$ entspricht H der Hesse-Matrix, die Mengen $\mathcal{I}$ und $\mathcal{J}$ indizieren die Ungleichheits- und Gleichheitsbedingungen. Oft wird dabei gefordert, dass die Matrix H positiv semidefinit ist, da dann das Optimierungsproblem konvex ist.

Active Set Eine Nebenbedingung $i \in \mathcal{I}$ ist *aktiv* an einem Punkt x, wenn $a_i^T x = u_i$ gilt.

Das Active Set $\mathcal{A}(x)$ ist die Menge aller aktiven Bedingungen an einem gültigen Punkt x:

$$\mathcal{A}(x) := \left\{ i \in \mathcal{I} \ : \ a_i^T x = u_i \right\} \cup \{ j \in \mathcal{J} \}$$

Algorithmus

Active-Set-Methoden setzen eine initiale gültige Lösung $x^{[0]}$ voraus. Die Algorithmen berechnen dann in jeder Iteration einen gültigen Punkt $x^{[k]}$, bis ein Optimum erreicht ist. Dabei wird eine Menge $\mathcal{A}_k$ verwaltet, die angibt, welche Nebenbedingungen in der aktuellen Iteration aktiv sein sollen.

```
1  INPUT: gültiger Punkt x^[0], A_0 ⊆ A(x^[0])
2  for (k = 0, 1, ...)
3    Berechne eine Suchrichtung d^[k]
4    if (d^[k] == 0)
5      Berechne Lagrange-Multiplikatoren μ_i
6      if (∀i : μ_i ≥ 0)
7        STOP und OUTPUT: x^[k]
8      else
9        Finde Ungleichheitsbedingung i ∈ A_k ∩ I mit μ_i < 0
10         A_{k+1} = A_k \ {i}
11     endif
12   else (d^[k] ≠ 0)
13     Berechne Schrittlänge α_k
14     if (α_k < 1)
```

15 Finde Nebenbedingung i die α_k beschränkt

16 $\mathcal{A}_{k+1} = \mathcal{A}_k \cup \{i\}$

17 endif

18 $x^{[k+1]} = x^{[k]} + \alpha_k \cdot d^{[k]}$

19 endif

20 endfor

Berechnung der Suchrichtung $d^{[k]}$ Die Nebenbedingungen in $\mathcal{A}_k$ definieren einen Unterraum. Wenn $\hat{x}$ in der optimalen Lösung der Zielfunktion in diesem Unterraum ist, kann man die Suchrichtung als $d^{[k]} = \hat{x} - x^{[k]}$ definieren. Substituiert man dies in die Zielfunktion, erhält man die Suchrichtung $d^{[k]}$ durch Lösen eines quadratischen Subproblems:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2}(d^{[k]})^T H d^{[k]} + g_k^T d_k \\ \text{s.t.} \quad & A^T d^{[k]} = 0 \qquad \forall i \in \mathcal{A}_k \end{aligned}$$

wobei $g_k = Hx^{[k]} + c$ der Gradient an der aktuellen Lösung ist und die Spalten der Matrix A die Vektoren $a_i,\ i \in \mathcal{A}_k$ sind.

Berechnung der Lagrange-Multiplikatoren μ_i Falls die Suchrichtung $d^{[k]} = 0$ ist, ist $x^{[k]}$ bereits optimal im aktuellen Unterraum. Man muss dann eine geeignete Ungleichheitsbedingung aus $\mathcal{A}_k$ entfernen. Die Lagrange-Multiplikatoren μ_i erhält man durch Lösen eines linearen Gleichungssystems:

$$\sum_{i \in \mathcal{A}_k \cap \mathcal{I}} a_i \mu_i = g^{[k]} = Hx^{[k]} + c$$

Falls alle $\mu_i \geq 0$ sind, erfüllen $x^{[k]}$ und μ die Karush-Kuhn-Tucker-Bedingungen, welche notwendige Kriterien für die Optimalität sind. Wenn zudem die Hesse-Matrix H positiv semi-definit ist, sind diese Bedingungen hinreichend und $x^{[k]}$ ist die optimale Lösung des Problems. Entfernt man eine Ungleichheitsbedingung mit negativem Lagrange-Multiplikator aus $\mathcal{A}_k$ erhält man in der nächsten Iteration eine Suchrichtung.

Berechnung der Schrittlänge α_k Hat man eine Suchrichtung $d^{[k]}$, muss man die maximale Schrittlänge α_k berechnen. Eine volle Schrittlänge mit $\alpha_k = 1$ führt direkt zum Minimum im durch $\mathcal{A}_k$ definierten Unterraum. Die Schrittlänge ist jedoch häufig durch eine Nebenbedingung $i \notin \mathcal{A}_k$ beschränkt.

Alle Nebenbedingungen in $i \notin \mathcal{A}_k$ mit $a_i^T d^{[k]} \geq 0$ sind auch am Punkt $x^{[k]} + \alpha_k \cdot d^{[k]}$ für alle $\alpha_k \geq 0$ erfüllt, da dann die Ungleichung

$$a_i^T (x^{[k]} + \alpha_k \cdot d^{[k]}) = a_i^T x^{[k]} + \alpha_k \cdot a_i^T d^{[k]} \geq a_i^T x^{[k]} \geq u_i$$

gilt. Alle Nebenbedingungen $i \notin \mathcal{A}_k$ mit $a_i^T d^{[k]} < 0$ werden am neuen Punkt nur dann eingehalten, wenn für diese Nebenbedingungen die Ungleichung

$$a_i^T x^{[k]} + \alpha_k \cdot a_i^T d^{[k]} \geq u_i$$

gilt. Dies ist äquivalent mit der Bedingung

$$\alpha_k \leq \frac{u_i - a_i^T x^{[k]}}{a_i^T d^{[k]}} \quad \forall i \notin \mathcal{A}_k \ : \ a_i^T d^{[k]} < 0$$

Um so nah wie möglich an das Optimum im aktuellen Unterraum zu kommen, kann man die maximale Schrittlänge durch diese Formel berechnen:

$$\alpha_k = \min \left\{ 1, \min_{i \notin \mathcal{A}_k,\ a_i^T d^{[k]} < 0} \frac{u_i - a_i^T x^{[k]}}{a_i^T d^{[k]}} \right\}$$

Die Nebenbedingung, die diese Länge beschränkt, wird in die Menge $\mathcal{A}_{k+1}$ aufgenommen, da diese Nebenbedingung nun aktiv ist.

1.4.6 Verfahren der konjugierten Gradienten

Das CG-Verfahren (von engl. conjugate gradients) ist eine effiziente numerische Methode zur Lösung von großen linearen Gleichungssystemen der Form $Ax = b$ mit symmetrischer, positiv-definiter Systemmatrix A.

Das Verfahren liefert, in exakter Arithmetik, nach spätestens m Schritten die exakte Lösung, wobei m die Größe der quadratischen Matrix $A \in \mathbb{R}^{m \times m}$ ist. Insbesondere ist es aber als iteratives Verfahren interessant, da der Fehler monoton fällt. Das CG-Verfahren kann in die Klasse der Krylow-Unterraum-Verfahren eingeordnet werden.

Algorithm 1.4.6.1. (CG-Verfahren) *zur Berechnung einer Lösung von $Ax = b$*

(i) Wähle Startwert $x^{[0]}$, berechne $r^{[0]} = b - Ax^{[0]}$, setze $d^{[0]} = r^{[0]}$, $\ell := 0$.

(ii) Ist $r^{[\ell]} = 0$ (bzw. $\|r^{[\ell]}\|_A < tol$): `STOP`.

(iii) Berechne

- *temp. Zwischenwert*

$$z^{[\ell]} \quad = \quad Ad^{[\ell]},$$

- *Finde von $x^{[\ell]}$ in Richtung $d^{[\ell]}$ den Ort $x^{[\ell+1]}$ des Minimums der Funktion $E(x) := \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$ und aktualisiere den Gradienten bzw. das Residuum*

$$\begin{aligned} \alpha_\ell &= \frac{(r^{[\ell]})^T r^{[\ell]}}{(d^{[\ell]})^T z^{[\ell]}}, \\ x^{[\ell+1]} &= x^{[\ell]} + \alpha_\ell \cdot d^{[\ell]}, \\ r^{[\ell+1]} &= r^{[\ell]} - \alpha_\ell \cdot z^{[\ell]}, \quad \left(= b - Ax^{[\ell+1]}\right) \end{aligned}$$

- *Korrigiere die Suchrichtung $d^{[\ell+1]}$ mit Hilfe von $d^{[\ell]}$ und $r^{[\ell+1]}$:*

$$\begin{aligned}
\beta_\ell &= \frac{(r^{[\ell+1]})^T r^{[\ell+1]}}{(r^{[\ell]})^T r^{[\ell]}} \quad \text{(Fletcher-Reeves)}, \\
\beta_\ell &= \frac{(r^{[\ell+1]})^T (r^{[\ell+1]} - r^{[\ell]})}{(r^{\ell})^T r^{[\ell]}} \quad \text{(\textit{alternative} Polak-Ribière)}, \\
\beta_\ell &= \frac{(r^{[\ell+1]})^T (r^{[\ell+1]} - r^{[\ell]})}{(d^{[\ell]})^T (r^{[\ell+1]} - r^{[\ell]})} \quad \text{(\textit{alternative} Hestenes-Stiefel)}, \\
d^{[\ell+1]} &= r^{[\ell+1]} + \beta_\ell \cdot d^{[\ell]},
\end{aligned}$$

(iv) Setze $\ell := \ell + 1$ und gehe zu (ii)

□

1.4.7 Methode der kleinsten Quadrate

Die Methode der kleinsten Quadrate (kurz: MKQ) oder KQ-Methode (englisch: method of least squares oder lediglich least squares, kurz: LS); zur Abgrenzung von daraus abgeleiteten Erweiterungen wie z. B. der *verallgemeinerten Methode der kleinsten Quadrate* oder der *zweistufigen Methode der kleinsten Quadrate* auch mit dem Zusatz 'gewöhnliche' bezeichnet, d. h. gewöhnliche Methode der kleinsten Quadrate (englisch: ordinary least squares, kurz: OLS; veraltet Methode der kleinsten Abweichungsquadratsumme) ist das mathematische Standardverfahren zur Ausgleichungsrechnung.

Dabei wird zu einer Menge von Datenpunkten eine Funktion bestimmt, die möglichst nahe an den Datenpunkten verläuft und somit die Daten bestmöglich zusammenfasst. Die am häufigsten verwendete Funktion ist die Gerade, die dann Ausgleichsgerade genannt wird. Um die Methode anwenden zu können, muss die Funktion mindestens einen Parameter enthalten. Diese Parameter werden dann durch die Methode bestimmt, so dass, wenn die Funktion mit den Datenpunkten verglichen und der Abstand zwischen Funktionswert und Datenpunkt quadriert wird, die Summe dieser quadrierten Abstände möglichst gering wird. Die Abstände werden dann Residuen genannt.

Typischerweise werden mit dieser Methode reale Daten, etwa physikalische oder wirtschaftliche Messwerte, untersucht. Diese Daten beinhalten oft unvermeidbare Messfehler und Schwankungen. Unter der Annahme, dass die gemessenen Werte nahe an den zugrunde liegenden „wahren Werten“ liegen und zwischen den Messwerten ein bestimmter Zusammenhang besteht, kann die Methode verwendet werden, um eine Funktion zu finden, die diesen Zusammenhang der Daten möglichst gut beschreibt. Die Methode kann auch umgekehrt verwendet werden, um verschiedene Funktionen zu testen und dadurch einen unbekannten Zusammenhang in den Daten zu beschreiben.

Messpunkte und deren Abstand von einer nach der Methode der kleinsten Quadrate bestimmten Funktion. Hier wurde eine logistische Funktion als Modellkurve gewählt.

In der Stochastik wird die Methode der kleinsten Quadrate meistens als regressionsanalytische Schätzmethode benutzt, wo sie auch als Kleinste-Quadrate-Schätzung bzw. gewöhnliche Kleinste-Quadrate-Schätzung bezeichnet wird. Da die Kleinste-Quadrate-Schätzung die Residuenquadratsumme minimiert, ist es dasjenige Schätzverfahren, welches das Bestimmtheitsmaß maximiert. Angewandt als Systemidentifikation ist die Methode der kleinsten Quadrate in Verbindung mit Modellversuchen z. B. für Ingenieure ein Ausweg aus der paradoxen Situation, Modellparameter für unbekannte Gesetzmäßigkeiten zu bestimmen.

Das Verfahren

Voraussetzungen Man betrachtet eine abhängige Größe y, die von einer Variablen x oder auch von mehreren Variablen beeinflusst wird. So hängt die Dehnung einer Feder nur von der aufgebrachten Kraft ab, die Profitabilität eines Unternehmens jedoch von mehreren Faktoren wie Umsatz, den verschiedenen Kosten oder dem Eigenkapital. Zur Vereinfachung der Notation wird im Folgenden die Darstellung auf eine Variable x beschränkt. Der Zusammenhang zwischen y und den Variablen wird über eine Modellfunktion f, beispielsweise eine Parabel oder eine Exponentialfunktion

$$y(x) = f(x; \alpha_1, \ldots, \alpha_m),$$

die von x sowie von m Funktionsparametern α_j abhängt, modelliert. Diese Funktion entstammt entweder der Kenntnis des Anwenders oder einer mehr oder weniger aufwendigen Suche nach einem Modell, eventuell müssen dazu verschiedene Modellfunktionen angesetzt und die Ergebnisse verglichen werden. Ein einfacher Fall auf Basis bereits vorhandener Kenntnis ist beispielsweise die Feder, denn hier ist das Hookesche Gesetz und damit eine lineare Funktion mit der Federkonstanten als einzigem Parameter Modellvoraussetzung. In schwierigeren Fällen wie dem des Unternehmens muss der Wahl des Funktionstyps jedoch ein komplexer Modellierungsprozess vorausgehen.

Um Informationen über die Parameter und damit die konkrete Art des Zusammenhangs zu erhalten, werden zu jeweils n gegebenen Werten x_i der unabhängigen Variablen x entsprechende Beobachtungswerte y_i $(i = 1, \ldots, n)$ erhoben. Die Parameter α_j dienen zur Anpassung des gewählten Funktionstyps an diese beobachteten Werte y_i. Ziel ist es nun, die Parameter α_j so zu wählen, dass die Modellfunktion die Daten bestmöglich approximiert.

Gauss und Legendre hatten die Idee, Verteilungsannahmen über die Messfehler dieser Beobachtungswerte zu machen. Sie sollten im Durchschnitt Null sein, eine gleichbleibende Varianz haben und von jedem anderen Messfehler stochastisch unabhängig sein. Man verlangt damit, dass in den Messfehlern keinerlei systematische Information mehr steckt, sie also rein zufällig um Null schwanken. Außerdem sollten die Messfehler normalverteilt sein, was zum einen wahrscheinlichkeitstheoretische Vorteile hat und zum anderen garantiert, dass Ausreißer in y so gut wie ausgeschlossen sind.

Um unter diesen Annahmen die Parameter α_j zu bestimmen, ist es im Allgemeinen notwendig, dass deutlich mehr Datenpunkte als Parameter vorliegen, es muss also $n > m$ gelten.

Minimierung der Summe der Fehlerquadrate Das Kriterium zur Bestimmung der Approximation sollte so gewählt werden, dass große Abweichungen der Modellfunktion von den Daten stärker gewichtet werden als kleine. Sofern keine Lösung ganz ohne Abweichungen möglich ist, dann ist der Kompromiss mit der insgesamt geringsten Abweichung das beste allgemein gültige Kriterium.

Dazu wird die Summe der Fehlerquadrate, die auch Fehlerquadratsumme (genauer: Residuenquadratsumme) heißt, als die Summe der quadrierten Differenzen zwischen den Werten der Modellkurve $f(x_i)$ und den Daten y_i definiert.

In Formelschreibweise mit den Parametern $\vec{\alpha} = (\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m$ und

$$\vec{f} = (f(x_1, \vec{\alpha}), \ldots, f(x_n, \vec{\alpha})) \in \mathbb{R}^n$$

ergibt sich

$$\sum_{i=1}^{n} \left(f(x_i, \vec{\alpha}) - y_i\right)^2 = \|\vec{f} - \vec{y}\|_2^2.$$

Es sollen dann diejenigen Parameter α_j ausgewählt werden, bei denen die Summe der quadrierten Anpassungsfehler minimal wird:

$$\min_{\vec{\alpha}} \|\vec{f} - \vec{y}\|_2^2.$$

Wie genau dieses Minimierungsproblem gelöst wird, hängt von der Art der Modellfunktion ab.

Zusammenhang mit dem zentralen Grenzwertsatz Selbst wenn die Fehlerterme nicht normalverteilt sind, folgt aus dem zentralen Grenzwertsatz oft, dass der Schätzer der bedingten Erwartung $f(x, \alpha) = \hat{E}[Y|x]$ approximativ normalverteilt ist, solange die Stichprobe hinreichend groß ist. Aus diesem Grund ist die Verteilung des Fehlerterms bei großen Stichprobenumfängen oft kein gravierendes Problem in der Regressionsanalyse. Speziell ist es häufig nicht wichtig, ob der Fehlerterm einer Normalverteilung folgt, es sei denn es liegen beispielsweise folgende Punkte vor:

- die Stichprobengröße ist klein
- die Verteilung der Fehler ist eine Heavy-tailed-Verteilung, welche zur Erzeugung von Daten führt, welche weit weg von den anderen Daten liegen (Stichproben aus den Heavy tails werden dann oft als Ausreißer interpretiert)
- Multimodale Fehlerverteilungen
- große Schiefe der Fehlerverteilung

Lineare Modellfunktion

Lineare Modellfunktionen sind Linearkombinationen aus beliebigen, im Allgemeinen nicht-linearen Basisfunktionen. Für solche Modellfunktionen lässt sich das Minimierungsproblem auch analytisch über einen Extremwertansatz ohne iterative Annäherungsschritte lösen. Zunächst werden einige einfache Spezialfälle und Beispiele gezeigt.

Spezialfall einer einfachen linearen Ausgleichsgeraden

Herleitung und Verfahren Eine einfache Modellfunktion mit zwei linearen Parametern stellt das Polynom erster Ordnung

$$f(x) = \alpha_0 + \alpha_1 \cdot x$$

dar. Gesucht werden zu n gegebenen Messwerten $(x_1, y_1), \ldots, (x_n, y_n)$ die Koeffizienten α_0 und α_1 der bestangepassten Geraden. Die Abweichungen r_i zwischen der gesuchten Geraden und den jeweiligen Messwerten

$$\begin{aligned} r_1 &= \alpha_0 + \alpha_1 \cdot x_1 - y_1 \\ r_2 &= \alpha_0 + \alpha_1 \cdot x_2 - y_2 \\ &\vdots \\ r_n &= \alpha_0 + \alpha_1 \cdot x_n - y_n \end{aligned}$$

nennt man Anpassungsfehler oder Residuen. Gesucht sind nun die Koeffizienten α_0 und α_1 mit der kleinsten Summe der Fehlerquadrate

$$\min_{\alpha_0,\, \alpha_1} \sum_{i=1}^{n} r_i^2.$$

Der große Vorteil des Ansatzes mit diesem Quadrat der Fehler wird sichtbar, wenn man diese Minimierung mathematisch durchführt: Die Summenfunktion wird als Funktion der beiden Variablen α_0 und α_1 aufgefasst (die eingehenden Messwerte sind dabei numerische Konstanten), dann die Ableitung (genauer: partielle Ableitungen) der Funktion nach diesen Variablen (also α_0 und α_1) gebildet und von dieser Ableitung schließlich die Nullstelle gesucht. Es ergibt sich das lineare Gleichungssystem

$$\begin{aligned} n \cdot \alpha_0 + \left(\sum_{i=1}^{n} x_i\right) \alpha_1 &= \sum_{i=1}^{n} y_i \\ \left(\sum_{i=1}^{n} x_i\right) \alpha_0 + \left(\sum_{i=1}^{n} x_i^2\right) \alpha_1 &= \sum_{i=1}^{n} x_i y_i \end{aligned}$$

mit der Lösung

$$\alpha_1 = \frac{\sum_{i=1}^{n} x_i(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{SP_{xy}}{SQ_x} \text{ und } \alpha_0 = \overline{y} - \alpha_1 c\overline{x},$$

wobei SP_{xy} die Summe der Abweichungsprodukte zwischen x und y darstellt, und SQ_x die Summe der Abweichungsquadrate von x darstellt. Dabei ist $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ das arithmetische Mittel der x-Werte, $\overline{y}$ entsprechend. Die Lösung für α_1 kann mit Hilfe des Verschiebungssatzes auch in nichtzentrierter Form

$$\alpha_1 = \frac{\sum_{i=1}^{n}(x_i \cdot y_i) - n \cdot \overline{x} \cdot \overline{y}}{(\sum_{i=1}^{n} x_i^2) - n \cdot \overline{x}^2}$$

angegeben werden. Diese Ergebnisse können auch mit Funktionen einer reellen Variablen, also ohne partielle Ableitungen, hergeleitet werden.

Aus der Lösung von α_0 wird zudem eine Eigenschaft der linearen Ausgleichsgerade ersichtlich: Die Ausgleichsgerade verläuft stets durch den Punkt $(\overline{x}, \overline{y})$. Das ist hilfreich, falls die Ausgleichsgerade sehr steil oder gar senkrecht verläuft und der Achsenabschnitt dadurch sehr groß wird oder gar nicht berechnet werden kann. In diesem Fall kann dieser Punkt als Stützpunkt einer Vektordarstellung der Ausgleichsgerade verwendet werden.

Spezialfall einer linearen Ausgleichsfunktion mit mehreren Variablen Ist die Modellfunktion ein mehrdimensionales Polynom erster Ordnung, besitzt also statt nur einer Variablen x mehrere unabhängige Modellvariablen $x_1, \ldots, x_N$, erhält man eine lineare Funktion der Form

$$f(x_1, \ldots, x_N; \alpha_0, \alpha_1, \ldots, \alpha_N) = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_N x_N,$$

die auf die Residuen

$$\begin{array}{rccccccc} r_1 = & \alpha_0 + \alpha_1 x_{1,1} + & \cdots & + \alpha_j x_{j,1} + & \cdots & + \alpha_N x_{N,1} - y_1 \\ r_2 = & \alpha_0 + \alpha_1 x_{1,2} + & \cdots & + \alpha_j x_{j,2} + & \cdots & + \alpha_N x_{N,2} - y_2 \\ \vdots & \vdots & & \vdots & & \vdots \\ r_i = & \alpha_0 + \alpha_1 x_{1,i} + & \cdots & + \alpha_j x_{j,i} + & \cdots & + \alpha_N x_{N,i} - y_i \\ \vdots & \vdots & & \vdots & & \vdots \\ r_n = & \alpha_0 + \alpha_1 x_{1,n} + & \cdots & + \alpha_j x_{j,n} + & \cdots & + \alpha_N x_{N,n} - y_n \end{array}$$

führt und über den Minimierungsansatz

$$\min_{\alpha} \sum_{i=1}^{n} r_i^2$$

gelöst werden kann.

Der allgemeine lineare Fall Im Folgenden soll der allgemeine Fall von beliebigen linearen Modellfunktionen mit beliebiger Dimension gezeigt werden. Zu einer gegebenen Messwertfunktion

$$y(x_1, x_2, \ldots, x_N)$$

mit N unabhängigen Variablen sei eine optimal angepasste lineare Modellfunktion

$$f(x_1, \ldots, x_N; \alpha_1, \ldots, \alpha_m) = \sum_{j=1}^{m} \alpha_j \varphi_j(x_1, \ldots, x_N)$$

gesucht, deren quadratische Abweichung dazu minimal sein soll. x_i sind dabei die Funktionskoordinaten, α_j die zu bestimmenden linear eingehenden Parameter und φ_j beliebige zur Anpassung an das Problem gewählte linear unabhängige Funktionen.

Bei n gegebenen Messpunkten

$$(x_{1,1}, x_{2,1}, \ldots, x_{N,1}; y_1), (x_{1,2}, x_{2,2}, \ldots, x_{N,2}; y_2), \ldots, (x_{1,n}, x_{2,n}, \ldots, x_{N,n}; y_n)$$

erhält man die Anpassungsfehler

$$\begin{array}{rcccl}
r_1 = & \alpha_1\varphi_1(x_{1,1}, \ldots, x_{N,1}) & + & \cdots & + \alpha_m\varphi_m(x_{1,1}, \ldots, x_{N,1}) - y_1 \\
\vdots & \vdots & & & \vdots \\
r_i = & \alpha_1\varphi_1(x_{1,i}, \ldots, x_{N,i}) & + & \cdots & + \alpha_m\varphi_m(x_{1,i}, \ldots, x_{N,i}) - y_i \\
\vdots & \vdots & & & \vdots \\
r_n = & \alpha_1\varphi_1(x_{1,n}, \ldots, x_{N,n}) & + & \cdots & + \alpha_m\varphi_m(x_{1,n}, \ldots, x_{N,n}) - y_n
\end{array}$$

oder in Matrixschreibweise

$$r = A\alpha - y,$$

wobei der Vektor $r \in \mathbb{R}^n$ die r_i zusammenfasst, die Matrix $A \in \mathbb{R}^{n \times m}$ die Basisfunktionswerte $A_{ij} := \varphi_j(x_{1,i}, \ldots, x_{N,i})$, der Parametervektor $\alpha \in \mathbb{R}^m$ die Parameter α_j und der Vektor $y \in \mathbb{R}^n$ die Beobachtungen y_i, wo $n \geq m$.

Der beste Schätzer wird durch die Lösung des Minimierungsproblems bestimmt. Das Minimierungsproblem, das sich mithilfe der euklidischen Norm durch

$$\min_{\alpha} \sum_{i=1}^{n} r_i^2 = \min_{\alpha} \|f(\alpha) - y\|_2^2 = \min_{\alpha} \|A\alpha - y\|_2^2$$

formulieren lässt, kann im regulären Fall (d. h. A hat vollen Spaltenrang, somit ist A^TA regulär und damit invertierbar) mit der Formel

$$\hat{\alpha} = (A^TA)^{-1}A^Ty$$

eindeutig analytisch gelöst werden (siehe nächster Abschnitt). Im generalisierten Fall der gewichteten kleinsten Quadrate muss zudem noch die inverse Kovarianzmatrix V^{-1} berücksichtigt werden

$$\hat{\alpha} = (A^T V^{-1} A)^{-1} A^T V^{-1} y.$$

Im singulären Fall, wenn A nicht von vollem Rang ist, ist das Normalgleichungssystem nicht eindeutig lösbar, d.h. der Parameter α nicht identifizierbar.

Jedoch ist in vielen praktischen Anwendungen die Modellfunktionen $y(x_1, x_2, \ldots, x_N)$ nicht analytisch bekannt, sondern kann nur für verschiedene diskrete Werte $(x_1, x_2, \ldots, x_N)$ bestimmt werden. In diesem Fall kann die Modellfunktion mithilfe einer linearen Regression näherungsweise bestimmt werden, und der beste Schätzer wird direkt mit der Gleichung des linearen Template Fits bestimmt:

$$\hat{\alpha} = \left((Y\tilde{M})^\intercal V^{-1} Y\tilde{M}\right)^{-1} (Y\tilde{M})^\intercal V^{-1}(d - Y\bar{m}).$$

Dabei ist Y die Matrix mit den bekannten Werten der Modellfunktion (Template Matrix) für alle x, und der Vektor d bezeichnet die Zufallsvariablen (bspw. eine Messung). Die Matrix $\tilde{M}$ und der Vektor $\tilde{m}$ werden mithilfe der Stützstellen x (zusammengefasst in der Matrix Y) berechnet.

Lösung des Minimierungsproblems

Herleitung und Verfahren Das Minimierungsproblem ergibt sich, wie im allgemeinen linearen Fall gezeigt, als

$$\min_\alpha \|A\alpha - y\|_2^2 = \min_\alpha (A\alpha - y)^T (A\alpha - y) = \min_\alpha (\alpha^T A^T A\alpha - 2y^T A\alpha + y^T y).$$

Dieses Problem ist immer lösbar. Hat die Matrix A vollen Rang, so ist die Lösung sogar eindeutig. Zum Bestimmen des extremalen Punktes ergibt Nullsetzen der partiellen Ableitungen bezüglich der α_j,

$$\nabla \|A\alpha - y\|_2^2 = 2(A\alpha - y)^T A,$$

ein lineares System von Normalgleichungen (auch Gaußsche Normalgleichungen oder Normalengleichungen)

$$A^T A\alpha = A^T y,$$

welches die Lösung des Minimierungsproblems liefert und im Allgemeinen numerisch gelöst werden muss. Hat A vollen Rang und ist $n \geq m$, so ist die Matrix $A^T A$ positiv definit, so dass es sich beim gefundenen Extremum in der Tat um ein Minimum handelt. Damit kann das Lösen des Minimierungsproblems auf das Lösen eines Gleichungssystems reduziert werden. Im einfachen Fall einer Ausgleichsgeraden kann dessen Lösung, wie gezeigt wurde, sogar direkt als einfache Formel angegeben werden.

Alternativ lassen sich die Normalgleichungen in der Darstellung

$$A^T A\alpha - A^T y = \begin{pmatrix} \langle\varphi_1, \varphi_1\rangle & \langle\varphi_1, \varphi_2\rangle & \cdots & \langle\varphi_1, \varphi_m\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle\varphi_m, \varphi_1\rangle & \langle\varphi_m, \varphi_2\rangle & \cdots & \langle\varphi_m, \varphi_m\rangle \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix} - \begin{pmatrix} \langle y, \varphi_1\rangle \\ \vdots \\ \langle y, \varphi_m\rangle \end{pmatrix} = 0.$$

ausschreiben, wobei $\langle\cdot,\cdot\rangle$ das Standardskalarprodukt symbolisiert und auch als Integral des Überlappens der Basisfunktionen verstanden werden kann. Die Basisfunktionen φ_i sind als Vektoren

$\vec{\varphi_i} = (\varphi_i(x_{1,1}, \ldots, x_{N,1}), \varphi_i(x_{1,2}, \ldots, x_{N,2}), \ldots, \varphi_i(x_{1,n}, \ldots, x_{N,n}))$ zu lesen mit den n diskreten Stützstellen am Ort der Beobachtungen $y = \vec{y} = (y_1, y_2, \ldots, y_n)$.

Ferner lässt sich das Minimierungsproblem mit einer Singulärwertzerlegung gut analysieren. Diese motivierte auch den Ausdruck der Pseudoinversen, einer Verallgemeinerung der normalen Inversen einer Matrix. Diese liefert dann eine Sichtweise auf nichtquadratische lineare Gleichungssysteme, die einen nicht stochastisch, sondern algebraisch motivierten Lösungsbegriff erlaubt.

Numerische Behandlung der Lösung Zur numerischen Lösung des Problems gibt es zwei Wege. Zum einen können die Normalgleichungen

$$A^T A\alpha = A^T y$$

gelöst werden, die eindeutig lösbar sind, falls die Matrix A vollen Rang hat. Ferner hat die Produktsummenmatrix $A^T A$ die Eigenschaft, positiv definit zu sein, ihre Eigenwerte sind also alle positiv. Zusammen mit der Symmetrie von $A^T A$ kann dies beim Einsatz von numerischen Verfahren zur Lösung ausgenutzt werden: beispielsweise mit der Cholesky-Zerlegung oder dem CG-Verfahren. Da beide Methoden von der Kondition der Matrix stark beeinflusst werden, ist dies manchmal keine empfehlenswerte Herangehensweise: Ist schon A schlecht konditioniert, so ist $A^T A$ quadratisch schlecht konditioniert. Dies führt dazu, dass Rundungsfehler so weit verstärkt werden können, dass sie das Ergebnis unbrauchbar machen. Durch Regularisierungsmethoden kann die Kondition allerdings verbessert werden.

Eine Methode ist die sog. Ridge-Regression, die auf Hoerl und Kennard (1970) zurückgeht. Das englische Wort ridge heißt soviel wie Grat, Riff, Rücken. Hier wird anstelle der schlecht konditionierten Matrix $A^T A$ die besser konditionierte Matrix $A^T A + \delta I_m$ benutzt. Dabei ist I_m die m-dimensionale Einheitsmatrix. Die Kunst besteht in der geeigneten Wahl von δ. Zu kleine δ erhöhen die Kondition nur wenig, zu große δ führen zu verzerrter Anpassung.

Zum anderen liefert das ursprüngliche Minimierungsproblem eine stabilere Alternative, da es bei kleinem Wert des Minimums eine Kondition in der Größenordnung der Kondition von A, bei großen Werten des Quadrats der Kondition von A hat. Um die Lösung zu berechnen wird eine QR-Zerlegung verwendet, die mit Householdertransformationen oder Givens-Rotationen erzeugt wird. Grundidee ist, dass orthogonale Transformationen die euklidische Norm eines Vektors nicht verändern. Damit ist

$$\|A\alpha - y\|_2 = \|Q(A\alpha - y)\|_2$$

für jede orthogonale Matrix Q. Zur Lösung des Problems kann also eine QR-Zerlegung von A berechnet werden, wobei man die rechte Seite direkt mittransformiert. Dies führt auf eine Form

$$\|R\alpha - Q^T y\|_2$$

mit $R = \binom{\tilde{R}}{0}$, wobei $\tilde{R} \in \mathbb{R}^{m \times m}$ eine rechte obere Dreiecksmatrix ist. Die Lösung des Problems ergibt sich somit durch die Lösung des Gleichungssystems

$$\tilde{R} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix} = \begin{pmatrix} (Q^T y)_1 \\ \vdots \\ (Q^T y)_m \end{pmatrix}$$

Die Norm des Minimums ergibt sich dann aus den restlichen Komponenten der transformierten rechten Seite $(Qy)_{m+1}, \ldots, (Qy)_n$, da die dazugehörigen Gleichungen aufgrund der Nullzeilen in R nie erfüllt werden können.

In der statistischen Regressionsanalyse spricht man bei mehreren gegebenen Variablen $x_1, \ldots, x_n$ von multipler linearer Regression. Der gebräuchlichste Ansatz ein multiples lineares Modell zu schätzen ist als die gewöhnliche Kleinste-Quadrate-Schätzung bzw. gewöhnliche Methode der kleinsten Quadrate (englisch ordinary least squares, kurz OLS) bekannt. Im Gegensatz zur gewöhnlichen KQ-Methode wird die verallgemeinerte Methode der kleinsten Quadrate, kurz VMKQ (englisch generalised least squares, kurz GLS) bei einem verallgemeinerten linearen Regressionsmodell verwendet. Bei diesem Modell weichen die Fehlerterme von der Verteilungsannahme wie Unkorreliertheit und/oder Homoskedastizität ab. Dagegen liegen bei multivariater Regression für jede Beobachtung $(i = 1, \ldots, n)$ r viele y-Werte vor, so dass statt eines Vektors eine $n \times r$-Matrix Y vorliegt (siehe Allgemeines lineares Modell). Die linearen Regressionsmodelle sind in der Statistik wahrscheinlichkeitstheoretisch intensiv erforscht worden. Besonders in der Ökonometrie werden beispielsweise komplexe rekursiv definierte lineare Strukturgleichungen analysiert, um volkswirtschaftliche Systeme zu modellieren.

Probleme mit Nebenbedingungen Häufig sind Zusatzinformationen an die Parameter bekannt, die durch Nebenbedingungen formuliert werden, die dann in Gleichungs- oder Ungleichungsform vorliegen. Gleichungen tauchen beispielsweise auf, wenn bestimmte Datenpunkte interpoliert werden sollen. Ungleichungen tauchen häufiger auf, in der Regel in der Form von Intervallen für einzelne Parameter. Im Einführungsbeispiel wurde die Federkonstante erwähnt, diese ist immer größer Null und kann für den konkret betrachteten Fall immer nach oben abgeschätzt werden.

Im Gleichungsfall können diese bei einem sinnvoll gestellten Problem genutzt werden, um das ursprüngliche Minimierungsproblem in eines einer niedrigeren Dimension umzuformen, dessen Lösung die Nebenbedingungen automatisch erfüllt.

Schwieriger ist der Ungleichungsfall. Hier ergibt sich bei linearen Ungleichungen das Problem

$$\min_{\alpha} \|\vec{f} - \vec{y}\|_2 \text{ mit } l \leq C\alpha \leq u,\ C \in \mathbb{R}^{n \times n},$$

wobei die Ungleichungen komponentenweise gemeint sind. Dieses Problem ist als konvexes und quadratisches Optimierungsproblem eindeutig lösbar und kann beispielsweise mit Methoden zur Lösung solcher angegangen werden.

Quadratische Ungleichungen ergeben sich beispielsweise bei der Nutzung einer Tychonow-Regularisierung zur Lösung von Integralgleichungen. Die Lösbarkeit ist hier nicht immer gegeben. Die numerische Lösung kann beispielsweise mit speziellen QR-Zerlegungen erfolgen.

Nichtlineare Modellfunktionen

Grundgedanke und Verfahren Mit dem Aufkommen leistungsfähiger Rechner gewinnt insbesondere die nichtlineare Regression an Bedeutung. Hierbei gehen die Parameter nichtlinear in die Funktion ein. Nichtlineare Modellierung ermöglicht im Prinzip die Anpassung von Daten an jede Gleichung der Form $y = f(\alpha)$. Da diese Gleichungen Kurven definieren, werden die Begriffe nichtlineare Regression und *curve fitting* zumeist synonym gebraucht.

Manche nichtlineare Probleme lassen sich durch geeignete Substitution in lineare überführen und sich dann wie oben lösen. Ein multiplikatives Modell von der Form

$$y = \alpha_0 \cdot x^{\alpha_1}$$

lässt sich beispielsweise durch Logarithmieren in ein additives System überführen. Dieser Ansatz findet unter anderem in der Wachstumstheorie Anwendung.

Im Allgemeinen ergibt sich bei nichtlinearen Modellfunktionen ein Problem der Form

$$\min_{\alpha} \|f(\alpha) - y\|_2,$$

mit einer nichtlinearen Funktion f. Partielle Differentiation ergibt dann ein System von Normalgleichungen, das nicht mehr analytisch gelöst werden kann. Eine numerische Lösung kann hier iterativ mit dem Gauß-Newton-Verfahren erfolgen.

Aktuelle Programme arbeiten häufig mit einer Variante, dem Levenberg-Marquardt-Algorithmus. Dabei wird durch eine Regularisierung die Monotonie der Näherungsfolge garantiert. Zudem ist das Verfahren bei größerer Abweichung der Schätzwerte toleranter als die Ursprungsmethode. Beide Verfahren sind mit dem Newton-Verfahren verwandt und konvergieren unter geeigneten Voraussetzungen (der Startpunkt ist genügend nahe beim lokalen Optimum) meist quadratisch, in jedem Schritt verdoppelt sich also die Zahl der korrekten Nachkommastellen.

Wenn die Differentiation auf Grund der Komplexität der Zielfunktion zu aufwendig ist, stehen eine Reihe anderer Verfahren als Ausweichlösung zu Verfügung, die keine Ableitungen benötigen, siehe bei Methoden der lokalen nichtlinearen Optimierung.

Fehlverhalten bei Nichterfüllung der Voraussetzungen

Die Methode der kleinsten Quadrate erlaubt es, unter bestimmten Voraussetzungen die wahrscheinlichsten aller Modellparameter zu berechnen. Dazu muss ein korrektes Modell gewählt worden sein, eine ausreichende Menge Messwerte vorliegen und die Abweichungen der Messwerte gegenüber dem Modellsystem müssen eine Normalverteilung bilden. In der Praxis kann die Methode jedoch auch bei Nichterfüllung dieser Voraussetzungen für diverse Zwecke eingesetzt werden. Dennoch sollte beachtet werden, dass die Methode der kleinsten Quadrate unter bestimmten ungünstigen Bedingungen völlig unerwünschte Ergebnisse liefern kann. Beispielsweise sollten keine Ausreißer in den Messwerten vorliegen, da diese das Schätzergebnis verzerren. Außerdem ist Multikollinearität zwischen den zu schätzenden Parametern ungünstig, da diese numerische Probleme verursacht. Im Übrigen können auch Regressoren, die weit von den anderen entfernt liegen, die Ergebnisse der Ausgleichsrechnung stark beeinflussen. Man spricht hier von Werten mit großer Hebelkraft (englisch: High Leverage Value).

Multikollinearität Das Phänomen der Multikollinearität entsteht, wenn die Messreihen zweier gegebener Variablen x_i und x_j sehr hoch korreliert sind, also fast linear abhängig sind. Im linearen Fall bedeutet dies, dass die Determinante der Normalgleichungsmatrix $A^T A$ sehr klein und die Norm der Inversen umgekehrt sehr groß ist; die Kondition von $A^T A$ ist also stark beeinträchtigt. Die Normalgleichungen sind dann numerisch schwer zu lösen. Die Lösungswerte können unplausibel groß werden, und bereits kleine Änderungen in den Beobachtungen bewirken große Änderungen in den Schätzwerten.

Ausreißer Als Ausreißer sind Datenwerte definiert, die 'nicht in eine Messreihe passen'. Diese Werte beeinflussen die Berechnung der Parameter stark und verfälschen das Ergebnis. Um dies zu vermeiden, müssen die Daten auf fehlerhafte Beobachtungen untersucht werden. Die entdeckten Ausreißer können beispielsweise aus der Messreihe ausgeschieden werden oder es sind alternative ausreißerresistente Berechnungsverfahren wie gewichtete Regression oder das Drei-Gruppen-Verfahren anzuwenden.

Im ersten Fall wird nach der ersten Berechnung der Schätzwerte durch statistische Tests geprüft, ob Ausreißer in einzelnen Messwerten vorliegen. Diese Messwerte werden dann ausgeschieden und die Schätzwerte erneut berechnet. Dieses Verfahren eignet sich dann, wenn nur wenige Ausreißer vorliegen.

Bei der gewichteten Regression werden die abhängigen Variablen y in Abhängigkeit von ihren Residuen gewichtet. Ausreißer, d.h. Beobachtungen mit großen Residuen, erhalten ein geringes Gewicht, das je nach Größe des Residuums abgestuft sein kann. Beim Algorithmus nach Mosteller und Tukey (1977), der als 'biweighting' bezeichnet wird, werden unproblematische Werte mit 1 und Ausreißer mit 0 gewichtet, was die Unterdrückung des Ausreißers bedingt. Bei der gewichteten Regression sind in der Regel mehrere Iterationsschritte erforderlich, bis sich die Menge der erkannten Ausreißer nicht mehr ändert.

Heteroskedastische Fehler Liegen heteroskedastische Fehler vor, so liefert die Minimierung des Mittelwertes der kleinsten Quadrate keinen effizienten Schätzer des (bedingten) Mittelwertes, obwohl dieser immer noch unverzerrt ist. Die Minimierung der Gausschen Negativen Log-Likelihood kann in diesem Fall eine Alternative sein.

Verallgemeinerte Kleinste-Quadrate-Modelle

Weicht man die starken Anforderungen im Verfahren an die Fehlerterme auf, erhält man so genannte verallgemeinerte Kleinste-Quadrate-Ansätze. Wichtige Spezialfälle haben dann wieder eigene Namen, etwa die gewichtete Methode der kleinsten Quadrate (englisch weighted least squares, kurz WLS), bei denen die Fehler zwar weiter als unkorreliert angenommen werden, aber nicht mehr von gleicher Varianz. Dies führt auf ein Problem der Form

$$\|D(A\alpha - y)\|_2,$$

wobei D eine Diagonalmatrix ist. Variieren die Varianzen stark, so haben die entsprechenden Normalgleichungen eine sehr große Kondition, weswegen das Problem direkt gelöst werden sollte.

Nimmt man noch weiter an, dass die Fehler in den Messdaten auch in der Modellfunktion berücksichtigt werden sollten, ergeben sich die *totalen kleinsten Quadrate* in der Form

$$\min_{E,r} \|(E, r)\|_F, \quad (A + E)\alpha = b + r,$$

wobei E der Fehler im Modell und r der Fehler in den Daten ist.

Schließlich gibt es noch die Möglichkeit, keine Normalverteilung zugrunde zu legen. Dies entspricht beispielsweise der Minimierung nicht in der euklidischen Norm, sondern der Summennorm. Solche Modelle sind Themen der Regressionsanalyse.

Chapter 2

ufPro

2.1 Pending Tasks / ToDos

- elevation, superelevation, outer chainage, design speed, ...
- try to implement the angle-value-handling vector-based (sin- and cos-values), preferred instead of cycling radiant-values. TICK
- Check the capabilities of coord.system transformations like the epsg.io service.

2.2 Basics and Preview

Die hier dargelegten Überlegungen bedienen sich strikt der mathematischen Nomenklatur, welche sich nicht nur punktuell von der vermessungstechnischen oder auch alltäglichen Vorstellung unterscheidet. Gegebenenfalls wird an passender Stelle explizit darauf hingewiesen.

Wohin geht die Reise? Wir wollen die typischen Möglichkeiten des Trassen-basierten Rechnens zusammen stellen, darauf aufbauende Dokumentationsmöglichkeiten andocken (Trassenplan, Weichenhöhenplan, Geschwindigkeits-Wege-Band, et. al.), sowie grafische Komponenten bereit stellen und aktuelle Schnittstellen (BIM-konform) implementieren.

2.2.1 2D-Kurven

Eine Trasse verstehen wir als stetig glatte Kurve im $\mathbb{R}^2$ (Ebene), bestehend aus Segmenten - wahlweise Gerade, Bogen oder Übergangsbogen (Clothoide, S-Form, Bloss, Sinusoide, Cosinoide ...) — entsprechend den Darstellungen z.B. in den landXML-Schemata.

Mit Startpunkt und -richtung, sowie Krümmungsverlauf ist der Kurven- respektive Trassenverlauf festgelegt (Anfangswertproblem). Üblicherweise sind Trassen bogenlängen-parametrisiert, was die Herleitung und Darstellung grundlegender Hilfsmittel (Formeln) stark vereinfacht. Schönes Beispiel: die Elemente des Tangentialfelds (jeder Tangentialvektor entlang der Kurve) haben alle Länge 1, und sind somit eindeutig mit Sinus- und Cosinuswert des zugrundeliegenden Richtungswinkels zu beschreiben.

$$\dot{c}(s) = \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix}(s) = \begin{pmatrix} \cos \\ \sin \end{pmatrix}(t(s))$$

Weiterhin ergibt sich unter Hinzunahme des Normalenfeldes ein Zugang zur Krümmung als Faktor zur Verbindung von Normaler und zweiter Ableitung.

$$\ddot{c}(s) = \dot{t}(s) \cdot \begin{pmatrix} -\sin \\ \cos \end{pmatrix} (t(s)) = \dot{t}(s) \cdot \begin{pmatrix} \cos \\ \sin \end{pmatrix}^{\perp} (t(s)) = k(s) \cdot N(s)$$

Woraus sich augenscheinlich die Lösbarkeit des oben erwähnten Anfangswertproblems ergibt. □

2.2.2 Anfangswertproblem

Sei

$$\kappa(\sigma) : [0,1] \to [0,1]$$

stetig mit $\kappa(0) = 0$ und $\kappa(1) = 1$. Des Weiteren seien k_a, k_e und L bekannt (und passabel!), ebenso t_a und (x_a, y_a). Die *normalisierte Krümmungsfunktion* κ skaliert nun mittels $k_a + (k_e - k_a) \cdot \kappa(s/L)$ für jedes $s \in [0, L]$ zu $k(s) : [0, L] \to [k_a, k_e]$, für welches gilt:

$$\begin{aligned}
k(s) &= k_a + (k_e - k_a) \cdot \kappa(s/L) \\
\int k(s)\,\mathrm{d}s &= k_a + (k_e - k_a) \cdot \int \kappa(s/L)\,\mathrm{d}s \\
&= k_a + (k_e - k_a) \cdot \int \kappa(\sigma) \cdot L\,\mathrm{d}\sigma, \qquad \text{mit } \sigma = \frac{s}{L} \\
\Delta t(s) &= \int_0^s k(s)\,\mathrm{d}s \\
&= k_a \cdot s + (k_e - k_a) \cdot L \cdot \int_0^\sigma \kappa(\sigma)\,\mathrm{d}\sigma \\
\text{setze } \Delta\theta(\sigma) &:= \int_0^\sigma \kappa(\sigma)\,\mathrm{d}\sigma, \text{ also} \\
\Delta t(s) &= k_a \cdot s + (k_e - k_a) \cdot L \cdot \Delta\theta(s/L)
\end{aligned}$$

Hmm ... Mal sehn wohin das führt ...

$$\begin{aligned}
\begin{pmatrix} x \\ y \end{pmatrix}(s) &= \begin{pmatrix} x_a \\ y_a \end{pmatrix} + \int_0^s \begin{pmatrix} \cos \\ \sin \end{pmatrix} (t_a + \Delta t(s))\,\mathrm{d}s \\
&. = \begin{pmatrix} x_a \\ y_a \end{pmatrix} + \int_0^s \begin{pmatrix} \cos t_a \cos \Delta t(s) - \sin t_a \sin \Delta t(s) \\ \sin t_a \cos \Delta t(s) + \cos t_a \sin \Delta t(s) \end{pmatrix}\,\mathrm{d}s \\
&. = \begin{pmatrix} x_a \\ y_a \end{pmatrix} + \cos t_a \cdot \int_0^s \begin{pmatrix} \cos \\ \sin \end{pmatrix} \Delta t(s)\,\mathrm{d}s + \sin t_a \cdot \int_0^s \begin{pmatrix} -\sin \\ \cos \end{pmatrix} \Delta t(s)\,\mathrm{d}s
\end{aligned}$$

Zusätzliche Anforderungen an κ wie Glattheit auch in Anfang und Ende (mit Klothoide nicht erfüllt!) oder Schiefsymmetrie ergeben Festlegungen in simplifizierter Form, etwa erster oder *nur* zweiter Ableitung, auch auf anderen Definitionsbereichen. Glattheit in Anfang und Ende heisst $\kappa'(0) = \kappa'(1) = 0$, ggf. mehrfach. Womit in polynomialer Form gilt: $\kappa'(\sigma) = (0-\sigma)^m \cdot (1-\sigma)^n \cdot \rho(\sigma)$, für Antisymmetrie in 0.5 zumindest auch $(0.5 - \sigma)^k = 0$ als Teil von $\kappa'' \ldots$

Remark 2.2.2.1 (Substitutionsregel)**.**

$$\int f(x)\,\mathrm{d}x = \int f(\varphi(\sigma)) \cdot \varphi'(\sigma)\,\mathrm{d}\sigma$$

Für $\varphi:\ \sigma \in [0,1] \to x \in [a,b]$ dann $a + (b-a)\cdot\sigma = x$, also $\varphi' = (b-a)$

...

2.2.3 Die Trassenelemente

Trassen bzw. Achsen als Bezugssystem linienförmiger Infrastrukur bestehen üblicherweise aus einer geordneten Abfolge von Elementen konstanter Krümmung (Gerade, Bogen) und jeweils zwischenliegender (krümmungs-)vermittelnder Elemente, den Übergangsbögen. (Anm.: Jedes Element für sich ist schon eine simple Kurve.)

Im folgenden gibt es einen kurzen Abriss zur expliziten AWP-Lösung für Gerade und Bogen (mit erwartbarem[1] Ergebnis), anschliessend gleiches für verschiedene Arten Übergangsbögen mit einem eigenen umfangreicheren Abschnitt.

	Krümmung κ	Richtung $\Delta\tau(s)$ $= \int_{s_0}^{s} \kappa(s)\,\mathrm{d}s$	Easting $Y(s)$ $= \int_{s_0}^{s} \cos(\Delta\tau(s))\,\mathrm{d}s$	Northing $X(s)$ $= \int_{s_0}^{s} \sin(\Delta\tau(s))\,\mathrm{d}s$
Gerade	0	$t = t_0 = const$	$s \cdot \cos t$	$s \cdot \sin t$
Bogen	const	$k \cdot s$ (oBdA $t_0 = 0$)	$\sin(k\cdot s)/k$	$(1-\cos(k\cdot s))/k$

Table 2.1: elementare Berechnung für Geraden und Bögen (mit Startpunkt im Koordinatenursprung)

Parametrisierungen

- Bogenlängenparametrisierung
- Polarkoordinaten
- algebraisch / implizite Beschreibung
- Funktionsgraphen (explizit)

Übergangsbögen

In den folgenden Betrachtungen beschränken wir uns auf Clothoiden, S-Form, Blossbögen, Cosinoiden und Sinusoiden. Spezialfälle wie 'unvermittelter Krümmungswechsel' (Übergangsbogenlänge gleich Null) oder Raritäten (z.B Lemniskate, Radioide, ...) bleiben vorerst aussen vor, ebenso wie 'simple' Approximationen (*kubische Parabel* für Clothoiden, *Sine Half-Wavelength Diminishing Tangent Curve* für Cosinoiden).

Ein Übergangsbogen (engl.: transition curve) vermittelt die Krümmungsänderung zwischen Segmenten konstanter Krümmung und ist damit einigen weiteren Anforderungen unterworfen (z.B. Geschwindigkeitsrelevanz und Fahrkomfort) - dazu eventuell später.

Zur einfachen Illustration und Herleitung betrachten wir den Fall Segmentlänge gleich 1 ($s \in [0,1]$) und stetiger (teils auch in den Anschlüssen glatter höherer Ordnung) Änderung der Krümmung

[1] oder auch nicht, siehe die vorangestellten Bemerkung hinsichtlich mathematischer Nomenklatur. Beispiel Krümmung (respektive invers Radius): einen Rechtsbogen durchfährt man nach rechts, nur hat dieser aber mathematisch eine negative Krümmung ... Oops!

von Null ($\kappa(0) = 0$) auf Eins ($\kappa(1) = 1$), nennen es die normalisierte Betrachtung. Die genannten Übergangsbögen unterscheiden sich, wie gleich explizit angegeben in ihrer polynomialen Ordnung, von erstem Grad (Clothoide) bis unendlich (Sinusoide). Im unserem Normalo-Schaufenster $[0,1] \times [0,1]$ sind es also Funktionsgraphen, welche beginnend mit schlicht gerade bis mehr und mehr s-förmig die gegenüber liegenden Eckpunkte $(0,0)$ und $(1,1)$ verbinden.

Bei der S-Form offensichtlich schon per Definition, so ist sogar bei allen Nicht-Clothoiden, also den Ü-Bögen mit geschwungenem Krümmungsverlauf, gegebenenfalls der halbe Verlauf (1. Schwung, half-wave) bereits von Interesse - Stichwort Übergang zwischen Gegenbögen (werden in der Praxis nicht mehr in 'einem Stück' gefertigt, liegen im Gleisbestand aber vor!)

In [**kufver:1997**] is given a very helpful overview!

Clothoid .

$$\kappa(\sigma) = \sigma$$

Biquadratic / Helmert curve / S-Form .

HELMERT (1872) discussed a type of transition curve which also is known as the bi-quadratic parabola and in Germany as the SCHRAMM curve (SCHUHR, 1985). According to SCHRAMM (1934), the advantage of the HELMERT curve is the lower value of the second derivative of cant and corresponding vertical acceleration. The first time these curves were used was in 1934–1935, when more than 300 were built on the Berlin – Hamburg line (SCHRAMM 1934, 1975), but they are also used in Sweden. In the HELMERT curve, the curvature has the shape of two second-degree parabolas such that the curvature, $k(s)$, and its derivative, $k'(s)$, are continuous functions.

$$\begin{aligned} \kappa \quad : \quad & [0, 0.5] \to [0, 0.5], \\ & \sigma \mapsto 2\sigma^2 \end{aligned}$$

Ruch curve (Clothoid and two parts of HELMERT curve)

This type of transition curve was suggested by RUCH (1903). The purpose with the RUCH curve was to create a smooth ride concerning lateral position of the mass centre of a standard vehicle, rolling and yaw velocity. The transition curve consists of three parts. The curvature in the first part has the shape of a second-degree parabola, the curvature in the second part is a linear function and the curvature in the last part again has the shape of a second-degree parabola. In the RUCH curve, the curvature $k(s)$ and the derivative of curvature $k'(s)$ for the trajectory of the mass centre of a standard vehicle, are continuous functions.

SCHUHR (1984) neglected the linear acceleration of the mass centre of the vehicle which follows an angular acceleration, where the instantaneous axis of rotation is located in the track centre line (or one of the rails). In this case, the clothoid and the HELMERT curve can be regarded as special cases of the RUCH curve. If the lengths of the first and third parts of the RUCH curve are zero, the curve is identical to a clothoid. If the lengths of the first and third parts are equal and the length of the second part is zero, the curve is identical to a HELMERT curve.

The equations given here correspond to the simplifications of SCHUHR (1984). If the original RUCH curve is to be calculated, the curvature in the first and third part must be adjusted

with constant terms. These terms equal the second derivative of cant multiplied by the height of the mass centre above the rails divided by 1500 mm. (The two correction terms will create a small reverse curve and will reduce the necessary lateral shift.) SCHUHR also assumed that the three parts of the RUCH curve have the same lengths.

...

Bloss curve .

This type of transition curve was suggested by BLOSS (1936). The curvature of the BLOSS curve consists of a third-degree parabola. The derivative of curvature, but not the second derivative, is a continuous function at the tangent points. The argumentation for this type of transition curve was that the curvature function involves only one equation (contrary to the HELMERT curve) and that the function is simpler than the equation for the cosine curve.

$$\kappa(\sigma) = 3\sigma^2 - 2\sigma^3$$

Sine .

The sinusoidal transition curve was suggested in 1937 by KLEIN (SCHUHR 1985, WEIGEND 1975). The curvature function is built up of one period of a sine function so that both its first and second derivatives are continuous functions at the tangent points.

$$\kappa(\sigma) = \sigma - \sin(2\pi\sigma)/(2\pi)$$

Cosine .

The transition curve with the curvature formed as a cosine function was suggested by VOJACEC (1868). It has been used in Japan and Spain and on test tracks in Germany (KICK 1969). The curvature consists of a half-period of a cosine function. The derivative of curvature, but not the second derivative, is a continuous function at the tangent points.

$$\kappa(\sigma) = (1 - \cos(\pi\sigma))/2$$

Gubar curve (Clothoid and two parts of Cosine)

GUBAR (1990) suggested a transition curve analogous to the RUCH curve, but consisting of a clothoid and two parts of a cosine curve. The background to this suggestion was the following: According to a limit on maximum jerk, the required length of a cosine curve may lead to such a large angle (the difference in direction between the starting point and ending point of the transition curve) that the connected circle must have a negative length. In order to reduce the required length of the transition curve (and hence the angle) the middle part is formed as a clothoid and only the ends of the transition curve have curvature functions where the second derivative is non-zero. (However, it should be noted that the RUCH curve also solves this problem.)

The clothoid and the cosine curve can be regarded as special cases of the GUBAR curve. If the lengths of the first and third parts of the GUBAR curve are zero, the curve is identical to a clothoid. If the lengths of the first and third parts are equal and the length of the second part is zero, the curve is identical to a cosine curve.

.

Watorek curve .

This transition curve was suggested by Watorek (1907). The curvature function is built up by a polynomial so that both its first and second derivatives are continuous functions at the tangent points. The main reason for having a continuous second derivative of curvature was to achieve a continuous second derivative of cant, which gives a smoother lateral ride for the mass centre of the vehicle.

$$\begin{aligned} \kappa \quad : \quad & [0,1] \to [0,1], \\ & \sigma \mapsto 10\sigma^3 - 15\sigma^4 + 6\sigma^5 \end{aligned}$$

Mieloszyk/Koc .

Mieloszyk and Koc (1991) suggested a new type of transition curve, with a polynomial curvature function such that both its first and second derivatives are continuous functions at the tangent points. At one of the tangent points, also the third derivative is a continuous function.

$$20\sigma^3 - 45\sigma^4 + 36\sigma^5 - 10\sigma^6$$

It is not clear which end should have the continuous third derivative of curvature when the transition has non-zero curvature at both ends (for example on reverse curves).

p-curve .

The p-curve was suggested by Broman (1982). The p-curve has a polynomial curvature function such that the first derivative is a continuous function at the tangent points. The p-curve is derived to solve alignment problems when a transition curve is desired between two points with defined coordinates, directions and curvatures. This problem is mathematically over-determined for conventional types of transition curves.

$$k(s) = k_0 + \left(3(s/L)^2 - 2(s/L)^3\right) \cdot (k_1 - k_0) + p_4 s^2 (L-s)^2 + p_5 s^2 \left(\frac{L}{2} - s\right)(L-s)^2$$

It should be noted that the two first terms correspond to the Bloss curve. The third term is needed to correct for a possible mismatch in direction and the fourth term is needed to correct for a possible mismatch in lateral position. It is not clear why the tangent points must coincide with the fixed points. If these two conditions are relaxed, by allowing the transition curve to start before or at least after the fixed points, but still requiring the alignment chain to pass through the two fixed points, the problem may be solved with the two first terms, the Bloss curve, alone.

With computer programs for automatic optimisation of alignments, a global optimum cannot be guaranteed since the number and types of elements must be specified by the user (Hupfeld 1970). The problem defined by Broman may be solved without the p-curve if a combination of two elements is used.

Wiener Bogen / Vienna curve .

Remark 2.2.3.1. Vienna curves / mass center alignment

$$\begin{aligned} k(s) &= \frac{k_c}{\Psi_c} \cdot \Psi(s) - h \cdot \frac{d^2\Psi}{ds^s} \\ k_H(s) &= k_1 + (k_2 - k_1) \cdot f(\frac{s}{l}) - h \cdot \Delta\Psi \cdot \frac{d^2 f}{ds^2} \end{aligned}$$

(V2) mit $\sigma := \frac{s}{l} \in [0,\ 1]$:

$$f(\sigma) = \sigma^4 \left(35 - 84\sigma + 70\sigma^2 - 20\sigma^3\right)$$

(!!! Fortschreibung von Bloss, Watorek nach 7. Ordnung)

(V3) mit $\sigma := \frac{s}{l} \in [0,\ 1]$ und $\tan\frac{Z}{2} = \frac{Z}{2} \approx 4.49340946$:

$$f(\sigma) = \underbrace{\sigma^2(3-2\sigma)}_{\text{Bloss}} + \underbrace{\frac{6}{Z^2}\left(+2\sigma - 1 + \cos(Z\sigma) - \frac{2}{Z}\sin(Z\sigma)\right)}_{\text{!!!}}$$

(V4) mit $\sigma := \frac{s}{l} \in [0,\ 1]$:

$$f(\sigma) = \frac{1}{4}\left(\frac{1}{12-\pi^2}\left(\pi^2 \underbrace{\sigma^2(2\sigma-3)}_{\text{Bloss}} + 6\underbrace{(1-\cos(\pi\sigma))}_{\text{Cosine}}\right)\right) + \frac{3}{4}\underbrace{\left(\sigma - \frac{1}{2\pi}\sin(2\pi\sigma)\right)}_{\text{Sine}}$$

(V5) mit $\sigma := \frac{s}{l} \in [0,\ 1]$:

$$f(\sigma) = \frac{15}{15-\pi^2}\underbrace{\left(\sigma - \frac{1}{2\pi}\sin(2\pi\sigma)\right)}_{\text{Sine}} - \frac{\pi^2}{15-\pi^2}\underbrace{\sigma^3(6\sigma^2 - 15\sigma + 10)}_{\text{Watorek}}$$

(V6) mit $\sigma := \frac{s}{l} \in [0,\ 1]$:

$$f(\sigma) = \frac{5}{10-\pi^2}\underbrace{(1-\cos(\pi\sigma))}_{\text{Cosine}} - \frac{\pi^2}{2(10-\pi^2)}\underbrace{\sigma^2(2\sigma^3 - 5\sigma^2 + 5)}_{\text{!!!}}$$

(V7) mit $\sigma := \frac{s}{l} \in [0,\ 1]$:

$$f(\sigma) = \sigma^5(126 - 420\sigma + 540\sigma^2 - 315\sigma^3 + 70\sigma^4)$$

(!!! Fortschreibung von Bloss, Watorek, (V2) nach 9. Ordnung)

Klauder curves .

(pol_m_n) mit $\sigma := \frac{s}{l} \in [-a,\ a]$:

$$k''(\sigma) = \text{const}(a+\sigma)^m \sigma^n (a-\sigma)^m$$

By this, you can easily figure out, that: Bloss is pol_0_1, and Watorek is pol_1_1.

(pol_2_1) A 5th order polynomial of form (constant) $(a+s)^2 s(a–s)^2$ That is like (pol_1_1) except that at each end the 2nd derivative of the curvature has a zero of order 2 rather than of order 1. This makes the slope of 2nd derivative of curvature zero at each end. Spirals of this form are longer and gentler than spirals based on any of the other forms examined here.

(pol_2_5) A 9th order polynomial of form (constant) $(a+s)^2 s^5(a–s)^2$ like (pol_2_1) but with the zero at the midpoint changed from 1st to 5th order to push acceleration of curvature toward the ends of the spiral and thus reduce spiral length for given curve offset.

(pol_2_9) A 13th order polynomial of form (constant) $(a+s)^2 s^9(a–s)^2$ like (pol_2_5) but with the zero at the midpoint changed from 5th to 9th order to push acceleration of curvature further toward the ends of the spiral and further reduce spiral length for given curve offset. Further increase of the central zero order would further increase dynamic disturbance at the ends of the spiral. In the limit of large odd central zero order this form would become equivalent to a linear spiral.

mentioned in landXML

```
<xs:simpleType name="spiralType">
<xs:restriction base="xs:string">
<xs:enumeration value="biquadratic"/>
<xs:enumeration value="bloss"/>
<xs:enumeration value="clothoid"/>
<xs:enumeration value="cosine"/>
<xs:enumeration value="cubic"/>
<xs:enumeration value="sinusoid"/>
<xs:enumeration value="revBiquadratic"/>
<xs:enumeration value="revBloss"/>
<xs:enumeration value="revCosine"/>
<xs:enumeration value="revSinusoid"/>
<xs:enumeration value="sineHalfWave"/>
<xs:enumeration value="biquadraticParabola"/>
<xs:enumeration value="cubicParabola"/>
<xs:enumeration value="japaneseCubic"/>
<xs:enumeration value="radioid"/>
<xs:enumeration value="weinerBogen"/>
</xs:restriction>
</xs:simpleType>
```

Lemniskate

Remark 2.2.3.2. lemniscate

$$\begin{aligned} scale &= 2/(3-\cos(2\cdot t)) \\ x &= scale\cdot\cos(t) \\ y &= scale\cdot\sin(2\cdot t)/2 \end{aligned}$$

...

Werfen wir einen Blick auf die Ableitungen erster Ordnung auf $[0,\ 1]$ ...

	$\tau(\sigma) = \int \kappa(\sigma)$	$\kappa(\sigma)$	const $\kappa'(\sigma)$	
Clothoid		σ	1	
Biquadratic		$2\sigma^2$	σ	
Bloss pol_0_1	$x^3 - x^4/2$	$3\sigma^2 - 2\sigma^3$	$\sigma(1-\sigma)$	
part of V6		$2.5\sigma^2 - 2.5\sigma^4 + \sigma^5$	$\sigma(1+\sigma-\sigma^2)(1-\sigma)$	
Watorek pol_1_1		$10\sigma^3 - 15\sigma^4 + 6\sigma^5$	$\sigma^2(1-\sigma)^2$	
MieKoc		$20\sigma^3 - 45\sigma^4 + 36\sigma^5 - 10\sigma^6$	$\sigma^2(1-\sigma)^3$	
Vienna2 pol_2_1		$35\sigma^4 - 84\sigma^5 + 70\sigma^6 - 20\sigma^7$	$\sigma^3(1-\sigma)^3$	
Vienna7 pol_3_1		$126\sigma^5 - 420\sigma^6 + 540\sigma^7 - 315\sigma^8 + 70\sigma^9$	$\sigma^4(1-\sigma)^4$	
pol_2_3				
pol_4_1			$\sigma^5(1-\sigma)^5$	
pol_2_5				
pol_5_1			$\sigma^6(1-\sigma)^6$	
pol_2_7				
pol_6_1			$\sigma^7(1-\sigma)^7$	
pol_2_9				
Sine		$\sigma - \frac{\sin(2\pi\sigma)}{2\pi}$	$1 - \cos(2\pi\sigma)$	
Cosine		$\frac{1}{2}(1 - \cos(\pi\sigma))$	$\frac{1}{2}\pi\sin(\pi\sigma)$	$\frac{1}{2}\pi^2\cos(\pi\sigma)$
pV3				

... und zweiter Ordnung auf $[-1,\ 1]$:

			const $\kappa'(\sigma)$	const $\kappa''(\sigma)$
Clothoid			$(1+\sigma)^0(1-\sigma)^0$	0
Biquadratic			$(1+\sigma)^1(1-\sigma)^0$	σ^0
Bloss pol_0_1			$(1+\sigma)^1(1-\sigma)^1$	$(1+\sigma)^0\sigma^1(1-\sigma)^0$
pV6			$(1+\sigma)(\sigma^2-5)(1-\sigma)$	$(\sqrt{3}+\sigma)\sigma(\sqrt{3}-\sigma)$
Watorek pol_1_1			$(1+\sigma)^2(1-\sigma)^2$	$(1+\sigma)^1\sigma^1(1-\sigma)^1$
MieKoc			$(1+\sigma)^2(1-\sigma)^3$	$(1+\sigma)(5\sigma+1)(1-\sigma)^2$
Vienna2 pol_2_1			$(1+\sigma)^3(1-\sigma)^3$	$(1+\sigma)^2\sigma^1(1-\sigma)^2$
Vienna7 pol_3_1			$(1+\sigma)^4(1-\sigma)^4$	$(1+\sigma)^3\sigma^1(1-\sigma)^3$
pol_2_3			$(1+\sigma)^3(3\sigma^2+1)(1-\sigma)^3$	$(1+\sigma)^2\sigma^3(1-\sigma)^2$
pol_4_1			$(1+\sigma)^5(1-\sigma)^5$	$(1+\sigma)^4\sigma^1(1-\sigma)^4$
pol_2_5			$(1+\sigma)^4(6\sigma^4+3\sigma^2+1)(1-\sigma)^4$	$(1+\sigma)^2\sigma^5(1-\sigma)^2$
pol_5_1			$(1+\sigma)^6(1-\sigma)^6$	$(1+\sigma)^5\sigma^1(1-\sigma)^5$
pol_2_7			$(1+\sigma)^3(10\sigma^6+6\sigma^4+3\sigma^2+1)(1-\sigma)^3$	$(1+\sigma)^2\sigma^7(1-\sigma)^2$
pol_6_1			$(1+\sigma)^7(1-\sigma)^7$	$(1+\sigma)^6\sigma^1(1-\sigma)^6$
pol_2_9			$(1+\sigma)^3(15\sigma^8+10\sigma^6+6\sigma^4+3\sigma^2+1)(1-\sigma)^3$	$(1+\sigma)^2\sigma^9(1-\sigma)^2$
Sine				
Cosine				
pV3				

$$
\begin{aligned}
\int (1+\sigma)^2\sigma^9(1-\sigma)^2 d\sigma &= \int (1-s^2)^2 s^9 ds \\
&= \int (1-2s^2+s^4)s^9 ds \\
&= \int (s^9 - 2s^{11} + s^{13}) ds \\
&= \frac{s^{10}}{10} - 2\frac{s^{12}}{12} + \frac{s^{14}}{14} + const \\
21s^{10} - 35s^{12} + 15s^{14} + const &= s^{10}(15s^4 - 35s^2 + 21) - 1 \\
15s^{14} - 35s^{12} + 21s^{10} - 1 &= (15s^{12} - 20s^{10} + s^8 + s^6 + s^4 + s^2 + 1)(s^2 - 1) \\
&= (15s^{10} - 5s^8 - 4s^6 - 3s^4 - 2s^2 - 1)(s^2 - 1)^2 \\
&= (15s^8 + 10s^6 + 6s^4 + 3x^2 + 1)(s^2 - 1)^3
\end{aligned}
$$

lesson learned

- wie schon bei den Vienna's ersichtlich, sind Kombinationen der Funktionen miteinander, aber auch mit Null-Funktionen denkbar.
- die Änderung des Wertebereichs weg von $[0,\ 1]$ auf $[-1,\ 1]$, und besser sogar auf $[a,\ b]$ ist für die Analyse offensichtlich hilfreich! a, b? Mit $\kappa''(\sigma) = (a+\sigma)^{m_a}\sigma^n(b-\sigma)^{m_b}$ sind alle polynomialen Fälle von oben redundant erfasst $(a < 0 < 1)$
- direkte Anschlussfrage: Ist Symmetrie $a + b = 0$ hilfreich???
- Gubar und Ruch curve führen uns zum Berlin-Dogma: Ein Übergangsbogen setzt sich aus den drei Teilen halfWave1 - clothoid - halfWave2 zusammen, mit jeweils glatten Übergängen (außer ...). Und: Erster und dritter Teil müssen nicht zwingend gleicher Herkunft sein, sogar der mittlere Wendepunkt ist wie bei Biquadratic verzichtbar. Damit ist jedes Polynom mit mindestens Grad zwei Transition-Atom. Sine und Cosine als C^∞-Polynome bedürfen hier weiterer Betrachtungen.
- Plus Schwerpunkttrassierung ... falls gwollt.

Der letzte Punkt ist recht entspannt - SPT an oder aus. Fertig. Na gut, wir brauchen dazu die Originalfunktion plus zweiter Ableitung - Haben wir ja.

Schauen wir uns die Waves an. Eigentlich ja nur halfWave1, weil halfWave2 sich per Drehung / Verschiebung aus einer halfWave1 ergibt (siehe Biquadratic). Eine passende halfWave1 f hat am Startwert a einen Scheitel- oder Wendepunkt mit $f'(a) = 0$, und mündet stetig wachsend im Endwert b (ohne weitere Anforderungen)

...

...

ADDITIONAL SPIRAL CONCEPTS:

An important spiral design conceptual advance of a different kind was achieved by the German engineer Donges (Donges 1968). His insight was that the forces needed to rotate a vehicle between its pre- and post-spiral bank angles will be less if the vehicle is rotated not about a longitudinal axis in the plane of the track but rather about a longitudinal axis through its so-called center-of-percussion with respect to a transverse impulse applied by the rails. (The center-of-percussion is above the center of mass.) To implement Donges' idea computationally one can choose a track roll axis height above a typical vehicle center of mass height, calculate the coordinates of each point on the path of that roll axis using any of the spiral types from Table 1 (although the last 5 that have second derivative of curvature zero at each end of the spiral may be preferred), and then note that the corresponding track point is shifted laterally by the roll axis height times the sine of the bank angle at the point. Spiral geometry of this type (but using a curvature function not illustrated in Table 1) was studied and successfully tested by the Austrian Railways as reported in Presle & Hasslinger (1998).

There is another conceptual advance that is worth noting even though with the small track bank angles that are used in conventional railroads its effects are quite small. That is to begin one's thinking about a spiral not by thinking of a shape on the ground but rather by thinking about how to rotate a vehicle between its pre- and post-spiral bank angles. The pol_1_1, Sine, and pol_2_1 entries in Table 1 for track curvature versus distance are gentle functions, and those

functions can be chosen to define how the roll angle should vary with distance. Ideally that would be the vehicle roll angle, but since a track designer cannot control vehicle roll angle directly that function can be used as a specification for the track bank angle $R(s)$ as a function of distance. To obtain the corresponding path of the track on the ground one obtains the curvature, $C(s)$, of the path of the vehicle roll axis from the roll $R(s)$ via the balance equation. That equation stipulates that the component of centripetal acceleration in the plane of the track should be provided by the component of gravity in the plane of the track and can be written:

$$\mathcal{C}(s) \cdot \mathcal{V}^2 \cdot \cos(\mathcal{R}(s)) = \mathcal{G} \cdot \sin(\mathcal{R}(s))$$

where $\mathcal{V}$ denotes a balancing speed and $\mathcal{G}$ denotes the acceleration of gravity. With the formula for curvature as a function of path length along the spiral established, integrations to obtain the compass bearings and coordinates of successive points along the spiral are done as already outlined.

...

...

2.3 alignment data transfer

	fix			zero			approx			transitions									
	fixElement	line	arc	Km-Sprung	kink	immidiate	cubParabola	???	???	clotho	biquadratic	bloss	kauder	sine	cose	? radioid ?	viennese	...	selfmade
AIMv0	x	-	-	x	x	x	-	-	-	x	x	x	x	x	x	-	x		x
Verm.Esn	-	x	x	x	x	(x)				x	x	x	-	x	x		-		-
landXML	-	x	x	(x)	(x)	(x)				x	x	x	-	x	x		-		
infraGML	-	x	x	(x)	(x)	(x)				x									
ifcAlignment	-	x	x	(x)	(x)	(x)				x									
railML	-	x	x	?			?			x									
...																			

...

2.4 Numerik der genutzten Wertebereiche

... als da wären: Länge, Krümmung, und Winkel!

2.4.1 Längenangaben und darauf aufbauende Werte

Die 1-dim. Länge wird schlicht als Float32 resp. Float64 behandelt. Höher-dimensionales entsprechend als Vektoren (bzw. arrays) mit Längen-Komponenten.

Eine Erweiterung des damit zugrunde-gelegten Zahlraums $\mathbb{R}$ um $\pm\infty$ und die Kehrseite $\neg\mathbb{R}$ in Hinblick auf orientierte homogene Koordinaten erfolgt vielleicht später.

2.4.2 Krümmung

... und Radius ... Letzteres ganz klar eine Länge, ersteres aber nutzbringender ($r = \pm 0$?!)

...

2.4.3 Winkel

abendfüllend! Sinn und nochmehr Unsinn der unsäglichen Debatten und Unkenntnis zur Verwendung von Radiant-Werten, Alt-Grad, Neugrad oder sogar time-based-values, Orientierung clock- or counterclockwise, τ instead of 2π etc. et. al. — mal außen vor gelassen: Um ganz einfach den Perioden-Überlauf und damit verbundene Probleme bei simplen Berechnungen (siehe Mittelwert von -2 gon and +1 gon vs. 398 gon and 1 gon) zu verringern (der Optimist sagt: zu vermeiden), probieren wir die Handhabe per sin/cos-Vektor als Pendant zum mathematischen Radiant-Wert.

$$\tau \to \begin{pmatrix}\cos\tau\\ \sin\tau\end{pmatrix} \quad , \quad \begin{pmatrix}c\\ s\end{pmatrix} \to (s \geq 0)\ ?\ +\arccos c\ :\ -\arccos c$$

Standardberechnungen am Winkel:

Addition	$\begin{pmatrix}\cos(\alpha+\beta)\\ \sin(\alpha+\beta)\end{pmatrix}$	$\begin{pmatrix}\cos\alpha\cdot\cos\beta-\sin\alpha\cdot\sin\beta\\ \sin\alpha\cdot\cos\beta+\cos\alpha\cdot\sin\beta\end{pmatrix}$
Subtraktion	$\begin{pmatrix}\cos(\alpha-\beta)\\ \sin(\alpha-\beta)\end{pmatrix}$	$\begin{pmatrix}\cos\alpha\cdot\cos\beta+\sin\alpha\cdot\sin\beta\\ \sin\alpha\cdot\cos\beta-\cos\alpha\cdot\sin\beta\end{pmatrix}$
Halbierung	$\begin{pmatrix}\cos\frac{\alpha}{2}\\ \sin\frac{\alpha}{2}\end{pmatrix}$	...
Vielfachung	$\begin{pmatrix}\cos(n\cdot\alpha)\\ \sin(n\cdot\alpha)\end{pmatrix}$	$\begin{pmatrix}\sum_{j=0}^{\lfloor\frac{n}{2}\rfloor}(-1)^j\cdot\binom{n}{2j}\cdot\sin^{2j}\alpha\cdot\cos^{n-2j}\alpha\\ \sum_{j=0}^{\lfloor\frac{n-1}{2}\rfloor}(-1)^j\cdot\binom{n}{2j+1}\cdot\sin^{2j+1}\alpha\cdot\cos^{n-2j-1}\alpha\end{pmatrix}$

...

2.5 Notwendige vs. redundante Parametrisierung

2.5.1 AWP-Input

- Anfangskoordinaten und -richtung (x_A, y_A, t_A)
- Segmentlängen $L^{(i)}$
- Anfangs- & Endkrümmungen $k_A^{(i)}$, $k_E^{(i)}$;
  Krümmungsverläufe $k^{(i)}(s) : [0, L^{(i)}] \to [k_A^{(i)}, k_E^{(i)}]$
- Knickwinkel $\delta^{(i)}$
- Endkoordinaten und -richtung (x_E, y_E, t_E) → Randwertproblem???

...

2.5.2 Tangentenschnitte

DER Vorteil: unabhängig eventueller Berechnungsprobleme bleibt der grobe Trassenverlauf erhalten!

...

2.6 Segmentberechnung

2.6.1 a

2.6.2 b

2.7 Trassenfindung

2.7.1 a

2.7.2 b

2.8 Trassenoptimierung

Standardsituation: es gibt eine (rechnerische) Trasse im betrachteten Bereich, diese trifft aber die Lage-vor-Ort eher gering. Es gibt wohl nicht-akzeptable Gleislagefehler. Der Ad-Hoc-Ansatz: Anpassung der Trassenelemente in Länge und Krümmung, ggf. auch Krümmungsverlauf = TransitionTyp (diesen aber händisch).

Wir betrachten eine Trasse $T = (C_0, \ldots, C_N) : \mathbb{R} \to \mathbb{R}^2$, beginnend mit $(T, T')(0) = (p_A, t_A)$ und endend mit $(T, T')(\sum l_i) = (p_E, t_E)$. Die C_i sind Kurven (alternierend Fixx - Flex - Fixx), mit stetig-(f.ü.)-glatten Übergängen. Die Flex-Segmente 'erben' ihre Krümmungsparameter (k_A, k_E) von den sie umgebenden Fixx-Elementen (das macht die Übergänge dann f.ü. glatt 2. Ordnung!). Die praxisrelevanten Nebenbedingungen sind alle stations-unabhängig - ausser eben $T(\sum l_i)$!

Also betrachten wir eigentlich eine Trasse als bogenlängen-parametrisierte Multikurve nicht nur im Eingang *station*, sondern auch abhängig der Längen-, Krümmungs- und / oder Delta-Werte: $T[l_0, ..., l_N, k_0, k_2 \ldots, k_{N-2}, k_N, \delta_1, \ldots, \delta_{N-1}](s) \to \mathbb{R}^2$ ...

Welche 'praxis-relevanten Nebenbedingungen'? Das sind zum einen fixe Längen-, Krümmungs- und / oder Delta-Werte (z.B. schlicht = 0), und 'Abstände'. Punktabstände sind Abfallprodukt der System-Transformation - im Verständnis einer Trasse als gekrümmtem Koordinatensystem. Dabei ist die Transformation aus Trassensystem in das Grosskoordinatensystem recht simpel ($T.\texttt{trForm} : (s, q) \mapsto (x, y)$). Die umgekehrte Richtung dagegen, $T.\texttt{umForm} : (x, y) \mapsto (s, q)$, ist die Herausforderung!

Desweiteren sind die Abstandsbedingungen meist oder immer 'Box'-Beschränkungen, d.h. der vorgegebene Sollwert ist eigentlich ein Paar aus lowerBound und upperBound. Das wird vertieft werden müssen. Oder aber sie gehen analog Ausgleichsrechnung via kleinste-Quadrate o.ä. in die Kostenfunktion ein.

Wie 'funktioniert' die Minimierung der Nullfunktion $f(x) \equiv 0$, sprich es gibt nur Restriktionen??? $\to$ das wäre der Fall wenn es gilt die Punktabstände zu minimieren unter den 'üblichen' Trassen-inheränten Restriktionen, und oder aber es sind keine Punkte im Input (weil nur die Trasse wieder scharf gemacht werden soll!)

Was haben wir denn? Die Elementlängen sind entweder gleich Null oder echt größer, bei den Krümmungen etwas einfacher, entweder gleich Null resp. konstant oder egal. Und die Deltas wiederum entweder Null oder egal. In Summe: fixe Werte (Gleichheitsrestriktion) oder beliebig, ausser den Längen — dort dann echt ungleich.

Für die Anschlusspunktbedingung ganz klar Gleichheitsrestriktion!

Was aber mit den gegebenenfalls boxed Punktbeschränkungen. Für mehr Flexibilität sind diese besser in den Kosten untergebracht — schlussendlich eine Soße: als Restriktionen werden sie durch die Multiplikatoren ebenso quadriert in die Lagrange-Funktion und deren Minimierung eingebracht, nur dort als muss in der Box. Als kleinste Quadrate in der Original Kostenfunktion auch, aber nicht als muss in der Box. Schaun wa mal!

Nochmal $f(x) \equiv 0$, also keine Abstände, nur Trassenzwänge. Wie schaut das aus?

$$\begin{aligned} 0 &\rightarrow \min \\ g(x) &\leq \Theta \\ h(x) &= \Theta \end{aligned}$$

Dann

$$\mathcal{L}(x, \mu, \nu) = \mu \cdot g(x) + \nu \cdot h(x) \rightarrow \min$$

...

2.8.1 und noch mal

$$\begin{aligned} f(x) &\rightarrow \min \\ g(x) &\leq \Theta \\ h(x) &= \Theta \end{aligned}$$

mit $f(x) = \sum L_i$; g trägt die $L_i \geq 0$ sowie punktuelle Abstände; und h hält zum Einen die Systemforderung $T(0) - T(\sum L_i) = p_E - p_A$ und $T'(0) - T'(\sum L_i) = t_E - t_A$, und mindestens weitere direkte Bedingen wie $L_j = 0 \quad \forall j \in J$

Netterweise ist die Richtungsforderung recht simpel!

Im weiter unten beschriebenen SQP-Verfahren werden die Linearisierungen der beteiligten Funktionen in x_k benötigt. Die Hesse-Matrix der Lagrange'schen dagegen nicht - sie wird durch eine numerisch günstigere Form ersetzt.

Aktuell betrachten wir x als Vektor $(L_1, \ldots, L_{2n+1}; K_1, K_3, K_{2n+1})$, also ohne Knicke. Wie lauten dann die Jacobi's der mindest-nötigen f, g und h aus?

f: wohl kein Problem!

g: nett für die $L_i \geq 0$, blöd für die punktuellen Abstände!!!

h: falls alle genutzten Krümmungsfunktionen explizit integrierbar sind, dann ergibt dich $t_E - t_A$ aus der Summe der Richtungsänderungen über alle Elemente, in linearer Abhängigkeit der x-Einträge –¿ nett! Blöd: $p_E - p_A$ braucht wohl die explizite Neu-Integration der "gestörten" Elemente ... Schon das Bestimmen von z.B.

$$\frac{d}{dL_i} \int_0^{L_i} \begin{pmatrix} \cos \\ \sin \end{pmatrix} (\Delta t_i(s)) ds$$

scheint etwas tricky! Ableiten einer Integral-Grenze ... Wow! Eine erste Idee wäre die Umschreibung auf die normierten τ- oder sogar κ-Funktionen. Integralgrenzen sind dann 0 und 1, und L_i sowie ΔK tauchen 'nur' als Faktoren auf:

$$\frac{d}{dL_i} A(L_i, \Delta K) \cdot \int_0^1 \begin{pmatrix} \cos \\ \sin \end{pmatrix} (B(L_i, \Delta K) \cdot \Delta\tau_i(\sigma)) d\sigma$$

Wenn dann der Operatortausch simpel wäre ...

$$\int_0^1 \left(\frac{d}{dL_i} \left(A(L_i, \Delta K) \cdot \begin{pmatrix} \cos \\ \sin \end{pmatrix} (B(L_i, \Delta K) \cdot \Delta\tau_i(\sigma)) \right) \right) d\sigma$$

Verstecken wir ΔK in A und B, dann ergibt sich mit Produktregel:

$$\int_0^1 \left(A'_K(L_i) \cdot \begin{pmatrix} \cos \\ \sin \end{pmatrix} (B_K(L_i) \cdot \Delta\tau_i(\sigma)) + A_K(L_i) \cdot \frac{d}{dL_i} \left(\begin{pmatrix} \cos \\ \sin \end{pmatrix} (B_K(L_i) \cdot \Delta\tau_i(\sigma)) \right) \right) d\sigma$$

Der erste Summand

$$A'_K(L_i) \cdot \int_0^1 \left(\begin{pmatrix} \cos \\ \sin \end{pmatrix} (B_K(L_i) \cdot \Delta\tau_i(\sigma)) \right) d\sigma$$

scheint noch recht entspannt zu sein. Der zweite

$$A_K(L_i) \cdot B'_K(L_i) \cdot \int_0^1 \begin{pmatrix} -\sin \\ \cos \end{pmatrix} (B_K(L_i) \cdot \Delta\tau_i(\sigma)) \cdot \Delta\tau_i(\sigma) d\sigma$$

nicht mehr.

...

2.8.2 Anwendungsfälle

Einrechnen

Seien eine Gerade (am Anfang) und ein Bogen (am Ende) gegeben, gesucht sind alle Längen inklusive des vermittelnden Übergangsbogen zwischen Gerade und Bogen.

Prinzipiell sind drei Fälle in Ausdruck der Abrückung zu unterscheiden:

> 0**:** Der eigentlich interessante Fall mit eindeutiger Lösung!

$= 0$**:** Die Länge des Übergangsbogen beträgt Null, es ist also de facto ein Bogenwechsel.

< 0**:** keine Lösung!

Ersterer: Gerade und Bogen sind mit der nötigen Parametrisierung inklusive Anfangs- respektive Endpunkt und -richtung bekannt, ebenso der Typ des Übergangsbogen (vielleicht durch Wechsel Clothoide nach Bloss ...). Weitergehende Zwänge wie anzurechnende Punkte nicht nötig, da überflüssig.

$$
\begin{aligned}
0 &\rightarrow \min_{x=(l_g,l_t,l_b,k_g,d_t,k_b)} ! \\
\text{s.t.} & \\
l_g &> 0 \\
l_t &> 0 \\
l_b &> 0 \\
k_g &= 0 \\
d_t &= 0 \\
k_b &= \text{const} \\
tra.\texttt{trForm}(\sum l_*, q = 0) &= (p_E, t_E)
\end{aligned}
$$

[2] Formal

$$
g(x) = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix}(x)
$$

und

$$
h(x) = \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix}(x)
$$

mit

$$
\begin{aligned}
g_1(x) &= -l_g \\
g_2(x) &= -l_t \\
g_3(x) &= -l_b \\
h_1(x) &= k_g \\
h_2(x) &= d_t \\
h_3(x) &= k_b \\
H(x) &= tra.\texttt{trForm}(\sum l_*, q = 0) - (p_E, t_E)
\end{aligned}
$$

und

$$
g(x) \leq \Theta, \quad h(x) = \Theta, \quad H(x) = \Theta
$$

Die Lagrange'sche ergibt sich als

$$
\begin{aligned}
\mathcal{L}(x; \mu, \nu, \mathcal{N}) &= 0 \\
&+ (\mu_1, \mu_2, \mu_3) \cdot g(x) \\
&+ (\nu_1, \nu_2, \nu_3) \cdot h(x) \\
&+ (\mathcal{N}_x, \mathcal{N}_y, \mathcal{N}_t) \cdot H(x)
\end{aligned}
$$

[2] >??? Praktisch nehmen wir hier statt dessen $\geq$ und schließen das aktiv-Werden aus ...

Damit

$$\begin{aligned}\nabla_x \mathcal{L} &= \mu \cdot \nabla g + \nu \cdot \nabla h + \mathcal{N} \cdot \nabla H \\ \nabla_\mu \mathcal{L} &= g \\ \nabla_\nu \mathcal{L} &= h \\ \nabla_{\mathcal{N}} \mathcal{L} &= H\end{aligned}$$

...

first choosing the initial iterate $(x^{[0]}, \mu^{[0]}, \nu^{[0]})$, then calculating H_0 instead of $\nabla^2 \mathcal{L}(x^{[0]}, \mu^{[0]}, \nu^{[0]})$, and $\nabla \mathcal{L}(x^{[0]}, \mu^{[0]}, \nu^{[0]})$.

Then the (QP_k) subproblem

$$\begin{array}{lrcll} \min\limits_d & & & \nabla f(x^{[k]})^T d & + \quad \frac{1}{2} d^T H_k d \\ \text{s.t.} & g(x^{[k]}) & + & \nabla g(x^{[k]})^T d & \leq \quad 0 \\ & h(x^{[k]}) & + & \nabla h(x^{[k]})^T d & = \quad 0. \end{array}$$

is built and solved to find the Newton step direction $d^{[k]}$ which is used to update the parent problem iterate using $x^{[k+1]} = x^{[k]} + d^{[k]}$. This process is repeated for $k = 0, 1, 2, \ldots$ until the parent problem satisfies a convergence test.

Remark 2.8.2.1.

$$\begin{aligned} x &\in \mathbb{R}^6 \\ \mu &\in \mathbb{R}^3 \\ \nu &\in \mathbb{R}^3 \\ \mathcal{N} &\in \mathbb{R}^3 \\ d &\in \mathbb{R}^{6+3+3+3=15} \end{aligned}$$

Howto solve a problem (QP) like

$$\begin{array}{lrcll} & \frac{1}{2} d^T W d & + & c^T d & \rightarrow \quad ! \min_d \\ \text{s.t.} & Ad & - & u & \leq \quad \Theta \\ & Bd & - & v & = \quad \Theta. \end{array}$$

or the ...

...

Index