

Teil II

Nichtlineare Optimierung

Kapitel 1

Einleitung

Bemerkung 1.1 Problemklasse. In diesem Abschnitt wird die Optimierung von Funktionen

$$\min_{\mathbf{x} \in \Omega} \{f(\mathbf{x})\}$$

betrachtet, wobei $\Omega \subset \mathbb{R}^n$ eine abgeschlossene Menge und $f : \Omega \rightarrow \mathbb{R}$ eine gegebene Funktion ist. Die Funktion f heißt Zielfunktion und Ω sei durch endlich viele Nebenbedingungen beschrieben. Bei den betrachteten Problemen können sowohl die Zielfunktion als auch die Nebenbedingungen nichtlinear sein. \square

Beispiel 1.2 Ausgleichsrechnung. Ein konkretes Beispiel für eine Aufgabe der Nichtlinearen Optimierung ist die Ausgleichsrechnung. Zur mathematischen Formulierung von Gesetzmäßigkeiten, die ein technisches oder physikalisches Phänomen beschreiben, wird häufig eine Hypothese über den möglichen funktionalen Zusammenhang bekannter und beobachteter Variablen formuliert. Diese enthält im allgemeinen noch freie Parameter und hat etwa die Gestalt

$$z = g(\mathbf{x}, \boldsymbol{\lambda}), \quad \mathbf{x} \in \mathbb{R}^n, \quad \boldsymbol{\lambda} \in \mathbb{R}^p.$$

In diesem Modell ist $\boldsymbol{\lambda}$ ein unbekannter Parametervektor mit p Komponenten und z ist eine Variable, deren Abhängigkeit modelliert werden soll.

Nach Auswertung von m Experimenten kennt man m Paare von Variablenwerten (z_i, \mathbf{x}_i) , die unter Berücksichtigung möglicher Beobachtungsfehler ε_i die funktionale Abhängigkeit annähernd erfüllen sollen. Das heißt, bei passenden Parametern muss gelten

$$z_i = g(\mathbf{x}_i, \boldsymbol{\lambda}) + \varepsilon_i, \quad i = 1, \dots, m.$$

Um möglichst kleine Abweichungen ε_i zu erhalten, versucht man, die unbekannt Parameter $\boldsymbol{\lambda}$ optimal anzupassen, indem man entsprechende Optimierungsaufgaben löst, zum Beispiel

$$\min_{\boldsymbol{\lambda} \in \mathbb{R}^p} \sum_{i=1}^m (z_i - g(\mathbf{x}_i, \boldsymbol{\lambda}))^2$$

(Methode der kleinsten Quadrate) oder

$$\min_{\boldsymbol{\lambda} \in \mathbb{R}^p} \sum_{i=1}^m \omega_i |z_i - g(\mathbf{x}_i, \boldsymbol{\lambda})|^r,$$

wobei $r \geq 1$ eine ganze Zahl ist und $\boldsymbol{\omega} \geq \mathbf{0}$. \square

Beispiel 1.3 Standortplanung. Eine Firma möchte n neue Lager der Kapazität a_i , $i = 1, \dots, n$, errichten. Gegeben sind die Standorte der Abnehmer (α_j, β_j) sowie

der Bedarf b_j , $j = 1, \dots, m$. In einigen Gebieten darf zudem nicht gebaut werden (ε_k -Umgebungen von (γ_k, δ_k) , $k = 1, \dots, p$).

Gesucht sind die Standorte der Lager (x_i, y_i) , die den Lieferplan z optimieren.

Mit den Bezeichnungen

- z_{ij} – Transportmenge vom Lager i zum Abnehmer j ,
- d_{ij} – Entfernung Lager i und Abnehmer j ,
- Δ_{ik} – Entfernung Lager i zum Mittelpunkt (γ_k, δ_k) der Verbotzone k ,

lautet die Optimierungsaufgabe wie folgt:

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^m d_{ij} z_{ij} &\rightarrow \min ! \\ \sum_{j=1}^m z_{ij} &\leq a_i, \quad i = 1, \dots, n \\ \sum_{i=1}^n z_{ij} &= b_j, \quad j = 1, \dots, m \\ \Delta_{ik} &\geq \varepsilon_k, \quad i = 1, \dots, n, \quad k = 1, \dots, p, \\ z_{ij} &\geq 0, \quad i = 1, \dots, n, \quad j = 1, \dots, m. \end{aligned}$$

Diese Aufgabe kann für verschiedene Entfernungsmaße gestellt werden, etwa für

$$d_{ij} = |x_i - \alpha_j| + |y_i - \beta_j|$$

oder die Euklidische Entfernung

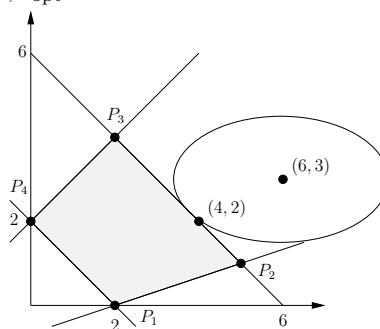
$$d_{ij} = \sqrt{(x_i - \alpha_j)^2 + (y_i - \beta_j)^2}.$$

Wird sowohl für d_{ij} als auch für Δ_{ik} die Euklidische Entfernung gewählt, so erhält man eine Optimierungsaufgabe mit quadratischer Zielfunktion und quadratischen Nebenbedingungen. \square

Beispiel 1.4 Quadratische Zielfunktion über konvexem Polyeder. Wir betrachten ein nichtlineares Programm mit quadratischer Zielfunktion und linearen Nebenbedingungen:

$$\begin{aligned} z &= (x_1 - 6)^2 + 2(x_2 - 3)^2 \rightarrow \min ! \\ x_1 + x_2 &\geq 2 \\ x_1 - x_2 &\geq -2 \\ x_1 + x_2 &\leq 6 \\ x_1 - 3x_2 &\leq 2 \\ \mathbf{x} &\geq \mathbf{0}. \end{aligned}$$

Durch die Zielfunktion werden konzentrische Ellipsen mit Mittelpunkt $(6, 3)^T$ beschrieben. Die optimale Lösung ist ein Punkt auf dem Rand $\partial\Omega$ von Ω , aber kein Eckpunkt: $\mathbf{x}_{\text{opt}} = (4, 2)^T$, $z_{\text{opt}} = 6$.



Betrachtet man eine andere Zielfunktion

$$z = (x_1 - 2)^2 + 2(x_2 - 2)^2 \rightarrow \min !,$$

dann ist das Optimum offensichtlich $\mathbf{x}_{\text{opt}} = (2, 2)^T$, $z_{\text{opt}} = 0$. Damit liegt das Optimum im Inneren von Ω .

Eine andere quadratische Zielfunktion ist

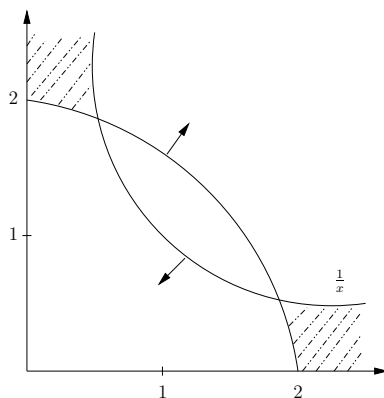
$$z = 2(x_1 - 2)^2 + (x_2 - 2)^2 \rightarrow \max !$$

Bei dieser Zielfunktion gibt es eine Ellipse, die gleichzeitig durch P_1 und P_3 geht. Beide Eckpunkte von Ω sind lokales Maximum mit $z = 4$. Ein weiteres lokales Maximum erhält man in P_4 mit $z = 8$. Das globale Maximum wird jedoch in P_2 mit $z = 19$ angenommen. Ein simplexartiges Vorgehen, wobei man von Eckpunkt zu Eckpunkt geht und in jedem Schritt den Zielfunktionswert verbessert (vergrößert), d.h. $P_1 \rightarrow P_4$ oder $P_3 \rightarrow P_4$ führt hier nicht zum Ziel. \square

Beispiel 1.5 Unzusammenhängender zulässiger Bereich. Der zulässige Bereich Ω sei definiert durch

$$\begin{aligned} x_1^2 + x_2^2 &\geq 4 \\ x_1 x_2 &\leq 1 \\ \mathbf{x} &\geq \mathbf{0}. \end{aligned}$$

Der zulässige Bereich besteht aus zwei getrennt liegenden Teilen. Beide sind nicht konvex. Selbst bei einer linearen Zielfunktion können lokale Minima auftreten, die keine globalen sind.



\square

Bemerkung 1.6 In der Vorlesung werden u.a. folgende Problemklassen innerhalb der nichtlinearen Programme nicht betrachtet:

- mehrere Zielfunktionen,
- unendlich viele Nebenbedingungen (semi-infinites Programme),
- stochastische Daten (stochastische Optimierung).

\square

Kapitel 2

Nichtlineare Optimierung ohne Nebenbedingungen

2.1 Minimierung nichtglatter Funktionen in einer Variablen

Bemerkung 2.1 Motivation. Die Lösung nichtlinearer Optimierungsaufgaben in einer Dimension tritt als Teilproblem in vielen Verfahren zur Lösung nichtlinearer Optimierungsprobleme in höheren Dimensionen auf.

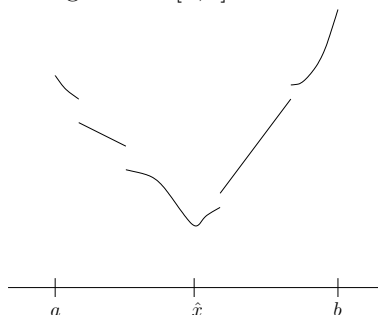
Seien $I \subset \mathbb{R}$ ein gegebenes abgeschlossenes Intervall und $f : I \rightarrow \mathbb{R}$ eine gegebene Funktion. Dann sucht man ein $\hat{x} \in I$, welches

$$f(\hat{x}) = \min_{x \in I} f(x) \quad (2.1)$$

erfüllt. Um ein solches Minimum mit Hilfe eines Verfahrens effizient bestimmen zu können, müssen einige einschränkende Bedingungen an $f(x)$ gestellt werden. Je nach Eigenschaften der Funktion, ist dann ein entsprechendes Verfahren zu wählen.

In diesem Abschnitt werden Verfahren zur Bestimmung des Minimums von nichtglatten Funktionen in einer Variablen im Detail vorgestellt, wobei der zulässige Bereich nicht durch Nebenbedingungen eingeschränkt ist. Die Minimierung eindimensionaler glatter Funktionen ohne Nebenbedingungen kann auf die Bestimmung von Nullstellen zurückgeführt werden. Verfahren zur Lösung dieser Aufgabe sind bereits aus der Grundvorlesung *Praktische Mathematik* bekannt. \square

Definition 2.2 Unimodale Funktion. Eine Funktion $f : I \rightarrow \mathbb{R}$ heißt unimodal (einwellig), falls für ein $\hat{x} \in I = [a, b]$ gilt, dass $f(x)$ streng monoton fallend auf $[a, \hat{x}]$ und streng monoton steigend auf $[\hat{x}, b]$ ist. \square



Bemerkung 2.3 Eigenschaften unimodaler Funktionen. Unimodale Funktionen erlauben nach Auswertung der Funktion an zwei Stellen $x < y$ mit $a < x <$

$y < b$ die Reduktion des Intervalls, in der man das Minimum zu suchen hat:

- 1. Fall: $f(a) \leq f(x)$, dann ist das Minimum in $[a, x]$.
- 2. Fall: $f(a) > f(x) > f(y)$. Dann kann das Minimum nicht in $[a, x]$ sein.
- 3. Fall: $f(a) > f(x)$ und $f(x) < f(y)$ und $f(y) < f(b)$. Dann kann das Minimum nicht in $[y, b]$ sein.
- 4. Fall: $f(y) \geq f(b)$. Dann ist das Minimum in $[y, b]$.

Alle anderen Fälle widersprechen der Definition einer unimodalen Funktion. \square

Bemerkung 2.4 Der goldene Schnitt. Wir wollen jetzt durch die rekursive Anwendung dieser Beobachtung einen Algorithmus zur Approximation des Punktes konstruieren, an dem das Minimum angenommen wird. Dafür stellen wir zunächst einige Forderungen an den Algorithmus:

- 1.) Im ersten Schritt sollen zwei Funktionswertauswertungen, in allen weiteren Schritten nur eine weitere Funktionswertauswertung im Restintervall verwendet werden.
- 2.) Die Punkte x und y sind stets symmetrisch im Restintervall zu wählen.
- 3.) Der Reduktionsfaktor σ für die Intervalllänge sei konstant, $\sigma \in (0.5, 1)$. Das heißt, der Quotient der Länge des neuen Intervalls und der Länge des vorherigen Intervalls ist $\sigma = \text{const.}$

Diese Forderungen besitzen gewisse Ähnlichkeiten zu den Eigenschaften des Bisektions-Verfahrens zur Berechnung von Nullstellen. Auch dort wird die Nullstelle in einem Intervall eingeschachtelt, man hat am Anfang zwei Funktionswerte und später pro Schritt einen zu berechnen und der Reduktionsfaktor ist konstant $\sigma = 0.5$.

Durch die symmetrische Wahl von x, y im Restintervall wird der Reduktionsfaktor unabhängig von der konkreten Funktion $f(x)$. Verlangt man einen konstanten Reduktionsfaktor in allen Iterationen, so ist dieser eindeutig bestimmt.

Um diesen von $f(x)$ unabhängigen Reduktionsfaktor zu bestimmen, genügt es eine streng monoton wachsende Funktion $f : [0, 1] \rightarrow \mathbb{R}$ zu betrachten. Damit ist $\hat{x} = 0$. O.B.d.A. sei $0 < x < y < 1$. Wegen der verlangten Symmetrie folgt $y = \sigma, x = 1 - \sigma$.

Im ersten Schritt wird das Intervall $[0, 1]$ auf das Restintervall $[0, y]$ reduziert. In diesem Restintervall wird dann ein Punkt z symmetrisch zu x gewählt. Damit ist gesichert, dass im nächsten Schritt der Punkt x eine der neuen Stützstellen ist, man den Funktionswert $f(x)$ noch einmal verwenden kann und nur den Funktionswert $f(z)$ neu berechnen muss. Je nach Wahl von y gilt $z < x$ oder $z \geq x$. Diese Fälle entsprechen der Wahl von σ aus unterschiedlichen Intervallen.

1. Fall: $\sigma < 2/3$. Dann gilt $x = 1 - \sigma > 1/3$ und damit

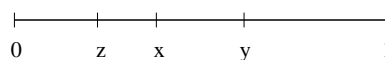
$$z = y - x = \sigma - (1 - \sigma) < 1/3 < x.$$

Bei der ersten Gleichung wurde die Symmetrie im Restintervall verwendet. Konstante Reduktion bedeutet nun

$$\frac{x}{y} = \frac{y}{1} \implies 0 = y^2 - x = \sigma^2 + \sigma - 1.$$

Die Lösung dieser quadratischen Gleichung ist

$$\sigma = \frac{1}{2} (\sqrt{5} - 1) \approx 0.618 < \frac{2}{3}.$$



2. Fall: $\sigma \geq 2/3$. Wird analog zum ersten Fall behandelt. Man erhält keine Lösung. *Übungsaufgabe*

Der gefundene Reduktionsfaktor ist also unter den obigen Annahmen der einzig mögliche und er legt den Algorithmus fest. \square

Algorithmus 2.5 Goldener Schnitt.

1. *Initialisierung.*

$i := 0; a_0 := a; b_0 := b;$
 $x_i := a_i + (1 - \sigma)(b_i - a_i); y_i := a_i + \sigma(b_i - a_i);$
 $fx := f(x_i); fy := f(y_i);$

2. *Iteration.*

Falls $fx < fy$, dann

$a_{i+1} := a_i; b_{i+1} := y_i;$
 $x_{i+1} := a_{i+1} + (1 - \sigma)(b_{i+1} - a_{i+1}); y_{i+1} := x_i;$
 $fy := fx; fx := f(x_{i+1});$

sonst

$a_{i+1} := x_i; b_{i+1} := b_i;$
 $x_{i+1} := y_i; y_{i+1} := b_{i+1} - (1 - \sigma)(b_{i+1} - a_{i+1});$
 $fx := fy; fy := f(y_{i+1});$

$i := i + 1;$

3. *Abbruch.*

Falls $(b_i - a_i)/2 > \varepsilon$, dann
gehe zu 2.

sonst

$\tilde{x} := (a_i + b_i)/2;$
stop

Man nimmt \tilde{x} als Approximation an \hat{x} . □

Bemerkung 2.6 Zum Algorithmus des Goldenen Schnittes. In der Praxis führt man erst den Vergleich für den Abbruch durch und berechnet dann den neuen Funktionswert. Das spart im letzten Schritt eine Funktionswertberechnung. Diese Einsparung kann sich akkumulieren, falls man das Verfahren im Rahmen eines komplexen Problems oft aufruft.

Algorithmus 2.5 bricht ab, sobald der Funktionswert, für den das Minimum angenommen wird, in einem Intervall der Länge 2ε eingeschachtelt ist. Da \tilde{x} der Mittelpunkt des Intervalls ist, gilt für den Fehlern $|\hat{x} - \tilde{x}| \leq \varepsilon$. Da nach $(n + 1)$ Funktionswertauswertungen (n Iterationen) die Länge des Restintervalls $\sigma^n(b - a)$ ist, kann man n in Abhängigkeit von ε a priori abschätzen. □

Bemerkung 2.7 Die Fibonacci-Suche. Wir wollen jetzt auf die Konstanz der Reduktion der Intervalllänge verzichten und untersuchen, ob es Modifikationen von Algorithmus 2.5 gibt, die bei gleicher Anzahl von Funktionswertauswertungen ein kleineres Restintervall liefern. Sei L_n die maximale Länge eines Intervalls, welches man mittels n Funktionswertauswertungen in L_n auf $L_1 = 1$ reduzieren kann.

Beim Bisektions-Verfahren ist beispielsweise $L_5 = 8$. Mit fünf Funktionsauswertungen kann man vier Iterationen durchführen (in der ersten Iteration braucht man zwei Funktionswerte). Halbiert man ein Intervall der Länge 8, so erhält man ein Intervall der Länge 4. Halbiert man dieses, so ist die Intervalllänge 2 und die letzte Halbierung führt auf ein Intervall der Länge 1.

Seien $x < y$ die beiden Stützstellen in L_n . Durch jede dieser Stützstellen wird L_n in zwei Teilintervalle zerlegt. Betrachtet man die Stützstelle x , so enthält das linke Teilintervall höchstens $(n - 2)$ Stützstellen des Gesamtprozesses (alle außer x und y) und das rechte Teilintervall höchstens $(n - 1)$ der Stützstellen (alle außer x). Entsprechend sind die Längen dieser Teilintervalle höchstens L_{n-2} beziehungsweise L_{n-1} . Damit gilt

$$L_n \leq L_{n-1} + L_{n-2}, \quad n = 2, 3, \dots \quad (2.2)$$

Da man mit einer Funktionsauswertungen noch keine Intervallreduktion durchführen

kann, gelten $L_0 = L_1 = 1$. Mit der obigen Abschätzung hat man das größtmögliche Intervall L_n , falls eine Lösung der Ungleichung (2.2) diese sogar als Gleichung erfüllt.

Die Lösung der Differenzgleichung

$$F_0 = F_1 = 1, \quad F_n = F_{n-1} + F_{n-2}$$

ist bekannt. Es sind die sogenannten Fibonacci-Zahlen. Die Darstellung dieser Zahlen in geschlossener Form ist

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

Für den Algorithmus der Fibonacci-Suche setzen wir noch $F_{-2} := 1, F_{-1} := 0$.

Bei der Fibonacci-Suche wird die geforderte Länge des Intervalls, in dem \hat{x} eingeschachtelt werden soll, vorgegeben. Die Iterationspunkte bei einem beliebigen Startintervall $[a, b]$ ergeben sich durch lineare Transformation mit dem Faktor

$$h = \frac{b - a}{F_n}. \quad (2.3)$$

Die Länge des Restintervalls in der Iteration $i = 0, 1, \dots, n - 1$ ist $F_{n-i}h$. Damit ist die Intervallreduktion

$$\frac{F_{n-i-1}h}{F_{n-i}h} = \frac{F_{n-i-1}}{F_{n-i}},$$

also der Quotient zweier aufeinanderfolgender Fibonacci-Zahlen. Außerdem besitzt das $(n - 1)$ -ste Intervall die Länge $F_1h = h$ und das vorletzten Restintervall $[a_{n-2}, b_{n-2}]$ die Länge $F_2h = 2h$. Die Endpunkte des $(n - 1)$ -sten Intervalls sind ein Endpunkt von $[a_{n-2}, b_{n-2}]$ und eine Stützstelle von $[a_{n-2}, b_{n-2}]$. Eine Stützstelle ist also der Mittelpunkt $a_{n-2} + h$ und wegen der Symmetrie der Stützstellen folgt $x_{n-2} = y_{n-2} = a_{n-2} + h$. Die Stützstellen fallen somit zusammen und die Iteration muss beendet werden. Für den Punkt \hat{x} , in dem das Minimum angenommen wird, gilt somit $\hat{x} \in [a_{n-2}, b_{n-2}]$, woraus $|\hat{x} - x_{n-2}| \leq h$ folgt. Damit ist durch (2.3) bei vorgegebenem h auch die Anzahl der Iterationen n gegeben. \square

Algorithmus 2.8 Fibonacci-Suche.

1. Initialisierung.

gebe h vor, bestimme n , berechne die Fibonacci-Zahlen
 F_0, \dots, F_n
 $i := 0$; $a_0 := a$; $b_0 := b$; $h := (b - a)/F_n$;
 $x_0 := a_0 + F_{n-2}h$; $y_0 := a_0 + F_{n-1}h$;
 $fx := f(x_0)$; $fy := f(y_0)$;

2. Iteration.

Falls $fx < fy$, dann
 $a_{i+1} := a_i$; $b_{i+1} := y_i$;
 $x_{i+1} := a_{i+1} + F_{n-i-3}h$; $y_{i+1} := x_i$;
 $fy := fx$; $fx := f(x_{i+1})$;
sonst
 $a_{i+1} := x_i$; $b_{i+1} := b_i$;
 $x_{i+1} := y_i$; $y_{i+1} := b_{i+1} - F_{n-i-3}h$;
 $fx := fy$; $fy := f(y_{i+1})$;
 $i := i + 1$;

3. Abbruch.


```

Falls  $i < n - 2$ , dann
    gehe zu 2.
sonst
     $\tilde{x} := x_i$ ;
    stop

```

Man nimmt \tilde{x} als Approximation an \hat{x} .

Analog wie beim Goldenen Schnitt kann man in der letzten Iteration eine Funktionswertberechnung sparen. \square

Beispiel 2.9 Fibonacci–Suche. Wir betrachten die Funktion $f(x) = \sin(x - 2)$ auf $[a, b] = [0, 2]$. Auf diesem Intervall ist die Funktion $f(x)$ unimodal und sie nimmt ihr Minimum in $\hat{x} = 2 - \pi/2 \approx 0.4292037$ an. Wir wollen die Fibonacci–Suche mit $n = 6$ durchführen. Die Fibonacci–Zahlen F_0, \dots, F_6 sind 1, 1, 2, 3, 5, 8, 13. Daraus folgt, dass man ein Restintervall der Länge

$$h = \frac{2}{13} \approx 0.1538462$$

findet.

Die im Verfahren auftretenden Stützstellen und Intervallgrenzen sind alle von der Form $a + kh$ mit $k \in \{0, 1, 2, 3, 5, 8, 13\}$. Für eine Stützstelle $t \in [a, b]$ nennt man die entsprechende ganze Zahl $k(t) := (t - a)/h$ den Fibonacci–Index von t . Beim Start gilt $k(a_0) = 0$, $k(x_0) = F_{n-2} = F_4 = 5$, $k(y_0) = F_{n-1}$, $k(b_0) = F_n = F_6 = 13$.

Während der Iteration werden die Funktionswerte $f(x_i)$ und $f(y_i)$ verglichen. Die Variable mit dem kleinerem Funktionswert bleibt Stützstelle, die mit dem größeren Funktionswert wird neue Intervallgrenze. Der Fibonacci–Index der neuen Stützstelle ergibt sich wegen der symmetrischen Lage der Stützstellen im Restintervall aus $k(x_{i+1}) - k(a_{i+1}) = k(b_{i+1}) - k(y_{i+1})$. Ordnet man alle Werte einer Iteration zeilenweise in einer Tabelle an, so verschieben sich diese Werte beim Übergang zur nächsten Iteration nach rechts beziehungsweise nach links ab der Position der neuen Stützstelle.

i	$k(a_i)$	x_i	$k(x_i)$	y_i	$k(y_i)$	$k(b_i)$	$f(x_i)$	$f(y_i)$
0	0	0.7692308	5	1.2307692	8	13	- 0.9427456	- 0.6955828
1	0	0.4615385	3	0.7692308	5	8	- 0.9994773	- 0.9427456
2	0	0.3076923	2	0.4615385	3	5	- 0.9926266	- 0.9994773
3	2	0.4615385	3	0.6153846	4	5	- 0.9994773	- 0.9827183
4	2	0.4615385	3	0.4615385	3	4	- 0.9994773	- 0.9994773

Mit $x_4 = 0 + 3h$ bricht das Verfahren ab und für das Minimum von $f(x)$ gilt

$$\hat{x} \in [0.4615385 - h, 0.4615385 + h].$$

\square

Bemerkung 2.10 Vergleich vom Goldenen Schnitt und Fibonacci–Suche, Anzahl der Iterationen. Aus

$$1 - \sigma = \sigma^2 \tag{2.4}$$

folgt induktiv (mit $F_{-2} = 1, F_{-1} = 0$)

$$\sigma^n = (-1)^n (F_{n-2} - F_{n-1}\sigma).$$

Übungsaufgabe Man rechnet direkt nach, dass

$$F_1 = 1 < \frac{1}{\sigma} < 2 = F_2, \quad F_2 < \frac{1}{\sigma^2} < F_3$$

gelten. Induktiv erhält man für $n > 1$ die Abschätzung

$$\frac{1}{F_n} > \sigma^n > \frac{1}{F_{n+1}}.$$

Außerdem findet man

$$\lim_{n \rightarrow \infty} \frac{F_n}{F_{n+1}} = \sigma.$$

Unter der Annahme, dass der rechte Grenzwert existiert, kann man diesen mit Hilfe von (2.4) berechnen. *Übungsaufgabe* Asymptotisch erhält man die gleiche Intervallreduktion und der Goldene Schnitt ist bei gleichem Aufwand demnach von kaum geringerer Genauigkeit. Auf der anderen Seite besitzt die Fibonacci-Suche den Nachteil, dass die gewünschte Genauigkeit a priori festgelegt werden muss, um n zu bestimmen.

Beide Verfahren reduzieren die Intervalllänge linear, das heißt es gilt

$$\frac{|b_{i+1} - a_{i+1}|}{|b_i - a_i|} \leq \lambda \quad \text{mit } \lambda \in (0, 1).$$

Dabei ist der Reduktionsfaktor λ beim Goldenen Schnitt durch σ gegeben und bei der Fibonacci-Suche durch eine Schranke für F_{n-1}/F_n , $n \geq 2$. Soll das n -te Intervall kleiner als ein gegebenes ε sein, so erhält man aus

$$\varepsilon \leq \lambda^n (b - a) = \lambda^n (b_0 - a_0)$$

die Abschätzung

$$n \geq \log \left(\frac{\varepsilon}{b - a} \right) / \log(\lambda).$$

□

2.2 Differenzierbare Funktionen in mehreren Dimensionen

Bemerkung 2.11 Notwendige Bedingung für ein Minimum, Berechnungsverfahren. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ mit $f \in C^1(\mathbb{R}^n)$ gegeben. Die notwendige Bedingung für ein lokales Minimum im Punkt $\hat{\mathbf{x}} \in \mathbb{R}^n$ ist

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

Zur Berechnung von möglichen Werten für $\hat{\mathbf{x}}$ hat man damit ein nichtlineares Gleichungssystem zu lösen. Verfahren zur Lösung nichtlinearer Gleichungssysteme wurden bereits in der Vorlesung *Praktische Mathematik* behandelt:

- Bisektion falls $n = 1$,
- Abstiegsverfahren (Gradientenverfahren, Verfahren des steilsten Abstiegs),
- Fixpunktiteration, wobei die nachfolgenden Verfahren Spezialfälle sind,
- Newton-Verfahren,
- vereinfachtes und Quasi-Newton-Verfahren.

Auf Abstiegsverfahren soll hier etwas näher eingegangen werden. □

2.2.1 Abstiegsverfahren

Im folgenden wird vorausgesetzt, dass $f(\mathbf{x})$ zweimal stetig differenzierbar ist.

Definition 2.12 Abstiegsverfahren, Abstiegsrichtung, Schrittweite. Abstiegsverfahren zur Berechnung von $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ besitzen folgende Gestalt:

- $\mathbf{x}^{(0)} \in \mathbb{R}^n$ sei gegeben,
- berechne für $k \geq 0$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}, \quad (2.5)$$

wobei $\mathbf{d}^{(k)}$ eine geeignete Suchrichtung, die sogenannte Abstiegsrichtung, ist und $\alpha_k > 0$ die Schrittweite.

Die Richtung $\mathbf{d}^{(k)}$ wird Abstiegsrichtung genannt, falls

$$\begin{aligned} \mathbf{d}^{(k)T} \nabla f(\mathbf{x}^{(k)}) &< 0 && \text{falls } \nabla f(\mathbf{x}^{(k)}) \neq \mathbf{0}, \\ \mathbf{d}^{(k)} &= \mathbf{0} && \text{falls } \nabla f(\mathbf{x}^{(k)}) = \mathbf{0}. \end{aligned} \quad (2.6)$$

□

Bemerkung 2.13 Abstiegsrichtung. Unter der Bedingung (2.6) existieren hinreichend kleine $\alpha_k > 0$, so dass

$$f(\mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}) < f(\mathbf{x}^{(k)}).$$

Aus der Taylor-Entwicklung folgt nämlich für $\boldsymbol{\xi} = \mathbf{x}^{(k)} + \theta \alpha_k \mathbf{d}^{(k)}$ mit $\theta \in (0, 1)$

$$f(\mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}) = f(\mathbf{x}^{(k)}) + \alpha_k \mathbf{d}^{(k)T} \nabla f(\boldsymbol{\xi}).$$

Falls α_k hinreichend klein ist, dann ist $\boldsymbol{\xi}$ hinreichend nahe an $\mathbf{x}^{(k)}$ und dann ist das Vorzeichen von $\mathbf{d}^{(k)T} \nabla f(\boldsymbol{\xi})$ das gleiche wie von $\mathbf{d}^{(k)T} \nabla f(\mathbf{x}^{(k)})$, woraus die Behauptung folgt. □

Beispiel 2.14 Abstiegsrichtungen. Einige Beispiele für Abstiegsrichtungen sind:

1. Newton-Verfahren

$$\mathbf{d}^{(k)} = -(H_f)^{-1}(\mathbf{x}^{(k)}) \nabla f(\mathbf{x}^{(k)}),$$

wobei $H_f(\mathbf{x}^{(k)})$ die Hesse-Matrix von $f(\mathbf{x})$ an der Stelle $\mathbf{x}^{(k)}$ ist.

2. Quasi-Newton-Verfahren, inexaktes Newton-Verfahren, vereinfachtes Newton-Verfahren

$$\mathbf{d}^{(k)} = -B_k^{-1} \nabla f(\mathbf{x}^{(k)}),$$

wobei B_k eine geeignete Approximation von $H_f(\mathbf{x}^{(k)})$ ist.

3. Gradienten-Verfahren, Verfahren des steilsten Abstieges

$$\mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)}),$$

also das Quasi-Newton-Verfahren mit $B_k = I$. Es gilt

$$\mathbf{d}^{(k)T} \nabla f(\mathbf{x}^{(k)}) = -\|\nabla f(\mathbf{x}^{(k)})\|_2^2 < 0 \quad \text{für } \nabla f(\mathbf{x}^{(k)}) \neq \mathbf{0}.$$

4. Verfahren der konjugierten Gradienten

$$\mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)}) + \beta_k \mathbf{d}^{(k-1)},$$

wobei die β_k Parameter sind, die so gewählt werden, dass die Suchrichtungen $\{\mathbf{d}^{(k)}\}$ bezüglich eines geeigneten Skalarproduktes orthogonal sind, siehe Bemerkung 2.26.

□

Bemerkung 2.15 Liniensuche. Nachdem man eine Abstiegsrichtung gewählt hat, bleibt das Problem, wie man α_k wählen soll.

- Eine Möglichkeit besteht in der Lösung des nichtlinearen Minimierungsproblems in einer Dimension: finde α_k , so dass

$$\phi(\alpha_k) = f\left(\mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}\right)$$

minimiert wird. *Übungsaufgabe Thm 7.3* Das geht nur in Spezialfällen mit vertretbarem Aufwand. Diese Spezialfälle sind aber durchaus wichtig, siehe Abschnitt 2.2.2.

- Man nutzt ein iteratives Verfahren, welches einen geeigneten Wert α_k bestimmt. Diese Herangehensweise wird Liniensuche genannt. □

Beispiel 2.16 Armijo–Liniensuche. Seien $f(\mathbf{x})$ eine zu minimierende Funktion, $\mathbf{x}^{(k)}$ die gegenwärtige Iterierte, $\mathbf{d}^{(k)}$ eine irgendwie berechnete Abstiegsrichtung, $\mu \in (0, 1)$ und $\gamma > 0$ eine Konstante. Man wählt $\alpha^{(0)} \geq \gamma \|\nabla f(\mathbf{x}^{(k)})\|_2$ und bestimme unter den Zahlen $\alpha^{(j)} = \beta^j \alpha^{(0)}$, zum Beispiel $\beta = 1/2$, die erste, für die

$$f\left(\mathbf{x}^{(k)} + \alpha^{(j)} \mathbf{d}^{(k)}\right) \leq f\left(\mathbf{x}^{(k)}\right) + \mu \alpha^{(j)} \nabla f\left(\mathbf{x}^{(k)}\right)^T \mathbf{d}^{(k)}$$

gilt. Die Motivation für diese Herangehensweise kommt von der nach dem linearen Glied abgebrochenen Taylor–Entwicklung

$$f\left(\mathbf{x}^{(k)} + \alpha^{(j)} \mathbf{d}^{(k)}\right) \approx f\left(\mathbf{x}^{(k)}\right) + \alpha^{(j)} \nabla f\left(\mathbf{x}^{(k)}\right)^T \mathbf{d}^{(k)}.$$

Ist $\mathbf{d}^{(k)}$ eine Abstiegsrichtung, dann findet man in dieser Richtung einen Funktionswert von $f(\mathbf{x})$, der kleiner als der Funktionswert $f(\mathbf{x}^{(k)})$ ist, falls $\alpha^{(j)}$ nur hinreichend klein ist, siehe Bemerkung 2.13. Man fängt mit irgendeinem hinreichend großen $\alpha^{(0)}$ an. Dieses hängt von $\|\nabla f(\mathbf{x}^{(k)})\|_2$ ab. Ist $\|\nabla f(\mathbf{x}^{(k)})\|_2$ klein, kann man vermuten in der Nähe eines Minimums zu sein (besitzt $f(\mathbf{x})$ in $\mathbf{x}^{(k)}$ ein Minimum, so ist $\nabla f(\mathbf{x}^{(k)}) = \mathbf{0}$) und man braucht vielleicht nur einen kleinen Schritt. Man testet ob der Funktionswert sich verkleinert. Ist das nicht der Fall, wird der Parameter $\alpha^{(j)}$ sukzessive verkleinert, bis er klein genug ist. □

2.2.2 Abstiegsmethoden für quadratische Funktionen

Bemerkung 2.17 Quadratische Funktionen

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T A \mathbf{x} - \mathbf{b}^T \mathbf{x}, \tag{2.7}$$

mit einer symmetrischen und positiv definiten Matrix $A \in \mathbb{R}^{n \times n}$ und $\mathbf{b} \in \mathbb{R}^n$ sind ein wichtiger Spezialfall. Die notwendige Bedingung für ein Minimum in $\hat{\mathbf{x}} \in \mathbb{R}^n$ ist

$$\mathbf{0} = \nabla f(\hat{\mathbf{x}}) = A \hat{\mathbf{x}} - \mathbf{b}.$$

Man hat also das lineare Gleichungssystem $A \hat{\mathbf{x}} = \mathbf{b}$ zu lösen. Da $H_f(\hat{\mathbf{x}}) = A$ positiv definit ist, ist auch eine hinreichende Bedingung für ein Minimum erfüllt. □

Lemma 2.18 Optimale Schrittweite. *Das eindimensionale Optimierungsproblem finde α_k , so dass*

$$f\left(\mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}\right)$$

minimiert wird, besitzt für die quadratische Funktion (2.7) die Lösung

$$\alpha_k = \frac{\mathbf{d}^{(k)T} \mathbf{r}^{(k)}}{\mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)}}, \quad \mathbf{r}^{(k)} := \mathbf{b} - \mathbf{A} \mathbf{x}^{(k)}, \quad (2.8)$$

$r^{(k)}$ – Residuum.

Beweis: Es gilt

$$\begin{aligned} & \frac{d}{d\alpha_k} f(\mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}) \\ &= \frac{d}{d\alpha_k} \left[\frac{1}{2} (\mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)})^T \mathbf{A} (\mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}) - \mathbf{b}^T (\mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}) \right] \\ &= \frac{d}{d\alpha_k} \left[\frac{\alpha_k^2}{2} \mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)} + \alpha_k \left(\frac{1}{2} \mathbf{d}^{(k)T} \mathbf{A} \mathbf{x}^{(k)} + \frac{1}{2} \mathbf{x}^{(k)T} \mathbf{A} \mathbf{d}^{(k)} - \mathbf{b}^T \mathbf{d}^{(k)} \right) + \dots \right] \\ &= \alpha_k \mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)} + \left(\frac{1}{2} \mathbf{d}^{(k)T} \mathbf{A} \mathbf{x}^{(k)} + \frac{1}{2} \mathbf{x}^{(k)T} \mathbf{A} \mathbf{d}^{(k)} - \mathbf{b}^T \mathbf{d}^{(k)} \right) \\ &= \alpha_k \mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)} + \left(\mathbf{x}^{(k)T} \mathbf{A} \mathbf{d}^{(k)} - \mathbf{b}^T \mathbf{d}^{(k)} \right) = \alpha_k \mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)} - \mathbf{r}^{(k)T} \mathbf{d}^{(k)}. \end{aligned}$$

Dabei wurde die Symmetrie von A genutzt

$$\mathbf{d}^{(k)T} \mathbf{A} \mathbf{x}^{(k)} = \mathbf{x}^{(k)T} \mathbf{A}^T \mathbf{d}^{(k)} = \mathbf{x}^{(k)T} \mathbf{A} \mathbf{d}^{(k)}.$$

Die Ableitung muss im Optimum verschwinden, woraus durch Umstellen die Formel für α_k folgt. Die zweite Ableitung bezüglich α_k ist $\mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)}$. Wegen der positiven Definitheit von A ist dieser Ausdruck positiv für alle $\mathbf{d}^{(k)} \neq \mathbf{0}$, woraus folgt, dass man mit α_k ein Minimum bestimmt hat. ■

Bemerkung 2.19 Von A induziertes Skalarprodukt und Norm. Die positiv definite symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ induziert ein Skalarprodukt auf \mathbb{R}^n

$$(\mathbf{x}, \mathbf{y})_A := \mathbf{x}^T \mathbf{A} \mathbf{y}$$

und damit eine Norm

$$\|\mathbf{x}\|_A := (\mathbf{x}, \mathbf{x})_A^{1/2}.$$

In dieser Norm betrachtet man die Konvergenz eines Abstiegsverfahrens. □

Lemma 2.20 Fehlerabschätzung eines Abstiegsverfahrens. Für den Fehler eines Abstiegsverfahrens zur Minimierung der quadratischen Funktion (2.7) gilt

$$\|\mathbf{x}^{(k+1)} - \hat{\mathbf{x}}\|_A = \rho_k \|\mathbf{x}^{(k)} - \hat{\mathbf{x}}\|_A$$

mit

$$\rho_k = \left(1 - \frac{(\mathbf{d}^{(k)T} \mathbf{r}^{(k)})^2}{\mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)} \mathbf{r}^{(k)T} \mathbf{A}^{-1} \mathbf{r}^{(k)}} \right)^{1/2}. \quad (2.9)$$

Beweis: Der Beweis geht wie folgt:

1. Schreibe $\|\mathbf{x}^{(k+1)} - \hat{\mathbf{x}}\|_A^2$ Term für Term.
2. Setze (2.8) ein.
3. Zeige $\|\mathbf{x}^{(k)} - \hat{\mathbf{x}}\|_A^2 = \mathbf{r}^{(k)T} \mathbf{A}^{-1} \mathbf{r}^{(k)}$ und setze ein.

Details sind *Übungsaufgabe*. ■

Satz 2.21 Lineare Konvergenz des Gradientenverfahrens. *Das Gradientenverfahren zur Lösung von (2.7) konvergiert linear mit*

$$\left\| \mathbf{x}^{(k+1)} - \hat{\mathbf{x}} \right\|_A \leq \frac{\lambda_{\max}(A) - \lambda_{\min}(A)}{\lambda_{\max}(A) + \lambda_{\min}(A)} \left\| \mathbf{x}^{(k)} - \hat{\mathbf{x}} \right\|_A.$$

Beweis: Für das Gradientenverfahren gilt $\mathbf{d}^{(k)} = \mathbf{r}^{(k)}$. Somit kann der Nenner in (2.9) außerhalb des Minimums nicht Null werden, da er die Gestalt

$$\mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)} = \mathbf{r}^{(k)T} A^{-1} \mathbf{r}^{(k)} = \left(\mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)} \right)^2 > 0$$

besitzt.

Zur Abschätzung des Konvergenzfaktors nutzt man die Ungleichung

$$\frac{(\mathbf{y}^T \mathbf{y})^2}{(\mathbf{y}^T \mathbf{A} \mathbf{y})(\mathbf{y}^T A^{-1} \mathbf{y})} \geq \frac{4\lambda_{\max}(A)\lambda_{\min}(A)}{(\lambda_{\max}(A) + \lambda_{\min}(A))^2},$$

siehe Literatur. Einsetzen dieser Abschätzung und Nutzung von

$$\left(1 - \frac{4\lambda_{\max}(A)\lambda_{\min}(A)}{(\lambda_{\max}(A) + \lambda_{\min}(A))^2} \right)^{1/2} = \left(\frac{(\lambda_{\max}(A) + \lambda_{\min}(A))^2 - 4\lambda_{\max}(A)\lambda_{\min}(A)}{(\lambda_{\max}(A) + \lambda_{\min}(A))^2} \right)^{1/2},$$

sowie binomischer Formel beendet die Abschätzung. ■

Bemerkung 2.22 Zum Gradientenverfahren. Falls A schlecht konditioniert ist, das heißt

$$\kappa_2(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$$

sehr groß ist, dann ist die Konvergenzrate

$$\frac{\lambda_{\max}(A) - \lambda_{\min}(A)}{\lambda_{\max}(A) + \lambda_{\min}(A)} = \frac{\kappa_2(A) - 1}{\kappa_2(A) + 1}$$

nahe bei Eins und das Verfahren ist sehr langsam. Deshalb ist es nötig, bessere Verfahren zu entwickeln. □

Definition 2.23 A -konjugiert. Sei $A \in \mathbb{R}^{n \times n}$ symmetrisch und positiv definit. Die Vektoren $\mathbf{y}_1, \dots, \mathbf{y}_m \in \mathbb{R}^n$ heißen A -konjugiert (oder A -orthogonal), falls $\mathbf{y}_i \neq \mathbf{0}$ für $i = 1, \dots, m$, und $(\mathbf{y}_i, \mathbf{y}_j)_A = 0$ für $i \neq j$ gelten. □

Lemma 2.24 *Seien $\mathbf{y}_1, \dots, \mathbf{y}_m \in \mathbb{R}^n$ A -konjugierte Vektoren. Dann sind $\mathbf{y}_1, \dots, \mathbf{y}_m$ linear unabhängig.*

Beweis: Indirekt. Sei

$$\mathbf{0} = \sum_{i=1}^m \alpha_i \mathbf{y}_i.$$

Durch Multiplikation von links mit $\mathbf{y}_k^T A$, $k = 1, \dots, m$, folgt

$$0 = \mathbf{y}_k^T A \sum_{i=1}^m \alpha_i \mathbf{y}_i = \alpha_k \underbrace{\mathbf{y}_k^T A \mathbf{y}_k}_{>0}.$$

Somit muss $\alpha_k = 0$, $k = 1, \dots, m$, gelten. ■

Satz 2.25 Endlichkeit des Abstiegsverfahrens für A -konjugierter Abstiegsrichtungen. *Verwendet man zur Minimierung von (2.7) A -konjugierte Abstiegsrichtungen $\{\mathbf{d}^{(k)}\}$ und wählt man α_k gemäß (2.8), dann gilt für beliebigen Startwert $\mathbf{x}^{(0)} \in \mathbb{R}^n$*

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

Beweis: Betrachte zunächst einen beliebigen Vektor $\mathbf{y} \in \mathbb{R}^n$. Da die Vektoren $\{\mathbf{d}^{(i)}\}_{i=0}^{n-1}$ linear unabhängig sind, bilden sie eine Basis des \mathbb{R}^n . Deshalb lässt sich \mathbf{y} eindeutig in der Form

$$\mathbf{y} = \sum_{i=0}^{n-1} \beta_i \mathbf{d}^{(i)}$$

darstellen. Es folgt für jedes $k = 0, \dots, n-1$,

$$\mathbf{d}^{(k)T} \mathbf{A} \mathbf{y} = \sum_{i=0}^{n-1} \beta_i \mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(i)} = \beta_k \mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)},$$

woraus sich die Darstellung

$$\mathbf{y} = \sum_{i=0}^{n-1} \frac{\mathbf{d}^{(i)T} \mathbf{A} \mathbf{y}}{\mathbf{d}^{(i)T} \mathbf{A} \mathbf{d}^{(i)}} \mathbf{d}^{(i)} \quad (2.10)$$

ergibt.

Nun ist

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)} = \mathbf{x}^{(k-1)} + \alpha_{k-1} \mathbf{d}^{(k-1)} + \alpha_k \mathbf{d}^{(k)} = \mathbf{x}^{(0)} + \sum_{i=0}^k \alpha_i \mathbf{d}^{(i)}.$$

Damit folgt für die Parameter α_k

$$\begin{aligned} \alpha_k &= \frac{\mathbf{d}^{(k)T} \mathbf{r}^{(k)}}{\mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)}} = \frac{\mathbf{d}^{(k)T} (\mathbf{b} - \mathbf{A} \mathbf{x}^{(k)})}{\mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)}} = \frac{\mathbf{d}^{(k)T} (\mathbf{b} - \mathbf{A} (\mathbf{x}^{(0)} + \sum_{i=0}^{k-1} \alpha_i \mathbf{d}^{(i)}))}{\mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)}} \\ &= \frac{\mathbf{d}^{(k)T} (\mathbf{b} - \mathbf{A} \mathbf{x}^{(0)}) - \sum_{i=0}^{k-1} \alpha_i \mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(i)}}{\mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)}} = \frac{\mathbf{d}^{(k)T} (\mathbf{b} - \mathbf{A} \mathbf{x}^{(0)})}{\mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)}}. \end{aligned}$$

Also gilt

$$\mathbf{x}^{(n)} = \mathbf{x}^{(0)} + \sum_{i=0}^{n-1} \frac{\mathbf{d}^{(i)T} (\mathbf{b} - \mathbf{A} \mathbf{x}^{(0)})}{\mathbf{d}^{(i)T} \mathbf{A} \mathbf{d}^{(i)}} \mathbf{d}^{(i)} = \mathbf{x}^{(0)} + \sum_{i=0}^{n-1} \frac{\mathbf{d}^{(i)T} \mathbf{A} (\mathbf{A}^{-1} \mathbf{b} - \mathbf{x}^{(0)})}{\mathbf{d}^{(i)T} \mathbf{A} \mathbf{d}^{(i)}} \mathbf{d}^{(i)}.$$

Mit Darstellung (2.10) folgt

$$\mathbf{x}^{(n)} = \mathbf{x}^{(0)} + (\mathbf{A}^{-1} \mathbf{b} - \mathbf{x}^{(0)}) = \mathbf{A}^{-1} \mathbf{b}.$$

Das ist die Lösung von $\mathbf{A} \mathbf{x} = \mathbf{b}$, die gesucht ist. ■

Bemerkung 2.26

- Die Methode endet nach spätestens n Abstiegen mit dem Optimum. Für große n ist diese Eigenschaft allerdings praktisch nicht bedeutsam.
- Eine Realisierung der Methode ist das Verfahren der konjugierten Gradienten

$$\begin{aligned} \mathbf{d}^{(k+1)} &= \mathbf{r}^{(k)} + \beta_k \mathbf{d}^{(k)}, \\ \beta_k &= -\frac{\mathbf{r}^{(k+1)T} \mathbf{A} \mathbf{d}^{(k)}}{\mathbf{d}^{(k)T} \mathbf{A} \mathbf{d}^{(k)}} = \frac{\mathbf{r}^{(k+1)T} \mathbf{r}^{(k+1)}}{\mathbf{r}^{(k)T} \mathbf{r}^{(k)}}. \end{aligned}$$

Man muss nicht alle A-konjugierten Abstiegsrichtungen speichern, sondern nur die drei Vektoren $\mathbf{r}^{(k)}$, $\mathbf{r}^{(k+1)}$, $\mathbf{d}^{(k)}$. *Übungsaufgaben*

- Für das Verfahren der konjugierten Gradienten kann man die Fehlerabschätzung

$$\|\mathbf{x}^{(k)} - \hat{\mathbf{x}}\|_A \leq 2 \left(\frac{\sqrt{\kappa_2(A)} - 1}{\sqrt{\kappa_2(A)} + 1} \right)^k \|\mathbf{x}^{(0)} - \hat{\mathbf{x}}\|_A$$

beweisen, wobei $\kappa_2(A)$ die Spektralkonditionszahl von A ist.

- Das Verfahren der konjugierten Gradienten ist eines der beliebtesten Verfahren zur Lösung von linearen Gleichungssystemen mit symmetrischer und positiv definiten Matrix. □

Kapitel 3

Konvexität

3.1 Konvexe Mengen

Der Begriff der konvexen Menge ist bereits aus Definition 1.4, Teil I, bekannt.

Definition 3.1 Konvexer Kegel. Eine Menge $\Omega \subset \mathbb{R}^n$ heißt konvexer Kegel, wenn mit $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ auch $\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 \in \Omega$ für alle $\lambda_1, \lambda_2 \geq 0$. \square

Beispiel 3.2 Konvexe Kegel.

- Für $\lambda_1 = \lambda_2 = 0$ folgt, dass insbesondere $\mathbf{0}$ in einem konvexen Kegel enthalten sein muss.
- Nicht jede konvexe Menge ist ein konvexer Kegel, zum Beispiel sind Kreise keine konvexen Kegel.
- Der Winkelraum mit Scheitel $\mathbf{0}$ ist ein konvexer Kegel. Der Winkelraum mit Scheitel $\neq \mathbf{0}$ ist kein konvexer Kegel.
- Sei $\mathbf{a} \in \mathbb{R}^n, \mathbf{a} \neq \mathbf{0}, \alpha \in \mathbb{R}$ und definiere die Hyperebene

$$H(\mathbf{a}, \alpha) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T \mathbf{x} = \alpha\}.$$

Diese Hyperebene ist eine konvexe Menge. *Übungsaufgabe* Ist $\alpha \neq 0$, dann ist $\mathbf{0} \notin H(\mathbf{a}, \alpha)$ und $H(\mathbf{a}, \alpha)$ ist kein konvexer Kegel. Im Falle $\alpha = 0$ ist $H(\mathbf{a}, \alpha)$ ein konvexer Kegel $H(\mathbf{a}, \alpha) = \mathbf{a}^\perp$. *Übungsaufgabe*

- Analog gilt für Halbräume

$$H_+(\mathbf{a}, \alpha) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T \mathbf{x} \geq \alpha\}, \quad H_-(\mathbf{a}, \alpha) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T \mathbf{x} \leq \alpha\},$$

dass diese für $\alpha = 0$ konvexe Kegel sind und sonst nicht.

- Eine Menge, die durch Hyperebenen berandet ist, die allesamt den Punkt $\mathbf{0}$ enthalten, wird auch polyhedrischer Kegel genannt. Polyhedrische Kegel sind konvexe Kegel, insbesondere

$$\mathbb{R}_+^n = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \geq \mathbf{0}\}.$$

\square

Definition 3.3 Trennung von Mengen. Die Hyperebene $H(\mathbf{a}, \alpha)$ trennt die nichtleeren Mengen $\Omega_1, \Omega_2 \subset \mathbb{R}^n$, wenn gilt

$$\mathbf{a}^T \mathbf{x} \leq \alpha \quad \forall \mathbf{x} \in \Omega_1 \quad \text{und} \quad \mathbf{a}^T \mathbf{x} \geq \alpha \quad \forall \mathbf{x} \in \Omega_2.$$

Falls in beiden Relationen die echten Ungleichungen stehen, dann wird die Trennung streng genannt. \square

Lemma 3.4 Sei $\Omega \subset \mathbb{R}^n$ nichtleer, abgeschlossen und konvex und gelte $\mathbf{0} \notin \Omega$. Dann existiert eine Hyperebene $H(\mathbf{a}, \alpha)$ mit $\alpha > 0$, so dass $\mathbf{a}^T \mathbf{x} > \alpha$ für alle $\mathbf{x} \in \Omega$.

Beweis: Wir werden eine konkrete Hyperebene angeben.

Seien $\mathbf{x}_1 \in \Omega$ und $B_1 = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \leq \|\mathbf{x}_1\|_2\}$ eine Kugel mit Radius $\|\mathbf{x}_1\|_2$ und Mittelpunkt $\mathbf{0}$. Dann ist $\Omega \cap B_1 \neq \emptyset$ und der Durchschnitt ist kompakt (abgeschlossen und beschränkt). Damit nimmt die stetige Funktion $d(\mathbf{x}) = \|\mathbf{x}\|_2$ ihr Minimum über $\Omega \cap B_1$ in einem Punkt $\mathbf{x}_0 \in \Omega \cap B_1$ an (Satz von Bolzano–Weierstrass), dass heißt es gilt

$$\|\mathbf{x}\|_2 \geq \|\mathbf{x}_0\|_2 \quad \forall \mathbf{x} \in \Omega \cap B_1. \quad (3.1)$$

Wegen $\mathbf{0} \notin \Omega$ ist $\|\mathbf{x}_0\|_2 > 0$. Da die Elemente von Ω , die nicht in $\Omega \cap B_1$ liegen, eine größere Norm als $\|\mathbf{x}_1\|_2 \geq \|\mathbf{x}_0\|_2$ besitzen, gilt (3.1) für alle $\mathbf{x} \in \Omega$.

Wegen der Konvexität von Ω ist $\mathbf{x}_0 + \lambda(\mathbf{x} - \mathbf{x}_0) \in \Omega$ für alle $\mathbf{x} \in \Omega$ und $\lambda \in (0, 1]$ und es folgt

$$\|(1 - \lambda)\mathbf{x}_0 + \lambda\mathbf{x}\|_2 = \|\mathbf{x}_0 + \lambda(\mathbf{x} - \mathbf{x}_0)\|_2 \geq \|\mathbf{x}_0\|_2 > 0.$$

Quadratur dieser Ungleichung und Ausschreiben der Skalarprodukte ergibt

$$2\lambda(\mathbf{x} - \mathbf{x}_0)^T \mathbf{x}_0 + \lambda^2(\mathbf{x} - \mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0) \geq 0.$$

Division durch λ und dann $\lambda \rightarrow 0$ ergibt

$$\mathbf{x}^T \mathbf{x}_0 \geq \mathbf{x}_0^T \mathbf{x}_0 = \|\mathbf{x}_0\|_2^2 > \frac{\|\mathbf{x}_0\|_2^2}{2} > 0 \quad \forall \mathbf{x} \in \Omega.$$

Die Hyperebene mit $\mathbf{a} = \mathbf{x}_0$ und $\alpha = \|\mathbf{x}_0\|_2^2/2$ erfüllt nun die Bedingungen des Lemmas. ■

Eine Folgerung ist der nachfolgende Satz.

Satz 3.5 Erster Trennungssatz. Seien $\Omega_1 \neq \emptyset$ abgeschlossen und konvex und Ω_2 kompakt und konvex, $\Omega_1, \Omega_2 \subset \mathbb{R}^n$, mit $\Omega_1 \cap \Omega_2 = \emptyset$. Dann existiert eine Hyperebene $H(\mathbf{a}, \alpha)$, die Ω_1 und Ω_2 streng trennt.

Beweis: Seien

$$\begin{aligned} -\Omega_1 &= \{\mathbf{x} : \mathbf{x} = -\mathbf{y}, \mathbf{y} \in \Omega_1\}, \\ \Omega_2 - \Omega_1 &= \{\mathbf{x} : \exists \mathbf{x}_1 \in \Omega_1, \mathbf{x}_2 \in \Omega_2 \text{ mit } \mathbf{x} = \mathbf{x}_2 - \mathbf{x}_1\}. \end{aligned}$$

Diese Mengen sind konvex und abgeschlossen. Da Ω_1 und Ω_2 einen leeren Durchschnitt haben, gilt $\mathbf{0} \notin \Omega_2 - \Omega_1$. Nach Lemma 3.4 gibt es eine Hyperebene $H(\mathbf{a}, \alpha)$ mit $\alpha > 0$ und $\mathbf{a}^T(\mathbf{x}_2 - \mathbf{x}_1) > \alpha > 0$ für alle $\mathbf{x}_1 \in \Omega_1, \mathbf{x}_2 \in \Omega_2$. Durch Umstellen dieser Beziehung erhält man

$$\inf_{\mathbf{x}_2 \in \Omega_2} \mathbf{a}^T \mathbf{x}_2 \geq \sup_{\mathbf{x}_1 \in \Omega_1} \mathbf{a}^T \mathbf{x}_1 + \alpha > \sup_{\mathbf{x}_1 \in \Omega_1} \mathbf{a}^T \mathbf{x}_1.$$

Jede Hyperebene $H(\mathbf{a}, \beta)$ mit

$$\beta \in \left(\sup_{\mathbf{x}_1 \in \Omega_1} \mathbf{a}^T \mathbf{x}_1, \inf_{\mathbf{x}_2 \in \Omega_2} \mathbf{a}^T \mathbf{x}_2 \right)$$

trennt damit Ω_1 und Ω_2 streng. ■

Übungsaufgabe: Notwendigkeit der Kompaktheitsvoraussetzung für strenge Trennung an Beispiel zeigen.

Satz 3.6 Jede abgeschlossene konvexe Menge $\Omega \subset \mathbb{R}^n$ ist der Durchschnitt aller abgeschlossenen Halbräume, die Ω enthalten.

Beweis: Beweisskizze. Bezeichnen wir den Durchschnitt aller abgeschlossenen Halbräume, die Ω enthalten, mit D . Dann ist $\Omega \subset D$ klar (Durchschnitt beliebig vieler konvexer Mengen ist konvex). Zu zeigen bleibt $D \subset \Omega$. Das geschieht indirekt, wobei der Widerspruch mit einer trennenden Hyperebene hergeleitet wird (sogenannter zweiter Trennungssatz, Details siehe [ERSD77, Abschnitt 2.1.4]). ■

Bemerkung 3.7 Eckpunkt, Extrempunkt. Der Eckpunkt oder Extrempunkt einer konvexen Menge wurde bereits in Definition 1.10, Teil I, definiert als ein Punkt, für den es keine echte konvexe Linearkombination gibt. Dafür, dass $\mathbf{x}_0 \in \Omega$ ein Extrempunkt ist, reicht es bereits aus, dass man keine $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ findet, so dass

$$\mathbf{x}_0 = \frac{1}{2}\mathbf{x}_1 + \frac{1}{2}\mathbf{x}_2, \quad \mathbf{x}_1, \mathbf{x}_2 \neq \mathbf{x}_0. \quad \text{Übungsaufgabe}$$

□

Satz 3.8 Existenz eines Extrempunkts. Jede nichtleere, kompakte, konvexe Menge $\Omega \subset \mathbb{R}^n$ besitzt mindestens einen Extrempunkt.

Beweis: Da Ω kompakt ist, gibt es ein $\mathbf{x}_0 \in \Omega$ mit $\|\mathbf{x}_0\|_2 \geq \|\mathbf{x}\|_2$ für alle $\mathbf{x} \in \Omega$. Sei

$$\mathbf{x}_0 = \frac{1}{2}\mathbf{x}_1 + \frac{1}{2}\mathbf{x}_2 = \mathbf{x}_1 + \frac{1}{2}(\mathbf{x}_2 - \mathbf{x}_1) = \mathbf{x}_2 + \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_2)$$

für $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$, $\mathbf{x}_1, \mathbf{x}_2 \neq \mathbf{x}_0$. Damit folgt

$$\|\mathbf{x}_1\|_2^2 \leq \|\mathbf{x}_0\|_2^2 = \|\mathbf{x}_1\|_2^2 + \underbrace{\frac{1}{4}\|\mathbf{x}_2 - \mathbf{x}_1\|_2^2 + (\mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{x}_1}_{\geq 0, \text{ da } \|\mathbf{x}_0\|_2 \geq \|\mathbf{x}_1\|_2}.$$

Das heißt

$$\|\mathbf{x}_2 - \mathbf{x}_1\|_2^2 \geq -4(\mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{x}_1$$

und analog

$$\|\mathbf{x}_2 - \mathbf{x}_1\|_2^2 \geq -4(\mathbf{x}_1 - \mathbf{x}_2)^T \mathbf{x}_2 = 4(\mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{x}_2.$$

Addition der Ungleichungen ergibt

$$2\|\mathbf{x}_2 - \mathbf{x}_1\|_2^2 \geq 4(\mathbf{x}_2 - \mathbf{x}_1)^T (\mathbf{x}_2 - \mathbf{x}_1) = 4\|\mathbf{x}_2 - \mathbf{x}_1\|_2^2$$

und damit $\mathbf{x}_2 = \mathbf{x}_1 = \mathbf{x}_0$. Somit ist \mathbf{x}_0 ein Extrempunkt. ■

3.2 Konvexe und konkave Funktionen

Definition 3.9 Konvexe, konkave Funktion. Die Funktion $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ heißt konvex, wenn für beliebige $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ gilt

$$f((1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2) \leq (1-\lambda)f(\mathbf{x}_1) + \lambda f(\mathbf{x}_2) \quad \forall \lambda \in [0, 1].$$

Sie heißt konkav, wenn

$$f((1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2) \geq (1-\lambda)f(\mathbf{x}_1) + \lambda f(\mathbf{x}_2) \quad \forall \lambda \in [0, 1].$$

Gilt die echte Ungleichung spricht man von strenger Konvexität (Konkavität). □

Beispiel 3.10 Die lineare Funktion $f(\mathbf{x}) = \mathbf{c}^T \mathbf{x}$ mit $\mathbf{c} \in \mathbb{R}^n$ ist konvex und konkav in \mathbb{R}^n , jedoch nicht streng. □

Beispiel 3.11 Jede über \mathbb{R}^n positiv semidefinite Bilinearform $f(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$ mit $A = A^T \in \mathbb{R}^{n \times n}$ ist konvex. Sei nämlich

$$\mathbf{x} = (1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2 = \mathbf{x}_1 + \lambda(\mathbf{x}_2 - \mathbf{x}_1), \quad \mathbf{x}_1 \neq \mathbf{x}_2 \in \mathbb{R}^n, \quad \lambda \in [0, 1],$$

dann gilt

$$\begin{aligned} \mathbf{x}^T A \mathbf{x} &= \mathbf{x}_1^T A \mathbf{x}_1 + 2\lambda(\mathbf{x}_2 - \mathbf{x}_1)^T A \mathbf{x}_1 + \underbrace{\lambda^2}_{\leq \lambda} \underbrace{(\mathbf{x}_2 - \mathbf{x}_1)^T A (\mathbf{x}_2 - \mathbf{x}_1)}_{\geq 0} \\ &\leq \mathbf{x}_1^T A \mathbf{x}_1 + 2\lambda(\mathbf{x}_2 - \mathbf{x}_1)^T A \mathbf{x}_1 + \lambda(\mathbf{x}_2 - \mathbf{x}_1)^T A (\mathbf{x}_2 - \mathbf{x}_1) \\ &= \mathbf{x}_1^T A \mathbf{x}_1 + \lambda(\mathbf{x}_2 - \mathbf{x}_1)^T A (\mathbf{x}_1 + \mathbf{x}_2) \\ &= (1-\lambda)\mathbf{x}_1^T A \mathbf{x}_1 + \lambda\mathbf{x}_2^T A \mathbf{x}_2. \end{aligned}$$

Analog zeigt man, dass

- A ist positiv definit, dann ist $f(\mathbf{x})$ streng konvex,
- A ist negativ semidefinit, dann ist $f(\mathbf{x})$ konkav,
- A ist negativ definit, dann ist $f(\mathbf{x})$ streng konkav.

□

Beispiel 3.12 Sind die Funktionen $f_j(\mathbf{x}), j = 1, \dots, k$, über $\Omega \subset \mathbb{R}^n$ konvex, dann ist auch die Linearkombination

$$f(\mathbf{x}) = \sum_{j=1}^k \alpha_j f_j(\mathbf{x}), \quad \alpha_j \geq 0 \quad \forall j$$

konvex. Dies zeigt man durch direktes Nachrechnen.

□

Zur Stetigkeit von konvexen Funktionen gibt es folgende Aussage.

Satz 3.13 Stetigkeit konvexer Funktionen. Seien $\Omega \subset \mathbb{R}^n$ konvex und das Innere der Menge, $\text{int}(\Omega)$, nichtleer. Dann ist jede konvexe Funktion $f : \Omega \rightarrow \mathbb{R}$ stetig in $\text{int}(\Omega)$.

Beweis: Siehe Literatur, zum Beispiel [ERSD77, Satz 2.65]. Man prüft die ε - δ -Definition nach. ■

Beispiel 3.14 Nichtstetige konvexe Funktion über konvexer Menge. Stetigkeit bis auf den Rand muss bei einer konvexen Funktion im allgemeinen nicht vorliegen. Die konvexe Funktion $f : [1, 2] \rightarrow \mathbb{R}$ mit

$$f(x) = \begin{cases} 1/x & \text{für } x \in [1, 2) \\ 2 & \text{für } x = 2 \end{cases}$$

ist ein Beispiel.

□

Lemma 3.15 Monotonieaussage für den Differenzenquotienten. Seien $\Omega \subset \mathbb{R}^n$ eine offene konvexe Menge, $f : \Omega \rightarrow \mathbb{R}$ mit $f \in C^1(\Omega)$ eine konvexe Funktion, $\mathbf{x}_0 \in \Omega$, $\mathbf{h} \in \mathbb{R}^n$, $\mathbf{x}_0 - \nu \mathbf{h} \in \Omega$ und $\mathbf{x}_0 + \rho \mathbf{h} \in \Omega$ für gewisse $\nu, \rho > 0$. Dann gilt

$$\begin{aligned} \frac{f(\mathbf{x}_0) - f(\mathbf{x}_0 - \nu \mathbf{h})}{\nu} &\leq \frac{f(\mathbf{x}_0) - f(\mathbf{x}_0 - \mu \mathbf{h})}{\mu} \\ &\leq \frac{f(\mathbf{x}_0 + \lambda \mathbf{h}) - f(\mathbf{x}_0)}{\lambda} \leq \frac{f(\mathbf{x}_0 + \rho \mathbf{h}) - f(\mathbf{x}_0)}{\rho} \end{aligned}$$

für alle $\mu \in (0, \nu], \lambda \in (0, \rho]$. Damit ist für konvexe Funktionen der Differenzenquotient im Punkt \mathbf{x}_0 in Richtung \mathbf{h} eine monoton wachsende Funktion von λ .

Beweis: Übungsaufgabe ■

Jetzt werden einige Bedingung dafür gegeben, dass ein Funktion auf einer konvexen Menge konvex ist.

Satz 3.16 Hinreichende und notwendige Bedingung für Konvexität. Seien $\Omega \subset \mathbb{R}^n$ eine offene konvexe Menge und $f : \Omega \rightarrow \mathbb{R}$ mit $f \in C^1(\Omega)$. Dann ist $f(\mathbf{x})$ auf Ω genau dann konvex, wenn für alle $\mathbf{x}_0, \mathbf{x}_1 \in \Omega$ mit $\mathbf{x}_0 \neq \mathbf{x}_1$ gilt

$$f(\mathbf{x}_1) - f(\mathbf{x}_0) \geq (\mathbf{x}_1 - \mathbf{x}_0)^T \nabla f(\mathbf{x}_0). \quad (3.2)$$

Beweis: 1) *Notwendigkeit von (3.2).* Da $f(\mathbf{x})$ differenzierbar ist, gilt für die Richtungsableitung in Richtung $\mathbf{h} \in \mathbb{R}^n$ mit $\|\mathbf{h}\|_2 = 1$

$$\frac{\partial f}{\partial \mathbf{h}}(\mathbf{x}_0) = \mathbf{h}^T \nabla f(\mathbf{x}_0) = \lim_{\lambda \rightarrow 0} \frac{f(\mathbf{x}_0 + \lambda \mathbf{h}) - f(\mathbf{x}_0)}{\lambda} = - \lim_{\lambda \rightarrow 0} \frac{f(\mathbf{x}_0 - \lambda \mathbf{h}) - f(\mathbf{x}_0)}{\lambda},$$

wobei bei der letzten Gleichung λ durch $-\lambda$ ersetzt wurde. Nach Lemma 3.15 gilt alle $\rho > 0$ mit $\mathbf{x}_0 + \rho \mathbf{h} \in \Omega$ (nehme $\lambda \rightarrow 0$ in Lemma 3.15)

$$\frac{f(\mathbf{x}_0 + \rho \mathbf{h}) - f(\mathbf{x}_0)}{\rho} \geq \mathbf{h}^T \nabla f(\mathbf{x}_0)$$

oder

$$f(\mathbf{x}_0 + \rho \mathbf{h}) - f(\mathbf{x}_0) \geq \rho \mathbf{h}^T \nabla f(\mathbf{x}_0).$$

Wählt man $\mathbf{x}_1 = \mathbf{x}_0 + \rho \mathbf{h}$, so ergibt sich (3.2).

2) *Hinlänglichkeit von (3.2).* Sei $\lambda_0 > 0$ beliebig gewählt, setze $\lambda_1 = 1 - \lambda_0$. Dann gelten für $\mathbf{x}_2 = \lambda_0 \mathbf{x}_0 + \lambda_1 \mathbf{x}_1$ nach Voraussetzung

$$f(\mathbf{x}_0) - f(\mathbf{x}_2) \geq (\mathbf{x}_0 - \mathbf{x}_2)^T \nabla f(\mathbf{x}_2), \quad f(\mathbf{x}_1) - f(\mathbf{x}_2) \geq (\mathbf{x}_1 - \mathbf{x}_2)^T \nabla f(\mathbf{x}_2).$$

Damit folgt

$$\begin{aligned} \lambda_0 f(\mathbf{x}_0) + \lambda_1 f(\mathbf{x}_1) &\geq f(\mathbf{x}_2) + \left(\lambda_0 (\mathbf{x}_0 - \mathbf{x}_2) + \lambda_1 (\mathbf{x}_1 - \mathbf{x}_2) \right)^T \nabla f(\mathbf{x}_2) \\ &= f(\mathbf{x}_2) + \underbrace{\left(\lambda_0 \mathbf{x}_0 + \lambda_1 \mathbf{x}_1 - \mathbf{x}_2 \right)^T}_{=0} \nabla f(\mathbf{x}_2) \\ &= f(\mathbf{x}_2) = f(\lambda_0 \mathbf{x}_0 + \lambda_1 \mathbf{x}_1) \end{aligned}$$

für alle $\mathbf{x}_0, \mathbf{x}_1 \in \Omega$. ■

Bemerkung 3.17 Taylor–Entwicklung. Für eine zweimal stetig differenzierbare Funktion $f : \Omega \rightarrow \mathbb{R}, \Omega \subset \mathbb{R}^n$, offen, ist die Hesse–Matrix

$$H_f(\mathbf{x}) := \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{i,j=1,\dots,n}(\mathbf{x})$$

symmetrisch (Satz von Schwarz).

Liegt zusammen mit $\mathbf{x}_0, \mathbf{x}_1$ die Strecke $[\mathbf{x}_0, \mathbf{x}_1]$ vollständig in Ω , so gilt für eine geeignete Zahl $\lambda \in (0, 1)$ nach dem Satz von Taylor

$$f(\mathbf{x}_1) = f(\mathbf{x}_0) + (\mathbf{x}_1 - \mathbf{x}_0)^T \nabla f(\mathbf{x}_0) + \frac{1}{2} (\mathbf{x}_1 - \mathbf{x}_0)^T H_f(\mathbf{x}_0 + \lambda(\mathbf{x}_1 - \mathbf{x}_0)) (\mathbf{x}_1 - \mathbf{x}_0). \quad (3.3)$$

□

Satz 3.18 Hinreichende und notwendige Bedingung für Konvexität mit Hesse–Matrix. Seien $\Omega \subset \mathbb{R}^n$ eine offene konvexe Menge und $f : \Omega \rightarrow \mathbb{R}$ mit $f \in C^2(\Omega)$. Dann ist $f(\mathbf{x})$ auf Ω genau dann konvex, wenn $H_f(\mathbf{x})$ positiv semidefinit ist für alle $\mathbf{x} \in \Omega$.

Beweis: 1.) Sei $H_f(\mathbf{x})$ positiv semidefinit. Dann folgt aus (3.3) für beliebige $\mathbf{x}_0, \mathbf{x}_1 \in \Omega$ und eine geeignete Zahl $\lambda \in (0, 1)$

$$f(\mathbf{x}_1) - f(\mathbf{x}_0) - (\mathbf{x}_1 - \mathbf{x}_0)^T \nabla f(\mathbf{x}_0) = \frac{1}{2} (\mathbf{x}_1 - \mathbf{x}_0)^T H_f(\mathbf{x}_0 + \lambda(\mathbf{x}_1 - \mathbf{x}_0)) (\mathbf{x}_1 - \mathbf{x}_0) \geq 0.$$

Aus Satz 3.16 folgt, dass $f(\mathbf{x})$ konvex ist.

2) Sei $f(\mathbf{x})$ konvex auf Ω . Seien $\mathbf{x}_0 \in \Omega$ und $\mathbf{h} \in \mathbb{R}^n$ beliebig gegeben. Da Ω offen ist, gibt es eine Zahl $\theta_0 > 0$ so dass $\mathbf{x}_0 + \theta \mathbf{h} \in \Omega$ für alle $\theta \in (0, \theta_0]$. Wählt man in (3.2) $\mathbf{x}_1 = \mathbf{x}_0 + \theta \mathbf{h}$, erhält man

$$f(\mathbf{x}_0 + \theta \mathbf{h}) - f(\mathbf{x}_0) - \theta \mathbf{h}^T \nabla f(\mathbf{x}_0) \geq 0 \quad \forall \theta \in (0, \theta_0].$$

Aus (3.3) folgt nun

$$\mathbf{h}^T H_f(\mathbf{x}_0 + \lambda\theta\mathbf{h})\mathbf{h} \geq 0 \quad \forall \theta \in (0, \theta_0]$$

mit $\lambda \in (0, 1)$. Da die Hesse-Matrix nach Voraussetzung stetig ist, folgt für $\theta \rightarrow +0$

$$\mathbf{h}^T H_f(\mathbf{x}_0)\mathbf{h} \geq 0.$$

Da $\mathbf{x}_0 \in \Omega$ und $\mathbf{h} \in \mathbb{R}^n$ beliebig gewählt wurden, ist $H_f(\mathbf{x})$ in jedem Punkt $\mathbf{x} \in \Omega$ positiv semidefinit. ■

Beispiel 3.19 1D. Eine skalare Funktion einer skalaren Veränderlichen $f : (a, b) \rightarrow \mathbb{R}$ mit $f \in C^2(a, b)$ ist genau dann konvex, wenn $f''(x) \geq 0$ in (a, b) , zum Beispiel die Funktion $f(x) = x^2$. □

Bemerkung 3.20 Es gibt Verallgemeinerungen des Begriffs der konvexen Funktion, zum Beispiel quasi-konvex und explizit konvex, siehe Literatur. □

3.3 Ungleichungen und konvexe Mengen

Satz 3.21 Die Funktion $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ sei konvex, Ω sei konvex und sei $b \in \mathbb{R}$. Dann ist

$$V := \{\mathbf{x} \in \Omega : f(\mathbf{x}) \leq b\}$$

konvex.

Beweis: Zu zeigen ist: $\mathbf{x}_1, \mathbf{x}_2 \in V \implies \mathbf{x} = (1 - \lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2 \in V$ für alle $\lambda \in [0, 1]$. Aus $f(\mathbf{x}_1), f(\mathbf{x}_2) \leq b$ folgt aus der Konvexität von f

$$f(\mathbf{x}) \leq (1 - \lambda)f(\mathbf{x}_1) + \lambda f(\mathbf{x}_2) \leq b.$$

Damit ist $\mathbf{x} \in V$. ■

Bemerkung 3.22 Unterhalbmenge.

- Die Menge V aus Satz 3.21 heißt Unterhalbmenge von $f(\mathbf{x})$ in Ω zum Wert b .
- Die Niveaumenge $\{\mathbf{x} \in \Omega : f(\mathbf{x}) = b\}$ der konvexen Funktion $f(\mathbf{x})$ ist im allgemeinen nicht konvex. Zum Beispiel besteht die Niveaumenge zu $b = 1$ der konvexen Funktion $f(x) = x^2, x \in \mathbb{R}$, aus den Punkten -1 und 1 . □

Folgerung 3.23 Sei $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ konkav, sei Ω konvex und sei $b \in \mathbb{R}$. Dann ist $U := \{\mathbf{x} \in \Omega : f(\mathbf{x}) \geq b\}$ konvex.

Beweis: Nutze Satz 3.21 mit $-f(\mathbf{x})$, da diese Funktion konvex ist. ■

3.4 Extrema von konvexen Funktionen

Es wird die Frage untersucht, unter welchen Bedingungen eine konvexe Funktion ihr globales Minimum über einer gegebenen konvexen Menge Ω annimmt.

Satz 3.24 Lokales Minimum ist globales Minimum. Sei $f : \Omega \rightarrow \mathbb{R}$ eine konvexe Funktion über einer konvexen Menge Ω . Dann ist ein lokales Minimum von $f(\mathbf{x})$ in Ω zugleich das globale Minimum über Ω .

Beweis: Indirekt. Nehme $f(\mathbf{x})$ ein lokales Minimum in $\mathbf{x}_0 \in \Omega$ an. Sei weiter $\mathbf{x}^* \in \Omega$ mit $f(\mathbf{x}^*) < f(\mathbf{x}_0)$. Aus der Konvexität von $f(\mathbf{x})$ folgt

$$f((1-\lambda)\mathbf{x}_0 + \lambda\mathbf{x}^*) \leq (1-\lambda)f(\mathbf{x}_0) + \lambda f(\mathbf{x}^*) = f(\mathbf{x}_0) + \lambda(f(\mathbf{x}^*) - f(\mathbf{x}_0)) < f(\mathbf{x}_0)$$

für alle $\lambda \in (0, 1)$, da $f(\mathbf{x}^*) < f(\mathbf{x}_0)$. Aus der Konvexität von Ω folgt, dass der Funktionswert im Zwischenpunkt wohldefiniert ist. Wählt man λ beliebig klein, so ist der Zwischenpunkt beliebig nahe an \mathbf{x}_0 und die obige Ungleichung steht im Widerspruch dazu, dass $f(\mathbf{x})$ in \mathbf{x}_0 ein lokales Minimum annimmt. ■

Beispiel 3.25 Konvexe Funktion über konvexer Menge ohne Annahme eines lokalen Minimums. Der obige Satz kann nur angewendet werden, wenn es ein lokales Minimum gibt, das angenommen wird. Das muss nicht der Fall sein. Die konvexe Funktion aus Beispiel 3.14 nimmt ihr Minimum nicht an. □

Folgerung 3.26 Die Menge Ω_0 , über der eine konvexe Funktion $f : \Omega \rightarrow \mathbb{R}$ über einer konvexen Menge Ω das globale Minimum annimmt, ist konvex.

Beweis: Folgt aus Satz 3.21. ■

Satz 3.27 Bedingung für globales Minimum für stetig differenzierbare konvexe Funktionen. Sei $f(\mathbf{x}) \in C^1(\Omega)$ über einer konvexen Menge Ω konvex. Dann nimmt $f(\mathbf{x})$ in jedem $\mathbf{x}_1 \in \Omega$ mit $\nabla f(\mathbf{x}_1) = \mathbf{0}$ sein globales Minimum über Ω an.

Beweis: Es ist zu zeigen, dass alle Punkte $\mathbf{x}_1 \in \Omega$ mit $\nabla f(\mathbf{x}_1) = \mathbf{0}$ wirklich Extrempunkte sind und zudem Minima.

Sei $\mathbf{x}_1 \in \Omega$ mit $\nabla f(\mathbf{x}_1) = \mathbf{0}$. Da $f(\mathbf{x})$ konvex ist, gilt

$$f(\lambda\mathbf{x}_2 + (1-\lambda)\mathbf{x}_1) \leq (1-\lambda)f(\mathbf{x}_1) + \lambda f(\mathbf{x}_2) = f(\mathbf{x}_1) + \lambda(f(\mathbf{x}_2) - f(\mathbf{x}_1))$$

für alle $\mathbf{x}_2 \in \Omega$ und $\lambda \in [0, 1]$. Für $\lambda \in (0, 1]$ erhält man daraus

$$\frac{f(\mathbf{x}_1 + \lambda(\mathbf{x}_2 - \mathbf{x}_1)) - f(\mathbf{x}_1)}{\lambda} \leq f(\mathbf{x}_2) - f(\mathbf{x}_1).$$

Nach dem Mittelwertsatz der Differentialrechnung gibt es ein $\theta \in [0, 1]$ mit

$$(\mathbf{x}_2 - \mathbf{x}_1)^T \nabla f(\mathbf{x}_1 + \theta\lambda(\mathbf{x}_2 - \mathbf{x}_1)) \leq f(\mathbf{x}_2) - f(\mathbf{x}_1).$$

Für $\lambda \rightarrow 0$ folgt nun mit $\nabla f(\mathbf{x}_1) = \mathbf{0}$

$$0 = (\mathbf{x}_2 - \mathbf{x}_1)^T \nabla f(\mathbf{x}_1) \leq f(\mathbf{x}_2) - f(\mathbf{x}_1),$$

also $f(\mathbf{x}_1) \leq f(\mathbf{x}_2)$ für alle $\mathbf{x}_2 \in \Omega$. ■

Beispiel 3.28 Nichteindeutigkeit des globalen Minimums. Das globale Minimum kann an mehr als einem Punkt angenommen werden. Betrachte zum Beispiel die konvexe Funktion $f : [0.9, 2] \rightarrow \mathbb{R}$

$$f(x) = \begin{cases} \exp\left(-\frac{1}{1-x^2}\right) & \text{für } x \in [0.9, 1), \\ 0 & \text{für } x \in [1, 2]. \end{cases}$$

Es gilt $f \in C^\infty([0.9, 2])$. Das globale Minimum wird in $[1, 2]$ angenommen. In diesem Intervall gilt auch $f'(x) = 0$. □