7 Nichtlineare Optimierung

In Kapitel 6 behandelte Klasse von Funktionen war charakterisiert durch

- lineare Zielfunktion
- lineare Nebenbedingungen

es wurde eine Lösung in $W \subseteq \mathbb{R}^n$ gesucht

Beim Finden der Lösung kann die spezielle Struktur des zulässigen Bereichs (konvexes Polyeder), die Lage der Optima (in einer Ecke des zulässigen Bereichs) und das Zusammenfallen lokaler und globaler Optima genutzt werden

In diesem Abschnitt werden allgemeine Zielfunktionen und allgemeine Nebenbedingungen zugelassen.

Folgende Einschränkungen sollen allerdings gelten

- es gilt $W \subseteq \mathbb{R}^n$, wir betrachten keine ganzzahligen oder kombinatorischen Probleme
- die Zielfunktiontswerte sind exakt bestimmbar
  (keine stochastischen Schwankungen, Messfehler, ...)
- die Zielfunktion sei stetig differenzierbar und die ersten (und zweiten) Ableitungen seien bestimmbar (entweder exakt oder approximativ)

Man unterscheidet

- $\mathbf{x} \in \mathbb{R}^n$ unrestringiertes Problem
- $\mathbf{x} \in W \subset \mathbb{R}^n$ restringiertes Problem, Restriktionen als Nebenbedingungs-Ungleichungssystem

Generelle Beobachtungen

- nichtlineare Nebenbedingungen „schwieriger“ als nichtlineare Zielfunktion
- restringierte Probleme „schwieriger“ als unrestringierte Probleme
- zwischen lokalen/relativen und globalen/absoluten Optimum unterscheiden
  (zur Erinnerung globales Minimum $f(\mathbf{x}^*) \leq f(\mathbf{x})$ für alle $\mathbf{x} \in W$,
  lokales Minimum $f(\mathbf{x}^*) \leq f(\mathbf{x})$ für alle $\mathbf{x} \in N_\varepsilon(x^*)$
  mit $N_\varepsilon(\mathbf{x}^*) = \{\mathbf{x} \mid \mathbf{x} \in W \text{ und } \|\mathbf{x}-\mathbf{x}^*\| < \varepsilon\}$ für $\varepsilon > 0$)
- Hoffung auf „generelle Methode“ (angesichts der Vielfalt nichtlinearer Funktionen) offensichtlich müßig

Ziele des Abschnitts können deshalb nur sein

- generelller Überblick
- für die Lösung hilfreiche Strukturen herausarbeiten und damit eine Klassifizierung von Problemen ermöglichen
- robuste Methoden, die für größere Problemklassen anwendbar sind

Gliederung

7.1 Grundbegriffe und Optimalitätsbedingungen

7.2 Konvexe Optimierung

7.3 Verfahren für unrestringierte Probleme

7.4 Verfahren für restringierte Probleme

7.1 Grundbegriffe und Optimalitätsbedingungen

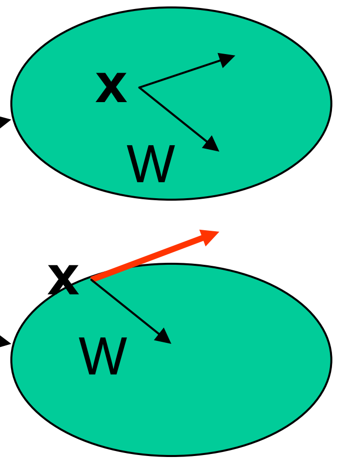
Zur Erinnerung Vorgehen des Simplexverfahrens:

Simplex bewegte sich von zulässigem Punkt (Ecke)

- über zulässige Punkte (entlang einer Kante)
- in Z-verbessernder Richtung (Z=f(**x**) monoton fallend)
- zu zulässigem Punkt (Ecke)

Wir versuchen die Idee zu übertragen, benötigen Festlegungen:

- „über zulässige Punkte", „in Z-verbessernder Richtung", u.a.m.
- sei $\mathbf{x} \in W$, $\mathbf{s} \in \mathbb{R}^n$, $\delta' > 0$: falls $(\mathbf{x} + \delta \mathbf{s}) \in W \; \forall \; 0 \leq \delta < \delta'$
  dann ist **s** zulässige Richtung
  - **x** im Innern von W: alle $\mathbf{s} \in \mathbb{R}^n$ zulässig
  - **x** auf Rand von W: nicht alle $\mathbf{s} \in \mathbb{R}^n$ zulässig



Es wurde gefordert f(**x**) partiell differenzierbar nach **x** ($\in \mathbb{R}^n$)

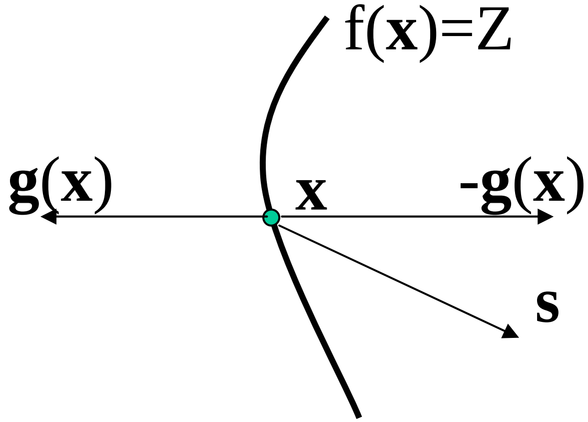
- $\mathbf{g}(\mathbf{x}) := \text{grad } f(\mathbf{x}) := (\partial f(\mathbf{x})/\partial x_1, \ldots, \partial f(\mathbf{x})/\partial x_n)^T$
- **g**(**x**) ist Gradient von f in **x**, weist in Richtung des stärksten Anstiegs von f (**x**)
- f(**x**)=Z -> **g**(**x**) orthogonal zu Hyperfläche $f(\mathbf{x}) \equiv Z$
  -> **-g** (**x**) ist Richtung größter Verbesserung (Verkleinerung) von f

$S(\mathbf{x}) := \{\mathbf{s} \mid \mathbf{s}$ zulässige Richtung in $\mathbf{x}\}$

Menge aller in $\mathbf{x}$ zulässigen Richtungen

Satz 7.1 (Verbessernde Richtungen)
Sei $f(\mathbf{x})$ stetig differenzierbar und $\mathbf{x}, \mathbf{s} \in \mathbb{R}^n$, $\mathbf{s}^T\mathbf{g}(\mathbf{x})<0$
(impliziert $\mathbf{s}^T(-\mathbf{g}(\mathbf{x}))>0$) => $\exists\ \delta'>0$: $f(\mathbf{x}+\delta\mathbf{s}) < f(\mathbf{x})\ \forall\ 0 < \delta \leq \delta'$.

Beweisskizze:



Wenn $\mathbf{s}^T\mathbf{g}(\mathbf{x})<0$, dann bilden $\mathbf{s}$ und $\mathbf{g}(\mathbf{x})$ stumpfen Winkel, $\mathbf{s}$ und $-\mathbf{g}(\mathbf{x})$ spitzen Winkel

Fortschreiten im spitzen Winkel zu $-\mathbf{g}(\mathbf{x})$ verkleinert Wert der Zielfunktion

$\mathbf{x} \in W$, $\mathbf{s} \in S(\mathbf{x})$ => zulässiges, verbesserndes Voranschreiten

Satz 7.2 (Notwendige Optimalitätsbedingungen)

Sei $f(\mathbf{x})$ stetig differenzierbar, $\mathbf{x}$ lokaler Minimalpunkt von $f(\mathbf{x})$ auf W

=> $\mathbf{s}^T\mathbf{g}(\mathbf{x}) \geq 0 \quad \forall\, \mathbf{s} \in S(\mathbf{x})$ (1)

Für einen inneren Punkt $\mathbf{x} \in W$ ist (1) nur zu erfüllen,

falls $g(\mathbf{x}) = \mathbf{0}$ (2).

Einen Punkt, der (1) erfüllt,
nennt man **stationären Punkt** von $f(\mathbf{x})$ auf W.
Für unrestringierte Probleme ist natürlich (2) bedeutsam!

Erinnerung an den eindimensionalen Fall:

„Kurvendiskussion“: $y = f(x)$

- Falls f auf $\mathbb{R}$ definiert, sind alle Punkte x aus $\mathbb{R}$ innere Punkte.
- Bedingungen für lokale Minimalstelle
  - notwendig: $f'(x)=0$
  - hinreichend: $f'(x)=0$ und $f''(x) > 0$

hier Erweiterung für n>1

- Für zweimal partiell differenzierbare Funktionen f(**x**) heißt die (symmetrische) Matrix der zweiten partiellen Ableitungen Hesse-Matrix

$$\mathbf{H} := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

und **H**(**x**) heißt Hesse-Matrix am Punkt **x**

Es wir im Folgenden auf positive Definitheit/Semidefinitheit ankommen

Daher zur Erinnerung

- Eine symmetrische n x n Matrix **A** heißt
  - positiv semidefinit wenn $\forall\, \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x} \neq \mathbf{0}$: $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$
  - positiv definit wenn $\forall\, \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x} \neq \mathbf{0}$: $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$

  ($\mathbf{x}^T \mathbf{A}\, \mathbf{x}$) heißt quadratische Form

Beispiel:

$$\mathbf{xAx}^T = (x_1, x_2)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1 - x_2, -x_1 + x_2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= (x_1 - x_2)x_1 + (-x_1 + x_2)x_2 = (x_1 - x_2)^2 \qquad \geq 0 \text{ für } \mathbf{x} \neq \mathbf{0}$$

Algorithmus um zu testen, ob symmetrische Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ positiv definit/semidefinit:

- **A** lässt sich als Produkt **A** = **LR** zweier Dreiecksmatrizen **L**, **R** darstellen
  (Berechnung von **L** und **R**, z.B. mittels Gauß-Algorithmus)

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & .. & 0 \\ l_{21} & 1 & 0 & : \\ & .. & . & 0 \\ l_{n1} & .. & l_{n,n-1} & 1 \end{pmatrix} \qquad \mathbf{R} = \begin{pmatrix} r_{11} & & .. & r_{1n} \\ 0 & r_{22} & & : \\ & .. & . & \\ 0 & .. & 0 & r_{nn} \end{pmatrix}$$

Da **A** symmetrisch gilt ferner $\mathbf{R} = \mathbf{DL}^T$ für eine Diagonalmatrix **D**

Aus $\mathbf{A} = \mathbf{LR} = \mathbf{LDL}^T$ folgt $\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T \mathbf{LDL}^T \mathbf{x}$

bzw. mit $\mathbf{y} := \mathbf{L}^T\mathbf{x}$ auch $\mathbf{z} = \mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{y}^T \mathbf{D}\, \mathbf{y}$ mit $z_i = \Sigma\, d_{ii}y_i^2$

=> **A** genau dann positiv definit, wenn alle $d_{ii}>0$

positiv semidefinit, wenn alle $d_{ii}\geq 0$

Folgender Satz (hier ohne Beweis) liefert Optimalitätsbedingungen für unrestringierte Probleme.

Für restringierte Probleme zusätzliche Probleme durch die Grenzen des zulässigen Bereichs

Satz 7.3 (Hinreichende Optimalitätsbedingung)
Sei f(**x**) zweimal stetig differenzierbar und **x** innerer Punkt von W, **g**(**x**)=0 und **H**(**x**) positiv definit
=> **x** ist lokaler Minimalpunkt von f(**x**) auf W.

Falls **H**(**x**) nur positiv semidefinit, so ist dies nur eine notwendige Bedingung

Nach Satz 7.2 ist $\mathbf{s}^T\mathbf{g}(\mathbf{x}) \geq 0$ für alle $\mathbf{s} \in S(\mathbf{x})$ eine notwendige Optimalitätsbedingung, die

- sich für unrestringierte Probleme auf $\mathbf{g}(\mathbf{x})=\mathbf{0}$ reduziert
- für restringierte Probleme (wo Optimalpunkte auf dem Rand von W liegen können) praktisch schlecht handhaben lässt (wegen der Schwierigkeit der expliziten Notierung von $S(\mathbf{x})$)

Abhilfe: Methode der Lagrange´schen Multiplikatoren

- Integration der Nebenbedingungen in die Zielfunktion
- Optimierung des unrestringierten Problems

Anmerkung: funktioniert auch bei Nebenbedingungsgleichungen durch Umformung auf Ungleichungen

Für Optimierungsproblem min $f(\mathbf{x})$ udN $h_i(\mathbf{x}) \leq 0$ $(i=1,...,m)$

- werden Lagrange-Multiplikatoren u_i $(i=1,...,m)$ eingeführt und Lagrange-Funktion $L: \mathbb{R}^{n+m} \to \mathbb{R}$ definiert durch
  $L(\mathbf{x},\mathbf{u}) := f(\mathbf{x}) + \Sigma_{i=1..m}\, u_i h_i(\mathbf{x}) = f(\mathbf{x}) + \mathbf{u}^T \mathbf{h}(\mathbf{x})$
  mit $\mathbf{h}(\mathbf{x}) = (h_1(\mathbf{x}),...,h_m(\mathbf{x}))^T$ und $\mathbf{u} = (u_1,...,u_m)^T$

Ein Punkt $(\mathbf{x}^* \in \mathbb{R}^n, \mathbf{u}^* \in \mathbb{R}_+^m)$ heißt Sattelpunkt von L, wenn

$L(\mathbf{x}^*,\mathbf{u}) \le L(\mathbf{x}^*,\mathbf{u}^*) \le L(\mathbf{x},\mathbf{u}^*)$ für alle $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{u} \in \mathbb{R}_+^m$

Satz 7.4 (Sattelpunkte und Optimalität)
Ist $(\mathbf{x}^*, \mathbf{u}^*)$ ein Sattelpunkt von L, so ist $\mathbf{x}^*$ eine optimale Lösung des Optimierungsproblems min $f(\mathbf{x})$ und $h_i(\mathbf{x}) \le 0$ $(i = 1,\dots,m)$.

Beweis:

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

- damit gilt $\mathbf{u}^T\mathbf{h}(\mathbf{x}^*) \le \mathbf{u}^{*T}\mathbf{h}(\mathbf{x}^*)$ für alle $\mathbf{u} \in R_+^m$
  - für $\mathbf{h}(\mathbf{x}^*) > 0$ und $\mathbf{u} > \mathbf{u}^*$ nicht erfüllbar
    damit $\mathbf{h}(\mathbf{x}^*) \le \mathbf{0} \Rightarrow \mathbf{x}^*$ zulässig für das Optimierungsproblem
  - $\mathbf{u} = \mathbf{0}$ in obiger Ungleichung bedingt $0 \le \mathbf{u}^{*T}\mathbf{h}(\mathbf{x}^*)$

Zusammen mit der zweiten Sattelpunktbedingung

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

folgt für alle $\mathbf{x} \in \mathbb{R}^n$: $f(\mathbf{x}^*) \le f(\mathbf{x}) + \mathbf{u}^{*T}\mathbf{h}(x)$

Da $\mathbf{u} \ge \mathbf{0}$ und für zulässige $\mathbf{x}$: $\mathbf{h}(\mathbf{x}) \le \mathbf{0}$ folgt $f(\mathbf{x}^*) \le f(\mathbf{x})$ ($\mathbf{x}^*$ ist optimal)

Umkehrung gilt nur für einige Problemklassen!

7.2 Konvexe Optimierung

Satz 7.3 legte Bedingungen für lokale Minimalpunkte fest.

- Bei linearen Optimierungsproblemen waren
  lokale Minimalpunkte gleichzeitig auch globale Minimalpunkte

Kann nicht allgemein gelten,
  aber möglicherweise über lineare Zielfunktionen hinaus ?

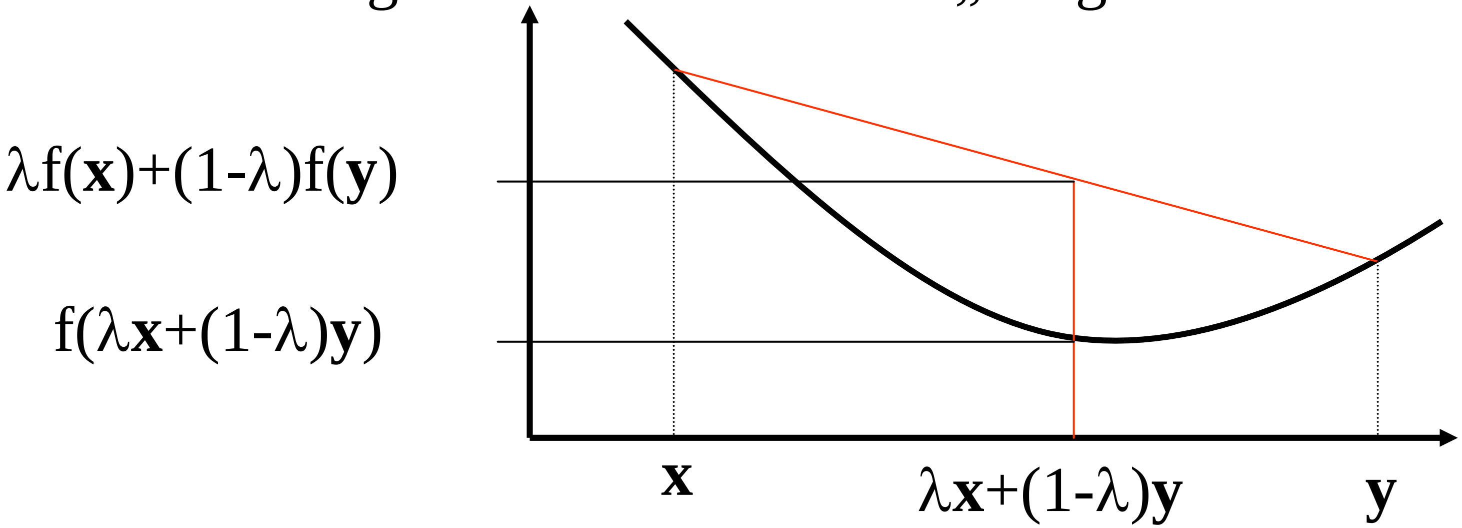
Zur Erinnerung:

- Menge $K \subseteq R^n$ heißt konvex, wenn für alle $\mathbf{x},\mathbf{y} \in K$ und $0 \leq \lambda \leq 1$
  $\lambda\mathbf{x}+(1-\lambda)\mathbf{y} \in K$ gilt
- Durchschnitt konvexer Mengen ist konvex
- Sei $K \subseteq \mathbb{R}^n$ konvex, eine Funktion $f: K \rightarrow \mathbb{R}$ heißt konvex,
  wenn für alle $\mathbf{x}$, $\mathbf{y} \in K$ und $0 \leq \lambda \leq 1$ gilt, dass
  $f(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x})+(1-\lambda)f(\mathbf{y})$

analog zu definieren:
  streng konvex für $<$, konkav für $\geq$, streng konkav für $>$

- Lineare Funktionen sind konkav und konvex (aber nicht streng)

Für K=ℝ: streng konvexe Funktionen „hängen nach unten durch“



Der folgende Satz folgt aus der Definition von Konvexität

Satz 7.5 (Eigenschaften konvexer Funktionen)

Seien $K \subseteq \mathbb{R}^n$ konvexe Menge, f und $f_1,\ldots,f_m$: $K \to \mathbb{R}$ konvexe Funktionen und $\alpha \in \mathbb{R}$, dann

- sind alle nichtnegativeen Linearkombinationen der f_i konvexe Funktionen auf K und
- die Mengen $\{\mathbf{x}\in K \mid f(\mathbf{x}) \leq \alpha\}$ (abgeschlossen) und $\{\mathbf{x}\in K \mid f(\mathbf{x}) < \alpha\}$ (offen) sind konvexe Mengen

Es gibt diverse nützliche Kriterien zur Feststellung der Konvexität einer Funktion, so gelten folgende Sätze
(hier ohne Beweis vorgestellt):

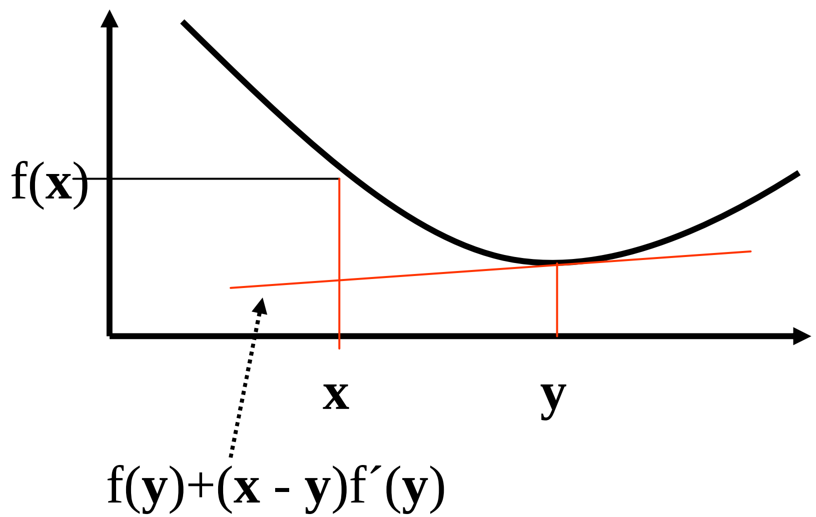
Satz 7.6 (Konvexitätskriterium I)
Sei $K \subseteq \mathbb{R}^n$ eine konvexe Menge mit inneren Punkten und $f : K \rightarrow \mathbb{R}$ zweimal stetig differenzierbar, dann gilt
f ist genau dann konvex (streng konvex) auf K, wenn die Hesse-Matrix $\mathbf{H}(\mathbf{x})$ für alle $\mathbf{x} \in K$ positiv semidefinit (positiv definit) ist.

Satz 7.7 (Konvexitätskriterium II)
Sei $K \subseteq \mathbb{R}^n$ konvexe Menge, $f : K \rightarrow \mathbb{R}$ stetig differenzierbar, dann gilt
f ist genau dann konvex auf K, wenn für alle $\mathbf{x}, \mathbf{y} \in K$:

$$f(\mathbf{x}) \geq f(\mathbf{y}) + (\mathbf{x} - \mathbf{y})^T \text{grad } f(\mathbf{y})$$

(f streng konvex bei „>“ statt „ ≥“).

Veranschaulichung (von Satz 7.7)



Konvexes Optimierungsproblem mit

- konvexer Zielfunktion $f(\mathbf{x})$
- konvexen Nebenbedingungen $h_i(\mathbf{x}) \leq 0$ (i=1,...,m)

und damit konvexer Lösungsmenge

(Durchschnitt von konvexen Mengen ist konvex siehe Folie 22)

Satz 7.8 (Hauptsatz der konvexen Optimierung)

Sei $W \subseteq \mathbb{R}^n$ konvexe (Lösungs-)Menge, $W \neq \emptyset$, $f: W \rightarrow \mathbb{R}$ konvexe Zielfunktion, dann ist

1. die Menge W^* aller globalen Minimalpunkte von f auf W konvex,
2. jeder lokale Minimalpunkt von f auf W auch globaler Minimalpunkt

Beweis:

zu 1.: Sei $\mathbf{x}^*$ globaler Minimalpunkt mit Minimum $Z^* := f(\mathbf{x}^*)$

nach Satz 7.5 ist die Menge $\{\mathbf{x} \in W | f(\mathbf{x}) \leq Z^*\}$ konvex

Beweis (Fortsetzung)

zu 2.: Sei $\mathbf{x}'$ lokaler Minimalpunkt mit Minimum $Z' := f(\mathbf{x}')$

Annahme: $\exists \mathbf{x}: f(\mathbf{x}) < Z'$ ($\Rightarrow \mathbf{x}'$ nicht globaler Minimalpunkt)

f ist konvex d.h. für $0<\lambda<1$ gilt

$$f(\mathbf{x}' + \lambda(\mathbf{x}-\mathbf{x}')) = f(\lambda\mathbf{x}+(1-\lambda)\mathbf{x}') \le \lambda\, f(\mathbf{x}) + (1-\lambda)\, f(\mathbf{x}') \le f(\mathbf{x}')$$

da $\lambda\mathbf{x}+(1-\lambda)\mathbf{x}' = \mathbf{x}'+\lambda(\mathbf{x}-\mathbf{x}')$

$\Rightarrow$ in ε-Umgebung von $\mathbf{x}'$ finden sich $\mathbf{x}''$ mit $f(\mathbf{x}'') < f(\mathbf{x}')$

(damit kann $\mathbf{x}'$ kein lokales Minimum sein!)

Insbesondere werden

- die notwendigen Bedingungen für lokalen Minimalpunkt aus Satz 7.2, nämlich $\mathbf{s}^T\mathbf{g}(\mathbf{x}) \ge 0$ für alle $\mathbf{s} \in S(\mathbf{x})$ und $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ für innere Punkte von W, für konvexe Optimierungsprobleme zu hinreichenden Bedingungen für globalen Minimalpunkt

Satz 7.9 (Hinreichende Optimalitätsbedingung der konv. Opt.)
Sei $W \subseteq \mathbb{R}^n$ konvexe (Lösungs-)Menge, f: $W \to \mathbb{R}$ konvexe, stetig differenzierbare (Ziel-)Funktion und $\mathbf{x}^* \in W$.
Gelte $\mathbf{s}^T\mathbf{g}(\mathbf{x^*}) \geq 0$ für alle $\mathbf{s} \in S(\mathbf{x^*})$ falls $\mathbf{x}^*$ kein innerer Punkt und $\mathbf{g}(\mathbf{x^*}) = \mathbf{0}$ falls $\mathbf{x}^*$ ein innerer Punkt ist
$\Rightarrow$ $\mathbf{x}^*$ ist globaler Minimalpunkt von f auf W

Beweis:

- $\mathbf{x} \neq \mathbf{x}^*$, $\mathbf{x} \in W$
  W konvex damit Verbindungsstrecke $\mathbf{x}^*$ nach $\mathbf{x}$ in W
- $\mathbf{x}$-$\mathbf{x}^* \in S(\mathbf{x}^*)$ (ist zulässige Richtung)
- für alle $\mathbf{s} \in S(\mathbf{x}^*)$: $\exists$ x $\in W$ und $\delta > 0$: $\delta\, \mathbf{s} = \mathbf{x}$-$\mathbf{x}^*$
  - damit gilt auch für alle $\mathbf{x} \in W$: $(\mathbf{x}-\mathbf{x}^*)^T\, \mathbf{g}(\mathbf{x}) \geq 0$
  - Satz 7.7 liefert dann $f(\mathbf{x}) \geq f(\mathbf{x}^*)+(\mathbf{x}-\mathbf{x}^*)^T\mathbf{g}(\mathbf{x}) \geq f(\mathbf{x}^*)$
    für alle $\mathbf{x} \in S(\mathbf{x}^*)$

damit ist $\mathbf{x}^*$ Minimalpunkt

Umkehrung von Satz 7.4 für konvexe Probleme:
Formulierung der sog. Slater-Bedingung:

$\exists\, \mathbf{x} \in \mathbb{R}^n$, so dass für alle nichtlinearen Nebenbedingungen $h_i(\mathbf{x})<0$ gilt

(bei ausschließlich nichtlinearen Nebenbedingungen impliziert Slater-Bedingung, dass W innere Punkte enthält)

Satz 7.10 (von Kuhn/Tucker)
Sei im Rahmen eines konvexen Optimierungsproblems die Slater-Bedingung erfüllt. Dann gilt:

$\mathbf{x}^*$ ist optimale Lösung des Optimierungsproblems

$$\Leftrightarrow$$

Lagrange-Funktion $L(\mathbf{x},\mathbf{u})$ besitzt Sattelpunkt $(\mathbf{x}^*,\mathbf{u}^*)$ mit $\mathbf{u}^*\geq 0$.

Globale Sattelpunkteigenschaft ist schlecht explizit nachprüfbar!

Für (hier gegebene) differenzierbare, konvexe Funktionen f und h_i kann die globale Sattelpunktbedingung durch äquivalente lokale Bedingungen ersetzt werden:

Karush-Kuhn-Tucker-Bedingungen (KKT-Bedingunen)

- stetig differenzierbare, konvexe, f, h_i (i=1,...,m)
- Slaterbedingung erfüllt

Dann ist **x*** eine optimale Lösung falls

$$
\begin{aligned}
\operatorname{grad} f(\mathbf{x}^*) + \sum_{i=1}^{m} u^*_i \operatorname{grad} h_i(\mathbf{x}^*) &= 0 \\
u^*_i\, h_i(\mathbf{x}^*) &= 0 \qquad i = 1,\dots,m \\
h_i(\mathbf{x}^*) &\le 0 \qquad i = 1,\dots,m \\
\mathbf{u}^* &\ge 0
\end{aligned}
$$

KKT-Bedingungen

- lassen sich für spezifische (Unter-)Klassen weiter konkretisieren
- so z.B. (breit untersucht) für quadratische Optimierungsprobleme d.h. quadratische Zielfunktion und lineare Nebenbedingungen (W ist konvexes Polyeder)

7.3 Verfahren für unrestringierte Probleme

Zur Klärung des Vorgehens betrachten wir zuerst eindimensionale Optimierungsprobleme
mit $f: W \rightarrow \mathbb{R}$ wobei $W = \mathbb{R}$ oder $\mathbb{R}_+$ oder $[a,b] \in \mathbb{R}$

- einfachste Optimierungsaufgabe mit eigenem Interesse
- Teilproblem in höherdimensionalen Optimierungsaufgaben (wird in komplexeren Lösungsverfahren als ein Lösungsschritt verwendet)

Problem besitzt genau eine optimale Lösung, wenn f im zulässigen Bereich unimodal ist

(d.h. $\exists x^* \in [a,b]$: f streng monoton fallend auf $[a,x^*]$
f streng monoton steigend auf $[x^*,b]$)

Es gilt: f streng konvex $\Rightarrow$ unimodal (aber nicht die Umkehrung)

Einfache Lösungsverfahren („ableitungsfrei“)

- nutzen nur Funktionswerte
- verkleinern „Suchintervall“ schrittweise

Typisch: 2 „Stützstellen“, x_1, x_2 mit $a < x_1 < x_2 < b$

- falls $f(x_1) \leq f(x_2)$ rekursiv weiter mit $[a,x_2]$,
- falls $f(x_1) > f(x_2)$ rekursiv weiter mit $[x_1,b]$
- Teilung z.B. „goldener Schnitt“ mit $\delta=(\sqrt{5}-1)/2 \approx 0{,}618$ wg
  $1/\delta= \delta/(1- \delta)$ also $\delta^2=1- \delta$
  Aufteilung $x_1=a+(1- \delta)(b-a)$ und $x_2=a+ \delta(b-a)$
  Mittelwert aus Intervall als Schätzwert
  durch Wahl der Punkte wir sichergestellt, dass pro Teilung des Intervalls nur eine Funktionsauswertung erfolgen muss

Beispiel x^* liege im Intervall $[x_1,b]$ also $f(x_1)>f(x_2)$

Dann gilt:

$x_3 = x_1 + (1-\delta)(b-x_1) = b - \delta(b-x_1)$

da $x_1 = b - \delta(b-a)$ gilt $x_3 = b - \delta^2(b-a) = b - (1-\delta)(b-a) = x_2$

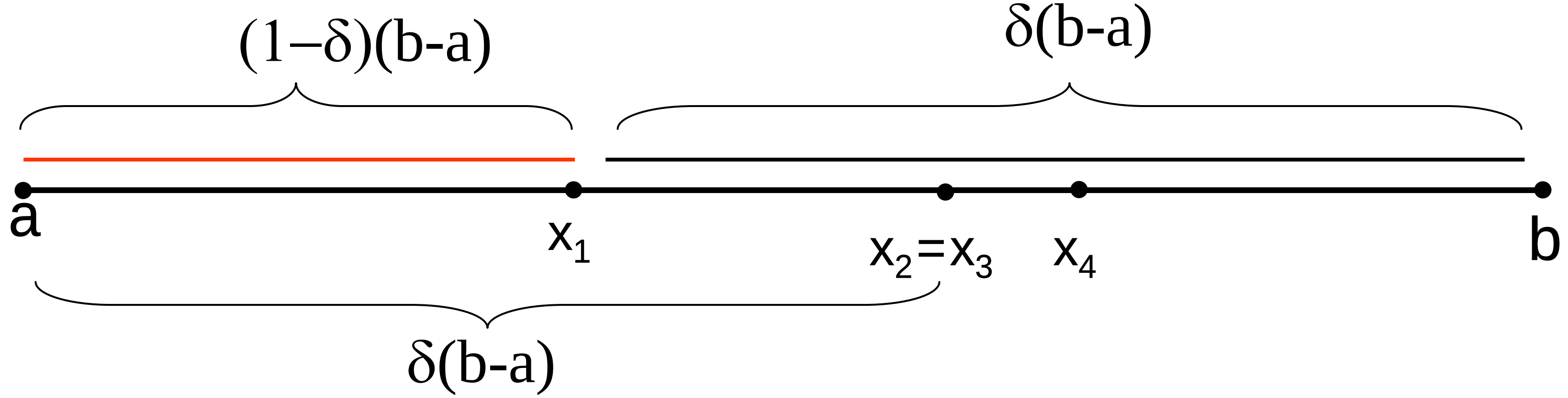
$x_4 = x_1 + \delta(b-x_1)$

$\Rightarrow$nur $f(x_4)$ ist neu zu berechnen

Approximation des Optimums durch Mittelpunkt des letzten Intervalls

$\Rightarrow$ obere Schranke für den maximalen Fehler nach k Iterationen

$\delta^k(b-a)/2 = (b-a)/(2d^k)$ mit $d = 1/\delta \approx 1.618$



Aufwändige (i.d.R. bessere) Lösungsverfahren

- nutzen Funktionswerte + Werte 1. (und höherer) Ableitungen
- konstruieren Folge von (hoffentlich) konvergierenden Punkten

Wir beginnen mit dem Newton Verfahren, ...

Newton-Verfahren:

Annahme: f sei zweimal stetig differenzierbar

- Minimalpunkt $x^* \in (a,b)$ und Näherungspunkt $x^i \in (a,b)$; i=0,1,2,...
  - Vorbereitung durch „einfaches“ Verfahren zur Intervallbestimmung,
  - Grenzpunkte a, b sind separat zu prüfen
- Ziel: Bestimmung x^*: $f'(x^*)=0$ d.h. Nullstelle von f'
  (falls f streng konvex: notwendig und hinreichend für globale Minimalstelle, ansonsten evtl. zweite Ableitung prüfen)

- Vorgang: Berechnung $f'(x^i)$, $f''(x^i)$,
  - Linearisierung der Ableitungsfunktion f' in x^i
  - Bestimmung der Tangentennullstelle zu $x^{i+1} := x^i - f'(x^i)/f''(x^i)$
- Newton verwendet quadratische (parabolische) Approximation von f
  $\Rightarrow$ bei quadratischem f ist f‘ linear und
  Optimalstelle wird in einem Schritt gefunden
- Konvergenz bei beliebigem Startpunkt x^0 nicht gesichert, fehlende Konvergenz an x_i Werten außerhalb [a,b] erkennbar (analytische Bedingungen für gesichertes x^0 kompliziert)
- bei der praktischen Anwendung Voranalyse mittels der Methode des Goldenen Schnittes, um „nicht zu großes“ Intervall [a,b] zu ermitteln, in dem x* liegt
- bei Divergenz des Newton-Verfahrens, Intervall [a,b] verkleinern

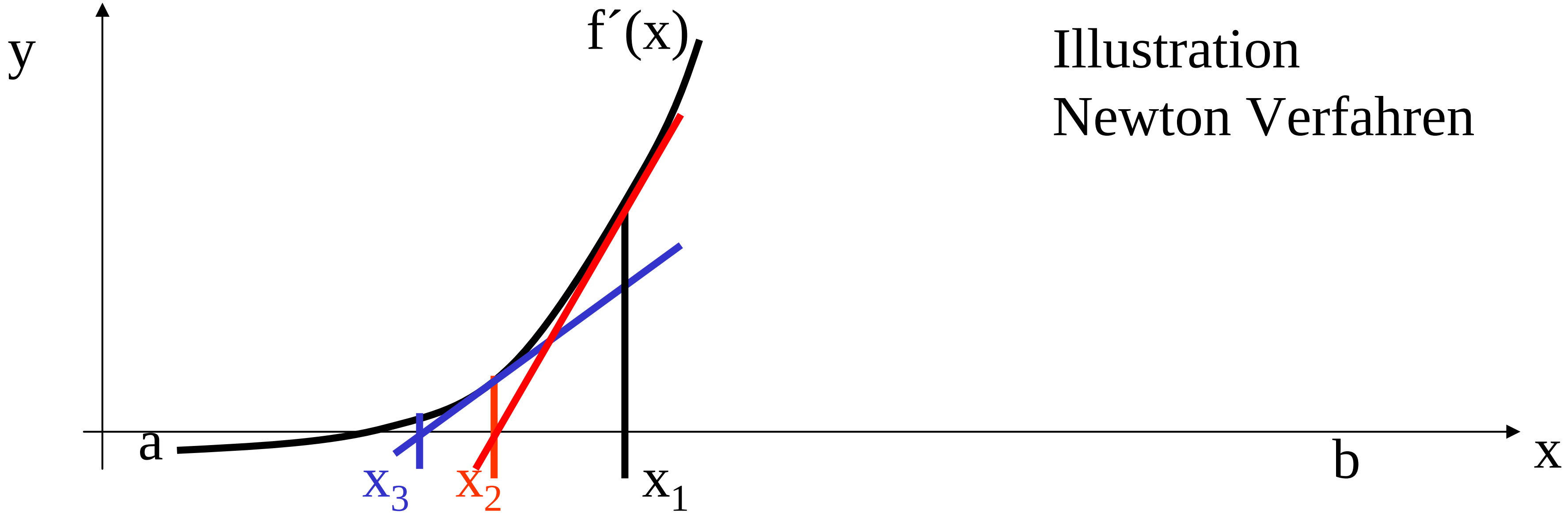


Illustration
Newton Verfahren

- sind f´,f´´ nicht analytisch bekannt, kann mit numerischen Näherungen auf Basis der Funktionswerte „benachbarter“ Punkte gearbeitet werden
  d.h. ersetze Differentenquotienten durch

$$f'(x) \approx \frac{f(x) - f(x-h)}{h} \text{ und } f''(x) = \frac{f'(x) - f'(x-h)}{h} = \frac{f(x) - f(x-2h)}{h^2}$$

Allgemeine unrestringierte Optimierungsprobleme im $\mathbb{R}^n$

sind generell iterativ:

also Berechnung $\mathbf{x}^0 \to \mathbf{x}^1 \ldots \to \mathbf{x}^i \to \ldots$

mit zulässigen $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^i \in W \subseteq \mathbb{R}^n$,

i.d.R. inhärent nicht abbrechend $\Rightarrow$ Näherungslösungen

Hoffnung auf Konvergenz zu optimalem Punkt $\mathbf{x}^*$:

$$\mathbf{s}^T \mathbf{g}(\mathbf{x}^*) \geq 0 \quad \forall\, \mathbf{s} \in S(\mathbf{x}^*)$$

bzw $\mathbf{g}(\mathbf{x}^*) = \mathbf{0}$ im unrestringierten Fall mit $W=\mathbb{R}^n$

bei nichtkonvexem Problem lediglich stationärer Punkt,

nicht notwendig Extrempunkt / optimaler Punkt erreicht

bezeichne LM die Lösungsmenge (stationäre Punkte)

$$LM := \{\, \mathbf{x}^* \in W \mid \mathbf{s}^T \mathbf{g}(\mathbf{x}^*) \geq 0 \;\forall\, \mathbf{s} \in S(\mathbf{x}^*) \,\}$$

bzw $LM := \{\, \mathbf{x}^* \in M \mid \mathbf{g}(\mathbf{x}^*) = \mathbf{0} \}$

Konvergenz und Konvergenzgeschwindigkeit im konkreten Fall zu prüfen!

Allgemeine Form min $f(\mathbf{x})$ udN $\mathbf{x} \in \mathbb{R}^n$:
f sei mindestens einmal stetig differenzierbar
LM:=$\{ \mathbf{x}^* \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}^*) = \mathbf{0} \}$

Behandlungsmöglichkeit mittels Abstiegsverfahren

- Initialisierung: wähle $\mathbf{x}^0 \in \mathbb{R}^n$, $i:=0$
- Iteration über
  - Abbruchprüfung:
    berechne $\mathbf{g}^i := \mathbf{g}(\mathbf{x}^i)$, terminiere falls $\mathbf{g}^i = \mathbf{0}$ (bzw $\|\mathbf{g}^i\| < \varepsilon$)
  - Abstiegsrichtung, Suchrichtung:
    bestimme $\mathbf{s}^i \in \mathbb{R}^n$ mit $(\mathbf{s}^i)^T \mathbf{g}^i < 0$
  - Schrittweite:
    berechne $d^i > 0 : f(\mathbf{x}^i + d^i \mathbf{s}^i) < f(\mathbf{x}^i)$
    optimale Schrittweite ist eindimensionales Subproblem
  - Fortschritt $\mathbf{x}^{i+1} := \mathbf{x}^i + d^i \mathbf{s}^i$; $i := i+1$

$$\min \ f(\mathbf{x}^i + d^i \mathbf{s}^i)$$
$$\text{udN} \ d^i \in R_+$$

Konkrete Verfahren unterscheiden sich durch die Wahl der Abstiegsrichtung und der Schrittweite

- Klassisches Gradientenverfahren, Verfahren des steilsten Abstiegs:
  $\mathbf{s}^i := - \mathbf{g}^i$ (oft langsame Konvergenz)
- i.d.R. konvergiert mehrdimensionales Newtonverfahren schneller:
  zusätzlich benötigt: zweite Ableitungen
  $\Rightarrow$ Forderung f mindestens 2-mal stetig differenzierbar
  - Notation: $\mathbf{H}^i := \mathbf{H}(\mathbf{x}^i)$ Hesse-Matrix
  - $\mathbf{H}^i$ positiv definit (Satz 7.3)
    $\Rightarrow$ f ist in Umgebung von $\mathbf{x}^i$ streng konvex

  Standard Newton setzt $\mathbf{s}^i := - (\mathbf{H}^i)^{-1} \mathbf{g}^i$ und $d^i := 1$
  so dass $\mathbf{x}^{i+1} := \mathbf{x}^i - (\mathbf{H}^i)^{-1} \mathbf{g}^i$

zur Erinnerung: eindimensionaler Fall: $\mathbf{x}^{i+1} := \mathbf{x}^i - f'(\mathbf{x}^i)/f''(\mathbf{x}^i)$

  - Anmerkungen
    - $d^i := 1$ i.a. nicht optimal, sollte optimiert werden!
    - $\mathbf{H}^i$ muss nicht explizit invertiert werden, da
      $\mathbf{H}^i \mathbf{x}^{i+1} = \mathbf{H}^i \mathbf{x}^i - \mathbf{g}^i \Rightarrow \mathbf{H}^i \mathbf{x}^{i+1} = \mathbf{b}^i$
      $\mathbf{x}^{i+1}$ als Lösung eines linearen Gleichungssystems

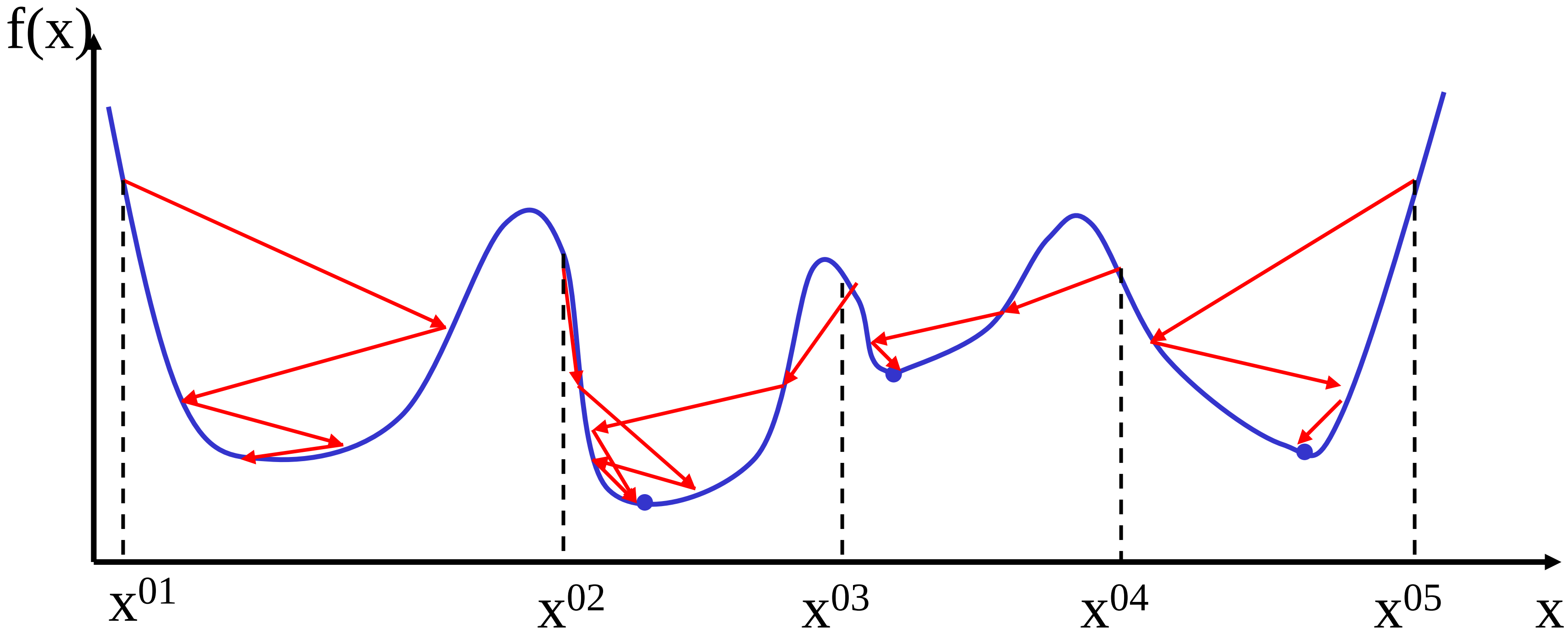
Nachteile des Newtonverfahrens: hoher Rechenaufwand
- Hesse-Matrix bestimmen
- lineares Gleichungssystem lösen

abgemildert in vereinfachtem Newton Verfahren,
- welches $\mathbf{H}^i$ für einige Schritte i, i+1, ... verwendet
  (verringert Aufwand aber reduziert Konvergenzgeschwindigkeit)

weitere Verfahren
- „zwischen“ Gradientenverfahren (schlechte Konvergenz)
- und Newtonverfahren (hoher Aufwand)

z.B. Verfahren der konjugierten Gradienten
- bzgl Aufwand:
  Verfahren mit konj. Gradienten < Newton
- bzgl Güte der Konvergenz:
  Gradientenverfahren < Verfahren mit konj. Gradienten

Was können wir von den vorgestellten Verfahren erwarten?



- Wenn überhaupt, nur lokale Konvergenz
- Unterschiedliche Startpunkte führen zu unterschiedlichen Minima, Sattelpunkten oder vielleicht auch zu Divergenz
- Heuristisches Vorgehen: Starten des Verfahrens an unterschiedlichen Punkten und Vergleich der Resultate

7.4 Verfahren für restringierte Probleme

min f($\mathbf{x}$) udN $\mathbf{x} \in W \subset \mathbb{R}^n$ wobei $W \neq \emptyset$ und f stetig differenzierbar

Menge der stationären Punkte

$$LM := \{ \mathbf{x}^* \in W \mid \mathbf{s}^T \mathbf{g}(\mathbf{x}^*) \geq 0 \ \forall \ \mathbf{s} \in S(\mathbf{x}^*) \}$$

und falls f und W konvex $\Rightarrow$ $\mathbf{x}^*$ globaler Minimalpunkt (Satz 7.9)

im allgemeinen Fall sind keine globalen Aussagen möglich

Ein robustes Verfahren für (fast) alle Problemklassen existiert nicht!

Praktisch eingesetzt werden folgende Verfahren(sklassen):

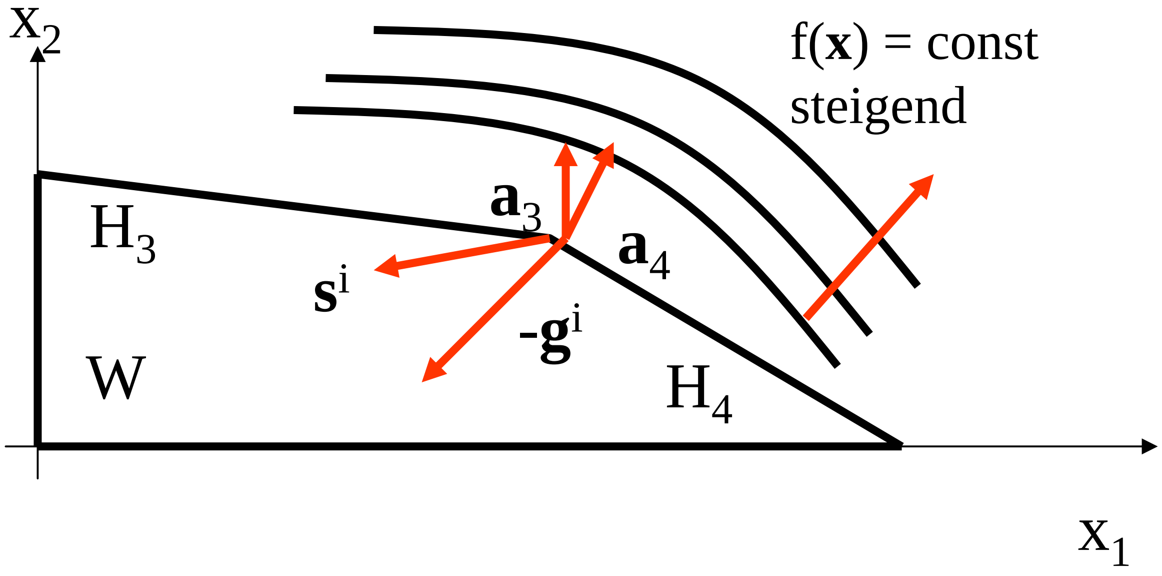
- Verfahren zulässiger Richtungen
- Verfahren der Straffunktionen (Barrierefunktionen)
- Schnittebenenverfahren
- Karush-Kuhn-Tucker-Verfahren
- ...

Verfahren zulässiger Richtungen

- Erweiterung des Vorgehens für unrestringierte Probleme (siehe Folie 27)
- Anwendung i. w. für lineare Restriktionen $\mathbf{Ax} \leq \mathbf{b}$ und $\mathbf{b} \in \mathbb{R}^m$
- Initialisierung: wähle $\mathbf{x}^0 \in W$, i:=0
- Iteration über
  - Abstiegsrichtung:
    bestimme $\mathbf{s}^i \in \mathbb{R}^n$ mit $(\mathbf{s}^i)^T \mathbf{g}^i < 0$
    falls erfolglos: Abbruch
  - Schrittweite:
    berechne $d^i > 0 : f(\mathbf{x}^i + d^i \mathbf{s}^i) < f(\mathbf{x}^i)$
    mit optimaler Schrittweite, aber unter Beachtung Restriktionen
    d.h. $\mathbf{x}^i + d^i \mathbf{s}^i \in W$
  - Fortschritt $\mathbf{x}^{i+1} := \mathbf{x}^i + d^i \mathbf{s}^i$; $i := i+1$

Iterationspunkt $\mathbf{x}^i \in W$ gilt für eine Ausgangslösung

- sei I^i Menge der Indizes $k \in \{1,\ldots,m\}$, so dass
  $\Sigma_{j=1,\ldots,n}\, a_{kj} x^i_j = \mathbf{a}_k^T \mathbf{x}^i = b_k$
  - falls $I^i \neq \emptyset$, liegt $\mathbf{x}^i$ auf den Hyperebenen H_k ($k \in I^i$), die durch $\mathbf{a}_k^T \mathbf{x}^i$ gegeben sind ($\mathbf{a}_k$ ist orthogonal zu H_k „nach außen“)
    H_k ($k \in I^i$) heißen aktive Hyperebenen
  - falls $I^i = \emptyset$, so ist ein Punkt im Inneren von W



- Bedingungen Fortschrittsrichtung $\mathbf{s}^i$:
- im stumpfen Winkel zu $\mathbf{a}_k$, $k \in I^i$ „zulässig“, $(\mathbf{a}_k)^T(\mathbf{x}^i + \delta \mathbf{s}^i) \leq b_k$
- im spitzen Winkel zu **-g** „Zielfunktionswert besser / kleiner“

Bedingungen formal

- stumpfe, äußerstenfalls rechte Winkel: $(\mathbf{a}_k)^T\mathbf{s}^i \leq 0$ für $k \in I^i$
- spitzer Winkel: $(-\mathbf{g}^i)^T\mathbf{s}^i > 0 \Rightarrow (\mathbf{g}^i)^T\mathbf{s}^i < 0$

Verschiedene Regeln zur Wahl von $\mathbf{s}^i$

- Rosen-Verfahren
  - berechnet $\mathbf{s}^i$ als Projektion von $\mathbf{g}^i$ auf $\cap(H_k;\, k \in I^i)$ für $I^i \neq \emptyset$
  - verwendet $-\mathbf{g}^i$ als $\mathbf{s}^i$ für $I^i = \emptyset$

  in beiden Fällen ist $\|\mathbf{s}^i\| < \varepsilon$ (Nullvektor) Abbruchbedingung (stationärer Punkt erreicht)
- Zusätzlich zu beachten:
  - Schrittweite „d^i“ so dass W nicht verlassen wird, wegen linearer Nebenbedingungen leicht formalisierbar, u.U. zu optimieren (eindimensionale Optimierung)
  - Begrenzung Schrittweite führt implizit auch zur Entdeckung „unbegrenzten Problems“

Verfahren der Straffunktionen (Barrierefunktionen)

Optimierungsproblem min $f(\mathbf{x})$ udN $h_i(\mathbf{x}) \leq 0$ $(i=1,...,m)$

- Restringiertes Optimierungsproblem durch unrestringiertes Optimierungsproblem approximiert derart dass
  1. Verlassen des zulässigen Bereichs mit „Strafkosten" belastet Veränderung von $f(\mathbf{x})$ außerhalb von W oder
  2. Rand des zulässigen Bereichs in „Barriere" verwandelt Veränderung $f(\mathbf{x})$ bei Annäherung an Rand W

Skizze Straffunktionen

Ursprüngliches Problem min $f(\mathbf{x})$ udN $h_i(\mathbf{x}) \leq 0$ $(i=1,...,m)$

wird ersetzt durch min $f'(\mathbf{x})$ udN $\mathbf{x} \in \mathbb{R}^n$

wobei $f'(\mathbf{x}) := f(\mathbf{x}) + p(\mathbf{x})$ mit $p(\mathbf{x}) = 0$ falls $\mathbf{x} \in W$, $p(\mathbf{x}) > 0$ sonst

verschiedene Straffunktionen, oft verwendet z.B. :

- $p(\mathbf{x}) := \mathbf{h}^+(\mathbf{x})^T \, \mathbf{h}^+(\mathbf{x})$ mit $h^+_i(\mathbf{x}) := \max(0, h_i(\mathbf{x}))$

Idee i.d.R. verfeinert in Form $f'(\mathbf{x}) := f(\mathbf{x}) + r^j \, p(\mathbf{x})$ (j Iterationsindex) mit $r^j > 0$, Vergrößerung der r^j im Iterationsverlauf

Skizze Barrierefunktionen

Voraussetzung: W besitzt innere Punkte $W^o \in W$

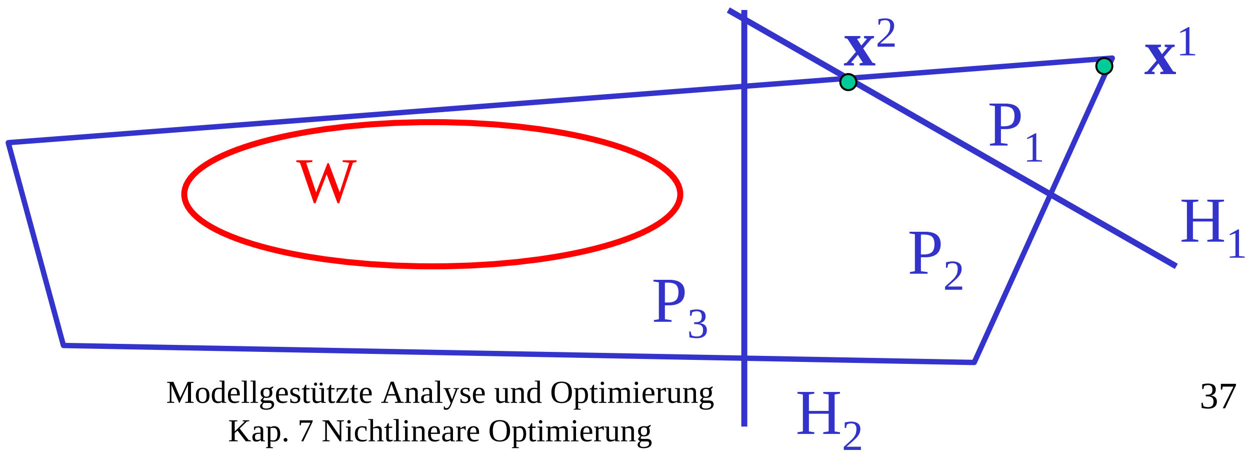
- Barrierefunktionen $b: W^o \to \mathbb{R}$ besitzen folgende Eigenschaften
  - b stetig auf W^o
  - $b(\mathbf{x}) \to \infty$ bei Annäherung von **x** an den Rand von W
- verschiedene Barrierefunktionen
  - im Falle von Nebenbedingungen $h_i(\mathbf{x})<0$
    (Rand also nicht erlaubt) oft verwendet
    $b_1(\mathbf{x}) := -\Sigma \ln(-h_i(\mathbf{x}))$ und $b_2(\mathbf{x}) := -\Sigma\, 1/h_i(\mathbf{x})$
    folgende Behandlung wie unrestringiertes Problem:
    - Minimierung der Ersatzfunktion
      $f'(\mathbf{x}) := f(\mathbf{x}) + 1/r^j\, b(\mathbf{x})$
      mit $r^j > 0$, Vergrößerung der r^j im Iterationsverlauf

Schnittebenenverfahren

Vorrangig bei nichtlinearen Restriktionen eingesetzt

Zunächst lineare Zielfunktion

- wobei W nichtleer, kompakt, konvex,
- aber kein Polytop (sonst wäre der Fall ja bereits bekannt)
- „um W" wird ein konvexes Polytop $P_1 \supset W$ gelegt
- Optimierung (z.B. mittels Simplex) dann auf P_1 statt W liefert Lösung $\mathbf{x}^1$
- falls $\mathbf{x}^1 \in W \Rightarrow \mathbf{x}^1$ ist optimale Lösung
- falls $\mathbf{x}^1 \notin W$
    - bestimme Hyperebene H_1, welche P_1 beschneidet,
    - so dass konvexes Polytop P_2 mit $P_2 \supset P_1 \supset W$ und $x^1 \notin P_2$ entsteht
- und so fort ...



Verschiedene Verfahren zur Konstruktion der Schnittebenen sind im Gebrauch

Eigentlich doch auch bei nichtlinearer Zielfunktion eingesetzt ???

Anmerkung:

- Die bisher angenommene Linearität von f ist nicht wesentlich!
- Wir betrachten min f(**x**) udN $\mathbf{x} \in W$
  wobei f nichtlinear, konvex und W nichtleer, kompakt, konvex
  - Einführung zusätzlicher Variabler $x_{n+1} \in \mathbb{R}$ und
  - zusätzlicher Nebenbedingung $f(\mathbf{x}) - x_{n+1} \leq 0$

  Betrachte Optimierungsproblem

  $$\min x_{n+1} \text{ udN } \mathbf{x} \in W \text{ und } f(\mathbf{x}) - x_{n+1} \leq 0$$

  mit linearer Zielfunktion
  - $\mathbf{x}^*$ ist genau dann optimale Lösung des Ursprungsproblems,
  - wenn $(\mathbf{x}^*, x^*_{n+1})$ Lösung des Ersatzproblems ist,
  - wobei $x^*_{n+1} = f(\mathbf{x}^*)$

Karush-Kuhn-Tucker-Verfahren

Analytische Verfolgung der KKT-Bedingungen (siehe Folie 19) (hinreichend für Optimalität)

- für konkrete Problemklassen(z.B. für quadratische Zielfunktion) ggf. einfache behandelbare Ersatzprobleme (z.B. mit Methoden der linearen Optimierung)

An diesem Punkt wollen wir jedoch die Betrachtung nichtlinearer Optimierungsprobleme beenden!

Ausblick:

- Es sind noch viele interessante Resultate bekannt: Lagrange Dualität, Separable Optimierung, Quotientenoptimierung, konkave Optimierung, B&B-Verfahren, Deflations- und Tunnel-Techniken, nichtdeterministische Verfahren ....

Im praktischen Einsatz mehrere Methoden „ausprobieren“, unterschiedliche Startpunkte wählen ... (viel Heuristik!)