Konvexe Optimierung

11. Mai 2017

Inhaltsverzeichnis

1. Einführung in die Optimierung

1.1. Grundlagen

Definition 1.1. Ein (mathematisches) **Optimierungsproblem** hat die folgende Form

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{F} \end{aligned} \tag{P}$$

- $x = (x_1, \cdots, x_n)^\mathsf{T} \in \mathbb{R}^n$ sind die **Variablen**
- $f\colon D \to \mathbb{R}, D \subset \mathbb{R}^n$, ist die **Zielfunktion**
- $\mathcal{F}$ ist die **zulässige Menge**; ihre Elemente heißen **zulässige Punkte** ◇

Die Nebenbedingungen

- Ist $\mathcal{F} = D$, dann spricht man von einem **unrestringierten Problem**.
- Wird $\mathcal{F}$ durch Nebenbedingungen (Restiktionen) definiert, dann heißt (P) **restringiertes Optimierungsproblem** oder **Optimierungsproblem mit Nebenbedingungen**; $\mathcal{F}$ ist hierbei typischerweise durch Gleichungen und Ungleichungen definiert:

$$\mathcal{F} = \{x \in \mathbb{R}^n \mid h(x) = 0,\ g(x) \leq 0\}$$

Einfache Beispiele

a) Mit $\mathcal{F} = D = \mathbb{R}$ sei $f\colon \mathbb{R} \to \mathbb{R}$ durch $f(x) = x^2$ definiert. Dann ist $x^* = 0$ die eindeutig bestimmte Lösung von (P).

b) Mit $\mathcal{F} = D = \mathbb{R}$ sei $f\colon \mathbb{R} \to \mathbb{R}$ durch $f(x) = \sin(x)$ definiert. Dann hat (P) unendlich viele Lösungen $x^* = 2k\pi - \frac{\pi}{2}, k \in \mathbb{Z}$.

c) Mit $\mathcal{F} = D = \mathbb{R}$ sei $f\colon \mathbb{R} \to \mathbb{R}$ durch $f(x) = x$ definiert. In diesem Fall ist f auf $\mathcal{F}$ nicht nach unten beschränkt und (P) hat keine Lösung.

d) Mit $\mathcal{F} = D = \mathbb{R}$ sei $f\colon \mathbb{R} \to \mathbb{R}$ durch

$$f(x) = (2x - 2)^2(3x + 3)^2 + 10x$$

definiert. Dann ist der Punkt $x^* = -1$ globale Lösung von (P). Der Punkt $\tilde{x} = 1$ ist eine lokale Lösung.

e) Wir betrachten das Problem

$$\begin{aligned} \min_x \quad & f(x) = x^3 \\ \text{s.t.} \quad & x \geq 1 \end{aligned} \tag{P2}$$

Hier ist $D = \mathbb{R}$ und $\mathcal{F} = \{x \in \mathbb{R} \mid x \geq 1\}$. Der Punkt $x^* = 1$ ist Lösung. Ohne die Restriktion hätte das Problem (P2) keine Lösung.

Definition 1.2. (lokaler Minimalpunkt)
Ein Punkt $x^* \in \mathcal{F}$ heißt **lokaler Minimalpunkt** von f auf $\mathcal{F}$ oder **lokale Lösung** von (P), falls es ein $r > 0$ mit

$$f(x) \geq f(x^*) \qquad \forall x \in \mathcal{F} \cap B(x^*, r)$$

gibt.

Ein Punkt $x^* \in \mathcal{F}$ heißt **strikter lokaler Minimalpunkt** von f auf $\mathcal{F}$ oder **strikte lokale Lösung** von (P), falls es ein $r > 0$ mit

$$f(x) > f(x^*) \qquad \forall x \in \mathcal{F} \cap B(x^*, r),\ x \neq x^*$$

gibt. ◇

Bemerkung 1.3. Eine offene Kugel mit Radius r um x ist definiert als,

$$B(x, r) = \{y \in \mathbb{R}^n \ \mid \|y - x\| < r\}.$$

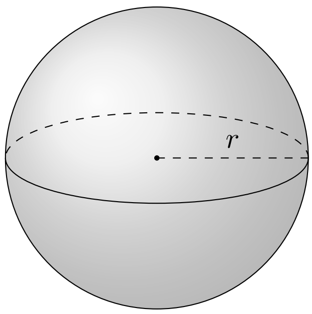

Abbildung 1.1.: Beispiel für $B(x, r)$ für $x \in \mathbb{R}^3$

◇

Definition 1.4. (globaler Minimalpunkt)
Ein Punkt $x^* \in \mathcal{F}$ heißt **globaler Minimalpunkt** von f auf $\mathcal{F}$ oder **globale Lösung** von (P), falls

$$f(x) \geq f(x^*) \qquad \forall x \in \mathcal{F}$$

gibt.

Ein Punkt $x^* \in \mathcal{F}$ heißt **strikter globaler Minimalpunkt** von f auf $\mathcal{F}$ oder **strikte globale Lösung** von (P), falls

$$f(x) > f(x^*) \qquad \forall x \in \mathcal{F},\ x \neq x^*$$

gibt. ◇

Einfache Beispiele

f) Die Funktion $f(x) = \sin(x)$ hat unendlich viele globale Minima.

g) Die Funktion

$$f(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2) - \cos(x_1^2) - \cos(x_2^2)$$

hat ein striktes globales Minimum und noch weitere strikte lokale Minima. Die Funktion besteht aus einem quadratischen Term und noch zwei weiteren Termen, die man als „Rauschen“ interpretieren kann.

Maximierung

- Oft soll die Zielfunktion f maximiert werden, d. h., wir suchen ein $x^* \in \mathcal{F}$ mit

$$f(x) \leq f(x^*) \qquad \forall x \in \mathcal{F}$$

- Diese Aufgabe ist äquivalent dazu, ein $x^* \in \mathcal{F}$ zu finden mit

$$-f(x) \geq -f(x^*) \qquad \forall x \in \mathcal{F}$$

$\Rightarrow$ Äquivalentes Problem:

$$\begin{array}{rl} \min & g(x) = -f(x) \\ \text{s.t.} & x \in \mathcal{F} \end{array} \tag{PMax}$$

- Es genügt daher, nur Minimierungsaufgaben zu betrachten.

1.2. Beispiele

Portfolio-Optimierung

- Variablen: Aufteilung des Portfolios
- Nebenbedingungen: Budget, Beschränkungen bei Investitionen
- Zielfunktion: maximale Rendite, minimales Risiko

Optimale Steuerung eines Raketenautos

- Variablen: Beschleunigung des Autos
- Nebenbedingungen: Dynamik, maximale Beschleunigung
- Zielfunktion: Zielort erreichen, Verbrauch an Treibstoff minimieren

Data Fitting

- Variablen: Modell-Parameter
- Nebenbedingungen: geg. Informationen, Parameterbeschränkungen
- Zielfunktion: Abweichung/Fehler minimieren

1.2.1. Portfolio-Optimierung

- Es werden n Wertpapiere gehandelt, wobei R_j die Rendite des $j-$ten Wertpapiers in der nächsten Zeitperiode darstellt (**Zufallsvariable**).
- Ein **Portfolio** besteht aus einer Zusammenstellung dieser Wertpapiere, dargestellt durch nichtnegative Zahlen $x_j \in \mathbb{R}^+, j = 1, \ldots, n$.
- Rendite eines gegebenen Portfolios und die erwartete Rendite:

$$R = \sum_{j=1}^{n} x_j Rj \qquad \mathbb{E}[R] = \sum_{j=1}^{n} x_j \mathbb{E}[R_j]$$

- Als Maß für das Risiko nutzen wir die durchschnittliche absolute Abweichung vom Erwartungswert

$$\mathbb{E}[R - \mathbb{E}[R]] = \mathbb{E}\left| \sum_{j=1}^{n} x_j (R_j - \mathbb{E}[R_j]) \right|$$

Es ergibt sich folgendes Optimierungsproblem:

$$\begin{aligned} \max_{x_1,\ldots,x_n} \quad & \mu \sum_{j=1}^{n} x_j \mathbb{E}[R_j] - \mathbb{E}\left| \sum_{j=1}^{n} x_j (R_j - \mathbb{E}[R_j]) \right| \\ \text{s.t.} \quad & \sum_{j=1}^{n} x_j = 1 \\ & x_j \geq 0, \qquad j = 1, 2, \ldots, n \end{aligned}$$

- Die beiden gegensätzlichen Ziele (erwartete Rendite maximieren vs. Risiko minimieren) werden durch den Parameter μ gewichtet
- Approximieren wir $\mathbb{E}[R_j]$ durch den Mittelwert der letzten beobachteten Renditen, lässt sich dieses Problem in ein **lineares Optimierungsproblem** transformieren

1.2.2. Optimale Steuerung - Raketenauto

Wir betrachten ein Auto mit Raketenantrieb, dass nur geradeaus fährt. Durch den Schub der Rakete kann das Auto sowohl nach links als auch nach rechts beschleunigen. Vom Startpunkt x_0 aus soll ein Ziel x_f möglichst genau in gegebener Zeit t_f erreicht werden.

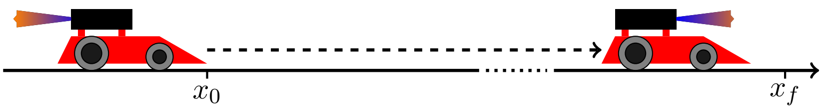

Abbildung 1.2.: Düsenauto

Starten wir am Punkt $(x_1(0), x_2(0) = (a_1, a_2)$ und wollen den Punkt $(0, 0)$ so gut wie möglich erreichen und zusätzlich den Energieverbrauch minimal halten, dann ergibt sich folgendes Optimierungsproblem:

$$\begin{aligned}
\min_{x_1, x_2, u} \quad & x_1(t_f)^2 + x_2(t_f)^2 + \alpha \int_0^{t_f} u(t)^2 dt \\
\text{s.t.} \quad & \dot{x_1}(t) = x_2(t) \quad \dot{x_2}(t) = u(t), \\
& x_1(0) = a_1, \quad x_2(0) = a_2, \\
& b_l \leq u(t) \leq b_u \quad \text{für fast alle } t \in [0, t_f]\,.
\end{aligned}$$

Der Regularisierungsparameter α gewichtet hier zum einen die beiden Optimierungsziele, zum anderen glättet er die optimale Steuerung. Gelöst wird diese Optimierungsproblem numerisch, indem man die Steuerung stückweise konstant approximiert und die gewöhnlichen Differentialgleichungen der Nebenbedingungen mit dem Euler-Verfahren diskretisiert. Man erhält hierbei ein **linear-quadratisches Optimierungsproblem**.

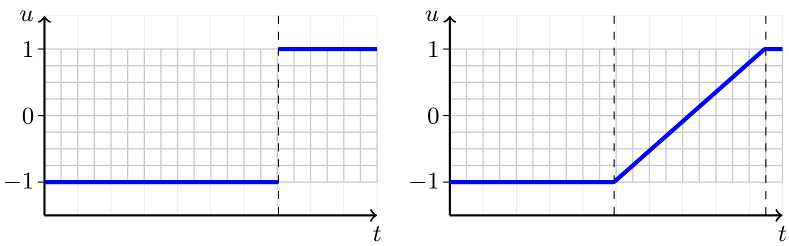

Abbildung 1.3.: Optimale Steuerung für $\alpha = 0$ (links) und $\alpha = 1$ (rechts)

1.2.3. Lineare Regression

Gegeben seien Messwerte $(\xi_i, \eta_i), i = 1, \dots, m$. Der funktionale Zusammenhang zwischen den $\xi-$ und den $\eta-$ Werten soll durch eine Gerade

$$\eta(\xi) = g(\xi; x_1, x_2) = x_1\xi + x_2$$

beschrieben werden. Aufgrund von Messfehlern liegen nicht alle Messwerte auf der Geraden. Ziel ist es den Parameter $x = (x_1, x_2)^\top \in \mathbb{R}^2$ so zu bestimmen, dass die zugehörige Gerade „optimal“ zu den Messwerten passt.

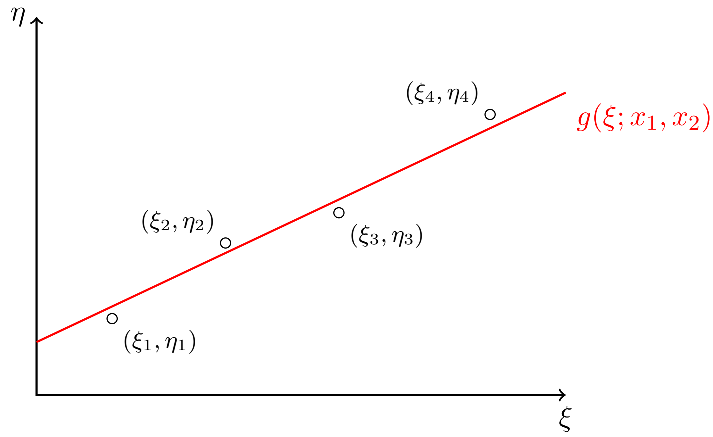

Abbildung 1.4.: Lineare Regression

Das übliche Kriterium für Optimalität ist die Minimierung der Summe der Fehlerquadrate in den Messpunkten. Dazu definieren wir die Zielfunktion $f : \mathbb{R}^2 \to \mathbb{R}$ durch

$$f(x) = \sum_{i=1}^{m} [g(\xi_i, x) - \eta_i]^2 = \sum_{i=1}^{m} [x_1\xi_i + x_2 - \eta_i]^2$$

Das resultierende Optimierungsproblem

$$\min_{x_1, x_2} \sum_{i=1}^{m} [x_1\xi_i + x_2 - \eta_i]^2$$

ist **unrestringiert** und die Zielfunktion ist **quadratisch**.

2. Konvexe Optimierungsprobleme

- Historisch standen lineare Probleme im Fokus der Optimierung.
- Ursprüngliche Unterscheidung in lineare und nichtlineare Probleme.
- Aber: Bestimmte nichtlineare Probleme können effizient gelöst werden.
- Daher unterscheidet man zwischen **konvexen und nichtkonvexen Problemen**.

Das auf Seite 5 vorgestellte Problem (P)

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{F} \end{aligned} \tag{P}$$

ist genau dann ein **konvexes Optimierungsproblem**, wenn die zulässige Menge $\mathcal{F}$ konvex ist und die Zielfunktion f konvex auf $\mathcal{F}$ ist.

Probleme mit Gleichungs- und Ungleichungsrestriktionen

Typischerweise ist die zulässige Menge durch Gleichungen und Ungleichungen definiert. Das Problem

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & h_i(x) = 0\,, \quad i = 1, \ldots, m\,, \\ & g_j(x) \leq 0\,, \quad j = 1, \ldots, p\,. \end{aligned} \tag{QP}$$

ist konvex, wenn f und $g_j, j = 1, \ldots, p$ konvexe Funktionen sind, und die Funktionen $h_i, i = 1, \ldots, m$ affin-linear sind, d.h.,

$$h_i(x) = a_i^\top x + b_i, \qquad i = 1, \ldots, m,$$

mit $a_i \in \mathbb{R}^n$ und $b_i \in \mathbb{R}$.

2.1. Konvexe Mengen

Definition 2.1. (konvexe Menge)
Eine Menge $C \subseteq \mathbb{R}^n$ heißt konvex, falls für beliebige $x, y \in C$ gilt

$$\{(1-t)x + ty \mid 0 \leq t \leq 1\} \subseteq C$$

d.h., für die Verbindungsstrecke $[x, y] = \{(1-t)x + ty \mid 0 \leq t \leq 1\}$ gilt $[x, y] \subseteq C$. ⋄

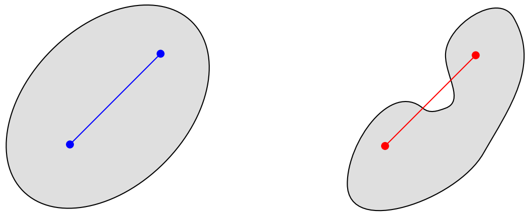

Abbildung 2.1.: links: konvexe Menge, rechts: keine konvexe Menge

2.1.1. Beispiele

a) Die konvexen Teilmengen des $\mathbb{R}$ sind, neben $\mathbb{R}$ selbst, die Intervalle. Beispielsweise $[a, b]$ mit $a, b \in \mathbb{R}, a \leq b$, oder $]a, \infty[$ mit $a \in \mathbb{R}$.

Beweis. $C = [a, b]$ mit $a, b \in \mathbb{R},\ a \leq b \Rightarrow C$ ist konvex.
Seien x, y beliebige Punkte aus C und $t \in [0, 1]$. Dann gilt einerseits die Abschätzung nach unten

$$(1-t)x + ty \overset{a \leq x,y}{\geq} (1-t)a + ta = a$$

und andererseits die Abschätzung nach oben durch

$$(1-t)x + ty \overset{x,y \leq b}{\leq} (1-t)b - tb = b$$

Das heißt zu zwei beliebigen Punkten x, y aus dem Intervall $[a, b]$ liegt das davon erzeugte Teilintervall $[x, y] \subseteq [a, b]$. □

b) Für $x \in \mathbb{R}^n$ und $r > 0$ bezeichnen wir mit

$$B(x, r) = \{y \in \mathbb{R}^n \mid \|y - x\| < r\}$$

die offene Kugel und mit

$$\bar{B}(x, r) = \{y \in \mathbb{R}^n \mid \|y - x\| \leq r\}$$

die abgeschlossene Kugel mit Radius r um den Punkt x. Beide Mengen sind konvex.[1]

c) Hyperebenen

$$\mathcal{H}(a, b) = \{x \in \mathbb{R}^n \mid a^\mathsf{T} x = b\}$$

mit $a \in \mathbb{R}^n, b \in \mathbb{R}$ und die durch sie definierten Halbräume $\{x \in \mathbb{R} \mid a^\mathsf{T} x \leq b\}$ sind konvexe Mengen.

d) Polyeder

$$P(A, b) = \{x \in \mathbb{R}^n \mid Ax \leq b\}$$

mit $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$ sind konvex.

[1] siehe Übungsblatt 1

Beweis. Seien $x_1, x_2 \in P(A,b)$ und $t \in [0,1] \Rightarrow Ax_1 \leq b \wedge Ax_2 \leq b$.
Dann gilt

$$\begin{aligned} A\left[(1-t)x_1 + tx_2\right] &= A(1-t)x_1 + A(tx_2) \\ &= (1-t)\underbrace{Ax_1}_{\geq b} + t\underbrace{Ax_2}_{\leq b} \\ &\leq (1-t)b + tb = b \\ &\Leftrightarrow (1-t)x_1 + tx_2 \in P(A,b) \end{aligned}$$

□

Bemerkung 2.2. Ein typischer Fall einer zulässigen Menge beschrieben durch Gleichungen und Ungleichungen, ist gegeben durch

$$\bar{P}(A,b,G,r) = \{x \in \mathbb{R}^n \mid Ax \leq b\ ,\ Gx = r\}$$

Wie kann man zeigen, dass diese Menge konvex ist? Es gilt:

$$Gx = r \Leftrightarrow (Gx \leq r) \wedge (\underbrace{Gx \geq r}_{-Gx \leq -r})$$

Damit ist

$$\Rightarrow \begin{bmatrix} & A \\ & G \\ - & G \end{bmatrix} x \leq \begin{bmatrix} & b \\ & r \\ - & r \end{bmatrix}$$

ein Spezialfall von oben. ◇

2.1.2. Schnitt, Skalierung und Verschiebung

Lemma 2.3. *Ist $(C_j)_{j \in J}$ eine Familie konvexer Mengen mit einer beliebigen Indexmenge J, dann ist auch $C = \bigcap\limits_{j \in J} C_j$ konvex.* ◇

Beweis. Folgt unmittelbar aus der Definition. □

Lemma 2.4. *Ist $C \subseteq \mathbb{R}^n$ konvex, dann ist auch*

$$aC + b = \{ax + b \mid x \in C\}$$

konvex für alle $a \in \mathbb{R}$ und $b \in \mathbb{R}^n$. ◇

Definition 2.5. (Konvexkombination)
Sind $x^{(1)}, \ldots, x^{(k)} \in \mathbb{R}^n$ und $\alpha_1, \ldots \alpha_k \in \mathbb{R}$, dann heißt ein Vektor

$$\sum_{i=1}^{k} \alpha_i x^{(i)} \text{ mit } \sum_{i=1}^{k} \alpha_i = 1 \text{ und } \alpha_i \geq 0, i = 1, \ldots, k,$$

Konvexkombination der Vektoren $x^{(1)}, \ldots, x^{(k)}$. ◇

Satz 2.6. *Eine Menge $C \subset \mathbb{R}^n$ ist genau dann konvex, wenn sie alle Konvexkombinationen von Punkten in C enthält.* ◇

Beweis. siehe Übungsblatt 1 □

Definition 2.7. (Konvexe Hülle)
Für $A \subset \mathbb{R}^n$ ist die konvexe Hülle von A, bezeichnet mit $\operatorname{co} A$, die kleinste konvexe Menge, die A umfasst, d.h.

$$\operatorname{co} A = \bigcap \{C \subseteq \mathbb{R}^n \mid C \text{ konvex}, C \supseteq A\}$$

◇

Lemma 2.8. *Für $A \subset \mathbb{R}^n$ ist*

$$\operatorname{co} A = \{x \in \mathbb{R}^n \mid x \textit{ ist Konvexkombination von Punkten in } A\}.$$

◇

Beweis.
Sei $B = \{x \in \mathbb{R}^n \mid x \text{ ist Konvexkombination von Punkten in } A\} \underbrace{\Rightarrow}_{\text{z.z.}} \operatorname{co} A = B$

B konvex und $A \subseteq B \Rightarrow \operatorname{co} A \subseteq \operatorname{co} B = B$.
Umgekehrt gilt $\forall C$ konvex mit $C \supseteq A \Rightarrow C \supseteq B \underbrace{\Rightarrow}_{C = \operatorname{co} A} \operatorname{co} A \supseteq B$ □

Definition 2.9. (Kegel)
Eine nichtleere Menge $K \subseteq \mathbb{R}^n$ heißt **Kegel**, wenn mit $x \in K$ auch $tx \in K$ für alle $t > 0$ gilt, d.h., wenn mit $x \in K$ auch der offene Halbstrahl $\{tx \mid t > 0\} \subseteq K$ ist. ◇

Lemma 2.10. *Ein Kegel $K \subseteq \mathbb{R}^n$ ist genau dann konvex, wenn $K + K \subseteq K$ ist.* ◇

Beweis. s. Übungsblatt 1 □

Beispiele für Kegel

a) Der Kegel $\{x \in \mathbb{R}^n \mid x \geq 0_n\}$ ist konvex. Es gilt $K + K = K$.

b) Die Menge
$$\{x \in \mathbb{R}^n \mid x \geq 0_n\} \cup \{x \in \mathbb{R}^n \mid x \leq 0_n\}$$
ist ein Kegel, aber nicht konvex, denn es gilt $K + K = \mathbb{R}^n \not\subseteq K$.

c) Der Kegel der positiv semidefiniten Matrizen
$$\mathcal{S}_+^n = \{X \in \mathcal{S} \mid X \succcurlyeq 0\}$$
ist konvex, wobei $\mathcal{S}^n$ die Menge der symmetrischen $n \times n$-Matrizen ist.

Wichtige Kegel

Definition 2.11. Ist $S \subseteq \mathbb{R}^n$ und $x \in S$, dann heißt die Menge

$$K(S,x) = \{t(s-x) \mid s \in S,\ t > 0\}$$

der von $S - x$ **erzeugte Kegel** oder **konische Hülle** von $S - x$. $\diamond$

Die Menge $K(S,x)$ besteht aus Halbstrahlen, die durch die Vektoren $S - x$ erzeugt werden, s.d. die Menge nach Definition ein Kegel ist. Wegen $x \in S$ ist immer auch $0_n \in K(S,x)$ (für $s = x$) und es gilt $K(S,x) = K(s-x, 0_n)$.

Lemma 2.12. *Sei $C \subseteq \mathbb{R}^n$ konvex und $x \in C$. Dann ist $K(C,x)$ konvex.* $\diamond$

Beweis. cf. undergraduate convexity $\square$

Definition 2.13. Ist $C \subseteq \mathbb{R}^n$ konvex und $x \in C$. Dann heißt $s \in \mathbb{R}^n$ **Normalenrichtung** von C in x, wenn

$$\langle s, y - x\rangle \leq 0 \qquad \forall y \in C$$

gilt. Die Menge

$$N(C,x) = \{s \in R^n \mid s \text{ ist Normalenrichtung von } C \text{ in } x\}$$

heißt **Normalenkegel** von C in x. $\diamond$

Bemerkung 2.14. Der Normalenkegel ist konvex und abgeschlossen. $\diamond$

Ein Beispiel für einen Normalenkegel lässt sich wie folgt konstruieren. Seien C ein Unterraum und $x \in C$ beliebig. Dann ist die konische Hülle $K(C,x) = C$ und der Normalenkegel $N(C,x) = C^{\perp}$, also das orthogonale Komplement von C. Der Beweis der Aussage ist eine gute Übung.

2.2. Konvexe Funktionen

Definition 2.15. (konvexe Funktion)
Ist $D \subseteq \mathbb{R}^n$ und ist $\emptyset \neq C \subseteq D$ konvex, dann heißt die Funktion $f\colon D \to \mathbb{R}$ konvex auf C, wenn

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$

für alle $x, y \in C$ und alle $t \in [0,1]$ gilt. $\diamond$

Definition 2.16 (Epigraph)**.** Der *Epigraph* einer Funktion $f : \mathbb{R}^n \to \mathbb{R}$ ist definiert als

$$\{(x,r)^T \mid x \in D, r \geq f(x)\} \subseteq \mathbb{R}^{n+1} =: \text{epi}\, f \qquad \diamond$$

Bemerkung 2.17 (Epigraphcharakterisierung von Funktionen). Die obige Definition erlaubt uns nun eine konvexe Funktion als eine Funktion aufzufassen, deren Epigraph eine konvexe Menge ist. In der Tat lassen sich die meisten Klassen von Funktionen, die in der Optimierung eine Rolle spielen, durch äquivalente Charakterisierungen des Epigraphen in Mengensprache fassen. $\diamond$

Definition 2.18 (strikte und gleichmäßige Konvexität)**.** Gilt für eine Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$

$$\forall t \in (0,1) \forall x, y \in D : f((1-t)x + ty) < (1-t)f(x) + tf(y)$$

so sprechen wir von einer *stikt*, *streng* oder *stark* konvexen Funktion. Gilt außerdem auch

$$\forall t \in (0,1) \forall x, y \in D : f((1-t)x + ty) + \frac{\lambda}{2}\|x\|_2^2 \leq (1-t)f(x) + tf(y)$$

so sprechen wir auch von einer *gleichmäßig* konvexen Funktion zum Parameter λ. ◇

Bemerkung 2.19. Wenn die beiden Konzepte verbal formulieren wollen, so bedeutet *strikt* konvex, dass eine Funktion superlinear gekrümmt ist. *Gleichmäßige* Konvexität bedeutet dann, dass die Krümmung von f mindestens quadratisch ist. In der Tat gilt

glm. Konvexität ⇒ strikte Konvexität ⇒ Konvexität

während keine der Umkehrungen gilt. ◇

Beispiele konvexer Funktionen

a) $f(x) = x$ ist konvex auf $\mathbb{R}$, ebenso auch $-f(x)$ und f damit ebenfalls konkav.

b) $f(x) = x^2$ ist sogar strikt konvex auf $\mathbb{R}$

c) $f(x) = x^3$ ist nicht konvex auf $\mathbb{R}$ aber gleichmäßig konvex auf $C = [0, \infty)$

d) $f(x) = \exp(\alpha x)$ ist für alle α konvex auf $\mathbb{R}$

e) $f(x) = -\log(x)$ ist strikt konvex auf $(0, \infty)$

f) Eine beliebige Norm $\|\cdot\|$ auf $\mathbb{R}^n$ ist konvex, aber nicht strikt konvex (Übungsblatt 1)

g) Die max-Funktion $f(x) = \max\{x_1, \ldots, x_n\}$ ist konvex auf $\mathbb{R}^n$

2.2.1. Charakterisierungen konvexer Funktionen

Wir betrachten nun ausreichend schöne Funktionen, d. h., Funktionen für die die Voraussetzungen der folgenden Theoreme gegeben sind. Dazu zählen insbesondere die differenzierbaren und die stetig differenzierbaren Funktionen.

Satz 2.20 (Charakterisierung erster Ordnung)**.** *Seien $D \subseteq \mathbb{R}^n$ offen, $\emptyset \neq \mathcal{F} \subseteq D$ konvex und $f : D \to \mathbb{R}$ differenzierbar auf D. Dann ist f auf $\mathcal{F}$ konvex genau dann, wenn $\forall x, y \in \mathcal{F}$ gilt:*

$$f(y) \geq f(x) + \nabla f(x)^\mathsf{T}(y - x) \tag{2.1}$$

◇

Satz 2.21 (Charakterisierung erster Ordnung)**.** *Seien $D \subseteq \mathbb{R}^n$ offen, $\emptyset \neq \mathcal{F} \subseteq D$ konvex und $f : D \to \mathbb{R}$ differenzierbar auf D. Dann ist f auf $\mathcal{F}$ strikt konvex genau dann, wenn $\forall x, y \in \mathcal{F}$ gilt:*

$$f(y) > f(x) + \nabla f(x)^\mathsf{T}(y - x) \tag{2.2}$$

Gilt außerdem für $\mu \in (0, \infty)$

$$f(y) - f(x) > \nabla f(x)^\mathsf{T}(y - x) + \mu\|y - x\|^2 \tag{2.3}$$

so heißt f gleichmäßig konvex. ◇

Satz 2.22 (Charakterisierung zweiter Ordnung)**.** *Seien $D \subseteq \mathbb{R}^n$ offen, $\emptyset \neq \mathcal{F} \subseteq D$ konvex und $f : D \to \mathbb{R}$ zweimal stetig differenzierbar auf D. Wenn $\forall x, y \in \mathcal{F}$ gilt:*

$$\nabla^2 f(x) \succcurlyeq 0 \tag{2.4}$$

so ist f konvex auf $\mathcal{F}$ und falls $\mathcal{F}$ offen ist, so gilt auch die Umkehrung. ◇

Satz 2.23 (Charakterisierung zweiter Ordnung)**.** *Seien $D \subseteq \mathbb{R}^n$ offen, $\emptyset \neq \mathcal{F} \subseteq D$ konvex und $f : D \to \mathbb{R}$ zweimal stetig differenzierbar auf D. Wenn $\forall x \in \mathcal{F}$ gilt:*

$$\nabla^2 f(x) \succ 0 \tag{2.5}$$

also $\forall x \in \mathcal{F}$ und $0 \neq d \in \mathbb{R}^n$

$$d^\mathsf{T}\nabla^2 f(x)d > 0 \tag{2.6}$$

so ist f stikt konvex auf $\mathcal{F}$. ◇

Bemerkung 2.24. Die Umkehrung von 2.23 gilt im Allgemeinen nicht. ◇

Satz 2.25 (Charakterisierung zweiter Ordnung)**.** *Seien $D \subseteq \mathbb{R}^n$ offen, $\emptyset \neq \mathcal{F} \subseteq D$ konvex und $f : D \to \mathbb{R}$ zweimal stetig differenzierbar auf D. Ist $\nabla^2 f(x)$ gleichmäßig positiv definit, also wenn $\forall x \in \mathcal{F}$, $d \in \mathbb{R}^n$ und $\beta \in (0, \infty)$*

$$d^\mathsf{T}\nabla^2 f(x)d > \beta\|d\|^2 \tag{2.7}$$

gilt, so ist f gleichmäßig konvex auf $\mathcal{F}$. Ist $\mathcal{F}$ offen, so gilt auch die Umkehrung. ◇

Rechenregeln für konvexer Funktionen

a) Seien $f_i, i \in [n]$ konvex und $\alpha_i \in \mathbb{R}$. Dann ist auch $\sum_{i=1}^n \alpha_i f_i$ konvex.

b) Seien f konvex, $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$. Dann ist auch $g(x) := f(Ax + b)$ konvex.

c) Seien J eine beliebige auch unendliche Indexmenge und f_j konvex für alle $j \in J$. Dann ist auch $\max_{j \in J} f_j(x)$ konvex.

d) Seien $g(x, y)$ konvex in x und y und die Menge C konvex. Dann ist auch $f(x) := \min_{y \in C} g(x, y)$ konvex.

3. Optimierungsverfahren für allgemeine Probleme

Allgemeine, **unrestringierte Probleme** mit **nichtlinearer**, **differenzierbarer** Zielfunktion vom Typ

$$\min_{x \in \mathbb{R}^n} \quad f(x) \tag{3.1}$$

lassen sich mit den im vorigen Kapitel behandelten Verfahren nicht lösen. Der negative Gradient als Richtung des steilsten Abstiegs ist immer noch als Abstiegsrichtung geeignet, allerdings kann keine exakte **Schrittweite** mehr bestimmt werden.

Sei $x^{(k)}$ ein Iterationspunkt mit einer Abstiegsrichtung $d^{(k)}$. Bei hinreichend kleiner Schrittweite σ_k gilt:

$$f(x^{(k+1)}) = f(x^{(k)} + \sigma_k d^{(k)}) < f(x^{(k)}) \tag{3.2}$$

, die Folge $f(x^{(k)})$ ist streng monoton fallend. Das muss jedoch nicht bedeuten, dass das Problem konvergiert.

Beispiel für nicht konvergierende Probleme

$f(x) = x^2$, $x^{(0)} = 1$

Grafik einfügen

$d^{(0)} = -1$

Abstiegsrichtung: $\nabla f(x^{(k)})^T d^{(k)} = -2x^{(k)} < 0 \forall x^{(k)} > 0$

Betrachte Schrittweite $\sigma_k = 1/2^{k+2} \forall k \geq 0$

$$x^{(k-1)} = x^k + \sigma_k d^{(k)} = x^{(k)} - \sigma_k = x^{(k)} - (1/2)^{(k+2)} = x^{(0)} - \sum_{i=1}^{k} (1/2)^{(i+2)} = 1 - \sum_{i=1}^{k} (1/2)^{(i+2)} = 1/2 + (1/2)^{(k+2)} \underbrace{\Rightarrow}_{k \Rightarrow \infty} \tag{3.3}$$

Die Folge $x^{(k)}$ konvergiert nicht gegen $x^* = 0$.

Mögliche Schrittweitenstrategien

- Konstante Schrittweite $\sigma_k = \sigma > 0$
- Kleiner werdende Schrittweiten, z.B. $\sigma_k = 1/k$
- Exakte Schrittweiten, z.B. $\sigma_k = arg \min \sigma \geq 0 f(x^{(k)} + \sigma d^{(k)})$
- Armijo-Verfahren

3.1. Das Schrittweitenverfahren von Armijo

Seien x und eine Abstiegsrichtung d von f in x gegeben. Weiter sei $c_1 \in]0,1[$ eine von x und d unabhängige Konstante. Zur Berechnung einer effizienten Schrittweite σ soll die Abstiegsbedingung $f(x + \sigma d) \leq f(x) + c_1 \sigma \nabla f(x)^T d$ erfüllt werden. Damit die Schrittweite nicht zu klein wird, fordert man zusätzlich mit einer von x und d unabhängigen Konstante $c_2 > 0$, dass

$$\sigma \geq -c_2(\nabla f(x)^T d)/||d||^2 \tag{3.4}$$

Abbildung einfügen
Gegeben seien von x und d unabhängige Konstanten: $\delta \in]0,1[$, $\gamma > 0$ und $0 < \beta_1 \leq \beta_2 < 1$

1. Wähle eine Startschrittweite σ_0, für die mit $c_2 = \gamma$ gilt:

$$\sigma \geq -\gamma(\nabla f(x)^T d)/||d||^2 \tag{3.5}$$

   Setze $j = 0$.

2. Ist die Abstiegsbedingung $f(x + \sigma d) \leq f(x) + \delta \sigma_j \nabla f(x)^T d$ erfüllt, dann setze $\sigma_A = \sigma_j$ und stoppe das Verfahren.

3. Wähle $\sigma_j + 1 \in [\beta_1 \sigma_j, \beta_2 \sigma_j]$.

4. Setze $j = j + 1$ und gehe zu 2.

Praktisch haben sich folgende Werte bewährt:

- δ sollte klein sein; Größenordnung: $\delta = 0.01$
- γ sollte so gewählt werden, dass die Schrittweite 1 und die exakte Schrittweite nicht ausgeschlossen werden. $\gamma = 10^{-4}$
- $\sigma_0 = 1$ oder als Approximation der exakten Schrittweite:

$$\sigma_0 = -(\nabla f(x)^T d)/(2(f(x + d) - f(x) - \nabla f(x)^T d)) \tag{3.6}$$

- Zur Berechnung von $\sigma_j, j \geq 1$ kann man $\beta_1 = \beta_2 = \beta$ wählen:

$$\sigma_j = \beta^j \sigma_0 \tag{3.7}$$

  $j = 1, 2, ...$

Oft wählt man $\beta = 1/2$

3.1.1. Konvergenz des Verfahrens

Die theoretische Version des Armijo-Verfahrens konvergiert nach endlich vielen Schritten, falls die Standard-Voraussetzung erfüllt ist und die Ableitung der Zielfunktion Lipschitz-stetig ist, d. h., es gibt ein $L > 0$ mit

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||\forall x, y \in N_0 \tag{3.8}$$

Bei einer praktischen Implementierung mit der diskutierten Parameterwahl ist jedoch endliche Konvergenz nicht sichergestellt. Daher sollte eine maximale Iterationszahl vorgegeben werden. Sollte keine Konvergenz eintreten, kann das Verfahren mit anderen Parametern neugestartet werden.

Beispiel zur Lipschitz-Stetigkeit $f(x_1, x_2) = (x_1)^2 + (x_2)^2$ $\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$

$\nabla f(x_1, x_2)$ ist ∞ mal stetig differenzierbar

$$||\nabla f(x) - \nabla f(y)|| = ||2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 2 \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} || = ||2 \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \end{bmatrix} || \leq 2||x - y|| \tag{3.9}$$

∇f ist lipschitz-stetig mit L=2

Die Lipschitz-Stetigkeit kann als ein Verhältnis zwischen Funktionswerten und Argumenten interpretiert werden.

3.2. Das Gradientenverfahren

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

- Wähle einen Startpunkt $x^{(0)} \in R^n$ und setze k = 0.
- Ist $\nabla f(x^{(k)}) = 0_n$, dann stoppe das Verfahren.
- Berechne eine effiziente Schrittweite σ_k (bspw. Armijo) und setze $x^{(k+1)} = x(k) - \sigma_k \nabla f(x^{(k)})$.
- Setze $k = k + 1$ und gehe zu 2.

Der negative Gradient ist die eindeutig bestimmte Lösung des quadratischen Problems

$$\min_{d \in \mathbb{R}^n} \underbrace{f(x^{(k)}) + \nabla f(x^{(k)})^T d}_{Taylor-Approximation 1. Ordnung} + \underbrace{\underbrace{1/2 d^T d}_{1/(2\sigma)||x^{(k+1)} - x^{(k)}||^2}}_{Abstand x^{(k+1)} zu x^{(k)}} \tag{3.10}$$

$x^{(k+1)} = x^{(k)} + \sigma d$
$d = (x^{(k+1)} - x^{(k)})/\sigma$
für: $Q = I_n$ $q = \nabla f(x^{(k)})$ $Qd + q = 0$ $Id + \nabla f(x^{(k)}) = 0$ ¡=¿ $d = -\nabla f(x^{(k)})$

3.2.1. Funktionenklassen

$\mathcal{F}_L^{k,l}(R^n)$
Menge aller konvexen Funktionen $\mathcal{F} : R^n \Rightarrow R$, die k mal stetig differenzierbar sind und deren l-te ABleitung lipschitz-stetig mit Konstante L ist, d.h.

$$||f^{(l)}(x) - f^{(l)}(y)|| \leq L||x - y||\forall x, y \in R^n \tag{3.11}$$

- Offensichtlich ist $k \geq l$
- $\mathcal{F}_L^{k_1,l} \leq \mathcal{F}_L^{k_2,l}, k_1 \geq k_2$
- $f_1 \in \mathcal{F}_{L_1}^{k,l}, f_2 \in \mathcal{F}_{L_2}^{k,l}, \alpha, \beta \geq 0$

$\Rightarrow \alpha f_1 + \beta f_2 \in \mathcal{F}_{\alpha L_1 + \beta L_2}^{k,l}$
$\mathcal{F}^k \Rightarrow$ Menge aller konvexen Funktionen, die k mal stetig differenzierbar sind.
Für Gradientenverfahren relevant: $\mathcal{F}_L^{1,1}$ Eigenschaft von $\mathcal{F}^1$: Für $f \in \mathcal{F}^1$ gilt:

$$f(y) \geq f(x + \nabla f(x)^T(y - x) \tag{3.12}$$

Eigenschaften von $\mathcal{F}_L^{1,1}$: Für $f \in \mathcal{F}_L^{1,1}$ gilt:

- $f(y) \leq f(x) + \nabla f(x)^T(y - x) + L/2||x - y||^2$
- $f(x) + \nabla f(x)^T(y - x) + 1/(2L)||\nabla f(x) - \nabla f(y)||^2 \leq f(y)$
- $1/L||\nabla f(x) - \nabla f(y)||^2 \leq (\nabla f(x) - \nabla f(y))^T(x - y)$
- $(\nabla f(x) - \nabla f(y))^T(x - y) \leq L||x - y||^2$
- $\alpha f(x) + (1 - \alpha)f(y) \geq f(\alpha x + (1 - \alpha)y) + (\alpha(1 - \alpha))/(2L)||\nabla f(x) - \nabla f(y)||^2$
- $\alpha f(x) + (1 - \alpha)f(y) \geq f(\alpha x + (1 - \alpha)y) + \alpha(1 - \alpha)L/2||x - y||^2$

Definition 3.1. Eine stetig differenzierbare Funktion f heißt gleichmäßig konvex mit Konvexitätsparameter $\mu > 0 (f \in S_\mu^1)$, wenn $f(y) \geq \nabla f(x)^T(y - x) + (1/2)\mu||y - x||^2$ ◇

Wir definieren den Raum $S_{\mu,L}^{k,l}$ analog zu$\mathcal{F}_L^{k,l}$
Eigenschaften: $f_1 \in S_{\mu_1}^1, f_2 \in S_{\mu_2}^1, \alpha, \beta \geq 0$
$\Rightarrow \alpha f_1 + \beta f_2 \in S_{\alpha mu_1 + \beta mu_2}^1$

Lemma 3.2. *Eine zweimal stetig differenzierbare Funktion f gehört zu* S_μ^2 ◇

$\Leftrightarrow f^n(x) \geq \mu * I_n \forall x \in R^n$
Für $f \in S_{\mu,L}^{2,1}$ gilt:
$\mu I_n \leq f''(x) \leq LI_n$

Definition 3.3. $Q = L/\mu$ ist die Kondition der Funktion f. ◇

Bei kleinen Q führen kleine Änderungen im Problem zu kleinen Änderungen im Funktionswert.

3.2.2. Konvergenz des Gradientenverfahrens

Der Einfachheit halber betrachten wir nur die konstante Schrittweitenstrategie $\sigma_k = \sigma$.

Theorem Sei $f : R^n \Rightarrow R$ konvex und stetig differenzierbar und die Ableitung ∇f Lipschitz-stetig mit Lipschitzkonstante $L > 0$. Dann gilt für das Gradientenverfahren mit konstanter Schrittweite $0 < \sigma \leq 1/L$

$$f(x^{(k)}) - f(x^*) \leq (||x^{(0)} - x^*||^2)/2\sigma k \tag{3.13}$$

aufgehört bei 08-15

A. Mathematische Grundlagen

Definition A.1. Eine symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ heißt *positiv definit*, wenn

$$x^\mathsf{T} Ax > 0 \quad \forall x \in \mathbb{R}^n\,,\; x \neq 0_n$$

gilt. Die Matrix A heißt *positiv semidefinit*, wenn

$$x^\mathsf{T} Ax \geq 0 \quad \forall x \in \mathbb{R}^n$$

gilt. ◇

Beispiel A.2. $A = I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$\begin{aligned} x^\mathsf{T} Ax &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^\mathsf{T} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^\mathsf{T} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= x_1^2 + x_2^2 > 0, \end{aligned}$$

wobei $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. ◇

Definition A.3 (Differenzierbarkeit)**.** Seien $f\colon D \to \mathbb{R}$ und $D \subseteq \mathbb{R}^n$ offen und existiere der Gradient

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}\right)^\mathsf{T}$$

für alle $x \in D$. Dann heißt f *differenzierbar* auf D. ◇

Definition A.4 (zweimalige Differenzierbarkeit)**.** Seien $f\colon D \to \mathbb{R}$ und $D \subseteq \mathbb{R}^n$ offen und existiere die Hessematrix $\nabla^2 f(x) \subseteq \mathcal{S}^n$ mit

$$\nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{i,j\in[n]}$$

für alle $x \in D$. Dann heißt f *zweimal differenzierbar* auf D. ◇

B. Optimisational English

. . . **program** Optimierungsproblem. Der Begriff *programming* für *mathematische Optimierung* rührt daher, dass Optimierungsmodule typischerweise die ersten Module sind, in denen Mathematikstudenten gezwungen sind, zu programmieren. Dies ist natürlich nicht richtig. Tatsächlich stammt der Begriff von der englischen Übersetzung von Einsatzplan, bzw. Einsatzplanung. Die ersten spezifischen Optimierungsmethoden wurden im Militär entwickelt, weswegen sich insbesondere im englischen Sprachraum der Begriff *programming* eingebürgert hat. Ein *program* ist daher ein *Optimierungsproblem*. In jüngerer Zeit wird allerdings *optimisation*, respektive *optimization* immer stärker verwendet.

convex konvex

set Menge

function Funktion, Abbildung

mapping, map Abbildung, Funktion

cone Kegel. *conic* – kegelförmig.

hull Hülle, Hüllenoperator.

continuous stetig. *continuity* – Stetigkeit.