I. Unrestringierte Probleme

Einführung

Lösbarkeit

Definition 1.2.3 (Lösbarkeit)**.**
Das Minimierungsproblem P heißt ***lösbar****, falls ein $\overline{x} \in M$ existiert mit*

$$\inf_{x \in M} f(x) = f(\overline{x})$$

Satz 1.2.5. *Das Minimierungsproblem P ist genau dann lösbar, wenn es einen globalen Minimalpunkt besitzt.*

Bemerkung. *Es können drei Fälle der Unlösbarkeit auftreten:*

- $\inf_{x \in M} f(x) = +\infty$
- $\inf_{x \in M} f(x) = -\infty$
- *Ein endliches Infimum wird nicht angenommen.*

Satz 1.2.6 (Satz von Weierstraß)**.**
Die Menge $M \subseteq \mathbb{R}^n$ sei nichtleer und kompakt, und die Funktion $f\colon M \to \mathbb{R}$ sei stetig. Dann besitzt f auf M (mindestens) einen globalen Minimalpunkt und einen globalen Maximalpunkt.

Definition 1.2.8 (Untere Niveaumenge)**.** *Für $X \subseteq \mathbb{R}^n, f\colon X \to \mathbb{R}$ und $\alpha \in \mathbb{R}$ heißt*

$$\operatorname{lev}^{\alpha}_{\leq}(f, X) = \big\{x \in X \mid f(x) \leq \alpha\big\}$$

untere Niveaumenge von f auf X zum Niveau α*. Im Fall $X = \mathbb{R}^n$ schreiben wir auch kurz*

$$f^{\alpha}_{\leq} \coloneqq \operatorname{lev}^{\alpha}_{\leq}(f, \mathbb{R}^n) = \big\{x \in \mathbb{R}^n \mid f(x) \leq \alpha\big\}$$

Übung 1.2.10. *Für eine abgeschlossene Menge $X \subseteq \mathbb{R}^n$ sei die Funktion $f\colon X \to \mathbb{R}$. Dann ist die Menge $\operatorname{lev}^{\alpha}_{\leq}(f, X)$ für alle $\alpha \in \mathbb{R}$ abgeschlossen.*

Übung 1.2.11. *Für eine abgeschlossene Menge $X \subseteq \mathbb{R}^n$ und endliche Indexmengen I und J seien die Funktion $g_i\colon X \to \mathbb{R}, i \in I$, und $h_j\colon X \to \mathbb{R}, j \in J$, stetig. Dann ist die Menge*

$$M = \big\{x \in X \mid g_i(x) \leq 0, i \in I,\ h_j(x) = 0, j \in J\big\}$$

abgeschlossen.

Definition. *Die Menge der globalen Minimalpunkte lautet:*

$$S = \{\overline{x} \in M \quad | \; \forall x \in M : f(x) \geq f(\overline{x})\}$$

Lemma 1.2.12. *Für ein $\alpha \in \mathbb{R}$ sei $\operatorname{lev}^{\alpha}_{\leq}(f, M) \neq \emptyset$. Dann gilt*

$$S \subseteq \operatorname{lev}^{\alpha}_{\leq}(f, M).$$

Satz 1.2.13 (Verschärfter Satz von Weierstraß)**.** *Für eine (nicht notwendigerweise beschränkte oder abgeschlossene) Menge $M \subseteq \mathbb{R}^n$ sei $f\colon M \to \mathbb{R}$ stetig, und mit einem $\alpha \in \mathbb{R}$ sei $\operatorname{lev}^{\alpha}_{\leq}(f, M)$ nichtleer und kompakt. Dann besitzt f auf M (mindestens) einen globalen Minimalpunkt.*

Definition 1.2.21 (Koerzivität)**.** *Gegeben seien eine abgeschlossene Menge $X \subseteq \mathbb{R}^n$ und eine Funktion $f\colon \mathbb{R}$ fall für alle Folgen $(x^k) \subseteq X$ mit $\lim_k \|x^k\| = +\infty$ auch*

$$\lim_k f(x^k) = +\infty$$

*gilt, dann heißt f **koerziv** auf X.*

Übung 1.2.24. *Gegeben sei die quadratische Funktion $q(x) = \frac{1}{2}x^T Ax + b^T x$ mit einer symmetrischen (n, n)-Matrix A (d.h. es gilt $A = A^T$) und $b \in \mathbb{R}^n$. Die Funktion q ist genau dann koerziv auf $\mathbb{R}^n$, wenn A positiv definit ist (d.h. wenn $d^T Ad > 0$ für alle $d \in \mathbb{R}^n \setminus \{0\}$ gilt).*

Beispiel 1.2.25. *Auf kompakten Mengen X ist jede Funktion f trivialerweise koerziv.*

Lemma 1.2.26. *Die Funktion $f\colon X \to \mathbb{R}$ sei stetig und koerziv auf der (nicht notwendigerweise beschränkten) abgeschlossenen Menge $X \subseteq \mathbb{R}^n$. Dann ist die Menge $\operatorname{lev}^{\alpha}_{\leq}(f, X)$ für jedes Niveau $\alpha \in \mathbb{R}$ kompakt.*

Korollar 1.2.27. *Es sei M nichtleer und abgeschlossen, aber nicht notwendigerweise beschränkt. Ferner sei die Funktion $f\colon M \to \mathbb{R}$ stetig und koerziv auf M. Dann besitzt f auf M (mindestens) einen globalen Minimalpunkt.*

Rechenregeln und Umformungen

Übung 1.2.28 (Skalare Vielfache und Summen)**.** *Gegeben seien $M \subseteq \mathbb{R}^n$ und $f, g\colon M \to \mathbb{R}$. Dann gilt*

a) $\forall \alpha \geq 0,\ \beta \in \mathbb{R}\colon\ \min_{x \in M} (\alpha f(x) + \beta) = \alpha (\min_{x \in M} f(x)) + \beta$

b) $\forall \alpha < 0,\ \beta \in \mathbb{R}\colon\ \min_{x \in M} (\alpha f(x) + \beta) = \alpha (\max_{x \in M} f(x)) + \beta$

c) $\min_{x \in M} (f(x) + g(x)) \geq \min_{x \in M} f(x) + \min_{x \in M} g(x)$

Übung 1.2.29 (Separable Zielfunktion auf kartesischem Produkt)**.** *Es seien $X \subseteq \mathbb{R}^n$, $Y \subseteq \mathbb{R}^m$, $f\colon X \to \mathbb{R}$ und $g\colon Y \to \mathbb{R}$. Dann gilt*

$$\min_{(x,y) \in X \times Y} (f(x) + g(y)) = \min_{x \in X} f(x) + \min_{y \in Y} g(y)$$

Übung 1.2.30 (Vertauschung von Minima und Maxima)**.** *Es seien* $X \subseteq \mathbb{R}^n$*,* $Y \subseteq \mathbb{R}^m$*,* $M = X \times Y$ *und* $f\colon M \to \mathbb{R}$ *gegeben. Dann gilt:*

a) $\min_{(x,y)\in M} f(x,y) = \min_{x\in X} \min_{y\in Y} f(x,y) = \min_{y\in Y} \min_{x\in X} f(x,y)$

b) $\max_{(x,y)\in M} f(x,y) = \max_{x\in X} \max_{y\in Y} f(x,y) = \max_{y\in Y} \max_{x\in X} f(x,y)$

c) $\min_{x\in X} \max_{y\in Y} f(x,y) \geq \max_{y\in Y} \min_{x\in X} f(x,y)$

Übung 1.2.31 (Monotone Transformation)**.** *Zu* $M \subseteq \mathbb{R}^n$ *und einer Funktion* $f\colon M \to Y$ *mit* $Y \subseteq \mathbb{R}$ *sei* $\psi\colon Y \to \mathbb{R}$ *eine streng monoton wachsende Funktion. Dann gilt*

$$\min_{x\in M} \psi\left(f(x)\right) = \psi\left(\min_{x\in M} f(x)\right),$$

und die lokalen bzw. globalen Minimalpunkte stimmen überein.

Übung 1.2.32 (Epigraphumformulierung)**.** *Gegeben seien* $M \subseteq \mathbb{R}^n$ *und eine Funktion* $f\colon M \to \mathbb{R}$*. Dann sind die Probleme*

$$P\colon \min_{x\in\mathbb{R}^n} f(x) \text{ s.t. } x \in M \quad \text{und} \quad P_{epi}\colon \min_{(x,\alpha)\in\mathbb{R}^n\times\mathbb{R}} \alpha \text{ s.t. } f(x) \leq \alpha, x \in M$$

äquivalent, d.h. die Minimalwerte stimmen überein und Minimalpunkte entsprechen sich.

Definition 1.2.33 (Parallelprojektion)**.** *Es sei* $M \subseteq \mathbb{R}^n \times \mathbb{R}^m$*. Dann heißt*

$$\mathrm{pr}_x M = \left\{x \in \mathbb{R}^n \mid \exists y \in \mathbb{R}^m : (x,y) \in M\right\}$$

Parallelprojektion *von* M *(den „x-Raum“)* $\mathbb{R}^n$*.*

Übung 1.2.34 (Projektionsumformulierung)**.** *Gegeben seien* $M \subseteq \mathbb{R}^n \times \mathbb{R}^m$ *und eine Funktion* $f\colon \mathbb{R}^n \to \mathbb{R}$*, die nicht von den Variablen aus* $\mathbb{R}^m$ *abhängt. Dann sind die Probleme*

$$P\colon \min_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^m} f(x) \quad \text{s.t.} (x,y) \in M \quad \text{und} \quad P_{proj}\colon \min_{x\in\mathbb{R}^n} f(x) \text{ s.t. } x \in \mathrm{pr}_x M$$

äquivalent, d.h. die Minimalwerte stimmen überein und Minimalpunkte entsprechen sich.

Optimalitätsbedingungen

Abstiegsrichtung

Definition 2.1.35. *Es seien $f\colon \mathbb{R}^n \to \smallsetminus$ und $\overline{x} \in \mathbb{R}^n$. Ein Vektor $d \in \mathbb{R}^n$ heißt **Abstiegsrichtung** für f in $\overline{x}$, falls*

$$\exists \hat{f} > 0 \ \forall t \in (0, \hat{t})\colon f(\overline{x} + td) < f(\overline{x}).$$

Übung 2.1.36. *Für $f\colon \mathbb{R}^n \to \mathbb{R}$ sei $\overline{x}$ ein lokaler Minimalpunkt, dann kann keine Abstiegsrichtung für f in $\overline{x}$ existieren.*

Definition 2.1.37. *Gegeben seien $f\colon \mathbb{R}^n \to \mathbb{R}$, ein Punkt $\overline{x} \in \mathbb{R}^n$ und ein Richtungsvektor $d \in \mathbb{R}^n$. Die Funktion*

$$\varphi_d\colon \mathbb{R}^1 \to \mathbb{R}^1, \ t \mapsto f(\overline{x} + td)$$

*heißt **eindimensionale Einschränkung** von f auf die durch $\overline{x}$ in Richtung d verlaufende Gerade.*

Bemerkung. *Es gilt $\varphi_d(0) = f(\overline{x})$ für jede Richtung $d \in \mathbb{R}^n$. Daher ist d genau dann Abstiegsrichtung für f in $\overline{x}$, wenn*

$$\exists \hat{t} > 0 \ \forall t \in (0, \hat{t})\colon \varphi_d(t) < \varphi_d(0)$$

Optimalitätsbedingung erster Ordnung

Definition 2.1.38. *Eine Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ heißt an $\overline{x} \in \mathbb{R}^n$ in eine Richtung $d \in \mathbb{R}^n$ **einseitig richtungsdifferenzierbar**, wenn der Grenzwert*

$$f'(\overline{x}, d) \coloneqq \lim_{t \searrow 0} \frac{f(\overline{x} + td) - f(\overline{x})}{t}$$

*existiert. Der Wert $f'(\overline{x}, d)$ heißt dann **einseitige Richtungsableitung**. Die Funktion f heißt an $\overline{x}$ einseitig richtungsdifferenzierbar, wenn f an $\overline{x}$ in jede Richtung $d \in \mathbb{R}^n$ einseitig richtungsdifferenzierbar ist, und f heißt **einseitig richtungsdifferenzierbar**, wenn f an jedem $\overline{x} \in \mathbb{R}^n$ einseitig richtungsdifferenzierbar ist.*

Lemma 2.1.39. *Die Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ sei an $\overline{x} \in \mathbb{R}^n$ in Richtung $d \in \mathbb{R}^n$ einseitig richtungsdiffrenzierbar mit $f'(\overline{x}, d) < 0$. Dann ist d Abstiegsrichtung für f in $\overline{x}$.*

Lemma 2.1.40. *Die Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ sei an einem lokalen Minimalpunkt $\overline{x} \in \mathbb{R}^n$ einseitig richtungsdifferenzierbar. Dann gilt $f'(\overline{x}, d) \geq 0$ für jede Richtung $d \in \mathbb{R}^n l$*

Definition 2.1.41 (Abstiegsrichtung erster Ordnung)**.** *Für eine am Punkt $\overline{x} \in \mathbb{R}^n$ in Richtung $d \in \mathbb{R}^n$ einseitig richtungsdifferenzierbare Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ heißt d **Abstiegsrichtung erster Ordnung**, falls $f'(\overline{x}, d) < 0$ gilt.*

Definition 2.1.42 (Stationärer Punkt - unrestringierter Fall)**.** *Die Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ sei an $\overline{x} \in \mathbb{R}^n$ einseitig richtungsdifferenzierbar. Dann heißt $\overline{x}$ **stationärer Punkt** von f, falls $f'(\overline{x}, d) \geq 0$ für jede Richtung $d \in \mathbb{R}^n$ gilt.*

Satz 2.1.43 (Kettenregel). *Es seien $g\colon \mathbb{R}^n \to \mathbb{R}^m$ differenzierbar an $\overline{x} \in \mathbb{R}^n$ und $f\colon \mathbb{R}^m \to \mathbb{R}^k$ differenzierbar an $g(\overline{x}) \in \mathbb{R}^m$. Dann ist $f \circ g\colon \mathbb{R}^n \to \mathbb{R}^k$ differenzierbar an $\overline{x}$ mit*

$$D(f \circ g)(\overline{x}) = Df(g(\overline{x})) \cdot Dg(\overline{x}).$$

Lemma 2.1.44. *Die Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ sei am Punkt $\overline{x} \in \mathbb{R}^n$ differenzierbar, und für die Richtung $d \in \mathbb{R}^n$ gelte $\langle \nabla f(\overline{x}), d \rangle < 0$. Dann ist d Abstiegsrichtung für f in $\overline{x}$.*

Bemerkung 2.1.45. *Ein Vektor d ist genau dann eine Abstiegsrichtung erster Ordnung für f in $\overline{x}$, wenn d einen stumpfen Winkel mit dem Gradienten $\nabla f(\overline{x})$ bildet. Wobei das Skalarprodukt genau dann negativ, wenn sie einen stumpfen Winkel miteinander bilden, und analog ist das Skalarprodukt genau für einen spitzen Winkel bildende Vektoren positiv.*

Übung 2.1.46. *Gegeben seien $\overline{x} \in \mathbb{R}^n$, eine endliche Indexmenge K und an $\overline{x}$ differenzierbare Funktionen $f_k\colon \mathbb{R}^n \to \mathbb{R}$, $k \in K$. Die Funktion $f(x) := \max_{k \in K} f_k(x)$ ist an $\overline{x}$ einseitig richtungsdifferenzierbar und dass mit $K_*(\overline{x}) = \{k \in K \mid f_k(\overline{x}) = f(\overline{x})\}$*

$$f'(\overline{x}, d) = \max_{k \in K_x(\overline{x})} \langle \nabla f_k(\overline{x}), d \rangle$$

für jede Richtung $d \in \mathbb{R}^n$ gilt.

Satz 2.1.47 (Notwendige Optimalitätsbedingung erster Ordnung – Fermat'sche Regel). *Die Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ sei differenzierbar an einem lokalen Minimalpunkt $\overline{x} \in \mathbb{R}$. Dann gilt $\nabla f(\overline{x}) = 0$.*

Definition 2.1.48 (Kritischer Punkt). *Die Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ sei an $\overline{x} \in \mathbb{R}^n$ differenzierbar. Dann heißt $\overline{x}$ kritischer Punkt von f, wenn $\nabla f(\overline{x}) = 0$ gilt.*

Übung 2.1.49. *Die Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ sei differenzierbar an einem Punkt $\overline{x} \in \mathbb{R}^n$. Der Punkt $\overline{x}$ ist genau dann stationärer Punkt von f, wenn er kritischer Punkt von f ist.*

Definition 2.1.17 (Sattelpunkt). *Die Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ sei an $\overline{x} \in \mathbb{R}^n$ differenzierbar. Dann heißt $\overline{x}$ **Sattelpunkt** von f, falls $\overline{x}$ zwar kritischer Punkt von f, aber weder lokaler Minimal- noch lokaler Maximalpunkt ist.*

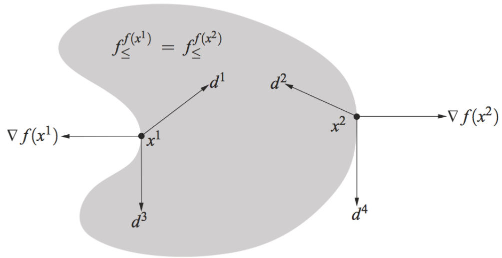
Geometrische Eigenschaften von Gradienten

Algorithmus 2.1: Konzeptioneller Algorithmus zur unrestringierten nichtlinearen Minimierung mit Informationen erster Ordnung

Input : Lösbares unrestringiertes differenzierbares Optimierungsproblem P
Output : Globaler Minimalpunkt $x^\star$ von f über $\mathbb{R}^n$

```
begin
    Bestimme alle kritischen Punkte von f, d. h. die Lösungsmenge K der Gleichung
    ∇f(x) = 0.
    Bestimme einen Minimalpunkt x* von f in K.
end
```

Abb. 2.2 Gradienten und Abstiegsrichtungen



Bemerkung. *Man kann zeigen, dass jeder Vektor $d \in \mathbb{R}^n$ mit $\langle \nabla f(\overline{x}), d \rangle > 0$ eine Anstiegsrichtung erster Ordnung ist. Da für einen nichtkritischen Punkt $\overline{x}$ die Gradientenrichtung $d = \nabla f(\overline{x})$ die strikte Ungleichung*

$$\langle \nabla f(\overline{x}), \nabla f(\overline{x}) \rangle = \{\nabla f(\overline{x})\|_2^2 > 0$$

erfüllt, ist $d = \nabla f(\overline{x})$ also eine Anstiegsrichtung erster Ordnung für f in $\overline{x}$, und man kann zeigen, dass $\nabla f(\overline{x})$ senkrecht auf dem Rand von $f_{\leq}^{f(\overline{x})}$ steht.

Optimalitätsbedingungen zweiter Ordnung

Bemerkung. *Für normierte Richtungen d liefert die Cauchy-Schwarz-Ungleichung*

$$-\|\nabla f(\overline{x})\|_2 = -\|\nabla f(\overline{x})\|_2 \cdot \|d\|_2 \leq \langle \nabla f(\overline{x}), d \rangle \leq \|\nabla f(\overline{x})\|_2 \cdot \|d\|_2 = \|\nabla f(x)\|_2$$

und die Unter- und Oberschranken werden genau für linear abhängige d und $\nabla f(\overline{x})$ angenommen. Wegen $\nabla f(\overline{x})$ wird die kleinst- und größtmögliche Steigung daher folgend realisiert

$$d_{min} = -\frac{\nabla f(\overline{x})}{\|\nabla f(\overline{x})\|_2} \quad \textit{und} \quad d_{max} = \frac{\nabla f(\overline{x})}{\|\nabla f(\overline{x})\|_2}$$

In der Praxis arbeitet man aber nicht mit der negativen Gradientenrichtung, denn gerade in der Nähe der gesuchten kritischen Punkte - nahe bei null - ist die Division $\frac{\nabla f(\overline{x})}{\|\nabla f(\overline{x})\|_2}$ numerisch instabil.

Satz 2.1.19 (Entwicklungen 1. und 2. Ordnung per univariatem Satz von Taylor)**.**

a) *Es sei $\varphi\colon \mathbb{R} \to \mathbb{R}$ differenzierbar an $\overline{t}$. Dann gilt für alle $t \in \mathbb{R}$*

$$\varphi(t) = \varphi(\overline{t}) + \varphi'(\overline{t})(t - \overline{t}) + o(|t - \overline{t}|),$$

wobei $o(|t - \overline{t}|)$ einen Ausdruck der Form $\omega(t) \cdot |t - \overline{t}|$ mit $\lim_{t \to \overline{t}} \omega(t) = \omega(\overline{t}) = 0$ bezeichnet.

b) *Es sei $\varphi\colon \mathbb{R} \to \mathbb{R}$ zweimal differenzierbar an $\overline{t}$. Dann gilt für alle $t \in \mathbb{R}$*

$$\varphi(t) = \varphi(\overline{t}) + \varphi'(\overline{t})(t - \overline{t}) + \frac{1}{2}\varphi''(\overline{t})(t - \overline{t})^2 + o(|t - \overline{t}|^2),$$

wobei $o(|t - \overline{t}|^2)$ einen Ausdruck der Form $\omega(t) \cdot |t - \overline{t}|^2$ mit $\lim_{t \to \overline{t}} \omega(t) = \omega(\overline{t}) = 0$ bezeichnet.

Lemma 2.1.20. *Für $f\colon \mathbb{R}^n \to \mathbb{R}$, einen Punkt $\overline{x} \in \mathbb{R}^n$ und eine Richtung $d \in \mathbb{R}^n$ seien $\varphi_d'(0) = 0$ und $\varphi_d''(0) < 0$. Dann ist d Abstiegsrichtung für f in $\overline{x}$.*

Lemma 2.1.21. *Für $f\colon \mathbb{R}^n \to \mathbb{R}$ sei $\overline{x}$ ein lokaler Minmalpunkt. Dann gilt $\nabla f(\overline{x}) = 0$, und jede Richtung $d \in \mathbb{R}^n$ erfüllt $\varphi_d''(0) \geq 0$.*

Bemerkung. *Es gilt für eine Richtung d gilt, dass $\varphi_d''(0) = d^T D^2 f(\overline{x}) d$*

Lemma 2.1.22. *Für $f\colon \mathbb{R}^n \to \mathbb{R}$, einen Punkt $\overline{x} \in \mathbb{R}^n$ und eine Richtung $d \in \mathbb{R}^n$ seien $\langle \nabla f(\overline{x}), d\rangle = 0$ und $d^T D^2 f(\overline{x}) d < 0$. Dann ist d Abstiegsrichtung für f in $\overline{x}$.*

Definition 2.1.23 (Abstiegsrichtung zweiter Ordnung)**.** *Zu $f\colon \mathbb{R}^n \to \mathbb{R}$ und $\overline{x} \in \mathbb{R}^n$ heißt jeder Richtungsvektor $d \in \mathbb{R}^n$ mit $\langle \nabla f(\overline{x}), d\rangle = 0$ und $d^T D^2 f(\overline{x}) d < 0$* ***Abstiegsrichtung zweiter Ordnung*** *für f in $\overline{x}$.*

Satz 2.1.27 (Notwendige Optimalitätsbedingung zweiter Ordnung)**.** *Die Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ sei zweimal differenzierbar an einem lokalen Minimalpunkt $\overline{x} \in \mathbb{R}^n$. Dann gilt $\nabla f(\overline{x}) = 0$ und $D^2 f(\overline{x}) \succeq 0$.*

Algorithmus 2.2: Konzeptioneller Algorithmus zur unrestringierten nichtlinearen Minimierung mit Informationen zweiter Ordnung

---

Input : Lösbares unrestringiertes zweimal stetig differenzierbares Optimierungsproblem P
Output : Globaler Minimalpunkt $x^\star$ von f über $\mathbb{R}^n$

begin
  Bestimme alle kritischen Punkte mit positiv semidefiniter Hesse-Matrix von f, d. h. die Lösungsmenge K der beiden Bedingungen $\nabla f(x) = 0$ und $D^2 f(x) \succeq 0$.
  Bestimme einen Minimalpunkt $x^\star$ von f in K.
end

Satz 2.1.30 (Entwicklungen 1. und 2. Ordnung per univariatem Satz von Taylor)**.**

a) *Es sei $f\colon \mathbb{R}^n \to \mathbb{R}$ differenzierbar an $\overline{x}$. Dann gilt für alle $x \in \mathbb{R}^n$*

$$f(x) = f(\overline{x}) + \langle \nabla f(\overline{x}), x - \overline{x}\rangle + o(\|x - \overline{x}\|),$$

wobei $o(\|x-\overline{x}\|)$ einen Ausdruck der Form $\omega(x)\cdot \|x-\overline{x}|$ mit $\lim_{x\to\overline{x}} \omega(x) = \omega(\overline{x}) = 0$ bezeichnet.

b) *Es sei $f\colon \mathbb{R}^n \to \mathbb{R}$ zweimal differenzierbar an $\overline{x}$. Dann gilt für alle $x \in \mathbb{R}^n$*

$$f(x) = f(\overline{x}) + \langle f(\overline{x}), x - \overline{x}\rangle + \frac{1}{2}(x - \overline{x})^T D^2 f(\overline{x})(x - \overline{x}) + o(\|x - \overline{x}\|^2),$$

wobei $o(\|x - \overline{x}\|^2)$ einen Ausdruck der Form $\omega(x) \cdot \|x - \overline{x}\|^2$ mit $\lim_{x\to\overline{x}} \omega(x) = \omega(\overline{x}) = 0$ bezeichnet.

Definition.

- $B_{\leq}(\overline{x}, r) = \big\{x \in \mathbb{R}^n \mid \ \|x - \overline{x}\| \leq r\big\}$

- $B_=(\overline{x}, r) = \{x \in \mathbb{R}^n \mid \|x - \overline{x}\| = r\}$

Satz 2.1.31 (Hinreichende Optimalitätsbedingung zweiter Ordnung)**.** *Die Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ sei an $\overline{x} \in \mathbb{R}^n$ zweimal differenzierbar, und es gelte $\nabla f(\overline{x}) = 0$ und $D^2 f(\overline{x}) \succ 0$. Dann ist $\overline{x}$ ein strikter lokaler Minimalpunkt von f.*

Definition 2.1.35 (Nichtdegenerierte kritische und Minimalpunkt)**.** *Die Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ sei an $\overline{x}$ zweimal differenzierbar mit $\nabla f(\overline{x}) = 0$. Dann heißt $\overline{x}$*

a) ***nichtdegenerierter kritischer Punkt****, falls $D^2 f(\overline{x})$ nichtsingulär ist,*

b) ***nichtdegenerierter lokaler Minimalpunkt****, falls $\overline{x}$ lokaler Minimalpunkt und nichtdegenerierter kritischer Punkt ist.*

Lemma 2.1.36. *Der Punkt $\overline{x}$ ist genau dann nichtdegenerierter lokaler Minimalpunkt von f, wenn $\nabla f(\overline{x}) = 0$ und $D^2 f(\overline{x}) \succ 0$ gilt.*

Definition. $\mathcal{F} = \{f \in C^2(\mathbb{R}^n,\ \mathbb{R}) \mid \textit{alle kritischen Punkte von } f \textit{ sind nichtdegeneriert}\}$

Satz 2.1.37. *$\mathcal{F}$ ist C^2_s-offen und -dicht in $C^2(\mathbb{R}^n, \mathbb{R})$.*

Übung 2.1.38. *In einem nichtdegeneriertem Sattelpunkt existiert sowohl eine Ab- als auch eine Anstiegsrichtung zweiter Ordnung.*

Konvexe Optimierungsprobleme

Definition 2.1.39 (Konvexe Mengen und Funktionen)**.**

a) Eine Menge $X \subseteq \mathbb{R}^n$ heißt ***konvex****, falls*

$$\forall x, y \in X, \lambda \in (0,1): \quad (1-\lambda)x + \lambda y \in X$$

gilt (d.h. die Verbindungsstrecke von je zwei beliebigen Punkten in X gehört komplett zu X).

b) Für eine konvexe Menge $X \subseteq \mathbb{R}^n$ heißt eine Funktion $f\colon X \to \mathbb{R}$ ***konvex*** *(auf X), falls*

$$\forall x, y \in X, \lambda \in (0,1): \quad f((1-\lambda)x + \lambda y) \le (1-\lambda)f(x) + \lambda f(y)$$

gilt (d.h. der Funktionsgraph von f verläuft unter jeder seiner Sekanten).

Bemerkung. *Während die Konvexität einer Funktion geometrisch dadurch definiert ist, dass ihr Graph unter jeder ihrer Sekanten verläuft, lässt sich Konvexität einer stetig differenzierbaren Funktion f dadurch charakterisieren, dass ihr Graph über den Graphen jeder ihrer Linearisierungen verläuft.*

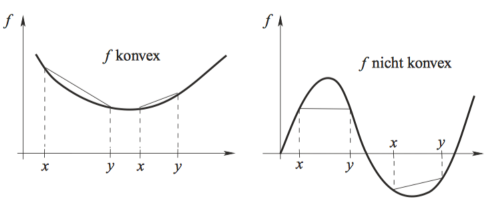


Abb. 2.4 Konvexität von Funktionen auf $\mathbb{R}$

Satz 2.1.40 (C^1-Charakterisierung von Konvexität)**.** *Auf einer konvexen Menge $X \subseteq \mathbb{R}^n$ ist eine Funktion $f \in C^1(X, \mathbb{R})$ genau dann konvex, wenn folgendes gilt:*

$$\forall x, y \in X : \quad f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

Korollar 2.1.41. *Die Funktion $f \in C^1(\mathbb{R}^n, \mathbb{R})$ sei konvex. Dann sind die kritischen Punkte von f genau die globalen Minimalpunkte von f.*

Satz 2.1.42 (C^2-Charakterisierung von Konvexität)**.** *Eine Funktion $f \in C^2(\mathbb{R}^n, \mathbb{R})$ ist genau dann konvex, wenn folgendes gilt:*

$$\forall x \in \mathbb{R}^n : \quad D^2 f(x) \succeq 0$$

Übung 2.1.43. *Gegeben sei die quadratische Funktion $q(x) = \frac{1}{2} x^T A x + b^T x$ mit $A = A^T \succ 0$ und $b \in \mathbb{R}^n$. Die Funktion q ist eine auf $\mathbb{R}^n$ (gleichmäßige) konvexe Funktion und ihr eindeutiger Minimalpunkt*

$$x^* = -A^{-1} b$$

mit Minimalwert $q(x^*) = -\frac{1}{2} b^T A^{-1} b$.

Numerische Verfahren

2.2.1 Abstiegsverfahren

Zunächst betrachten wir Verfahren, die in jedem Iterationsschritt einen Abstieg im Zielfunktionswert erzeugen, für die also

$$\forall k \in \mathbb{N}_0 : \quad f(x^{k+1}) < f(x^k)$$

gilt. Solche Verfahren können nur „unter sehr unglücklichen Umständen“ gegen lokale Maximalpunkte konvergieren und aus geometrischen Überlegungen heraus ist die Konvergenz gegen Sattelpunkte unwahrscheinlich.

Neben der Stetigkeit der Zielfunktion f werden wir im gesamten Abschn. 2.2 fordern, dass die untere Niveaumenge $f_{\leq}^{f(x^0)}$ zum Startpunkt $x^0 \in \mathbb{R}^n$ beschränkt ist.

Übung. *Als erste algorithmische Idee könnte man versuchen, die Gleichung $\nabla f(x) = 0$ mit dem aus der Numerik bekannten Newton-Verfahren*

$$x^{k+1} = x^k - \left(D^2 f(x^k) \right)^{-1} \nabla f(x^k), \quad k = 0, 1, 2, \ldots$$

Vorteil wäre eine hohe Konvergenzgeschwindigkeit, falls x^0 nahe genug an einer Lösung liegt. Nachteilig ist, dass x^0 nicht in der Nähe einer Lösung zu liegen braucht, dass die Hesse-Matrix $D^2 f(x^k)$ nicht notwendig invertierbar sein muss und dass das Newton-Verfahren auch gegen lokale Maximalpunkte und Sattelpunkte konvergieren kann.

Algorithmus 2.3: Allgemeines Abstiegsverfahren

Input : C^1-Optimierungsproblem P
Output : Approximation $\bar{x}$ eines kritischen Punkts von f (falls das Verfahren terminiert; Korollar 2.2.10)

1 **begin**
2   Wähle einen Startpunkt x^0, eine Toleranz $\varepsilon > 0$ und setze $k = 0$.
3   **while** $\|\nabla f(x^k)\| > \varepsilon$ **do**
4   Wähle x^{k+1} mit $f(x^{k+1}) < f(x^k)$.
5   Ersetze k durch $k + 1$.
6   **end**
7   Setze $\bar{x} = x^k$.
8 **end**

Lemma 2.2.3. *Für beschränktes $f_{\leq}^{f(x^0)}$ bricht die von Algorithmus 2.3 mit $\epsilon = 0$ erzeugte Folge (x^k) entweder nach endlich vielen Schritten mit einem kritischen Punkt ab, oder sie besitzt mindestens einen Häufungspunkt in $f_{\leq}^{f(x^0)}$, und die Folge der Funktionswerte $\big(f(x^k)\big)$ ist konvergent.*

Definition 2.2.5 (Effiziente Schrittweiten)**.** *Es sei (d^k) eine Folge von Abstiegsrichtungen erster Ordnung, und (t^k) erfülle*

$$\exists c > 0 \ \forall k \in \mathbb{N}: \quad f(x^k + t^k d^k) - f(x^k) \leq -c \left(\frac{\langle \nabla f(x^k), d^k \rangle}{\|d^k\|_2} \right)^2$$

Dann heißt (t^k) ***effiziente Schrittweitenfolge*** *für (d^k).*

Satz 2.2.6. *Die Menge $f_{\leq}^{f(x^0)}$ sei beschränkt, (d^k) sei eine Folge von Abstiegsrichtungen erster Ordnung, und (t^k) sei eine effiziente Schrittweitenfolge. Dann gilt (2.6):*

$$\lim_k \frac{\langle \nabla f(x^k), d^k \rangle}{\|d^k\|_2} = 0.$$

Definition 2.2.7 (Gradientenbezogene Suchrichtungen)**.** *Die Folge von Suchrichtungen (d^k) heißt* ***gradientenbezogen****, falls folgendes gilt:*

$$\exists c > 0 \ \forall k \in \mathbb{N}: \quad \frac{\langle \nabla f(x^k), d^k \rangle}{\|d^k\|_2} \leq -c \cdot \|\nabla f(x^k)\|_2$$

Übung 2.2.8. *Die Suchrichtungen $d^k = -\nabla f(x^k)$, $k \in \mathbb{N}$ sind gradientenbezogen.*

Satz 2.2.9. *Die Menge $f_{\leq}^{f(x^0)}$ sei beschränkt, und in Zeile 4 von Algorithmus 2.3 sei $x^{k+1} = x^k + t^k d^k$ mit einer gradientenbezogenen Suchrichtungsfolge (d^k) und einer effizienten Schrittweitenfolge (t^k) gewählt. Für $\epsilon = 0$ stoppt dann das Verfahren entweder nach endlich vielen Schritten mit einem kritischen Punkt, oder die Folge (x^k) besitzt einen Häufungspunkt, und für jeden solchen Punkt x^* gilt $\nabla f(x^*) = 0$.*

Korollar 2.2.10. *Die Menge $f_{\leq}^{f(x^0)}$ sei beschränkt, und in Zeile 4 von Algorithmus 2.3 sei $x^{k+1} = x^k + t^k d^k$ mit einer gradientenbezogenen Suchrichtungsfolge (d^k) und einer effizienten Schrittweitenfolge (t^k) gewählt. Dann terminiert das Verfahren für jedes $\epsilon > 0$ nach endlich vielen Schritten.*

Schrittweitensteuerung

Definition. *Eine Funktion $F\colon D \to \mathbb{R}^m$ heißt* ***Lipschitz-stetig*** *auf $D \subseteq \mathbb{R}^n$, falls*

$$\exists L > 0 \ \forall x, y \in D: \quad \|F(x) - F(y)\|_2 \leq L \cdot \|x - y\|_2$$

C^1-Funktionen sind auf kompakten Mengen immer Lipschitz-stetig sind, damit ist ∇f bei beschränkter Menge $f_{\leq}^{f(x^0)}$ zum Beispiel für jede C^2-Funktion f Lipschitz-stetig auf $f_{\leq}^{f(x^0)}$.

Bemerkung 2.2.12. *Bei beschränktem (und daher kompaktem) $f_{\leq}^{f(x^0)}$ ist die Menge $\operatorname{conv}(f_{\leq}^{f(x^0)})$ ebenfalls kompakt, so dass die Forderung der Lipschitz-Stetigkeit von ∇f auch auf $\operatorname{conv}(f_{\leq}^{f(x^0)})$ eine schwache Voraussetzung ist.*

Lemma 2.2.13. *Auf einer konvexen Menge $D \subseteq \mathbb{R}^n$ sei f differenzierbar mit Lipschitz-stetigem Gradienten ∇f und zugehöriger Lipschitz-Konstante $L > 0$. Dann gilt*

$$\nabla \overline{x}, x \in D: \quad |f(x) - f(\overline{x}) - \langle \nabla f(\overline{x}), x - \overline{x}\rangle| \leq \frac{L}{2}\|x - \overline{x}\|_2^2$$

Exakte Schrittweiten

Zu $x \in f_{\leq}^{f(x^0)}$ sei eine Abstiegsrichtung erster Ordnung d für f in x gegeben. Wegen $\varphi_d'(0) = \langle \nabla f(x), d\rangle < 0$ gilt $\varphi_d(t) < \varphi_d(0)$ für kleine positive t. Für beschränktes $f_{\leq}^{f(x^0)}$ besitzt φ_d nach dem Satz von Weierstraß sogar globale Minimalpunkte $t_e > 0$, die exakte Schrittweiten genannt werden. Per Definition der eindimensionalen Einschränkung φ_d erfüllen sie

$$f(x + t_e d) = \min_{t>0} f(x + td)$$

Eine exakte Schrittweite zu berechnen, um den größtmöglichen Abstieg von x aus entlang d zu erzielen, ist im Allgemeinen sehr aufwendig, so dass wir dieses Konzept meist nur für theoretische Zwecke benutzen werden und später stattdessen zu inexakten Schrittweiten übergehen werden.

Übung 2.2.14. *Gegeben sei die quadratische Funktion $q(x) = \frac{1}{2}x^T Ax + b^T x$ mit $A = A^T \succ 0$ und $b \in \mathbb{R}^n$, die nach Übung 1.2.24 koerziv und nach Übung 2.1.43 konvex ist. Für jedes $x \in$*
$mathbbR^n$ und jede Abstiegsrichtung erster Ordnung d für q in x die exakte Schrittweite eindeutig bestimmt zu

$$t_e = -\frac{\langle Ax + b, d\rangle}{d^T Ad}$$

Satz 2.2.15. *Die Menge $f_{\leq}^{f(x^0)}$ sei beschränkt, die Funktion ∇f sei Lipschitz-stetig auf $\operatorname{conv}(f_{\leq}^{f(x^0)})$, und (d^k) sei eine Folge von Abstiegsrichtung erster Ordnung. Dann ist jede Folge von exakten Schrittweiten (t_e^k) effizient.*

Konstante Schrittweiten

Falls die Funktion f keine besondere Struktur aufweist, lohnt sich der Aufwand nicht, in jedem Iterationsschritt eine exakte Schrittweite t_e^k zu berechnen. Daher bnutzt man dann lieber inexakte Schrittweiten, die ebenfalls effizient, aber erheblich leichter zu berechnen sind.

Eine zunächst naheliegend erscheinende Möglichkeit dafür besteht darin

$$t_c^k = -\frac{\langle \nabla f(x^k), d^k \rangle}{L \cdot \|d^k\|_2^2}$$

Auch diese Schrittweitre ist effizient, genauso wie die exakte. Im speziellen Fall $d^k = -\nabla f(x^k)$ gilt sogar

$$t_c^k = \frac{1}{L}$$

Armijo-Schrittweiten

Eine in modernen Implementierungen von Optimierungsverfahren sehr beliebte inexakte Schrittweitensteuerung geht auf eine Idee von Armijo zurück: Zu $x \in f_{\leq}^{f(x^0)}$ seien d eine Abstiegsrichtung erster Ordnung und $\sigma \in (0,1)$. Dann existiert ein $\hat{t} > 0$, so dass für alle $t \in (0, \hat{t})$ die Werte $\varphi_d(t)$ unter der „nach oben gedrehten Tangente“ $\varphi_d(0) + t\sigma\varphi_d'(0)$ liegen, so dass also gilt:

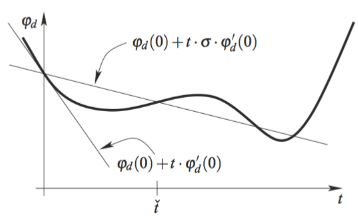
$$f(x + td) \leq f(x) + t\sigma\langle \nabla f(x), d \rangle$$

Offensichtlich erfüllt jedes solche $t££$ die Bedingung (2.3):

$$\exists c_1 > 0 \ \forall k \in \mathbb{N} : \quad f(x^k + t^k d^k) - f(x^k) \leq c_1 \cdot t^k \langle \nabla f(x^k), d^k \rangle$$

mit $c_1 = \sigma$.

Abb. 2.5 Armijo-Regel



Algorithmus 2.4: Armijo-Regel

Input : C^1-Funktion f und $x, d \in \mathbb{R}^n$ mit $\langle \nabla f(x), d \rangle < 0$
Output : Armijo-Schrittweite t_a

begin
  Wähle $\sigma, \rho \in (0,1)$ sowie $\gamma > 0$ (alle unabhängig von x und d).
  Wähle eine Startschrittweite $t^0 \geq -\gamma \, \langle \nabla f(x), d \rangle / \|d\|_2^2$ und setze $\ell = 0$.
  while $f(x + t^\ell d) > f(x) + t^\ell \sigma \langle \nabla f(x), d \rangle$ **do**
    Setze $t^{\ell+1} = \rho \, t^\ell$.
    Ersetze ℓ durch $\ell + 1$.
  end
  Setze $t_a = t^\ell$.
end

Satz 2.2.16. *Die Menge $f_{\leq}^{f(x^0)}$ sei beschränkt, die Funktion ∇f sei Lipschitz-stetig auf $\mathrm{conv}(f_{\leq}^{f(x^0)})$, und (d^k) sei eine Folge von Abstiegsrichtungen erster Ordnung. Dann ist die Folge der Armijo-Schrittweiten (t_a^k) aus Algorithmus 2.4 (mit unabhängig von k gewählten Parametern σ, ρ und γ) wohldefiniert und effizient.*

Übung 2.2.17. *Zeigen Sie für die Funktion $f(x) = \frac{1}{2}x^2$, den Startpunkt $x^0 = -3$, die Richtungen $d^k = 2^{-k}$ sowie $\sigma = \frac{1}{2}$, dass der durch die Wahl $t^0 := 1$ modifizierte Algorithmus 2.4 nicht zu einer effizienten Schrittweitenfolge führt.*

Man sollte t^0 also so initialisieren, wie in Algorithmus 2.4 angegeben, wobei sich die Wahl $\gamma = 10^{-4}$ bewährt hat. Es ist außerdem nicht schwer zu sehen, dass sich die Armijo-Regel auch für nur einseitig richtungsdifferenzierbare Funktionen einsetzen lässt, indem man das Skalarprodukt $\lambda \nabla f(\overline{x}), d \, \rangle$ durch $f'(\overline{x}, d)$ ersetzt.

2.2.2 Gradientenverfahren

Aufgrund seiner geometrischen Grundidee ist dies das Verfahren des steilsten Abstiegs.

Algorithmus 2.5: Gradientenverfahren

Input : C^1-Optimierungsproblem P
Output : Approximation $\bar{x}$ eines kritischen Punkts von f (falls das Verfahren terminiert; Satz 2.2.18)

1 **begin**
2   Wähle einen Startpunkt x^0, eine Toleranz $\varepsilon > 0$ und setze $k = 0$.
3   **while** $\|\nabla f(x^k)\| > \varepsilon$ **do**
4   Setze $d^k = -\nabla f(x^k)$.
5   Bestimme eine Schrittweite t^k.
6   Setze $x^{k+1} = x^k + t^k d^k$.
7   Ersetze k durch $k + 1$.
8   **end**
9   Setze $\bar{x} = x^k$.
10 **end**

Satz 2.2.18. *Die Menge $f_{\leq}^{f(x^0)}$ sei beschränkt, die Funktion ∇f sei Lipschitz-stetig auf $\mathrm{conv}(f_{\leq}^{f(x^0)})$, und in Zeile 5 seien exakte Schrittweiten (t_e^k) oder Armijo-Schrittweiten (t_a^k)*

gewählt. Dann terminiert Algorithmus 2.5 für jedes $\epsilon > 0$ nach endlich vielen Schritten. Falls eine Lipschitz-Konstante $L > 0$ zur Lipschitz-Stetigkeit von ∇f auf $\operatorname{conv}(f_{\leq}^{f(x^0)})$ bekannt ist, dann gilt dieses Ergebnis auch für die dann berechenbaren konstanten Schrittweiten $t_c^k = L^{-1}, k \in \mathbb{N}$.

Definition. *Es ist* $\|A\|_2 := \max\big\{\|Ad\|_2 \mid \|d\|_2 = 1\big\}$.

Übung 2.2.19. *Gegeben sei die quadratische Funktion $q(x) = \frac{1}{2}x^TAx + b^Tx$ mit $A = A^T$ und $b \in \mathbb{R}^n$. Zeigen Sie, dass der Gradient ∇q auf ganz $\mathbb{R}^n$ Lipschitz-stetig mit $L = \|A\|_2$ ist.*

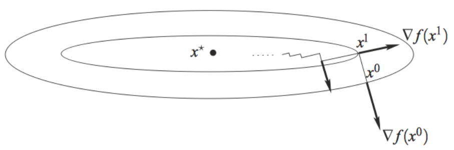
Beispiel 2.2.20. *Nach Übung 2.2.19 erzeugt das Gradientenverfahren eine sogar gegen den globalen Minimalpunkt von q konvergente Folge von Iterierten (x^k), wenn entweder exakte, konstante oder Armijo-Schrittweiten gewählt werden.*

Nach Übung 2.2.14 ist bei jeder Abstiegsrichtung erster Ordnung für q in x die (eindeutige) exakte Schrittweite beim Gradientenverfahren

$$t_e = \frac{\|\nabla q(x)\|_2^2}{Dq(x)A\nabla q(x)}$$

Falls die Höhenlinien von f die Form lang gezogener Ellipsen mit einem Minimalpunkt x^* in deren gemeinsamem Zentrum besitzen, dann zeigt $-\nabla f(x^k)$ typischerweise nicht in die Richtung von x^*. Die Iterierten springen dadurch entlang einer Zickzacklinie, weshalb man in Anlehnung an die englischsprachige Literatur auch vom Zigzagging-Effekt spricht.

Abb. 2.6 Zigzagging-Effekt



Definition 2.2.21 (Konvergenzgeschwindigkeiten). *Es sei (x^k) eine konvergente Folge mit Grenzpunkt x^*. Sie heißt*

a) ***linear konvergent**, falls $\exists 0 < c < 1,\ k_0 \in \mathbb{N}\ \forall k \geq k_0$:*

$$\|x^{k+1} - x^*\| \leq c \cdot \|x^k - x^*\|,$$

b) ***superlinear konvergent**, falls $\exists c^k \searrow 0, k_0 \in \mathbb{N}\ \forall k \geq k_0$:*

$$\|x^{k+1} - x^*\| \leq c^k \cdot \|x^k - x^*\|,$$

c) ***quadratisch konvergent**, falls $\exists c > 0, k_0 \in \mathbb{N}\ \forall k \geq k_0$:*

$$\|x^{k+1} - x^*\| \leq c \cdot \|x^k - x^*\|^2.$$

Der folgende Satz zeigt, dass das Gradientenverfahren schon für sehr angenehme Funktionen nur linear konvergente Funktionswerte der Iterierten besitzt, und zwar mit einer Konstante c, die sehr nahe bei eins liegen kann. Konkret betrachten wir die konvex-quadratische Funktion $q(x) = \frac{1}{2}x^T Ax + b^T x$ mit $A = A^T \succ 0$ sowie $b \in \mathbb{R}^n$ und bezeichnen den größten und den kleinsten Eigenwert der Matrix A mit λ_{max} bzw. λ_{min} - nach Beispiel 2.2.20 konvergieren dabei die Iterierten des Gradientenverfahrens mit exakten Schrittweiten gegen den globalen Minimalpunkt $x = -A^{-1}b$ von q.

Lemma 2.2.22 (Kantorowitsch-Ungleichung)**.** *Es sei $A = A^T \succ 0$ mit maximalem und minimalem Eigenwert λ_{max} bzw. λ_{min}. Dann gilt für jedes $v \in \mathbb{R}^n \setminus \{0\}$*

$$\frac{v^T A^{-1} v \cdot v^T A v}{\|v\|_2^4} \leq \frac{(\lambda_{max} + \lambda_{min})^2}{4\lambda_{max}\lambda_{min}}$$

Satz 2.2.23. *Auf die konvex-quadratische Funktion $q(x) = \frac{1}{2}x^T Ax + b^T x$ mit $A = A^T \succ 0$ und $b \in \mathbb{R}^n$ werde das Gradientenverfahren mit exakten Schrittweiten und $\epsilon = 0$ angewendet. Dann gilt für alle $k \in \mathbb{N}$*

$$|q(x^{k+1}) - q(x^*)| \leq \left(\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}}\right)^2 |q(x^k) - q(x^*)|.$$