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

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

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

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

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

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

abgeschlossen.

Definition. Die Menge der globalen Minimalpunkte lautet:

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

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

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

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 $\mathrm{lev}^{\alpha}_{\leq}(f, M)$ 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 $\mathrm{lev}^{\alpha}_{\leq}(f, X)$ 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 1.3.1 (Skalare Vielfache und Summen)**.** Gegeben seien $M \subseteq \mathbb{R}^n$ und $f, g\colon M \to \mathbb{R}$. Dann gilt

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

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

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

Übung 1.3.2 (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 1.3.3 (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 1.3.4 (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

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

und die lokalen bzw. globalen Minimalpunkte stimmen überein.

Übung 1.3.5 (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 1.3.6 (Parallelprojektion)**.** Es sei $M \subseteq \mathbb{R}^n \times \mathbb{R}^m$. Dann heißt

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

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

Übung 1.3.7 (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

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

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

Optimalitätsbedingungen

Abstiegsrichtung

Definition 2.1.1 (Abstiegsrichtung)**.** Es seien $f\colon \mathbb{R}^n \to \mathbb{R}$ und $\overline{x} \in \mathbb{R}^n$. Ein Vektor $d \in \mathbb{R}^n$ heißt **Abstiegsrichtung** für f in $\overline{x}$, falls

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

Übung 2.1.2. 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 2.1.3. 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 2.1.4. 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 2.1.5. 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 2.1.6. 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 2.1.7 (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 2.1.8 (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 2.1.9 (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 2.1.10. 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 2.1.11. 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 2.1.12. 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 $f(x) \coloneqq \max_{k\in K} f_k(x)$ 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 2.1.13 (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 2.1.14 (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 2.1.15. 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.

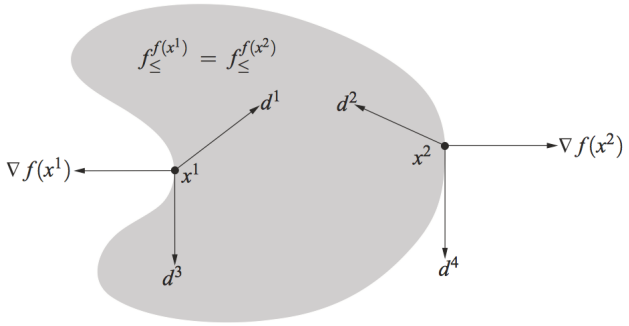
Geometrische Eigenschaften von Gradienten

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 $\mathbb{R}^n$

1 **begin**
2 Bestimme alle kritischen Punkte von f, d. h. die Lösungsmenge K der Gleichung $\nabla f(x) = 0$.
3 Bestimme einen Minimalpunkt $x^\star$ von f in K.
4 **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

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

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 $\bar{t}$. Dann gilt für alle $t \in \mathbb{R}$

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

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

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

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

wobei $o(|t - \bar{t}|^2)$ einen Ausdruck der Form $\omega(t) \cdot |t - \bar{t}|^2$ mit $\lim_{t \to \bar{t}} \omega(t) = \omega(\bar{t}) = 0$ 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 $\mathbb{R}^n$

1 **begin**
2 Bestimme alle kritischen Punkte mit positiv semidefiniter Hesse-Matrix von f, d. h. die Lösungsmenge K der beiden Bedingungen $\nabla f(x) = 0$ und $D^2 f(x) \succeq 0$.
3 Bestimme einen Minimalpunkt $x^\star$ von f in K.
4 **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.

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

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

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

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

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.

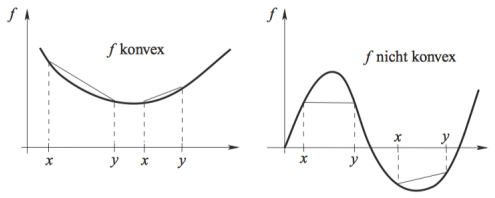


Abb. 2.4 Konvexität von Funktionen auf $\mathbb{R}$

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 $q(x) = \frac{1}{2}x^T Ax + b^T x$ 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

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 ε > 0 und setze k = 0.
3     while ‖∇f(x^k)‖ > ε do
4         Wähle x^{k+1} mit f(x^{k+1}) < f(x^k) .
5         Ersetze k durch k + 1.
6     end
7     Setze x̄ = 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 $\left(f(x^k)\right)$ 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 $\text{conv}(f_{\leq}^{f(x^0)})$ ebenfalls kompakt, so dass die Forderung der Lipschitz-Stetigkeit von ∇f auch auf $\text{conv}(f_{\leq}^{f(x^0)})$ 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. Es gilt

$$0 = \varphi_d(t_e) = \langle \nabla f(x + t_e d), d \rangle$$

Übung 2.2.14. 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 nach Übung 1.2.24 koerziv und nach Übung 2.1.43 konvex ist. Für jedes $x \in \mathbb{R}^n$ und jede Abstiegsrichtung erster Ordnung d für q in x die exakte Schrittweite eindeutig bestimmt zu

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

Satz 2.2.15. Die Menge $f_{\leq}^{f(x^0)}$ sei beschränkt, die Funktion ∇f sei Lipschitz-stetig auf $\text{conv}(f_{\leq}^{f(x^0)})$, 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 $t > 0$, 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:

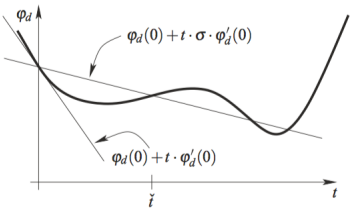
$$f(x + td) \leq f(x) + t\sigma\langle \nabla f(x), d \rangle$$

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 $\operatorname{conv}(f_{\leq}^{f(x^0)})$, 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.

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 $\operatorname{conv}(f_{\leq}^{f(x^0)})$, 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 $\|A\|_2 := \max\{\|Ad\|_2 \mid \|d\|_2 = 1\}$.

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

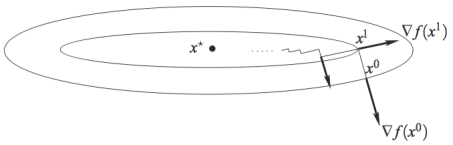
Beispiel 2.2.20. Nach Übung 2.2.19 erzeugt das Gradientenverfahren eine sogar gegen den globalen Minimalpunkt von q konvergente Folge von Iterierten (x^k), wenn entweder exakte, konstante oder Armijo-Schrittweiten gewählt werden.

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 $k \in \mathbb{N}$

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

Nach Satz 2.2.23 minimiert das Gradientenverfahren (mit exakten Schrittweiten) eine konvex-quadratische Funktion $q(x) = \frac{1}{2}x^T Ax + b^T x$ in einem einzigen Schritt, wenn der kleinste und größte Eigenwert λ_{min} bzw. λ_{max} von A übereinstimmen. Dann stimmen natürlich auch alle Eigenwerte von A miteinander überein, so dass q sphärenförmige Niveaumengen besitzt.

Variable-Metrik-Verfahren

Im Allgemeinen existiert stets ein rechtwinkliges Koordinatensystem, das zur Lage der ellipsodialen Niveaumengen von q „passend ausgerichtet“ ist, sodass in einem neuen Koordinatensystem die Niveaulinien sphärenförmig sind.

Für eine nicht notwendigerweise konvex-quadratische Funktion $f \in C^1(\mathbb{R}^n, \mathbb{R})$ beschränkt man sich darauf, approximativ eine Konstruktion wie bei konvex-quadratischen Funktionen zu benutzen:

Definition 2.2.27 (Gradient bezüglich einer positiv definiten Matrix)**.** Für $f \in C^1(\mathbb{R}^n, \mathbb{R})$ und eine (n, n)-Matrix $A = A^T \succ 0$ heißt

$$\nabla_A f(x) \coloneqq A^{-1} \nabla f(x)$$

Gradient von f bezüglich A an x.

Die verschiedenen Variable-Metrik-Verfahren unterscheiden sich durch die Wahl der Matrix A, mit deren Hilfe die Suchrichtung $-\nabla_A f(x)$ gebildet wird. Für jedes $A = A^T \succ 0$ ist diese Suchrichtung an einem nichtkritischen Punkt x jedenfalls eine Abstiegsrichtung erster Ordnung, denn da mit A auch A^{-1} positiv definit ist, gilt

$$\langle \nabla f(\overline{x}), -\nabla_A f(\overline{x}) \rangle = -\nabla f(\overline{x})^T A^{-1} \nabla f(\overline{x}) < 0$$

Übung 2.2.28. Für jedes $A = A^T \succ 0$ ist die Funktion $\langle x, y \rangle_A \coloneqq x^T Ay$ ein Skalarprodukt auf $\mathbb{R}^n$.

Übung 2.2.29. Für jedes $A = A^T \succ 0$ und für das von A induzierte Skalarprodukt $\langle \cdot, \cdot \rangle_A$ ist die folgende Funktion eine Norm auf $\mathbb{R}^n$:

$$\|x\|_A \coloneqq \sqrt{\langle x, x \rangle_A}$$

Übung 2.2.30. Es gilt unter für die konvex-quadratische Funktion $q(x) = \frac{1}{2} x^T A x + b^T x$ mit $A = A^T \succ 0$, $b \in \mathbb{R}^n$ und mit exakten Schrittweiten:

$$\frac{1}{2} \|x - x^*\|_A^2 = q(x) - q(x^*)$$

Da die quadrierten Abstände der Iterierten zum Grenzpunkt linear konvergieren, erhält man aus Übung 2.2.30 eine sogar noch langsamere als lineare (nämlich eine sog. sublineare) Konvergenzgeschwindigkeit der Iterierten selbst.

Übung 2.2.31. Gegeben sei die quadratische Funktion $q(x) = \frac{1}{2} x^T A x + b^T x$ mit $A = A^T$ und $b \in \mathbb{R}^n$. Für die exakte Schrittweite des Gradientenverfahrens gilt die Formel

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

Übung 2.2.32. Für das durch $A = A^T \succ 0$ induzierte Skalarprodukt $\langle \cdot, \cdot \rangle_A$ und die induzierte Norm $\| \cdot \|_A$ gilt die Cauchy-Schwarz-Ungleichung:

$$\forall x, y \in \mathbb{R}^n : \quad |\langle x, y \rangle_A| \leq \|x\|_A \cdot \|y\|_A$$

und die Abschätzung ist scharf.

Lemma 2.2.33. Es sei $\nabla f(x) \neq 0$. Dann löst der Vektor

$$d = -\frac{-\nabla_A f(x)}{\|\nabla_A f(x)\|_A}$$

das Problem $\min \langle \nabla f(x), d \rangle$ s.t. $\|d\|_A = 1$, und zwar mit optimalem Wert $-\|\nabla_A f(x)\|_A$.

Algorithmus 2.6: Variable-Metrik-Verfahren

Input : C^1-Optimierungsproblem P
Output : Approximation $\bar{x}$ eines kritischen Punkts von f (falls das Verfahren terminiert; Satz 2.2.37)

1 **begin**
2 Wähle einen Startpunkt x^0, eine Matrix $A^0 = (A^0)^\intercal \succ 0$, eine Toleranz $\varepsilon > 0$ und setze $k = 0$.
3 **while** $\|\nabla f(x^k)\|_2 > \varepsilon$ **do**
4 Setze $d^k = -\nabla_{A^k} f(x^k)$.
5 Bestimme eine Schrittweite t^k.
6 Setze $x^{k+1} = x^k + t^k d^k$.
7 Wähle $A^{k+1} = (A^{k+1})^\intercal \succ 0$.
8 Ersetze k durch $k + 1$.
9 **end**
10 Setze $\bar{x} = x^k$.
11 **end**

In Zeile 2 von Algorithmus 2.6 wählt man häufig $A^0 = E$, also als erste Suchrichtung die Gradientenrichtung $d^0 = -\nabla f(x^0)$. In Zeile 3 wäre ein konsistenteres Abbruchkriterium eigentlich

$$\|\nabla_{A^k} f(x^k)\|_{A^k} \leq \epsilon,$$

aber wegen

$$\|\nabla_{A^k} f(x^k)\|_{A^k} = \sqrt{Df(x^k)(A^k)^{-1}\nabla f(x^k)} = \|\nabla f(x^k)\|_{(A^k)^{-1}}$$

und der Äquivalenz von $\|\cdot\|_{(A^k)^{-1}}$ und $\|\cdot\|_2$ (d. h., es gibt Konstanten $c_1, c_2 > 0$, so dass alle $x \in \mathbb{R}^n$ die Abschätzungen $c_1\|x\|_{(A^k)^{-1}} \leq \|x\|_2 \leq c_2\|x\|_{(A^k)^{-1}}$ erfüllen) kann man ebensogut das angegebene und weniger aufwendigere Kriterium testen. In Zeile 4 berechnet man die Suchrichtung d^k numerisch nicht durch die eine Matrixinversion enthaltende Definition $-(A^k)^{-1}\nabla f(x^k)$, sondern weniger aufwendig als Lösung des linearen Gleichungssystems $A^k d = -\nabla f(x^k)$.

Möchte man die Konvergenz von Variable-Metrik-Verfahren im Sinne von Satz 2.2.9 garantieren, benötigt man neben der Effizienz der Schrittweiten auch die Gradientenbezogenheit der Suchrichtungen. Diese muss man noch fordern.

Definition 2.2.34 (Gleichmäßig positiv definite und beschränkte Matrizen)**.** Eine Folge (A^k) symmetrischer (n, n)-Matrizen heißt **gleichmäßig positiv definit** und **beschränkt**, falls folgendes gilt:

$$\exists 0 < c_1 \leq c_2 \; \forall d \in B_=(0,1),\; k \in \mathbb{N}: \quad c_1 \leq d^T A^k d \leq c_2$$

Übung 2.2.35. Die Folge (A^k) sei gleichmäßigpositiv definit und beschränkt mit Konstanten c_1 und c_2. Dann ist die Folge $\left((A^k)^{-1}\right)$ auch gleichmäßig positiv definit und beschränkt mit Konstanten $\frac{1}{c_2}$ und $\frac{1}{c_1}$. Außerdem ist die Folge $\left(\lambda_{max}\left((A^k)^{-1}\right)\right)$ der größten Eigenwerte von $\left((A^k)^{-1}\right)$ durch $\frac{1}{c_1}$ nach oben beschränkt.

Satz 2.2.36. Die Folge (A^k) sei gleichmäßig positivdefinit und beschränkt. Dann ist die Folge (d^k) mit $d^k = -\left(A^k\right)^{-1} \nabla f(x^k)$, $k \in \mathbb{N}$, gradientenbezogen.

Satz 2.2.37. Die Menge $f_{\leq}^{f(x^0)}$ sei beschränkt, die Funktion ∇f sei Lipschitz-stetig auf $\mathrm{conv}(f_{\leq}^{f(x^0)})$, die Folge (A^k) sei gleichmäßig positiv definit und beschränkt, und in Zeile 5 seien exakte Schrittweite (t_e^k) oder Armijo-Schrittweiten (t_a^k) gewählt. Dann terminiert Algorithmus 2.6 für jedes $\epsilon > 0$ nach endlich vielen Schritten.

Bemerkung 2.2.38 (Spektralnorm und Eigenwerte)**.** Durch die in Abschnitt 2.2.3 eingeführte Spektralnorm $\|A\|_2 \coloneqq \max\left\{\|Ad\|_2 \mid \|d\|_2 = 1\right\}$ können die Eigenwerte einer Matrix A durch die semidefinite Matrix $A^T A$ berechnet werden. Es ist

$$\|A\|_2 = \sqrt{\lambda_{max}(A^T A)} = \sqrt{\left(\lambda_{max}(A)\right)^2} = |\lambda_{max}(A)|$$

Damit ist nach Bemerkung 2.2.24 und Übung 2.2.5 die Länge der längsten Halbachse des Niveau-Ellipsoids

$$\frac{1}{\sqrt{\lambda_{min}(A^T A)^{-1})}} = \sqrt{\lambda_{max}(A^T A)}$$

Newton-Verfahren mit und ohne Dämpfung

Wählt man in Algorithmus 2.6 für $f \in C^2(\mathbb{R}^n, \mathbb{R})$ in jedem Schritt $A^k = D^2 f(x^k)$, so erhält man das Newton-Verfahren, sofern die Matrizen $D^2 f(x^k)$ positiv definit sind.

Ist x^* nichtdegenerierter lokaler Minimalpunkt von f, dann gilt aus den bereits am Ende von Abschnitt 2.1.5. aufgeführten Stetigkeitsgründen $D^2 f(x) \succ 0$ für alle x aus einer Umgebung von x^*. Für x^0 aus dieser Umgebung kann man also $A^k = D^2 f(x^k)$ setzen und erhält ein wohldefiniertes Abstiegsverfahren. Ferner sind die Suchrichtungen $d^k = -\left(D^2 f(x^k)\right)^{-1} \nabla f(x^k)$ gradientenbezogen, falls f und x gleichmäßig konvex ist, d. h. falls für eine Umgebung U von x gilt:

$$\exists c > 0 \; \forall x \in U, d \in B_{=}(0,1): \quad c \leq d^T D^2 f(x) d$$

Die für diese Folgerung nach Satz 2.2.36 noch erforderliche Beschränktheit der Folge $(D^2 f(x^k))$ resultiert dabei aus der Stetigkeit von $D^2 f$. Die Nichtdegeneriertheit des lokalen Minimalpunkts x^* gilt bei gleichmäßig konvexem f automatisch.

Die Dämpfung des Newton-Verfahrens hat den Vorteil, dass der Konvergenzradius (also der mögliche Abstand von x_0 zu x^*) etwas größer wird. Andererseits ist zunächst nicht klar, ob die Dämpfung nicht auch die lokale Konvergenz verlangsamt. Das ungedämpfte Newton-Verfahren konvergiert unter schwachen Voraussetzungen jedenfalls quadratisch.

Satz 2.2.39 (Quadratische Konvergenz des Newton-Verfahrens)**.** Die durch

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

definierte Folge (x^k) konveriere gegen einen nichtdegenerierten lokalen Minimalpunkt x^*, und $D^2 f$ sei Lipschitz-stetig auf einer konvexe Umgebung von x^*. Dann konvergiert die Folge (x^k) quadratisch gegen x^*.

Bemerkung 2.2.40. Die Voraussetzungen von Satz 2.2.39 lassen sich noch erheblich abschwächen. Erstens gilt die Aussage für jeden nichtdegenerierten kritischen Punkt x^* also nicht nur für lokale Minimalpunkte. Zweitens zeigt Satz 2.2.48, dass die Konvergenz der Folge (x^k) bereits impliziert, dass der Grenzpunt x^* ein nichtdegenerierter kritischer Punkt ist.

Die Konvergenzgeschwindigkeit überträgt sich aus Satz 2.2.39 natürlich auf das gedämpfte Newton-Verfahren, falls man mit einem $k_0 \in \mathbb{N}$ für alle $k \geq k_0$ nur $t_k = 1$ wählt. Die folgende Übung gibt eine natürliche Bedingung dafür an.

Übung 2.2.41. Für $f \in C^2(\mathbb{R}^n, \mathbb{R})$ liege x in einer genügend kleinen Umgebung eines nichtdegenerierten lokalen Minimalpunkts, und die Suchrichtung d werde mit dem gedämpften Newton-Verfahren per Armijo-Regel mit $t^0 = 1$ und $\sigma < \frac{1}{2}$ bestimmt. Dann gilt $\langle \nabla f(x), d \rangle < 0$ und dass die Armijo-Regel die Schrittweite $t_a = 1$ wählt.

Übung 2.2.42. Das ungedämpfte Newton-Verfahren liefert für die Funktion $q(x) = \frac{1}{2} x^T A x + b^T x$ mit $A = A^T \succ 0$ und $b \in \mathbb{R}^n$ von jedem Startpunkt $x^0 \in \mathbb{R}^n$ nach einem Schritt den globalen Minimalpunkt von q.

Übung 2.2.43. Für $f \in C^2(\mathbb{R}^n, \mathbb{R})$ sei eine Iterierte x^k mit $D^2 f(x^k) \succ 0$ gegeben. Die vom Newton-Verfahren erzeugte Suchrichtung d^k ist der eindeutige lokale Minimalpunkt der konvex-quadratischen Funktion

$$q(d) = f(x^k) + \langle \nabla f(x^k), d \rangle + \frac{1}{2} d^T D^2 f(x^k) d$$

Bemerkung. Für spezielle Funktionen, wie bei Kleinste-Quadrate-Problemen, also $f(x) = \frac{1}{2} \|r(x)\|_2^2$ mit einer glatten Funktion $r \colon \mathbb{R}^n \to \mathbb{R}^m$. Es ist

$$D^2 f(x) = \nabla r(x) Dr(x) + \sum_{j=1}^{m} r_j(x) D^2 r_j(x).$$

Als Modifikation, dem Gauß-Newton-Verfahren, wäre anstatt $A^k = D^2 f(x^k)$ in Zeile 4 $A^k = \nabla r(x^k) Dr(x^k)$ zu wählen. Dass die restlichen Summanden in der Darstellung von eine untergeordnete Rolle spielen, kann zum einen daran liegen, dass für $m \geq n$ überlicherweise ein Punkt x^* mit $r(x^*) = 0$ dass für $m \geq n$ approximiert wird, so dass die Werte $r_j(x^k)$ fast verschwinden, oder zum anderen daran, dass die Krümmung der Funktionen r_j an x^* vernachlässigbar sind, so dass sich die Matrizen $D^2 r_j(x^k)$ in der Nähe der Nullmatrix aufhalten.

Obwohl zum Aufstellen von A^k im Gauß-Newton-Verfahren also nur Ableitungsinformationen erster Ordnung (die Matrix $Dr(x^k)$) erforderlich sind, lässt sich unter bestimmten

Zusatzvoraussetzungen sogar quadratische Konvergenz zeigen. Zusätzlich sind die Suchrichtungen d^k im Gauß-Newton-Verfahren im Gegensatz zum allgemeinen Newton-Verfahren garantiert Abstiegsrichtungen (erster Ordnung), so dass eine Schrittweitensteuerung etwa per Armijo-Regel möglich ist.

Sollte die Jacobi-Matrix $Dr(x^k)$ nicht den vollen Rang besitzen oder zumindest schlecht konditioniert sein, so lässt das Gauß-Newton-Verfahren sich durch die Wahl $A^k = \nabla r(x^k)Dr(x^k) + \sigma^k E$ mit gewissen $\sigma^k > 0$ und der Einheitsmatrix E passender Dimension stabilisieren, was auf das Levenberg-Marquardt-Verfahren führt.

Superlineare Konvergenz

Falls im Newton-Verfahren x^0 zu weit von einem nichtdegenerierten Minimalpunkt entfernt liegt, ist $D^2 f(x^k)$ nicht notwendigerweise positiv definit und die Newton-Richtung entweder nicht definiert oder nicht notwendigerweise eine Abstiegsrichtung. Man versucht daher, das Newton-Verfahren zu globalisieren, d. h. Konvergenz im Sinne von Satz 2.2.9 gegen einen lokalen Minimalpunkt von jedem Startpunkt $x^0 \in \mathbb{R}^n$ aus zu erzwingen. Ein erster Ansatz dazu besteht darin, in Zeile 2 von Algorithmus 2.6 $A^0 = E$ zu wählen sowie in Zeile 7 (ähnlich wie im Levenberg-Marquardt-Verfahren)

$$A^{k+1} = D^2 f(x^{k+1}) + \sigma^{k+1} \cdot E$$

mit einem so großen Skalar σ^{k+1}, dass A^{k+1} positiv definit ist, und bei hinreichend großen k wieder $\sigma^k = 0$ (d. h. das Verfahren startet als Gradientenverfahren und geht nach endlich vielen Schritten in das gedämpfte Newton-Verfahren über). Ein Nachteil des Verfahrens besteht darin, dass die Bestimmung von σ_k sehr aufwendig sein kann: Man halbiert oder verdoppelt σ^k so lange, bis ein Test auf positive Definitheit von erfolgreich ist.

Im Folgenden werden wir Verfahren kennenlernen, die nicht nach endlich vielen Schritten, sondern nur asymptotisch in das gedämpfte Newton-Verfahren übergehen. Für diese lässt sich immerhin noch superlineare Konvergenz zeigen. Der entsprechende Konvergenzsatz erfordert einige Vorbereitungen.

Zunächst besitzt die Folge der Iterierten (x^k) nach Satz 2.2.9 einen Häufungspunkt, und jeder solche Häufungspunkt ist kritisch, sofern die Menge $f_{\leq}^{f(x^0)}$ beschränkt ist und gradientenbezogene Suchrichtungen sowie effiziente Schrittweiten benutzt werden. Die Gradientenbezogenheit der Suchrichtungen wird durch Satz 2.2.36 für gleichmäßig positiv definite und beschränkte (A^k) garantiert. Sei dazu

$$H^k := t^k \left(A^k\right)^{-1},$$

sodass $x^{k+1} = x^k - H^k \nabla f(x^k)$. Daraus folgt die Definition der superlinearen Konvergenz und diese ist äquivalent zu $\limsup_k \frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|} = 0$.

Lemma 2.2.46. Die Folge (x^k) sei nach obiger Vorschrift gebildet und gegen x^* konvergent. Ferner seien die Folgen $(\|H^k\|_2)$ und $(\|(H^k)^{-1}\|_2)$ beschränkt. Dann gilt

a) $\nabla f(x^*) = 0$

b) $\limsup_k \|x^{k+1} - x^*\|_2 / \|x^k - x^*\|_2 \leq \limsup_k \|E - H^k D^2 f(x^*)\|_2$

Lemma 2.2.47. Für zwei (n,n)-Matrizen A und B sei $L := \|E - AB\|_2 < 1$. Dann gilt

a) A und B sind nichtsingulär

b) $\|A\|_2 \leq (1+L) \cdot \|B^{-1}\|_2$

c) $\|A^{-1}\| \leq \frac{\|B\|_2}{(1-L)}$

Satz 2.2.48. Die Folge (x^k) sei nach obiger Vorschrift gebildet und gegen x^* konvergent. Ferner sei $L := \limsup_k \|E - H^k D^2 f(x^*)\|_2 < 1$. Dann gelten die folgenden Aussagen:

a) $D^2 f(x^*)$ ist nichtsingulär.

b) $\nabla f(x^*) = 0$

c) (x^k) konvergiert mindestens linear gegen x^*

d) Es gilt $L = 0$ genau im Fall von $\lim_k H^k = \left(D^2 f(x^*)\right)^{-1}$, und in diesem Fall konvergiert (x^k) superlinear gegen x^*

Nach Satz 2.2.48 sollte Algorithmus 2.6 also asymptotisch in das ungedämpfte Newton-Verfahren übergehen, um superlineare Konvergenz zu garantieren. Wegen

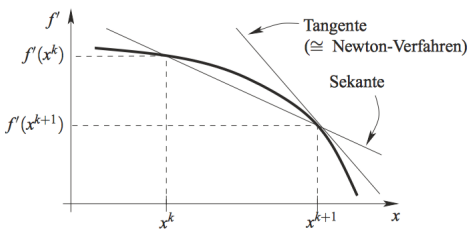
$$H^k = t^k \cdot \left(A^k\right)^{-1}$$

sind natürliche Bedingungen dafür $\lim_k t^k = 1$ und $\lim_k A^k = D^2 f(x^*)$. Das zu Beginn dieses Kapitels vorgeschlagene Verfahren erreicht dies mit hohem Aufwand bereits nach endlich vielen Schritten, ist in diesem Sinne also nicht effizient.

Quasi-Newton-Verfahren

Ein möglicher Ansatz Matrizen A^k zu finden mit $\lim_k A^k = D^2 f(x^*)$ besteht darin, zunächst das Sekantenverfahren zur Nullstellensuche einer Funktion von $\mathbb{R}$ nach $\mathbb{R}$ zu betrachten.

Abb. 2.8 Grundidee des Sekantenverfahrens



Hieraus erhält man für $n \geq 1$ di Gleichung

$$\nabla f(x^{k+1}) - \nabla f(x^k) = A^{k+1} \cdot \left(x^{k+1} - x^k\right)$$

die als Sekantengleichung oder Quasi-Newton-Bedingung an die (n,n)-Matrix A^{k+1} bezeichnet wird. Man zählt leicht nach, dass (2.13) n Gleichungen für die n^2 Einträge von A^{k+1} liefert. Selbst wenn man A^{k+1} als symmetrisch voraussetzt, sind noch immer $n(n+1)/2$ Einträge zu bestimmen. Aus diesem Grunde existieren viele Möglichkeiten, verschiedene Quasi-Newton-Verfahren anzugeben.

Die Grundidee der folgenden Verfahren besteht darin, die Matrix A^{k+1} nicht in jedem Iterationsschritt komplett neu zu berechnen, sondern sie als möglichst einfaches Update der Matrix A^k aus dem vorherigen Schritt aufzufassen. Als erfolgreicher Ansatz hat sich dabei erwiesen, mit $A^0 \succ 0$ zu starten und in Zeile 7 von Algorithmus 2.6 die Matrix A^{k+1} aus A^k durch Addition einer symmetrischen Matrix vom Rang eins oder zwei zu gewinnen:

$$A^{k+1} = A^k + \alpha_k (u^k)(u^k)^T + \beta_k (v^k)(v^k)^T$$

mit Skalaren $\alpha_k, \beta_k \in \mathbb{R}$ und Vektoren $u^k, v^k \in \mathbb{R}^n$, die so gewählt sind, dass A^{k+1} die Sekantengleichung erfüllt. Mit $s^k := x^{k+1} - x^k$ und $y^k := \nabla f(x^{k+1}) - \nabla f(x^k)$ heißt

$$A^{k+1} = A + \frac{yy^T}{y^T s} - \frac{Ass^T A}{s^T As}$$

BFGS-Update. Da man allerdings in Zeile 4 von Algorithmus 2.6 die Suchrichtung

$$d^k = -\left(A^k\right)^{-1} \nabla f(x^k)$$

wählt, wäre es günstig, die Matrix $(A^k)^{-1}$ explizit angeben zu können.

Übung 2.2.49 (Sherman-Morrison-Woodbury-Formel)**.**

a) Für eine nichtsinguläre (n,n)-Matrix A und Vektoren $b, c \in \mathbb{R}^n$ gelte, dass $A + bc^T$ genau dann nichtsingulär ist, wenn $1 + c^T A^{-1} b$ nicht verschwindet.

b) Es gilt die Sherman-Morrion-Woodbury-Formel für eine (n,n)-Matrix A und Vektoren $b, c \in \mathbb{R}^n$, wobei A und $A + bc^T$ nichtsingulär seien:

$$(A + bc^T)^{-1} = A^{-1} - \frac{A^{-1} bc^T A^{-1}}{1 + c^T A^{-1} b}$$

Übung 2.2.49 liefert eine Update-Formel für die inversen Matrizen $B := A^{-1}$ und $B^+ := (A^+)^{-1}$, nämlich

$$B^+_{BFGS} = B + \frac{ss^T}{s^T y} - \frac{Byy^T B}{y^T By} + rr^T$$

mit $r \coloneqq \sqrt{y^T B y} \cdot \left(\frac{s}{s^T y} - \frac{By}{y^T B y} \right)$. In Zeile 2 wählt man daher eine Matrix $B^0 \succ 0$ anstelle von A^0. In Zeile 4 setzt man

$$d^k = -B^k \cdot \nabla f(x^k)$$

und in Zeile 7 wählt man $B^{k+1} = B^{k+1}_{BFGS}$. Analog kann man auch das DFP-Update definieren, und zwar mit s und y vertauscht

$$D^+_{DFP} = B + \frac{ss^T}{s^T y} - \frac{Byy^T B}{y^T B y}$$

Dies unterscheidet sich vom BFGS-Update lediglich durch den Term rr^T. Die Einführung eines zusätzlichen Parameters $\theta \in \mathbb{R}$ liefert die Updates der Broyden-Familie

$$B^+_\theta = B^+_{DFP} + \theta \cdot rr^T$$

Offenbar gilt $B^+_0 = B^+_{DFP}$ und $B^+_1 = B^+_{BFGS}$. Außerdem gilt für die Wahl $\theta = \frac{s^T y}{s^T y - y^T B y}$:

$$B^+_{SR1} \coloneqq B^+_\theta = B + \frac{(s - By)(s - By)^T}{(s - By)^T y}$$

so dass die Update-Matrix nur den Rang 1 besitzt. Man spricht dann vom SR1-Update (SR1 = symmetric rank 1).

Übung 2.2.50. Sei $B^k \succ 0$ in einer Iteration eines Quasi-Newton-Verfahrens. Unter der Wahl von exakten Schrittweiten t^k_e gilt die Ungleichung

$$(y^k)^T s^k > 0$$

Lemma 2.2.51. Es sei $\theta \geq 0$ beliebig. Dann gilt unter den Bedingungen $B \succ 0$ und $s^T y > 0$ auch $B^+_\theta \succ 0$

Zur Division durch die Zahlen $s^T y$ und $y^T B y$ in den Update-Formeln lässt sich also anmerken, dass für mit $B \succ 0$ auch alle iterierten Matrizen B^k positiv definit sind, sofern $(s^k)^T y^k > 0$ positiv ist. Insbesondere gilt dann $y_k \neq 0$ und $(y^k)^T B^k y^k > 0$. Angemerkt sei, dass die Voraussetzung θ aus Lemma 2.2.51 zwar für den SR1-Update nicht garantiert ist, er in der Praxis aber dennoch häufig gute Ergebnisse liefert.

Wählt man exakte Schrittweiten t^k_e, so hängt nur der Koeffizient des Vektors r^k von θ ab, und somit ist die Suchrichtung für jedes $\theta \in \mathbb{R}$ identisch. Weil man aber entlang dieser Richtung exakt eindimensional minimiert, liefern alle Verfahren der Broyden-Familie identische Lösungsfolgen (x^k).

Dieses überraschende Ergebnis wird dadurch relativiert, dass man in der Praxis meist nicht exakt, sondern inexakt eindimensional minimiert, etwa per Armijo-Schrittweitensteuerung mit Backtracking Line Search. Während zum Beispiel das DFPUpdate dazu tendiert,

schlecht konditionierte Matrizen B^k zu erzeugen, verhält sich das BFGS-Update für Probleme mittlerer Größe numerisch oft sehr robust.

Leider lässt sich nicht zeigen, dass die Matrizen B^k stets gegen $(D2f(x^*))^{-1}$ streben, wie es zur Anwendung von Satz 2.2.48 zur superlinearen Konvergenz wünschenswert wäre. Mit einer recht technischen Verallgemeinerung von Satz 2.2.48 lässt sich für $\lim_k t^k = 1$ trotzdem die superlineare Konvergenz der BFGS- und DFP-Verfahren nachweisen, falls $(B^k)^{-1}$ und $D^2f(x^*)$ wenigstens entlang der Suchrichtungen d^k asymptotisch gleich sind.

Konjugierte Richtungen

Für viele Anwendungsprobleme ist die Anzahl der Variablen so hoch, dass sich zwar Vektoren der Länge n wie x^k und d^k noch gut abspeichern lassen, die Speicherung der $n(n+1)/2$ Einträge von Matrizen wie B^k aber zu einem Platzproblem führt.

Definition 2.2.52 (Konjugiertheit bezüglich einer positiv definiten Matrix)**.** Es sei A eine (n,n)-Matrix mit $A = A^T \succ 0$. Zwei Vektoren $v, w \in \mathbb{R}^n$ heißen **konjugiert bezüglich** A, falls $\langle v, w\rangle = 0$ gilt.

Im Folgenden betrachten wir das allgemeine Abstiegsverfahren $x^{k+1} = x^k + t_e^k d^k$ mit exakten Schrittweiten t_e^k und Abstiegsrichtungen erster Ordnung d_k für die konvex-quadratische Funktion

$$q(x) = \frac{1}{2}x^T Ax + b^T x$$

mit $A = A^T \succ 0$ und $b \in \mathbb{R}^n$.

Übung 2.2.53. Für $k \in \mathbb{N}$ seien $d^0, \ldots, d^k$ paarweise konjugiert bezüglich A und sämtlich ungleich null. Es gilt

a) Die Vektoren $d^0, \ldots, d^k$ sind linear unabhängig. insbesondere gilt $k < n$

b) Für $k = n-1$ gilt

$$A^{-1} = \sum_{l=0}^{n-1} \frac{(d^l)(d^l)^T}{(d^l)^T A (d^l)}$$

Lemma 2.2.54. Für $k \in \mathbb{N}$ seien $d^0, \ldots, d^k$ paarweise konjugiert bezüglich A. Dann gilt

$$\forall 0 \le l \le k: \quad \langle \nabla q(x^{k+1}), d^l\rangle = 0.$$

Satz 2.2.55. Die Vektoren $d^0, \ldots, d^{n-1}$ seien paarweise konjugiert bezüglich A und sämtlich ungleich null. Dann ist x^n der globale Minimalpunkt von q.

Satz 2.2.55 besagt, dass ein Abstiegsverfahren für die konvex-quadratische Funktion q bei exakter Schrittweitensteuerung und paarweise konjugierten Suchrichtungen nach höchstens n Schritten den globalen Minimalpunkt von q findet. Da für ein Abstiegsverfahren wegen

$f(x^{k+1}) < f(x^k)$ stets $t_e^k \cdot d^k = x^{k+1} - x^k \neq 0$ gilt, kann insbesondere keiner der Vektoren d^k verschwinden.

Im nächsten Schritt suchen wir nach Möglichkeiten, konjugierte Suchrichtungen explizit zu erzeugen. Der folgende Satz besagt, dass man konjugierte Richtungen zum Beispiel aus den Quasi-Newton-Verfahren der Broyden-Familie erhält.

Satz 2.2.56. Für $\theta \geq 0$ werde Algorithmus 2.6 mit $t^k = t_e^k$ und $B^{k+1} = B_\theta^{k+1}$ auf $q() = \frac{1}{2}x^T Ax + b^T x$ mit $A = A^T \succ 0$ angewendet, und für ein $k \in \mathbb{N}$ seien die Iterierten $x^0, \ldots, x^k$ paarweise verschieden. Dann sind die Richtungen $d^0, \ldots, d^{k-1}$ paarweise konjugiert bezüglich A und sämtlich von null verschieden.

Bei Wahl exakter Schrittweiten minimieren die Quasi-Newton-Verfahren der Broyden-Familie konvex-quadratische Funktionen also in höchstens n Schritten. Für eine beliebige C^2-Funktion f lässt sich das dahingehend interpretieren, dass sie die lokale quadratische Approximation an f in n Schritten minimieren, im Hinblick auf Übung 2.2.43 also einen Schritt des Newton-Verfahrens simulieren. Unter geeigneten Voraussetzungen und mit Neustarts nach jeweils n Schritten konvergieren sie daher „n-Schritt-quadratisch“.

Konjugierte-Gradienten-Verfahren

Wir betrachten weiterhin das Abstiegsverfahren $x^{k+1} = x^k + t_e^k d^k$ mit exakten Schrittweiten t_e^k und Abstiegsrichtungen erster Ordnung d_k für die konvex-quadratische Funktion

$$q(x) = \frac{1}{2}x^T Ax + b^T x$$

mit $A = A^T \succ 0$ und $b \in \mathbb{R}^n$. Gesucht sind Möglichkeiten, konjugierte Suchrichtungen (d^k) zu erzeugen.

Die Grundidee wird im Folgenden sein, die Suchrichtungen d^k rekursiv zu wählen, nämlich als Kombination des aktuellen negativen Gradienten $-\nabla q(x^k)$ und der letzten Suchrichtung d^{k-1} mit Hilfe eines noch zu bestimmenden „Gewichts“ $\alpha^k \in \mathbb{R}$ zu

$$d^k = -\nabla q(x^k) + \alpha_k \cdot d^{k-1}, k = 1, 2, \ldots$$

Zu Beginn dieser Rekursion setzen wir $d^0 = -\nabla q(x^0)$. Das folgende Lemma wird zur Bestimmung der Werte α_k wesentlich sein.

Lemma 2.2.57. Es seien $d^0, \ldots, d^{k-1}$ paarweise konjugiert bezüglich A und $x^1, \ldots, x^k$ schon generiert mit $x^l \neq x^{l-1}$ für $1 \leq l \leq k$. Dann ist d^k genau dann konjugiert zu einem d^l mit $0 \leq l \leq k-1$, wenn folgendes erfüllt ist

$$\langle \nabla q(x^{l+1}) - \nabla q(x^l), d^k \rangle = 0$$

Satz 2.2.58. Unter den Voraussetzungen von Lemma 2.2.58 ist die Richtung $d^k = -\nabla q(x^k) + \alpha_k \cdot d^{k-1}$ genau für

$$\alpha_k = \frac{\|\nabla q(x^k)\|_2^2}{\|\nabla q(x^{k-1})\|_2^2}$$

konjugiert zu den Vektoren $d^0, \ldots, d^{k-1}$.

Satz 2.2.58 motiviert den Algorithmus 2.7, da er für $f(x) = q(x) = \frac{1}{2}x^T Ax + b^T x$ mit $A = A^T \succ 0$ nach höchstens n Schritten den globalen Minimalpunkt liefert. Man benutzt dieses Verfahren zum Beispiel zur Lösung hochdimensionaler linearer Gleichungssysteme $Ax = b$ durch den Kleinste-Quadrate-Ansatz, also per Minimierung von $\|r(x)\|_2^2$ mit dem Residuum $r(x) = Ax - b$.

Wegen Rundungsfehlern bricht das Verfahren aber selten tatsächlich nach n Schritten ab, so dass auch seine Konvergenzgeschwindigkeit untersucht wurde. Es stellt sich heraus, dass sie von der Wurzel der Konditionszahl (also dem Quotienten aus größtem und kleinstem Eigenwert) der Matrix $A^T A$ abhängt. Es bietet sich daher an, das Gleichungssystem $Ax = b$ zunächst so äquivalent umzuformen, dass diese Konditionszahl sinkt. Dies ist als Präkonditionierung bekannt.

Algorithmus 2.7: CG-Verfahren von Fletcher-Reeves

Input : C^1-Optimierungsproblem P
Output : Approximation $\bar{x}$ eines kritischen Punkts von f (falls das Verfahren terminiert [25, Th. 5.7])

begin
  Wähle einen Startpunkt x^0, eine Toleranz $\varepsilon > 0$ und setze $d^0 = -\nabla f(x^0)$ sowie $k = 0$.
  while $\|\nabla f(x^k)\| > \varepsilon$ **do**
    Setze $x^{k+1} = x^k + t_e^k d^k$.
    Setze $d^{k+1} = -\nabla f(x^{k+1}) + \left(\|\nabla f(x^{k+1})\|_2^2 / \|\nabla f(x^k)\|_2^2\right) \cdot d^k$.
    Ersetze k durch $k + 1$.
  end
  Setze $\bar{x} = x^k$.
end

Bemerkung 2.2.59. Der Kleinste-Quadrate-Ansatz per CG-Verfahren zur Lösung linearer Gleichungssysteme $Ax = b$ lässt sich auch auf überbestimmte Gleichungssysteme anwenden, die keine Lösung besitzen.

Entscheidend für die Einsetzbarkeit von Algorithmus 2.7 ist, dass für $f = q$ nirgends explizit die Matrix A eingeht, aber trotzdem bezüglich$A$ konjugierte Suchrichtungen erzeugt werden. Man kann das Verfahren also auch für beliebige C^1-Funktionen formulieren, wobei die Armijo-Schrittweite für gewöhnlich verwendet wird. Unter geeigneten Voraussetzungen erhält man wieder, dass n CG-Schritte einen Newton-Schritt simulieren, also „n-Schrittquadratische Konvergenz“.

Trust-Region-Verfahren

Im Gegensatz zu klassischen Suchrichtungsverfahren wählen Trust-Region-Verfahren erst den Suchradius t und dann die Suchrichtung d. Dazu benutzt man in Iteration k des allgemeinen Abstiegsverfahrens aus Algorithmus 2.3 wie folgt ein quadratisches Modell für f.

Nach dem Satz von Taylor (Satz 2.1.30b) gilt für $f \in C^2(\mathbb{R}^n, \mathbb{R})$

$$f(x^k + d) \approx f(x^k) + \langle \nabla f(x^k), d \rangle + \frac{1}{2} d^T D^2 f(x^k) d$$

Mit $c^k := f(x^k)$, $b^k = \nabla f(x^k)$ und einer symmetrischen Matrix A^k (zum Beispiel, aber nicht notwendigerweise, $A^k = D^2 f(x^k)$) nennt man die Funktion

$$m^k(d) := c^k + \langle b^k, d \rangle + \frac{1}{2} d^T A^k d$$

ein lokales quadratisches Modell für f um x^k.

Man betrachtet daher m^k nur für $\|d\|_2 \leq t^k$ und bestimmt man einen optimalen Punkt d^k des Trust-Region-Hilfsproblems

$$TR^k: \quad \min_{d \in \mathbb{R}^n} m^k(d) \text{ s.t. } \|d\|_2 \leq t^k$$

Der folgende Quotient misst dabei die Güte der Approximation

$$r^k := \frac{f(x^k) - f(x^k + d^k)}{m^k(0) - m^k(d^k)}$$

- Ein Wert $r^k < 0$ impliziert daher $f(x^k + d^k) > f(x^k)$, d. h., $x^{k+1} = x^k + d^k$ würde einen Anstieg im Zielfunktionswert liefern. Folglich ist die Trust Region zu groß, und ihr Radius t^k muss verkleinert werden.
- Liegt andererseits r^k nahe bei eins, dann beschreibt das lokale Modell die Funktion f sehr gut; man setzt $x^{k+1} = x^k + d^k$ und vergrößert in der nächsten Iteration probeweise den Trust-Region-Radius t^k.
- Insbesondere für $r^k \geq \frac{1}{4}$ wird t^k dort nicht verkleinert, und der Schritt wird angenommen, für $r^k < 0$ wird t^k verkleinert, und der Schritt wird abgelehnt, und für $r^k \in [0, \frac{1}{4})$ wird t^k verkleinert, und der Schritt wird dann abgelehnt, wenn $r^k \leq \eta$ gilt.

Ein entscheidender Vorteil von Trust-Region-Verfahren gegenüber Variable-Metrik-Verfahren besteht allerdings darin, dass die Matrizen A^k nicht positiv definit zu sein brauchen. Insbesondere für $A^k \equiv 0$ erhält man als „Trust-Region-Gradientenverfahren“ lediglich ein übliches Gradientenverfahren mit einer speziellen Schrittweitensteuerung. Von einem solchen Verfahren ist wegen Satz 2.2.23 keine schnelle Konvergenz zu erwarten.

Allerdings auch hier eine inexakte Lösung von TR^k, um globale Konvergenz zu gewährleisten. Eine Möglichkeit dafür besteht darin, die zulässige Menge von TR k stark zu verkleinern und beispielsweise nur nichtnegative Vielfache der beim Verfahren gefundenen Suchrichtung zuzulassen:

$$TR_C^k: \quad \min_{s\in\mathbb{R}} m^k\left(-\frac{s\cdot t^k}{\|\nabla f(x^k)\|_2}\cdot\nabla f(x^k)\right) \text{ s.t. } 0\le s\le 1$$

Die Lösung zu diesem Problem ergibt sich $d_C^k := -\frac{s^k t^k}{\|\nabla f(x^k)\|_2}\nabla f(x^k)$ mit

$$s^k := \begin{cases} 1, & \text{falls } Df(x^k)A^k\nabla f(x^k)\le 0 \\ \min\left\{\frac{\|\nabla f(^k)\|_2^3}{t^k Df(x^k)A^k\nabla f(x^k)}, 1\right\}, & \text{sonst} \end{cases}$$

Algorithmus 2.8: Trust-Region-Verfahren

Input : C^1-Optimierungsproblem P
Output : Approximation $\bar{x}$ eines kritischen Punkts von f (falls das Verfahren terminiert; Satz 2.2.63)

begin
  Wähle einen Startpunkt x^0, eine Matrix $A^0 = (A^0)^\intercal$, eine Toleranz $\varepsilon > 0$, einen Maximalradius $\check{t} > 0$, einen Startradius $t^0 \in (0,\check{t})$, einen Parameter $\eta \in [0, 1/4)$ und setze $k = 0$.
  while $\|\nabla f(x^k)\|_2 > \varepsilon$ **do**
    Berechne einen (inexakten) Optimalpunkt d^k von TR^k und setze

$$r^k = \frac{f(x^k) - f(x^k + d^k)}{m^k(0) - m^k(d^k)}.$$

    if $r^k < \frac{1}{4}$ **then**
      Setze $t^{k+1} = \frac{1}{4}\|d^k\|_2$.
    else
      if $r^k > \frac{3}{4}$ **and** $\|d^k\|_2 = t^k$ **then**
        Setze $t^{k+1} = \min\{2t^k, \check{t}\}$.
      else
        Setze $t^{k+1} = t^k$.
      end
    end
    if $r^k > \eta$ **then**
      Setze $x^{k+1} = x^k + d^k$.
    else
      Setze $x^{k+1} = x^k$.
    end
    Wähle $A^{k+1} = (A^{k+1})^\intercal$.
    Ersetze k durch $k + 1$.
  end
  Setze $\bar{x} = x^k$.
end

Definition 2.2.60 (Cauchy-Punkt)**.** Der Punkt $x_C^{k+1} = x^k + d_C^k$ heißt **Cauchy-Punkt** zu x^k und t^k.

Übung 2.2.61. Der Vektor $d^k = d_C^k$ erfüllt die folgende Ungleichung mit $c = 0.5$

$$m^k(0) - m^k(d^k) \geq c \cdot \|\nabla f(x^k)\|_2 \cdot \min\left\{t^k, \frac{\|\nabla f(x^k)\|_2}{\|A^k\|_2}\right\} \tag{2.24}$$

Bemerkung 2.2.62. Die exakte Lösung d_e^k von TR^k erfüllt wegen der Zulässigkeit von d_C^k für TR^k die Ungleichung $m^k(d_e^k) \leq m^k(d_C^k)$ und damit nach Übung 2.2.61 ebenfalls (2.24) mit $c = 0.5$.

Satz 2.2.63. Die Menge $f_{\leq}^{f(x^0)}$ sei beschränkt, die Funktion ∇f sei Lipschitz-stetig auf $\operatorname{conv}(f_{\leq}^{f(x^0)})$, die Folge $(\|A^k\|_2)$ sei beschränkt, und die Folge (d^k) der inexakten Lösungen von TR^k erfülle (2.24) mit $c > 0$. Dann gilt in Algorithmus 2.8:

a) Für $\eta = 0$ ist $\liminf_k \|\nabla f(x^k)\|_2 = 0$ (d.h. (x^k) besitzt einen Häufungspunkt x^* mit $\nabla f(x^*) = 0$).

b) Für $\eta \in (0, 1)$ ist $\lim_k \nabla f(x^k) = 0$ (d.h. alle Häufungspunkte von (x^k) sind kritisch).

Nach Übung 2.2.61, Bemerkung 2.2.62 und Satz 2.2.63 liefern sowohl die inexakten Lösungen d_C^k als auch die exakten Lösungen d_e^k von TR^k globale Konvergenz. Während die exakte Lösung d_e^k wie erwähnt schwer berechenbar sein kann, ist das Ausweichen auf die inexakte Lösung d_C^k selten ratsam, da die Matrix A^k lediglich die Länge von d_C^k beeinflusst und man so im Wesentlichen nach wie vor das Gradientenverfahren erhält.

Dogleg-Methode

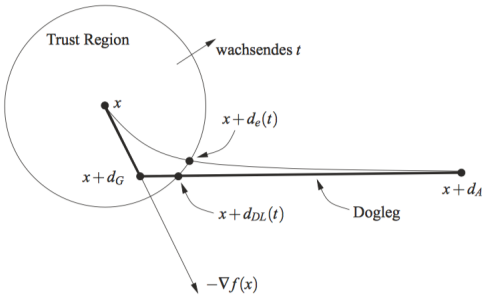
Sei A positiv definit und $d_A := -A^{-1}\nabla f(x)$. Die Dogleg-Methode approximiert diese Kurve durch einen Polygonzug von x nach $x + d_A$ mit zwei Segmenten, wobei als Zwischenpunkt $x + d_G$ mit dem exakten Minimalpunkt d_G von m entlang $-\nabla f(x)$ gewählt wird, den man zu

$$d_G = -\frac{\|\nabla f(x)\|_2^2}{Df(x)A\nabla f(x)}\nabla f(x)$$

berechnet. Formal lautet der Polygonzug damit $\{x + \hat{d}(s) \mid s \in [0, 2]\}$ mit

$$\hat{d}(s) = \begin{cases} s \cdot d_G, & 0 \leq s \leq 1 \\ d_G + (s-1)(d_A - d_G), & 1 \leq s \leq 2 \end{cases}$$

Abb. 2.10 Approximation der Kurve $\{x + d_e(t) | \, t \geq 0\}$ per Dogleg



Übung 2.2.64. Es gilt für $A = A^T \succ 0$:

a) $\|\hat{d}(s)\|_2$ ist monoton wachsend in s.

b) $m(\hat{d}(s))$ ist monoton fallend in s.

Somit ergibt die Dogleg-Moethode die inexakte Lösung $x + d_{DL}(t)$ mit

$$d_{DL} = \begin{cases} d_A, & \text{falls } \|d_A\|_2 < t \\ \|s \cdot d_G\|_2 = t & \text{falls } t < \|d_G\|_2 \\ \|d_G + (s-1)(d_A - d_G)\|_2 = t, & \text{sonst} \end{cases}$$

Minimierung auf einem zweidimensionalen Teilraum

Die inexakte Lösung von TR durch die Dogleg-Methode kann verbessert werden, indem man TR nicht auf den eindimensionalen Polygonzug einschränkt, sondern auf den zweidimensionalen Teilraum, der von d G und d A aufgespannt wird. In diesem Raum liegen insbesondere alle Punkte des Polygonzugs. Man erhält das Hilfsproblem

$$\min_{d \in \mathbb{R}^n} m(d) \text{ s.t. } \|d\|_2 \leq t, \; d \in \text{bild}(\nabla f(x), A^{-1}\nabla f(x))$$

Ein Hauptvorteil dieses Ansatzes besteht darin, dass er sich im Gegensatz zur Dogleg-Methode sinnvoll auf indefinite Matrizen A erweitern lässt. Für Details sei auf [25] verwiesen.

II. Restringierte Probleme

Wir betrachten im folgenden Problem

$$P: \quad \min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } g_i(x) \leq 0, i \in I,\ h_j(x) = 0, j \in J$$

Mit von vektorwertigen Funktionen lässt sich die Menge M der für P zulässigen Punkte schreiben als

$$M = \left\{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\right\}$$

Topologische Eigenschaften

In Übung 1.2.11 haben wir bereits gesehen, dass M unter der Stetigkeitsvoraussetzung an die Funktionen g und h eine abgeschlossene Menge ist.

Definition 3.1.1 (Aktive-Index-Menge)**.** Zu $\overline{x} \in M$ heißt

$$I_0(\overline{x}) = \left\{i \in I \mid g_i(\overline{x}) = 0\right\}$$

Menge der aktiven Indizes oder auch **Aktive-Index-Menge**.

Satz 3.1.3. Für jedes $\overline{x} \in M$ existiert eine Umgebung U von $\overline{x}$ mit

$$U \cap M = U \cap \left\{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i \in I_0(\overline{x}), h_j(x) = 0, j \in J\right\}$$

Definition 3.1.4 (Zulässge Abstiegsrichtung)**.** Gegeben sei das Problem

$$P: \quad \min f(x) \text{ s.t. } x \in M$$

mit (nicht notwendigerweise in funktionaler Beschreibung vorliegender) zulässiger Menge $M \subseteq \mathbb{R}^n$. Dann heißt ein Vektor $d \in \mathbb{R}^n$ **zulässige Abstiegsrichtung** für P in $\overline{x} \in M$, falls folgendes gilt

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

Übung 3.1.5. Für das Problem P aus Definition 3.1.4 sei $\overline{x}$ ein lokaler Minimalpunkt. Dann existiert keine zulässige Abstiegsrichtung für P in $\overline{x}$.

Übung 3.1.6. Gegeben sei das Problem

$$P: \quad \min f(x) \text{ s.t. } g_i(x) \leq 0, i \in I, h_j(x) = 0, j \in J$$

Ein Vektor $d \in \mathbb{R}^n$ ist genau dann zulässige Abstiegsrichtung für P in $\overline{x} \in M$, wenn folgendes gilt

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

Definition 3.1.7 (Äußerer Linearisierungskegel)**.** Für $\overline{x} \in \mathbb{R}^n$ heißt

$$L_{\leq}(\overline{x}, M) = \left\{ d \in \mathbb{R}^n \mid \langle \nabla g_i(\overline{x}), d \rangle \leq 0, i \in I_0(\overline{x}) \right\}$$

äußerer Linearisierungksegel an M in $\overline{x}$.

Definition. Eine Menge $A \subseteq \mathbb{R}^n$ wird als Kegel bezeichnet, wenn

$$\forall a \in A, \lambda > 0 : \quad \forall \cdot a \in A$$

Übung 3.1.8. Am Punkt $\overline{x} \in M$ seien die Funktionen g_i, $i \in I_0(\overline{x})$, differenzierbar, dann ist $L_{\leq}(\overline{x}, M)$ ein konvexer Kegel.

Die funktionale Beschreibung einer zulässigen Menge kann so ungeschickt sein kann, dass ein äußerer Linearisierungskegel die lokale Struktur der Menge nicht notwendigerweise gut wiedergibt.

Definition 3.1.11 (Innerer Linearisierungskegel)**.** Für $\overline{x} \in \mathbb{R}^n$ heißt

$$L_{<}(\overline{x}, M) = \left\{ d \in \mathbb{R}^n \mid \langle \nabla g_i(\overline{x}), d \rangle < 0, i \in I_0(\overline{x}) \right\}$$

innerer Linearisierungskegel an M in $\overline{x}$.

Definition 3.1.12 (Nichtdegenerierte funktionale Beschreibung einer Menge)**.** Die funktionale Beschreibung von M heißt an $\overline{x}$ **nichtdegeneriert**, wenn $\operatorname{cl} L_{<}(\overline{x}, M) = L_{\leq}(\overline{x}, M)$ gilt. Ansonsten heißt sie **degeneriert**. Diese Gleichheit ist auch als Cottle-Bedingung (Cottle constraint qualification) bekannt.

Satz 3.1.15. Die funktionale Beschreibung von M ist an $\overline{x}$ genau dann nichtdegeneriert, wenn $L_{<}(\overline{x}, M) \neq \emptyset$ gilt.

Es gibt zudem Fälle, in denen schon die Geometrie der zulässigen Menge so ungünstig ist, dass keine funktionale Beschreibung die gewünschte „gute“ Approximation erster Ordnung liefert.

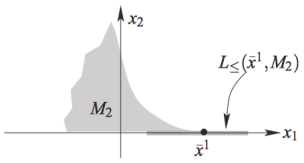


Abb. 3.4 Äußerer Linearisierungskegel in Beispiel 3.1.16

Definition 3.1.17 (Innerer und äußerer Tangentialkegel)**.** Es seien $\overline{x} \in \mathbb{R}^n$ und $M \subseteq \mathbb{R}^n$. Eine Richtung $\overline{d} \in \mathbb{R}^n$ liegt im

a) **inneren Tangentialkegel** $\Gamma(\overline{x}, M)$ an M in $\overline{x}$, fall ein $\hat{t} > 0$ und eine Umgebung von $\overline{d}$ existieren mit

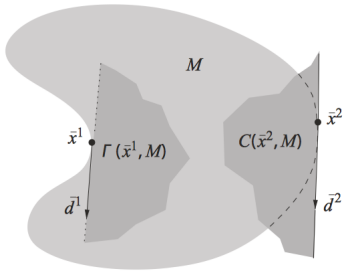
$$\forall t \in (0, \hat{t}), d \in D: \quad \overline{x} + td \in M,$$

b) **äußeren Tangentialkegel** $C(\overline{x}, M)$ an M in $\overline{x}$, falls Folgen (t^k) und (d^k) existieren mit

$$t^k \searrow 0, d^k \to \overline{d}, \ \forall k \in \mathbb{N}: \quad \overline{x} + t^k d^k \in M$$

Der äußere Tangentialkegel ist im Gegensatz zum äußeren Linearisierungskegel nicht notwendigerweise konvex.
Abb. 3.5 zeigt Beispiele für innere und äußere Tangentialkegel sowie Vektoren $\overline{d}^1 \notin \Gamma(\overline{x}^1, M)$ und $\overline{d}^2 \in C(\overline{x}^2, M)$. Würde man in Definition 3.1.17 keine variablen Richtungen zulassen, so resultierte dies hingegen in $\overline{d}^1 \in \Gamma(\overline{x}^1, M)$ und $\overline{d}^2 \notin C(\overline{x}^2, M)$.

Abb. 3.5 Innerer und äußerer Tangentialkegel



Lemma 3.1.18. Es seien $\overline{x} \in \mathbb{R}^n$ und $M \subseteq \mathbb{R}^n$. Dann gilt:

a) $\Gamma(\overline{x}, M) \subseteq C(\overline{x}, M)$.

b) $\Gamma(\overline{x}, M)^C = C(\overline{x}, M^c)$.

c) $\Gamma(\overline{x}, M)$ ist ein offener und $C(\overline{x}, M)$ ein abgeschlossener Kegel.

Definition 3.1.19 (Nichtdegenerierte Geometrie einer Menge)**.** Die Geometrie von M heißt an $\overline{x}$ **nichtdegeneriert**, wenn $\operatorname{cl}\Gamma(\overline{x}, M) = C(\overline{x}, M)$ gilt. Ansonsten heißt sie **degeneriert**.

Satz 3.1.24. Für alle $\overline{x} \in M$ gilt die Inklusions

$$L_<(\overline{x}, M) \subseteq \Gamma(\overline{x}, M) \subseteq C(\overline{x}, M) \subseteq L_\leq(\overline{x}, M).$$

Korollar 3.1.26. Die funktionale Beschreibung der Menge M sei an $\overline{x}$ nichtdegeneriert. Dann ist auch die Geometrie von M an $\overline{x}$ nichtdegeneriert.

Optimalitätsbedingungen

Stationarität

Da wir dabei nur die Geometrie von M benutzen wollen, definieren wir Stationarität mit Hilfe einer geometrischen Approximation erster Ordnung, nämlich des äußeren Tangentialkegels.

Definition 3.2.1 (Stationärer Punkt - restringierter Fall)**.** Die Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ sei an $\overline{x} \in M$ differenzierbar. Dann heißt $\overline{x}$ stationärer Punkt von P, falls $\langle \nabla f(\overline{x}), d\rangle \geq 0$ für jede Richtung $d \in C(\overline{x}, M)$ gilt.

Satz 3.2.2. Die Funktion $f\colon \mathbb{R} \to \mathbb{R}$ sei an einem lokalen Minimalpunkt $\overline{x}$ von P differenzierbar. Dann ist $\overline{x}$ stationärer Punkt im Sinne von Definition 3.2.1.

Hier nutzen wir aus, dass differenzierbare Funktionen auch einseitige richtungsdifferenzierbar im Sinne von Hadamard sind:

$$\langle \nabla f(x), d\rangle = \lim_k \frac{f(x + t^k d^k) - f(x)}{t^k}$$

Constraint Qualifications

Diese Beobachtungen führen zur Definition von zwei Regularitätsbedingungen (constraint qualifications).

Definition 3.2.3 (Abadie- und Mangasarian-Fromowitz-Bedingung für $J = \emptyset$)**.** An $\overline{x} \in M$ gilt

a) die **Abadie-Bedingung** (AB) für $J = \emptyset$, falls folgendes erfüllt ist

$$C(\overline{x}, M) = L_{\leq}(\overline{x}, M)$$

b) die **Mangasarian-Fromowitz-Bedingung** (MFB) für $J = \emptyset$, falls folgendes gilt

$$L_{<}(\overline{x}, M) \neq \emptyset$$

Nach Definition des inneren Linearisierungskegels ist die MFB an einem Punkt $\overline{x}$ genau dann erfüllt, wenn eine Richtung $d \in \mathbb{R}$ mit folgendem existiert

$$\langle \nabla g_i(\overline{x}), d\rangle < 0,\ i \in I_0(\overline{x})$$

Die MFB kann an einem Punkt bloß deshalb verletzt sein kann, weil die dort geometrisch nichtdegenerierte zulässige Menge degeneriert funktional beschrieben ist.

Korollar 3.2.4. An einem lokalen Minimalpunkt $\overline{x}$ von P seien f und die Funktionen g_i, $i \in I_0(\overline{x})$ differenzierbar.

a) Dann ist das folgende System mit keinem $d \in \mathbb{R}^n$ lösbar:

$$\langle \nabla f(\overline{x}), d\rangle < 0,\ \langle \nabla g_i(\overline{x}), d\rangle < 0 \quad i \in I_0(\overline{x}) \tag{3.1}$$

b) Falls an $\overline{x}$ die AB gilt, dann ist sogar das folgende System mit keinem $d \in \mathbb{R}^n$ lösbar:

$$\langle \nabla f(\overline{x}), d\rangle < 0,\ \langle \nabla g_i(\overline{x}), d\rangle \leq 0 \quad i \in I_0(\overline{x}) \tag{3.2}$$

Übung 3.2.5. Zu $\overline{x} \in M$ jede Richtung d mit (3.1) ist eine zulässige Abstiegsrichtung in $\overline{x}$ im Sinne von Definition 3.1.4, und zwar sowohl für P als auch für das linearisierte Problem $P_{lin}(\overline{x})$.

Übung 3.2.6. Zu $\overline{x} \in M$ ist jede Richtung d mit (3.2) eine zulässige Abstiegsrichtung für $P_{lin}(\overline{x})$ in $\overline{x}$ im Sinne von Definition 3.1.4 ist.

Stationarität schließt sogar die Existenz gewisser unzulässiger Abstiegsrichtungen aus. Da die notwendige Bedingung aus Korollar 3.2.4a bei verletzter MFB trivialerweise erfüllt sein kann, werden wir im Folgenden versuchen, die Bedingung aus Korollar 3.2.4b algorithmisch zu verwerten. Dazu ist es zunächst erforderlich, an einem Punkt $\overline{x} \in M$ algorithmisch nachprüfbare hinreichende Bedingungen für die Gültigkeit der abstrakt formulierten AB zu kennen.

Satz 3.2.8. An jedem $\overline{x} \in M$ impliziert die MFB die AB.

Übung 3.2.9. Die Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ sei an $\overline{x}$ differenzierbar mit $\nabla f(\overline{x}) \neq 0$, dann

$$C\left(\overline{x}, f_{\leq}^{f(\overline{x})}\right) = \left\{d \in \mathbb{R}^n \mid \langle \nabla f(\overline{x}), d\rangle \leq 0\right\}$$

Neben Satz 3.2.8 motiviert sich eine andere wichtige hinreichende Bedingung für die Gültigkeit der AB aus der Tatsache, dass die AB eine Linearisierungseigenschaft der zulässigen Menge fordert, die dann erfüllt sein sollte, wenn die Menge ohnehin schon durch endlich viele lineare Ungleichungen beschrieben ist. In diesem Fall nennt man M polyedrisch.

Beispiel 3.2.10. Für alle $1 \leq i \leq p$ sei

$$g_i(x) = a_i^T x + b_i$$

mit $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$. In Matrix-Vektor-Schreibweise ist $g(x) \leq 0$ dann gleichbedeutend mit der aus der linearen Optimierung bekannten Restriktion $Ax + b \leq 0$. Hier ist AB automatisch erfüllt.

Falls also $\overline{x} \in M$ ein lokaler Minimalpunkt einer dort differenzierbaren Funktion f über einer so definierten Menge M ist, dann lässt sich laut Korollar 3.2.4b das System

$$\langle \nabla f(\overline{x}), d\rangle < 0,\ \langle a_i, d\rangle \leq 0,\ i \in I_0(\overline{x})$$

mit keinem $d \in \mathbb{R}^n$ lösen.

Übung 3.2.12. In der Beschreibung der Menge $M = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i \in I\}$ seien alle Funktionen g_i, $i \in I$, konkav auf $\mathbb{R}^n$. Zeigen Sie, dass dann die AB an jedem Punkt von M gilt.

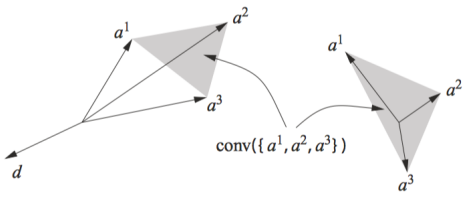
Alternativsätze

Geometrisch bedeutet die Bedingung in Korollar 3.2.4a

$$\langle a^k, d\rangle < 0 \quad 1 \leq k \leq r,$$

dass es keinen Vektor d gibt, der gleichzeitig mit allen Vektoren $a_1, \ldots, a_r$ einen stumpfen Winkel bildet.

Abb. 3.6 Stumpfe Winkel und konvexe Hüllen



In Abb. 3.6 sind rechts drei Vektoren eingezeichnet, für die dies der Fall ist, während links ein Vektor d existiert, der mit allen drei Vektoren gleichzeitig einen stumpfen Winkel bildet. Ebenfalls eingezeichnet ist die konvexe Hülle dieser drei Vektoren. Allgemein besteht die konvexe Hülle conv(A) einer Menge $A \subseteq \mathbb{R}^n$ aus der Menge aller Konvexkombinationen von Elementen in A. In Abb. 3.6 beobachtet man, dass die Unlösbarkeit des Ungleichungssystems (rechts) damit einhergeht, dass der Nullpunkt in der konvexen Hülle der drei Vektoren enthalten ist, während dies bei Lösbarkeit des Ungleichungssystems (links) nicht der Fall ist. Dass dies tatsächlich immer so ist, führt letztlich auf algorithmisch verwertbare Optimalitätsbedingungen und ist der Inhalt des folgenden zentralen Resultats, dessen vollständiger Beweis einiger Vorbereitung bedarf.

Lemma 3.2.13. Für Vektoren $a^k \in \mathbb{R}^n$, $1 \leq k \leq r$, mit $r \in \mathbb{N}$ gilt genau eine der beiden folgenden Alternativen.

a) Das System $\langle a^k, d\rangle < 0$, $1 \leq k \leq r$, hat eine Lösung $d \in \mathbb{R}^n$.

b) Es gilt $0 \in \text{conv}(\{a^1, \ldots, a^r\})$.

Satz 3.2.14 (Trennungssatz)**.** Es seien $X \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge sowie $z \in X^c$. Dann existieren ein $a \in \mathbb{R}^n \setminus \{0\}$ und ein $b \in \mathbb{R}$, so dass für alle $x \in X$ die folgende Ungleichungen erfüllt sind:

$$\langle a, x\rangle \leq b < \langle a, z\rangle$$

Definition. a) Konvexe Hülle $\text{conv}(A) = \left\{ \sum_{i=1}^s \lambda_i a^i \mid a^i \in A, \lambda_i \geq 0, \sum \lambda_i = 1, 1 \leq i \leq s, s \in \mathbb{N} \right\}$

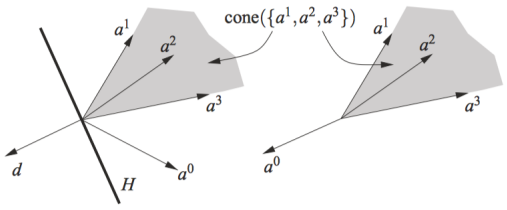
b) Konvexe Kegelhülle $\operatorname{cone}(A) = \left\{ \sum_{i=1}^{s} \lambda_i a^i \mid a^i \in A, \lambda_i \geq 0, 1 \leq i \leq s, s \in \mathbb{N} \right\}$

Satz 3.2.15 (Lemma von Farkas)**.** Für Vektoren $a^k \in \mathbb{R}^n$, $0 \leq k \leq r$, mit $r \in \mathbb{R}$ gilt genau eine der beiden folgenden Alternativen.

a) Das System $\langle a^0, d\rangle < 0, \langle a^k, d\rangle \leq 0$, $1 \leq k \leq r$ hat eine Lösung $d \in \mathbb{R}^n$.

b) Es gilt $-a^0 \in \operatorname{cone}(\{a^1, \ldots, a^r\})$

Abb. 3.8 Alternativen im Lemma von Farkas



Satz 3.2.16 (Satz von Carathéodory)**.** Für jede Menge $A \subseteq \mathbb{R}^n$ gelten die folgenden Aussagen:

a) Zu jedem $\overline{x} \in \operatorname{cone}(A) \setminus \{0\}$ existieren ein $r \leq n$ und linear unabhängige $x^k \in A$ sowie $\lambda_k > 0$, $1 \leq k \leq r$ mit $\overline{x} = \sum_{k=1}^{r} \lambda_k x^k$.

b) Zu jedem $\overline{x} \in \operatorname{conv}(A)$ existieren ein $r \leq n+1$ und $x^1, \ldots, x^r \in A$, so dass die Vektoren $x^2 - x^1, \ldots, x^r - x^1$ linear unabhängig sind und dass $\overline{x} \in \operatorname{conv}(\{x^1, \ldots, x^r\})$ gilt.

Die in Satz 3.2.16b auftretende konvexe Hülle $\operatorname{conv}(\{x_1, \ldots, x_r\})$ von r Punkten mit linear unabhängigen Vektoren $x^2 - x^1, \ldots, x^r - x^1$ wird auch $(r-1)$-Simplex genannt, und die Vektoren $x^1, \ldots, x^r$ heißen dann affin unabhängig.

Der Satz von Carathéodory liefert sofort die folgenden Verbesserungen der Aussagen im Lemma von Gordan und im Lemma von Farkas, wobei $|A|$ die Anzahl der Elemente einer Menge A bezeichne.

Korollar 3.2.17.

a) In Satz 3.2.13b lassen sich Gewichte λ_k mit $|\{1 \leq k \leq r \mid \lambda_k > 0\}| \leq n+1$ wählen.

b) In Satz 3.2.15b lassen sich Gewichte λ_k mit $|\{1 \leq k \leq r \mid \lambda_k > 0\}| \leq n$ wählen.

Optimalitätsbedingungen erster Ordnung ohne Gleichungsrestriktionen

Satz 3.2.18 (Satz von Fritz John für $J = \emptyset$)**.** Es sei $\overline{x}$ ein lokaler Minimalpunkt von P, an dem die Funktionen f und g_i, $i \in I_0(\overline{x})$, differenzierbar sind. Dann existieren Multiplikatoren $\kappa \geq 0$, $\lambda_i \geq 0$, $i \in I_0(\overline{x})$, nicht alle null, mit

$$\kappa \nabla f(\overline{x}) + \sum_{i \in I_0(\overline{x})} \lambda_i \nabla g_i(\overline{x}) = 0 \tag{3.3}$$

Dabei kann man κ und die λ_i so wählen, dass entweder $\kappa > 0$ und $|\{i \in I_0(\overline{x}) \mid \lambda_I > 0\}| \leq n$ gilt oder $\kappa = 0$ und $|\{i \in I_0(\overline{x}) \mid \lambda_i > 0\}| \leq n + 1$.

Die Beschränkung der Anzahl positiver Multiplikatoren in der zweiten Behauptung aus Satz 3.2.18 zieht nach sich, dass die $|I_0(x)| \leq n + 1$.

Zur Motivation der MFB in Definition 3.2.3 haben wir angeführt, dass die im Beweis zu Satz 3.2.18 benutzte Unlösbarkeit des Systems trivialerweise erfüllt sein kann, wenn nämlich $L_<(x, M) = \emptyset$ gilt.

Lemma 3.2.21. Es sei $\overline{x}$ ein lokaler Minimalpunkt von P, an dem die Funktionen f und g_i , $i \in I_0(\overline{x})$, differenzierbar sind. Dann ist (3.3), also

$$\kappa \nabla f(\overline{x}) + \sum_{i \in I_0(\overline{x})} \lambda_i \nabla g_i(\overline{x}) = 0,$$

genau dann mit $\kappa = 0$ erfüllbar, wenn die MFB an $\overline{x}$ verletzt ist.

Satz 3.2.22 (Satz von Karush-Kuhn-Tucker für $J = \emptyset$ unter MFB)**.** Es sei $\overline{x}$ ein lokaler Minimalpunkt von P, an dem die Funktionen f und g_i, $i \in I_0(\overline{x})$, differenzierbar sind, und an $\overline{x}$ gelte die MFB. Dann existieren Multiplikatoren $\lambda_i \geq 0$, $i \in I_0(\overline{x})$ mit

$$\nabla f(\overline{x}) + \sum_{i \in I_0(\overline{x})} \lambda_i \nabla g_i(\overline{x}) = 0$$

Dabei kann man die λ_i so wählen, dass $|\{i \in I_0(\overline{x}) \mid \lambda_i > 0\}| \leq n$ gilt.

Satz 3.2.23 (Satz von Karush-Kuhn-Tucker für $J = \emptyset$ unter AB)**.** Die Aussage von Satz 3.2.22 bleibt richtig, wenn man dort „MFB“ durch „AB“ ersetzt.

Korollar 3.2.24. Es seien $g_i(x) = a_i^T x + b_i$, $1 \leq i \leq p$, und $\overline{x}$ sei ein lokaler Minimalpunkt von P, an dem f differenzierbar ist. Dann existieren Multiplikatoren $\lambda_i \geq 0$, $i \in I_0(\overline{x})$ mit

Dabei kann man die λ_i so wählen, dass $|\{i \in I_0(\overline{x}) | \lambda_i > 0\}| \leq n$ gilt.

Trennungssatz

Lemma 3.2.25. Es seien $X \subseteq \mathbb{R}^n$ eine nichtleere abgeschlossene Menge und $z \in \mathbb{R}^n$. Dann ist das Projektionsproblem

$$Pr(z, X): \quad \min \|x - z\|_2 \text{ s.t. } x \in X$$

lösbar.

Satz 3.2.26. Es seien $X \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge sowie $z \in \mathbb{R}^n$. Dann besitzt das Problem $Pr(z, X)$ einen eindeutigen globalen Minimalpunkt.

Satz 3.2.27 (Variationsformulierung konvexer Probleme)**.** Die Menge $M \subseteq \mathbb{R}^n$ und die Funktion $f \in C^1(M, \mathbb{R})$ seien konvex. Dann gelten die folgenden Aussagen:

a) Der Punkt $\overline{x} \in \mathbb{R}^n$ ist genau dann globaler Minimalpunkt von

$$P: \quad \min f(x) \text{ s.t. } x \in M,$$

wenn $\overline{x}$ globaler Minimalpunkt von folgendem ist

$$\tilde{P}_{lin}(\overline{x}): \quad \min_x \langle \nabla f(\overline{x}), x - \overline{x} \rangle \text{ s.t. } x \in M$$

b) Die Menge der globalen Minimalpunkte von P stimmt mit der folgenden Menge überein

$$\left\{ \overline{x} \in M \mid \langle \nabla f(\overline{x}), x - \overline{x} \rangle \geq 0 \text{ für alle } x \in M \right\}$$

Definition 3.2.28 (Normalenkegel an konvexe Mengen)**.** Für eine konvexe Menge $X \subseteq \mathbb{R}^n$ und $\overline{x} \in X$ heißt

$$N(\overline{x}, X) \coloneqq \left\{ s \in \mathbb{R}^n \mid \langle s, x - \overline{x} \rangle \leq 0 \text{ für alle } x \in X \right\}$$

Normalenkegel an X in $\overline{x}$. Die Elemente s des **Normalenkegels** $N(\overline{x}, X)$ nennt man auch (äußere) Normalenrichtungen an X in $\overline{x}$.

Übung 3.2.29. für jede (nicht notwendigerweise konvexe) Menge $X \subseteq \mathbb{R}^n$ und jedes $\overline{x} \in X$, dass die Menge $N(\overline{x}, X)$ ein konvexer und abgeschlossener Kegel mit $0 \in N(\overline{x}, X)$ ist.

Geometrisch interpretiert liegen im Normalenkegel an X in x genau diejenigen Richtungen $s \in \mathbb{R}^n$, mit denen kein Vektor d einen spitzen Winkel bindet, für den $x + d$ in X liegt (was man mittels des Zusammenhangs $x = x + d$ sieht).

Korollar 3.2.30. Die Menge $M \subseteq \mathbb{R}^n$ und die Funktion $f \in C^1(M, \mathbb{R})$ seien konvex. Dann ist $\overline{x} \in \mathbb{R}^n$ genau dann globaler Minimalpunkt von

$$P: \quad \min f(x) \text{ s.t. } x \in M$$

wenn die Bedingungen $\overline{x} \in X$ und $-\nabla f(\overline{x}) \in N(\overline{x}, X)$ erfüllt sind.

Satz 3.2.31 (Projektionslemma)**.** Es seien $X \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge sowie $z \in \mathbb{R}^n$. Dann ist der eindeutige Minimalpunkt $pr(z, X)$ des Projektionsproblems

$$Pr(z, X): \quad \min \|x - z\|_2 \text{ s.t. } x \in X$$

gleichzeitig die eindeutige Lösung der Bedingungen

$$x \in X \text{ und } z \in x + N(x, X)$$

Normalenkegel

Definition 3.2.33 (Polarkegel)**.** Für eine Menge $A \subseteq \mathbb{R}^n$ heißt

$$A^\circ = \left\{ s \in \mathbb{R}^n \mid \langle s, d \rangle \leq 0 \text{ für alle } d \in A \right\}$$

Polarkegel von A.

Dabei gilt für eine konvexe Menge X, dass der Normalenkegel wie folgt lautet:

$$N(\overline{x}, X) = (X - \overline{x})^\circ.$$

Für eine beliebige Menge X benutzen wir stattdessen den Polarkegel des äußeren Tangentialkegels als Normalenkegel.

Definition 3.2.34 (Normalenkegel an beliebige Mengen)**.** Für eine Menge $X \subseteq \mathbb{R}^n$ und $\overline{x} \in X$ heißt

$$N(\overline{x}, X) = C^\circ(\overline{x}, X)$$

Normalenkegel an X in $\overline{x}$. Die Elemente s des Normalenkegels $N(\overline{x}, X)$ heißen wieder (äußere) Normalenrichtungen an X in $\overline{x}$.

Satz 3.2.35. Für jede konvexe Menge $X \subseteq \mathbb{R}^n$ und $\overline{x} \in X$ gilt

$$(X - \overline{x})^\circ = C^\circ(\overline{x}, X)$$

Satz 3.2.2 besagt in dieser Terminologie tatsächlich, dass an einem lokalen Minimalpunkt $\overline{x}$ von f auf M notwendigerweise die Beziehung

$$-\nabla f(\overline{x}) \in N(\overline{x}, M)$$

gilt. Die Stationaritätsbedingung kann man geometrisch also auch so interpretieren, dass an jedem lokalen Minimalpunkt $\overline{x}$ von P der Vektor $-\nabla f(\overline{x})$ (also die steilste Abstiegsrichtung für f in $\overline{x}$) eine äußere Normalenrichtung an M in $\overline{x}$ ist.

Da die funktionale Beschreibung der Menge M für dieses Ergebnis keine Rolle spielt (sondern nur ihre Geometrie), waren für seine Herleitung keine Constraint Qualifications erforderlich. Im Folgenden werden wir aber sehen, wie sie bei der Herleitung einer expliziten Darstellung der Elemente des Normalenkegels helfen.

Übung 3.2.36. Für alle $\overline{x} \in M$ gilt die Inklusion $L^\circ_\leq(\overline{x}, M) \subseteq N(\overline{x}, M)$.

Übung 3.2.37. Mit Hilfe des Lemmas von Farkas gilt die folgende Identität für alle $\overline{x} \in M$

$$L^\circ_\leq(\overline{x}, M) = \text{cone}(\{\nabla g_i(\overline{x}), i \in I_0(\overline{x})\})$$

Übung 3.2.38. An $\overline{x} \in M$ gelte die AB. Wegen Übung 3.2.36 und Übung 3.2.37 gilt die Identität

$$N(\overline{x}, M) = \text{cone}(\{\nabla g_i(\overline{x}), i \in I_0(\overline{x})\})$$

Somit gilt für jeden lokalen Minimalpunkt x von P, an dem die AB erfüllt ist, d.h.

$$-\nabla f(x) \in \text{cone}(\{\nabla g_i(x), i \in I_0(x)\}),$$

also gerade die Behauptung des Satzes von Karush-Kuhn-Tucker.

Optimalitätsbedingungen erster Ordnung mit Gleichungsrestriktionen

Von jetzt an sei die zulässige Menge von P wieder durch Ungleichungen und Gleichungen beschrieben, d. h., es gelte

$$M = \left\{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i \in I, h_j(x) = 0, j \in J\right\}.$$

Satz 3.2.39. Es sei $\overline{x}$ ein lokaler Minimalpunkt von P, an dem die Funktionen f, g_i, $i \in I_0(\overline{x})$, und h_j, $j \in J$, stetig differenzierbar sind. Dann existieren Multiplikatoren $\kappa \geq 0, \lambda_i \geq 0, i \in I_0(\overline{x}), \mu_j \in \mathbb{R}, j \in J$, nicht alle null, mit

$$\kappa \nabla f(\overline{x}) + \sum_{i \in I_0(\overline{x})} \lambda_i \nabla g_i(\overline{x}) + \sum_{j \in J} \mu_j \nabla h_j(\overline{x}) = 0$$

Dabei kann man κ und die λ_i so wählen, dass entweder $\kappa > 0$ und $|\{i \in I_0(\overline{x}) \mid \lambda_i > 0\}| \leq n - q$ gilt oder $\kappa = 0$ und $|\{i \in I_0(\overline{x}) \mid \lambda_i > 0\}| \leq n - q + 1$.

Definition 3.2.40 (Mangasarian-Fromowitz-Bedingung)**.** Der Punkt $\in M$ erfüllt die Mangasarian-Fromowitz-Bedingung (MFB), falls folgende Bedingungen gelten:

a) Die Vektoren $\nabla h_j(\overline{x}), j \in J$, sind linear unabhängig,

b) Es existiert ein $d \in \mathbb{R}^n$ mit $\langle \nabla g_i(\overline{x}), d\rangle < 0, i \in I_0(\overline{x}),\ \langle \nabla h_j(\overline{x}), d\rangle = 0, j \in J$

Satz 3.2.41 (Satz von Karush-Kuhn-Tucker)**.** Es sei $\overline{x}$ ein lokaler Minimalpunkt von P, an dem die Funktionen f, g_i, $i \in I_0(\overline{x})$, und h_j, $j \in J$, stetig differenzierbar sind, und an $\overline{x}$ gelte die MFB. Dann existieren Multiplikatoren $\lambda_i \geq 0$, $i \in I_0(\overline{x})$, $\mu_j \in \mathbb{R}$, $j \in J$, mit

$$\nabla f(\overline{x}) + \sum_{i \in I_0(\overline{x})} \lambda_i \nabla g_i(\overline{x}) + \sum_{j \in J} \mu_j \nabla h_j(\overline{x}) = 0$$

Dabei kann man die λ_i so wählen, dass $|\{i \in I_0(\overline{x}) \mid \lambda_i > 0\}| \leq n - q$ gilt.

Definition 3.2.42 (Lineare-Unabhängigkeits-Bedingung)**.** An $\overline{x} \in M$ gilt die Lineare-Unabhängigkeits-Bedingung (LUB), falls die Vektoren $\nabla g_i(\overline{x})$, $i \in I_0(\overline{x})$, $\nabla h_j(\overline{x})$, $j \in J$, also die Gradienten aller in $\overline{x}$ aktiven Restriktionen, linear unabhängig sind.

Das heißt insbesondere, dass falls nur eine Restriktion herrscht, diese im Punkt nicht identisch 0 ist.

Satz 3.2.43. Die LUB an $\overline{x}$ impliziert die MFB an $\overline{x}$.

Beispiel 3.2.45. Die Spektralnorm einer Matrix entspricht dem KKT-Multiplikator μ.

Bemerkung 3.2.46. Jede Zahl zu λ einem positiven Eigenwert λ der Matrix $A^T A$ heißt Singulärwert der (m, n)-Matrix A.

Karush-Kuhn-Tucker-Punkte

Der grundlegende Satz 3.2.41 erlaubt es, das aus dem unrestringierten Fall bekannte Konzept eines kritischen Punkts von P auf den restringierten Fall zu übertragen.

Definition 3.2.47 (Karush-Kuhn-Tucker-Punkt)**.** Zu $\overline{x} \in M$ gebe es $\lambda_i \geq 0$, $i \in I_0(\overline{x})$, $\mu_j \in \mathbb{R}$, $j \in J$ mit

$$\nabla f(\overline{x}) + \sum_{i \in I_0(\overline{x})} \lambda_i \nabla g_i(\overline{x}) + \sum_{j \in J} \mu_j \nabla h_j(\overline{x}) = 0$$

Dann heißt $\overline{x}$ Karush-Kuhn-Tucker-Punkt (KKT-Punkt) von P. Die Koeffizienten λ_i, $i \in I_0(\overline{x})$, und μ_j j $\in J$ heißen KKT-Multiplikatoren.

Aus Ergebnissen der linearen Algebra folgt, dass die Multiplikatoren eines KKT-Punkts $\overline{x}$ eindeutig bestimmt sind, wenn an $\overline{x}$ die LUB gilt. Satz 3.2.41 und Satz 3.2.43 liefern daher folgendes Ergebnis.

Korollar 3.2.48. Es sei $\overline{x}$ ein lokaler Minimalpunk von P, an dem die Funktionen f, g_i, $i \in I_0(\overline{x})$ und h_j, $j \in J$, stetig differenzierbar sind, und an $\overline{x}$ gelte die LUB. Dann ist $\overline{x}$ ein KKT-Punkt von P mit eindeutigen KKT-Multiplikatoren.

Algorithmus 3.1: Konzeptioneller Algorithmus zur restringierten nichtlinearen Minimierung mit Informationen erster Ordnung

Input : Lösbares restringiertes C^1-Optimierungsproblem P
Output : Globaler Minimalpunkt $x^\star$ von P

1 **begin**
2 Bestimme die Menge LA der Punkte in M, an denen die LUB verletzt ist.
3 Bestimme unter den Punkten in M, an denen die LUB erfüllt ist, die Menge KKT aller KKT-Punkte.
4 Bestimme einen Minimalpunkt $x^\star$ von f in $LA \cup KKT$.
5 **end**

Für kleine Werte von p löst man durch Fallunterscheidung der aktiven Indexmenge das System

$$\left.\begin{array}{ll} \nabla f(x) + \sum_{i \in I_0(x)} \lambda_i \nabla g_i(x) + \sum_{j \in J} \mu_j \nabla h_j(x) & = 0, \\ g_i(x) & \leq 0, i \in I, \\ h_j(x) & = 0, j \in J, \\ \lambda_i & \geq 0, i \in I_0(x). \end{array}\right\} \tag{3.8}$$

Ist p allerdings groß, schlägt man einen anderen Weg ein und löst das folgende System mit einer Komplementaritätsbedingung:

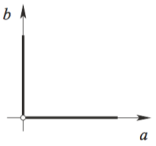
$$\left.\begin{array}{ll} \nabla f(x) + \sum_{i \in I} \lambda_i \nabla g_i(x) + \sum_{j \in J} \mu_j \nabla h_j(x) & = 0, \\ g_i(x) & \leq 0, i \in I, \\ h_j(x) & = 0, j \in J, \\ \lambda_i & \geq 0, i \in I. \\ \lambda_i \cdot g_i(x) & = 0, i \in I \end{array}\right\} \tag{3.9}$$

Man fasst die letzte Zeile nicht zu $\sum_{i \in I} \lambda_i g_i(x) = 0$ zusammen, da dies später zu Problemen führt.

Definition 3.2.49 (Strikte Komplementaritätsbedingung)**.** Am KKT-Punkt $\overline{x}$ gelte die LUB, und zu $\overline{x}$ sei $(\overline{\lambda}_{I_0}, \overline{\mu})$ die eindeutige Lösung von (3.8). Dann gilt an $\overline{x}$ die strikte Komplementaritätsbedingung (SKB), falls $\overline{\lambda}_i > 0$ für alle $i \in I_0(\overline{x})$ erfüllt ist.

Strikte Komplementarität bedeutet hier gerade, dass die Knickstelle dieser Menge, nämlich der Nullpunkt, aus der Menge entfernt wird.

Abb. 3.10 Durch (strikte) Komplementaritätsbedingung definierte Menge



Definition 3.2.52 (Lagrange-Funktion)**.** Die Funktion

$$L \colon \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R},\ (x, \lambda, \mu) \mapsto f(x) + \sum_{i \in I} \lambda_i g_i(x) + \sum_{j \in J} \mu_j h_j(x)$$

heißt **Lagrange-Funktion** von P.

Optimalitätsbedingung zweiter Ordnung

Um die Form der Optimalitätsbedingungen zweiter Ordnung in der restringierten Optimierung zu motivieren, betrachten wir die Gleichung, die aus (3.11) durch Unterschlagen der Ungleichung entsteht:

$$0 = \mathfrak{T}(x, \lambda, \mu) \coloneqq \begin{pmatrix} \nabla_x L(x, \lambda, \mu) \\ \operatorname{diag}(\lambda) \cdot \nabla_\lambda L(x, \lambda, \mu) \\ \nabla_\mu L(x, \lambda, \mu) \end{pmatrix}$$

Definition 3.2.55 (Kritischer Punkt eines restringierten Problems)**.** Ein Punkt $\overline{x} \in M$, der mit Multiplikatorem $\overline{\lambda}$ und $\overline{\mu}$ die Gleichung $\mathfrak{T}(\overline{x}, \overline{\lambda}, \overline{\mu}) = 0$ erfüllt, heißt**kritischer Punkt** von P.

Kritische Punkte unterscheiden sich von KKT-Punkten nur dadurch, dass keine Nichtnegativitätsbedingung an $\overline{x}$ gefordert wird.

Lemma 3.2.57. Es sei $\mathfrak{T}(\overline{x}, \overline{\lambda}, \overline{\mu}) = 0$. Dann ist $D\mathfrak{T}(\overline{x}, \overline{\lambda}, \overline{\mu})$ genau dann nichtsingulär, wenn $\overline{\lambda}_i \neq 0$ für alle $i \in I_0(\overline{x})$ gilt und wenn die folgende Matrix nichtsingulär ist

$$\begin{pmatrix} D_x^2 L(\overline{x}, \overline{\lambda}, \overline{\mu}) & \nabla g_{I_0}(\overline{x}) & \nabla h(\overline{x}) \\ Dg_{I_0}(\overline{x}) & 0 & 0 \\ Dh(\overline{x}) & 0 & 0 \end{pmatrix}$$

Definition 3.2.58 (Einschränkung einer Matrix)**.** Die $(n-m, n-m)-MatrixA|_{\operatorname{kern} B^T} \coloneqq V^T A V$ heißt **Einschränkung** von A auf $\operatorname{kern} B^T$.

Wir nennen

$$\operatorname{kern} B^T = \left\{ d \in \mathbb{R}^n \mid \langle \nabla g_i(\overline{x}), d \rangle = 0, i \in I_0(\overline{x}), \langle \nabla h_j(\overline{x}), d \rangle = 0, j \in J \right\} =: T(\overline{x}, M)$$

Tangentialraum an M in $\overline{x}$.

Lemma 3.2.60 (Strukturlemma)**.** Für eine (n, n)-Matrix A und ein (n, m)-Matrix B ist die Blockmatrix

$$\begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix}$$

genau dann nichtsingulär, wenn $\operatorname{rang}(B) = m$ gilt und wenn $A|_{\operatorname{kern} B^T}$ nichtsingulär ist.

Satz 3.2.61. Es sei $\mathfrak{T}(\overline{x}, \overline{\lambda}, \overline{\mu}) = 0$. Dann ist $\mathfrak{T}(\overline{x}, \overline{\lambda}, \overline{\mu})$ genau dann nichtsingulär, wenn die folgenden drei Bedingungen gleichzeitig gelten:

a) An $\overline{x}$ gilt die LUB.

b) Es gilt $\overline{\lambda}_i \neq 0$ für alle $i \in I_0(\overline{x})$

c) Die Matrix $D_x^2 L(\overline{x}, \overline{\lambda}, \overline{\mu})|_{T(\overline{x}, M)}$ ist nichtsingulär.

Es sei darauf hingewiesen, dass im Fall $n = p_0 + q$ die Bedingung c in Satz 3.2.61 trivialerweise erfüllt ist und entfallen kann.

Definition 3.2.62 (Nichtdegenerierter kritischer Punkt eines restringierten Problems)**.** Für einen kritischen Punkt $\overline{x}$ von P mit Multiplikatoren $\overline{\lambda}$, $\overline{\mu}$ seien die Bedingungen a, b und c aus Satz 3.2.61 erfüllt. Dann heißt $\overline{x}$ **nichtdegenerierter kritischer Punkt** von P.

Wie im unrestringierten Fall (Abschn. 2.1.4) kann man wieder zeigen, dass für „fast alle“ C^2-Optimierungsprobleme P jeder kritische Punkt nichtdegeneriert ist, dass die Bedingungen a, b und c aus Satz 3.2.61 also schwach sind. Weil in Definition 3.2.62 weder an die Multiplikatoren noch an die Eigenwerte der Matrix Vorzeichenbeschränkungen gefordert werden, können nichtdegenerierte kritische Punkte nicht nur Minimalpunkte, sondern auch Maximalpunkte oder Sattelpunkte sein.

Definition 3.2.65 (Nichtdegenerierter Minimalpunkt eines restringierten Problems)**.** Ein KKT-Punkt $\overline{x}$ von P mit Multiplikatoren $\overline{\lambda}$, $\overline{\mu}$ erfülle die folgenden Bedingungen:

a) An $\overline{x}$ gilt die LUB.

b) An $\overline{x}$gilt die SKB.

c) Die Matrix $D_x^2 L(\overline{x}, \overline{\lambda}, \overline{\mu})|_{T(\overline{x},M)}$ ist positiv definit.

Dann heißt $\overline{x}$ nichtdegenerierter lokaler Minimalpunkt von P.

Die Bedingung c in Definition 3.2.64 ist gleichbedeutend mit der Aussage

$$\forall d \in T(\overline{x}, M) \setminus 0 : \quad d^T D_x^2 L(\overline{x}, \overline{\lambda}, \overline{\mu}) d > 0$$

Im Fall $n = p_0 + q$ ist sie wieder trivialerweise erfüllt und kann entfallen.

Satz 3.2.66 (Hinreichende Optimalitätsbedingung zweiter Ordnung)**.** Jeder nichtdegenerierte lokale Minimalpunkt ist strikter lokaler Minimalpunkt von P.

Für die folgende Optimalitätsbedingung zweiter Ordnung, die ohne die Voraussetzungen der LUB und SKB auskommt, seien

$$\begin{aligned} I_{0+} &:= \big\{ i \in I_0(\overline{x}) \mid \ \overline{\lambda}_i > 0 \big\}, \\ I_{00}(\overline{x}) &:= \big\{ i \in I_0(\overline{x}) \mid \overline{\lambda}_i = 0 \big\} \end{aligned}$$

und

$$\begin{aligned} K(\overline{x}, M) = \big\{ d \in \mathbb{R}^n \mid \langle \nabla f(\overline{x}, d\rangle = 0, \langle \nabla g_i(\overline{x}, d\rangle = 0, i \in I_{0+}(\overline{x}), \\ \langle \nabla g_i(\overline{x}), d\rangle \leq 0, i \in I_{00}(\overline{x}), \langle \nabla h_j(\overline{x}), d\rangle = 0, j \in J \big\} \end{aligned}$$

Korollar 3.2.67. Ein KKT-Punkt $\overline{x}$ von P mit Multiplikatoren $\overline{\lambda}, \overline{\mu}$ erfülle $d^T D_x^2(\overline{x}, \overline{\lambda}, \overline{\mu}) d > 0$ für alle $d \in K(\overline{x}, M)$. Dann ist $\overline{x}$ strikter lokaler Minimalpunkt von P.

Es sei bemerkt, dass die fehlende Voraussetzung der SKB in Korollar 3.2.67 dadurch kompensiert wird, dass die Matrix auf einer gegebenenfalls größeren Menge als in Satz 3.2.66 positiv definit sein muss.

Korollar 3.2.68 (Hinreichende Optimalitätsbedingung erster Ordnung)**.** Ein KKT-Punkt $\overline{x}$ von P erfülle die folgenden Bedingungen:

a) An $\overline{x}$ gilt die LUB.

b) An $\overline{x}$ gilt die SKB.

c) Es gilt $p_0 + q = n$.

Dann ist $\overline{x}$ strikter lokaler Minimalpunkt von P.

Zur Vollständigkeit geben wir auch noch eine hinreichende Optimalitätsbedingung erster Ordnung an, die weder Constraint Qualifications noch Karush-Kuhn-Tucker-Punkte benutzt, sondern nur Fritz-John-Punkte (Satz 3.2.39) mit einer gewissen Rangeigenschaft voraussetzt. Die beteiligten Funktionen brauchen hier nur differenzierbar zu sein.

Satz 3.2.69. Zu $\overline{x} \in M$ gebe es Multiplikatoren $\overline{\kappa} \geq 0$, $\overline{\lambda}_i \geq 0$, $i \in I_0(\overline{x})$, $\overline{\mu}_j \in \mathbb{R}$, $j \in J$, mit

$$\overline{\kappa}\nabla f(\overline{x}) + \sum_{i \in I_0(\overline{x})} \overline{\lambda}_i \nabla g_i(\overline{x}) + \sum_{j \in J} \overline{\mu}_j \nabla h_j(\overline{x}) = 0$$

und so, dass die Vektoren

$$\overline{\kappa}\nabla f(\overline{x}), \quad \overline{\lambda}_i \nabla g_i(\overline{x}), i \in I_0(\overline{x}), \ \overline{\mu}_j \nabla h_j(\overline{x}), j \in J$$

gemeinsam den Rang n besitzen. Dann ist $\overline{x}$ ein strikter lokaler Minimalpunkt von P.

Es sei daran erinnert, dass hinreichende Optimalitätsbedingungen erster Ordnung für unrestringierte glatte Optimierungsprobleme nicht existieren können (Bemerkung 2.1.34), während sie im restringierten Fall wie gerade gesehen sinnvoll sind.

Neben den hinreichenden Optimalitätsbedingungen existiert analog zum unrestringierten Fall natürlich auch eine notwendige Optimalitätsbedingung zweiter Ordnung. Wie im unrestringierten Fall unterscheiden sich notwendige und hinreichende Bedingungen durch die Striktheit der auftretenden Ungleichungen.

Satz 3.2.70 (Notwendige Optimalitätsbedingung zweiter Ordnung)**.** Es sei $\overline{x}$ ein lokaler Minimalpunkt von P, und an $\overline{x}$ sei die LUB erfüllt. Dann gilt:

a) $\overline{x}$ ist KKT-Punkt von P mit eindeutigen Multiplikatoren $\overline{\lambda} \geq 0$ und $\overline{\mu}$.

b) Die Matrix $D_x 2L(\overline{x}, \overline{\lambda}, \overline{\mu})|_{T(\overline{x},M)}$ ist positiv semidefinit.

Konvexe Optimierungsprobleme

Das restringierte Optimierungsproblem

$$P: \quad \min f(x) \text{ s.t. } x \in M$$

heißt konvex, falls die Menge $M \subseteq \mathbb{R}^n$ und die Funktion $f: M \to \mathbb{R}$ konvex sind.

Übung 3.2.72. Die Funktionen $g_i: \mathbb{R}^n \to \mathbb{R}$, $i \in I$, seien konvex, und die Funktionen $h_j: \mathbb{R}^n \to \mathbb{R}$, $j \in J$, seien affin-linear, d. h., für alle $j \in J$ gelte

$$h_j(x) = a_j^T x + b_j$$

mit $a_j \in \mathbb{R}^n$ und $b_j \in \mathbb{R}$. Dann ist die Menge M konvex.

Definition 3.2.73 (Konvex beschriebene Menge)**.** Wir nennen eine mit beliebigen Indexmengen I und Jdurch Ungleichungen und Gleichungen gegebene Menge

$$M = \left\{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i \in I, h_j(x) = 0, j \in J\right\}$$

konvex beschrieben, wenn die Funktionen $g_i \coloneqq \mathbb{R}^n \to \mathbb{R}$, $i \in I$, konvex und die Funktionen $h_j: \mathbb{R}^n \to \mathbb{R}$, $j \in J$, affin-linear sind.

Übung 3.2.72 besagt in dieser Terminologie, dass konvex beschriebene Mengen konvex sind.

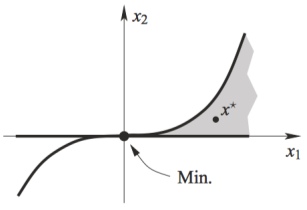
Definition 3.2.74 (Slater-Bedingung)**.** Die Menge $M \subseteq \mathbb{R}^n$ sei konvex beschrieben. Dann erfüllt M die Slater-Bedingung (SB), falls die folgenden beiden Bedingungen erfüllt sind:

a) Es gilt $\operatorname{rang}(A) = q$.

b) Es gibt einen Punkt $x^* \in \mathbb{R}^n$ mit $g(x^*) < 0$ und $h(x^*) = 0$.

Die SB ist eine globale Bedingung an M, denn der Slater-Punkt x^* braucht nichts mit einem Optimalpunkt von P zu tun zu haben. In der Tat können nichtkonvexe Probleme einen Slater-Punkt besitzen, während gleichzeitig in einem Optimalpunkt die MFB verletzt ist. Bei glatten konvexen Optimierungsproblemen kann dieser Effekt nicht auftreten.

Abb. 3.12 SB ohne MFB im Optimalpunkt



Satz 3.2.75. Die Menge $M \subseteq \mathbb{R}^n$ sei konvex beschrieben, nichtleer, und die Funktionen $g_i \colon \mathbb{R}^n \to \mathbb{R}$, $i \in I$, seien zusätzlich stetig differenzierbar. Dann sind die folgenden Aussagen äquivalent:

a) Die MFB gilt überall in M.

b) Die MFB gilt irgendwo in M.

c) M erfüllt die SB.

Dass die lokale Regularitätsbedingung MFB mit der globalen Regularitätsbedingung SB äquivalent sein kann, liegt an der Voraussetzung einer globalen Struktur an P, nämlich der Konvexität.

Laut Korollar 2.1.41 sind im glatten konvexen unrestringierten Fall die kritischen Punkte mit den globalen Minimalpunkten identisch. Zu vermuten ist also, dass im restringierten Fall die KKT-Punkte den globalen Minimalpunkten entsprechen. Im Folgenden werden wir sehen, dass dies zumindest unter der SB richtig ist.

Korollar 3.2.76. Das Problem P sei konvex, und M erfülle die SB. Dann ist jeder lokale Minimalpunkt von P KKT-Punkt.

Satz 3.2.77. Der Punkt $\overline{x}$ sei KKT-Punkt des konvexen Problems P. Dann ist $\overline{x}$ ein globaler Minimalpunkt von P.

Im Gegensatz zur notwendigen Optimalitätsbedingung aus Korollar 3.2.76 benötigt die hinreichende Bedingung in Satz 3.2.77 keinerlei Regularitätsvoraussetzung. Insgesamt erhalten wir folgende „Charakterisierung“ für globale Minimalpunkte konvexer Probleme P:

$$\overline{x} \text{ globaler Minimalpunkt} \quad \overset{SB}{\Longrightarrow} \overline{x} \text{ KKT-Punkt} \quad \Rightarrow \overline{x} \text{ globaler Minimalpunkt}$$

Numerische Verfahren

Dieser Abschnitt behandelt einige numerische Verfahren zur Lösung des Problems

$$P: \quad \min_{x \in \mathbb{R}^n} f(x) \;\text{ s.t. } g_i(x) \leq 0, i \in I, h_j(x) = 0, j \in J,$$

mit hinreichend glatten Funktionen f, g und h.

Straftermverfahren

Definition 3.3.1 (Straftermfunktion)**.** Eine Funktion $\alpha \colon \mathbb{R}^n \to \mathbb{R}$ heißt Straftermfunktion für $M \subseteq \mathbb{R}^n$, falls folgende Bedingungen erfüllt sind:

a) Für alle $x \in M$ gilt $\alpha(x) = 0$.

b) Für alle $x \in Mc$ gilt $\alpha(x) > 0$.

Das Straftermverfahren bedient sich des Skalarisierungsansatzes zur Lösung des Mehrzielproblems, d. h., minimiert wird eine gewichtete Summe der beiden Zielfunktionen f und α. Aufgrund von Übung 1.3.1a genügt es dabei, nur eine der beiden Funktionen mit einem Gewicht zu versehen. Dies führt auf den Ansatz, P durch ein unrestringiertes Problem mit Zielfunktion

$$P(t) : \min_{x \in \mathbb{R}^n} f(x) + t \cdot \alpha(x)$$

Die exakte Lösung von P würde man für jedes t mit der „idealen“ Straftermfunktion $\alpha(x) = +\infty$ für alle $x \in M^c$ erhalten. Da dies nicht numerisch realisierbar ist, folgt mit

$$g_i^+(x) \coloneqq \max\{0, g_i(x)\},\ i \in I$$

Wegen der Definitheit von Normen betrachten wir mit einem $r \in \mathbb{N}$ der Term

$$\alpha_r(x) \coloneqq \|(g_1^+(x), \ldots, g_p^+(x), h_1(x), \ldots, h_q(x))\|_r^r$$

Diese verschwindet genau dann, wenn auch das Argument eintragsweise verschwindet.

Übung 3.3.2. Zeigen Sie mit Hilfe von Übung 2.1.12, dass die Funktionen $\varphi(a) = \max\{0, a\}$ und $abs(a) = |a|$ an jedem $a \in \mathbb{R}$ einseitig richtungsdifferenzierbar sind, und zwar mit für alle $d \in \mathbb{R}$

$$\varphi'(a,d) = \begin{cases} 0, & \text{falls } a < 0 \\ \max\{0,d\}, & \text{falls } a = 0 \\ d, & \text{falls } a > 0 \end{cases}$$

$$abs'(a,d) = \begin{cases} -d, & \text{falls } a < 0 \\ |d|, & \text{falls } a = 0 \\ d, & \text{falls } a > 0 \end{cases}$$

Übung 3.3.3. Für auf $\mathbb{R}^n$ stetig differenzierbare Funktionen g und h die einseitige Richtungsdifferenzierbarkeit der ℓ_1-Straftermfunktion an jedem $x \in \mathbb{R}^n$ mit

$$\alpha_1'(x,d) = \sum_{i \in I_0(x)} \max\{0, \langle \nabla g_i(x), d\rangle\} + \sum_{i \in I_+(x)} \langle \nabla g_i(x), d\rangle - \sum_{j \in J_-(x)} \langle \nabla h_j(x), d\rangle$$
$$+ \sum_{j \in J_0(x)} |\langle \nabla h_j(x), d\rangle| + \sum_{j\ \in J_+(x)} \langle \nabla h_j(x), d\rangle$$

Übung 3.3.4. Die Funktion $\varphi^2(a) = (\max\{0, a\})^2$ ist an jedem $a \in \mathbb{R}$ stetig differenzierbar und für auf $\mathbb{R}^n$ stetig differenzierbare Funktionen g und h ist daher auch α_2 auf ganz $\mathbb{R}^n$ stetig differenzierbar.

Abb. 3.13 ℓ_1- und ℓ_2-Straftermfunktionen

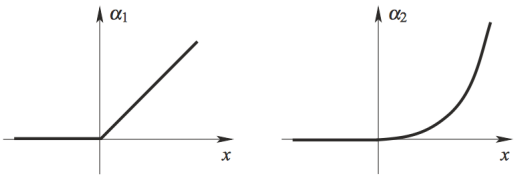
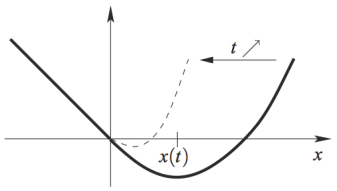


Abb. 3.14 Probleme $P_2(t)$ mit verschiedenen $t > 0$



Satz 3.3.6. Die Funktion $\alpha\colon \mathbb{R}^n \to \mathbb{R}$ sei eine stetige Straftermfunktion für M, (t^k) sei eine monoton wachsende Folge mit $\lim_k t^k = +\infty$, und für alle $k \in \mathbb{N}$ sei x^k ein globaler Minimalpunkt von $P(t^k)$. Dann ist jeder Häufungspunkt x^* von (x^k) ein globaler Minimalpunkt von P.

Algorithmus 3.2: Straftermverfahren

Input : Lösbares C^1-Optimierungsproblem P und stetige Straftermfunktion α
Output : Approximation $\bar{x}$ eines globalen Minimalpunkts von P (falls das Verfahren terminiert; Satz 3.3.6)

1 **begin**
2 Wähle einen Parameter $t^0 > 0$, einen Startpunkt $x^0 \in \mathbb{R}^n$, eine Toleranz $\varepsilon > 0$, einen Faktor $\rho > 1$ und setze $k = 0$.
3 **repeat**
4 Ersetze k durch $k + 1$.
5 Setze $t^k = \rho\, t^{k-1}$.
6 Bestimme vom Startpunkt x^{k-1} aus einen globalen Minimalpunkt x^k von

$$P(t^k): \quad \min f(x) + t^k\, \alpha(x).$$

7 **until** $t^k \alpha(x^k) < \varepsilon$
8 Setze $\bar{x} = x^k$.
9 **end**

Algorithmus 3.2 ist wegen seiner einfachen Implementierbarkeit zwar in der Praxis sehr beliebt, aber in dem Sinne nur konzeptionell, dass die Verfahren aus Abschn. 2.2 ohne weitere Voraussetzungen nicht *globale* Minimalpunkte der Hilfsprobleme $P(t^k)$ aus Zeile 6 identifizieren können. Die Anwendbarkeit der Verfahren aus Abschn. 2.2 setzt außerdem eine glatte Straftermfunktion voraus, also beispielsweise α_2, aber nicht α_1.

Selbst bei Nutzung einer glatten Straftermfunktion wird die numerische Behandlung der Probleme $P(t)$ für wachsende t zunehmend schwieriger, da die Zielfunktion $f(x) + t\alpha(x)$

dann am Rand von M eine immer stärkere Krümmung aufweist. Wird eine gewisse Schranke für t überschritten, ist die theoretisch glatte Zielfunktion von $P(t)$ numerisch nicht mehr von einer am Rand von M nichtglatten Funktion unterscheidbar, so dass die Verfahren aus Abschn. 2.2 gegebenenfalls nicht mehr funktionieren.
Ein weiterer Nachteil von Straftermverfahren besteht darin, dass die Punkte $x(t)$ mit $t > 0$ für P üblicherweise unzulässig sind. Beispiel 3.3.5 belegt außerdem, dass man einen exakten Optimalpunkt von P mit dem differenzierbaren ℓ_2-Strafterm im Allgemeinen tatsächlich nur für $t \to \infty$ erhält. Von Algorithmus 3.2 kann man also nur erwarten, einen „fast zulässigen“ Punkt x_k in dem Sinne zu generieren, dass sein Zulässigkeitsmaß $\alpha(x^k) < \epsilon/t^k$ (statt $\alpha(x^k) = 0$) erfüllt.

Mit der ℓ_1-Straftermfunktion α_1 findet man hingegen den exakten Optimalpunkt von P unter schwachen Voraussetzungen bereits für genügend großes t. Das Verfahren mit ℓ_-1-Straftermfunktion nennt man daher exaktes Straftermverfahren.

Satz 3.3.8. Es sei $\overline{x}$ ein nichtdegenerierter lokaler Minimalpunkt von P mit Multiplikatoren $\overline{\lambda}$, $\overline{\mu}$. Dann ist $\overline{x}$ für alle

$$t > \|(\overline{\lambda}_{I_0}, \overline{\mu})\|_\infty (= \max\{\overline{\lambda}_i, i \in I_0(\overline{x}), |\overline{\mu}_j|, j \in J.\})$$

auch lokaler Minimalpunkt von $P_1(t)$.

Ein wesentlicher Nachteil der ℓ_1-Straftermfunktion ist die mangelnde Glattheit der unrestringierten Hilfsprobleme $P_1(t)$, $t > 0$, so dass man dieses Verfahren überwiegend für Probleme anwendet, bei denen P selbst bereits nichtglatt ist.
Ein weiterer Nachteil besteht darin, dass die Schranke
$\|(\overline{\lambda}_{I_0}, \overline{\mu})\|_\infty$ nicht a priori bekannt ist. In Abschn. 3.3.6 werden wir die ℓ_1-Straftermfunktion trotzdem numerisch sinnvoll zum Einsatz bringen können.

Multiplikatorenverfahren

Einen glatten und exakten Strafterm liefert die folgende Überlegung, die wir wieder nur für den Fall $I = \emptyset$ darstellen, da sich Ungleichungen an nichtdegenerierten kritischen Punkten wie Gleichungen verhalten.

Satz 3.3.10. Es sei $\overline{x}$ ein nichtdegenerierter lokaler Minimalpunkt von P mit Multiplikator $\overline{\mu}$. Dann existiert ein $\overline{t} > 0$, so dass $\overline{x}$ für alle $t > \overline{t}$ ein strikter lokaler Minimalpunkt von $\pi(x, t, \overline{\mu}) \coloneqq L(x, \overline{\mu}) + t\|h(x)\|_2^2$ ist.

Der vom exakten Straftermverfahren bekannte Nachteil, dass die von t zu überschreitende Schranke a priori unbekannt ist, bleibt leider auch beim Multiplikatorenverfahren erhalten.

Barrierevrfahren

Barriereverfahren behandeln die Ungleichungsrestriktionen in P, weshalb wir uns hier auf den Fall ohne Gleichungen konzentrieren. Im Fall $J \neq \emptyset$ reduziert der Barriereansatz P

zumindest auf ein Problem mit $I = \emptyset$. Zum Problem

$$P: \quad \min_{x \in \mathbb{R}^n} f(x) \;\text{ s.t. }\; g_i(x) \leq 0, i \in I$$

mit

$$M_< \coloneqq \left\{ x \in \mathbb{R}^n \mid g_i(x) < 0, i \in I \right\} \neq \emptyset$$

d. h., die aus der konvexen Optimierung bekannte Slater-Bedingung (SB) sei erfüllt. Grundidee von Barriereverfahren ist es, P durch ein Problem zu ersetzen, bei dem die Ungleichungsrestriktionen nicht aktiv werden können. Im Folgenden bezeichnen wir mit $\operatorname{bd} A$ den (topologischen) Rand einer Menge $A \subseteq \mathbb{R}^n$ (d. h. die Menge aller Punkte, für die jede ihrer Umgebungen sowohl ein Element von A als auch ein Element von A^c enthält; bd = boundary).

Definition 3.3.11 (Barrierefunktion)**.** Die Funktion $\beta \colon M_< \to \mathbb{R}$ heißt **Barrierefunktion** für M, falls für alle Folgen $(x^k) \subseteq M_<$ mit $\lim_k x^k = \overline{x} \in \operatorname{bd} M_<$ folgendes gilt:

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

Die definierende Eigenschaft der Barrierefunktion lässt sich als „Strafe für zur große Nähe zum Rand von M“ interpretieren. Ein wichtiger Barriereterm ist (Frischs) logarithmische Barrierefunktion

$$\beta(x) = -\sum_{i \in I} \log(-g_i(x)).$$

Ein Barriereverfahren ersetzt P durch das Problem

$$P(t): \quad \min_{x \in \mathbb{R}^n} f(x) + t \cdot \beta(x) \;\text{ s.t. }\; x \in M$$

mit $t > 0$, wobei die die Menge M beschreibenden Ungleichungsrestriktionen aufgrund des Barriereterms nicht aktiv werden können. Die Zielfunktion von $P(t)$ ist nur auf $M_<$ definiert.

Satz 3.3.13. Es gelte $M = \operatorname{cl}(M_<)$, β sei stetig auf $M_<$, und für $t^k \searrow 0$ seien x^k globale Minimalpunkte von $P(t^k)$. Dann ist jeder Häufungspunkt x^* von (x^k) globaler Minimalpunkt von P.

Abb. 3.16 Logarithmische Barrierefunktion β

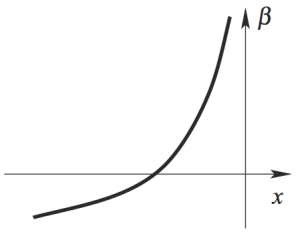
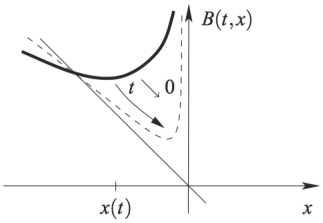


Abb. 3.17 Probleme $P(t)$ mit verschiedenen $t > 0$



Übung 3.3.14. Die Voraussetzung $M = \mathrm{cl}(M_<)$ aus Satz 3.3.13 lässt sich durch die Forderung der MFB in ganz M garantieren.

Zur numerischen Behandlung der Hilfsprobleme $P(t^k)$ in Algorithmus 3.3 durch die Verfahren aus Abschn. 2.2 ist es zunächst natürlich erforderlich, eine auf $M_<$ mindestens einmal stetig differenzierbare Barrierefunktion β zu wählen. Wie Algorithmus 3.2 ist Algorithmus 3.3 aber alleine schon deshalb nur konzeptionell, weil die Verfahren aus Abschn. 2.2 ohne weitere Voraussetzungen nicht globale Minimalpunkte der Hilfsprobleme $P(t^k)$ in Zeile 6 identifizieren können.

Algorithmus 3.3: Barriereverfahren

Input : Lösbares C^1-Optimierungsproblem P mit $J = \emptyset$ und $M_< \neq \emptyset$ sowie Barrierefunktion Lösbares β

Output : Approximation $\bar{x}$ eines globalen Minimalpunkts von P (falls das Verfahren terminiert; Satz 3.3.13)

1 **begin**
2 Wähle einen Parameter $t^0 > 0$, einen Startpunkt $x^0 \in M_<$, eine Toleranz $\varepsilon > 0$, einen Faktor $\rho \in (0, 1)$ und setze $k = 0$.
3 **repeat**
4 Ersetze k durch $k + 1$.
5 Setze $t^k = \rho\, t^{k-1}$.
6 Bestimme vom Startpunkt x^{k-1} aus einen globalen Minimalpunkt x^k von

$$P(t^k): \qquad \min f(x) + t^k \beta(x) \quad \text{s.t.} \quad x \in M.$$

7 **until** $t^k < \varepsilon$
8 Setze $\bar{x} = x^k$.
9 **end**

Außerdem wird hier die numerische Behandlung der Probleme $P(t^k)$ für fallende t^k zunehmend schwieriger, da die Barrierefunktion β dann am Rand von M eine immer stärkere Krümmung aufweist. Bei der numerischen Lösung des Hilfsproblems $P(t^k)$ mit einem der Verfahren aus Abschn. 2.2 ist darauf zu achten, dass eine zu große Schrittweite zur Verletzung der theoretisch inaktiven Restriktionen $g_i(x) \leq 0$, $i \in I$, führen kann. Daher muss man diese Nebenbedingung numerisch explizit berücksichtigen und gegebenenfalls hinreichend kleine Schrittweiten wählen.

Immerhin ist für $t > 0$ jeder optimale Punkt von $P(t$ ein innerer Punkt von$M$, Barriereverfahren sind also (primale) Innere-Punkte-Methoden. Falls man die numerische Iteration bei einer Inneren-Punkte-Methode nach wenigen Schritten beenden muss, ohne ein Abbruchkriterium erfüllt zu haben, liegt immerhin ein zulässiger Punkt mit üblicherweise verbessertem Zielfunktionswert vor.

Wir stellen das Wolfe-dual Problem auf:

$$D: \quad \max_{x,\lambda} L(x,\lambda) \;\text{ s.t. }\; \nabla_x L(x,\lambda) = 0, \lambda \geq 0$$

Ferner sei v_D das Supremum der Funktion L über der zulässigen Menge von D und v_P das Infimum der Funktion f über der zulässigen Menge von P. Dann gilt für jedes konvexe C^1-Problem P nach dem schwachen Dualitätssatz $v_D \leq v_P$. Die demnach nichtnegative Differenz $v_P - v_D$ wird Dualitätslücke genannt. Für jeden zulässigen Punkt $\overline{x}$ von P und jedes $\overline{\lambda}$, so dass $(\overline{x}, \overline{\lambda})$ zulässig für D ist, bildet die Differenz

$$f(\overline{x}) - L(\overline{x}, \overline{\lambda}) \geq v_P - v_D \geq 0$$

per Definition von v_P und v_D eine Oberschranke für die Dualitätslücke. Findet man also einen primal zulässigen Punkt $\overline{x}$ und einen zugehörigen dual zulässigen Punkt $(\overline{x}, \overline{\lambda})$ mit $f(\overline{x}) - L(\overline{x}, \overline{\lambda}) = 0$, so ist $\overline{x}$ ein global minimaler Punkt für P und $(\overline{x}, \overline{\lambda})$ ein global maximaler Punkt für D.

Da die Bedingung

$$0 = f(\overline{x}) - L(\overline{x}, \overline{\lambda}) = -\sum_{i \in I} \overline{\lambda}_i g_i(\overline{x})$$

gemeinsam mit der primalen und dualen Zulässigkeit genau bedeutet, dass $\overline{x}$ ein KKT-unkt von P mit Multiplikatoren $\overline{\lambda}$ ist, haben wir damit einerseits einen alternativen, auf Dualitätstheorie basierenden Beweis für Satz 3.2.77 erbracht. Andererseits werden wir im Folgenden sehen, dass auch das Abbruchkriterium $t^k < \epsilon$ aus Algorithmus 3.3 eine Oberschranke für die Dualitätslücke liefert.

Dazu schreiben wir auf, was die Fermat'sche Regel (Satz 2.1.13) für den unrestringierten Optimalpunkt x^k von $P(t^k)$ mit

$$\lambda_i^k := -\frac{t^k}{g_i(x^k)}, \; i \in I$$

erfüllt dieser also die Gleichung

$$\nabla_x L(c^k, \lambda^k) = 0$$

Wegen $t^k > 0$ und $g_i(x^k) < 0$, $i \in I$, gilt ferner $\lambda^k > 0$, so dass (x^k, λ^k) insgesamt ein zulässiger Punkt des Wolfe-dualen Problems D ist. Da außerdem x^k primal zulässig ist, erhalten wir aus den obigen Überlegungen

$$0 \leq v_p - v_D \leq f(x^k) - L(x^k, \lambda^k) = -\sum_{i \in I} \lambda_i^k g_i(x^k) = -\sum_{i \in I} \left(-\frac{t^k}{g_i(x^k)} \right) g_i(x^k) = pt^k$$

Eine Interpretation des Abbruchkriteriums $t^k < \epsilon$ in Algorithmus 3.3 besteht also darin, dass dann die Dualitätslücke beim Terminieren unter dem Wert $p\epsilon$ liegt.

Da das Barriereverfahren allerdings ein rein primales Verfahren ist, formulieren wir noch ein aus diesen Überlegungen folgendes rein primales Resultat zum Abbruchkriterium.

Satz 3.3.15. Bei Anwendung von Algorithmus 3.3 auf ein C^1-Problem P mit auf $\mathbb{R}^n$ konvexen Funktionen f und g_i, $i \in I$, sowie mit der logarithmischen Barrierefunktion $\beta(x) - \log(-g_i(x))$ erfüllt der generierte Punkt $\overline{x}$ die Ungleichungen

$$v_P \leq f(\overline{x}) < v_p + p\epsilon$$

Wegen der primalen Zulässigkeit des Punkts $\overline{x}$ handelt es sich bei ihm nach Satz 3.3.15 um einen sogenannten $(p\ \epsilon)$-optimalen Punkt von P. Instabilität folgt auch aus der Wahl von λ^k, da sowohl Nenner als auch Zähler gegen 0 streben.

Primal-duale Innere-Punkte-Methoden

Die im vorhergehenden Abschnitt hergeleitete Optimalitätsbedingung erster Ordnung für x^k in $P(t^k)$ ist äquivalent zur Lösbarkeit des Systems

$$\begin{aligned} \nabla f(x) + \sum_{i \in I} \lambda_i \nabla g_i(x) &= 0, \\ g_i(x) &< 0, i \in I \\ \lambda_i &> 0, i \in I \\ \lambda_i \cdot g_i(x) + t &= 0, i \in I \end{aligned}$$

durch (t^k, x^k, λ^k). Dieses System ist offenbar eng mit dem KKT-System von P verwandt, indem nämlich dessen Komplementaritätsbedingungen durch den Parameter t gestört werden. Entscheidend an dieser Beobachtung ist, dass beim Grenzübergang $t \searrow 0$ keine numerische Instabilität mehr zu erwarten ist. Dies führt auf die folgende Klasse von

Karush-Kuhn-Tucker-Verfahren. Grundidee dieser Verfahren ist es, für $t \searrow 0$ Nullstellen von

$$\mathfrak{T}(t, x, \lambda) = \begin{pmatrix} \nabla f(x) + \sum_{i \in I} \lambda_i \nabla g_i(x) \\ \operatorname{diag}(\lambda) \cdot g(x) + t \cdot e \end{pmatrix}$$

mit $g(x(t)) < 0$ und $\lambda(t) > 0$ zu berechnen, wobei e wieder den Einservektor bezeichnet. Im Fall der Konvergenz $x(t) \to \overline{x}$ und $\lambda(t) \to \overline{\lambda}$ ist $\overline{x}$ offenbar KKT-Punkt von P mit Multiplikator $\overline{\lambda}$. Vorteil dieses Ansatzes ist, dass weder Probleme mit numerischer Instabilität noch mit dem Definitionsbereich einer Barrierefunktion auftreten. In der Tat könnte man Nullstellen von $\mathfrak{T}(t, x, \lambda)$ numerisch sogar für $t < 0$ suchen, was inhaltlich natürlich nicht sinnvoll ist.

Satz 3.3.17. Es sei $\overline{x}$ ein nichtdegenerierter lokaler Minimalpunkt von P mit Multiplikator $\overline{\lambda}$. Dann gibt es eine Umgebung V von 0 und ($C^{k-1}-$)Funktionen $x(t)$, $\lambda(t)$ auf V mit $(x(0), \lambda(0)) = (\overline{x}, \overline{\lambda})$, $\mathcal{T}(t, x(t), \lambda(t)) = 0$ für alle $t \in V$ sowie $g(x(t)) < 0$ und $\lambda(t) > 0$ für alle positiven $t \in V$.

Definition 3.3.18 (Primal-dualer zentraler Pfad)**.** Mit den Bezeichnungen aus Satz 3.3.17 heißt die Menge

$$C_{CP} = \big\{(x(t), \lambda(t)) \mid t \in V, t > 0\big\}$$

primal-dualer zentraler Pfad an $(\overline{x}, \overline{\lambda})$.

Da der Rechenaufwand zur Identifizierung eines ϵ-genauen Minimalpunkts selbst im Worst Case nur polynomial in der Problemdimension anwächst, und dies insbesondere für lineare Optimierungsprobleme gilt, sind primal-duale InnerePunkte-Methoden in dieser Hinsicht dem Simplex-Algorithmus überlegen, der im Worst Case exponentiellen Rechenaufwand besitzt.

SQP-Verfahren

Verfahren des Sequential Quadratic Programming (SQP) nutzen die Idee des Newton-Verfahrens, eine Lösung eines nichtlinearen Problems durch die sukzessive Lösung von Linearisierungen anzunähern.

Wählt man als nichtlineares Problem dabei zunächst naiv das Optimierungsproblem P selbst, dann ist die Linearisierung ein lineares Optimierungsproblem, das sich etwa mit dem Simplex-Algorithmus oder mit primal-dualen Innere-Punkte-Methoden lösen lässt. Um die Lösung dieser Linearisierung wird P dann erneut linearisiert, die Linearisierung gelöst und so fort. Dieser Ansatz führt etwa auf das Verfahren der zulässigen Richtungen von Zoutendijk. Wegen schlechter Identifizierung aktiver Indizes können bei diesem Verfahren sogenanntes Jamming und Konvergenz gegen einen nichtkritischen Punkt auftreten. Für eine erfolgreichere Weiterentwicklung, die unter dem Namen Sequential Linear Programming (SLP) bekannt ist, sei auf verwiesen.

Die Linearisierung des Problems P selbst verallgemeinert allerdings gar nicht die Newton-Idee aus der unrestringierten Optimierung. Dort wird nicht die Zielfunktion linearisiert (was in Abwesenheit von Nebenbedingungen ja üblicherweise zu einem unbeschränkten Problem führen würde), sondern die Optimalitätsbedingung erster Ordnung $\nabla f(x) = 0$: Man setzt $x^{k+1} = x^k + d^k$ mit einer Lösung d^k der um x^k linearisierten Bedingung erster Ordnung

$$\nabla f(x^k) + D^2 f(x^k) \cdot d^k = 0$$

Wie wir bereits in Übung 2.2.43 gesehen haben, ist diese Gleichung sogar wieder eine Optimalitätsbedingung erster Ordnung, nämlich für das quadratische Optimierungsproblem

$$Q^k: \quad \min_{d \in \mathbb{R}^n} \langle \nabla f(x^k), d \rangle + \frac{1}{2} d^T D^2 f(x^k) d$$

Bei Konvergenz der Folge (x^k) gegen einen nichtdegenerierten lokalen Minimalpunkt ist die Matrix $D^2 f(x^k)$ für alle hinreichend großen k aus Stetigkeitsgründen positiv definit, so dass es sich bei Q^k zusätzlich um ein konvexes Optimierungsproblem handelt. Die Lösung d^k der linearisierten Bedingung erster Ordnung ist dann globaler Minimalpunkt von Q^k.

Das ungedämpfte Newton-Verfahren für unrestringierte Probleme besteht also im Wesentlichen darin, eine Folge quadratischer Optimierungsprobleme zu lösen, was zu der Bezeichnung Sequential Quadratic Programming (SQP) führt.Zu vermuten ist also, dass man für eine Verallgemeinerung des Newton-Verfahrens auf den restringierten Fall die Funktionen f, g_i, $i \in I$, h_j, $j \in J$, quadratisch approximieren und die entstehenden Probleme mit quadratischer Zielfunktion und quadratischen Nebenbedingungen lösen muss. Tatsächlich haben die Hilfsprobleme aber eine viel einfachere Struktur.

Um dies zu sehen, betrachten wir zunächst das Newton-Verfahren zur Lösung des KKT-Systems von P in Abwesenheit von Ungleichungsrestriktionen, d. h. für $I = \emptyset$. Gesucht ist also eine Nullstelle von

$$\mathfrak{T}(x, \mu) = \begin{pmatrix} \nabla_x L(x, \mu) \\ h(x) \end{pmatrix}$$

Zu gegebenen x^k und μ^k setzt das Newton-Verfahren dafür $\begin{pmatrix} x^{k+1} \mu^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \mu^k \end{pmatrix} + \begin{pmatrix} d^k \\ \sigma^k \end{pmatrix}$ mit einer Lösung (d^k, σ^k) von

$$0 = \mathfrak{T}(x^k, \mu^k) + D\mathfrak{T}(x^k, \mu^k) \begin{pmatrix} d \\ \sigma \end{pmatrix} \tag{3.27}$$

Lemma 3.3.19. Es sei $\overline{x}$ ein nichtdegenerierter lokaler Minimalpunkt von P mit Multiplikator $\overline{\mu}$, und der Startpunkt (x^0, μ^0) liege hinreichend nahe bei $(\overline{x}, \overline{\mu})$. Dann konvergiert (x^k, μ^k) quadratisch gegen $(\overline{x}, \overline{\mu})$.

Wir können festgestellen, dass man die Lösung (d^k, σ^k) von (3.27) auch als KKTPunkt und Multiplikator von

$$Q^k: \quad \min_{d \in \mathbb{R}^n} \langle \nabla f(x^k), d \rangle + \frac{1}{2} d^T D_x^2 L(x^k, \mu^k) d \;\text{ s.t. }\; h(x^k) + Dh(x^k) d = 0$$

gewinnen kann. Im Gegensatz zur obigen Vermutung ist also nur die Zielfunktion des Hilfsproblems quadratisch, während die Nebenbedingungen sogar linear sind. Probleme dieses Typs heißen quadratische Optimierungsprobleme, man erhält also wieder ein SQPVerfahren. Angemerkt sei ferner, dass man für $J = \emptyset$ gerade das Problem Q^k aus dem unrestringierten Fall erhält, so dass wir in der Tat eine Verallgemeinerung des NewtonVerfahrens vom unrestringierten auf den restringierten Fall entwickelt haben.