Technische Universität Dresden

Skript:

Optimierung I

Verfasser Franziska Kühn

Daten Prof. Dr. Andreas Fischer
Wintersemester 2010/11
Hauptstudium

Inhaltsverzeichnis

1 Einführung und Grundlagen

1.1 Optimierungsprobleme

1.1.1 Beispiele

1. Data Fitting:

   - gegeben: Datenpaare $(s_j, t_j) \in \mathbb{R} \times \mathbb{R}$ für $j = 1, \ldots, l$ sowie eine Funktion $\varphi : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$. Dann kann es sinnvoll und unter bestimmten Voraussetzungen möglich sein, einen Parametervektor $x^* \in \mathbb{R}^n$ so zu bestimmen, dass $f(x^*) \leq f(x)$ für alle $x \in \mathbb{R}^n$, wobei

   $$f(x) := \sum_{j=1}^{l} (t_j - \varphi(s_j, x))^2 \qquad (x \in \mathbb{R}^n)$$

   Im Allgemeinen ist nicht gesichert, dass ein $x^* \in \mathbb{R}^n$ existiert mit

   $$f(x^*) = \inf_{x \in \mathbb{R}^n} f(x)$$

2. Energiekostenoptimale Pumpensteuerung:

   - Zeitintervall $[0, T]$, Pumpe mit regelbarer Förderleistung $x : [0, T] \to [0, x_{\max}]$ (in $m^3 \cdot h^{-1}$) pumpt Wasser aus Brunnen durch eine Leitung in Hochbehälter. Aus dieser Leitung entnimmt ein Verbraucher Wasser mit der gegebenen Entnahmeleistung $v : [0, T] \to [0, v_{\max}]$ (in $m^3 \cdot h^{-1}$). Der Hochbehälter fasst höchstens $H_{\max}\ m^3$ und soll zur Sicherheit stets $H_{\min}\ m^3$ enthalten. Am Anfang des Zeitintervalls seien $H_0\ m^3$ Wasser im Behälter.
   - Die Energiekosten zum Betrieb im Zeitintervall $[0, T]$ seien gegeben durch

   $$f(x) := \int_{t=0}^{T} w(t, x(t))\, dt$$

   mit einer gegebenen Funktion $w : [0, T] \times [0, x_{\max}] \to [0, \infty)$. Die Funktion w bringt insbesondere

     – zeitabhängige Tarife
     – nichtlineare Abhängigkeit zwischen elektrischer Leistung und der Förderleistung der Pumpte

   zum Ausdruck.

   - Das Ziel besteht nun darin eine Funktion x^* aus einem geeigneten Funktionenraum X zu bestimmen, derart dass $f(x^*) \leq f(x)$ für alle $x \in X$, die einen ordnungsgemäßen Betrieb der Anlage sicherstellen. Letzteres ist gleichbedeutend damit, dass $x \in X$ die folgenden Nebenbedingungen erfüllt:

     (i) Die Förderleistung ist nichtnegativ und nach oben beschränkt, d.h.

     $$\forall t \in [0, T] : 0 \leq x(t) \leq x_{\max}$$

(ii) Die Wassermenge im Behälter liegt stets im Intervall $[H_{\min}, H_{\max}]$, d.h. für alle $t \in [0,T]$ gilt:

$$H_{\min} \leq H_0 + \int_{\tau=0}^{t} (x(\tau) - v(\tau))\, d\tau \leq H_{\max}$$

- Reduktion auf ein Optimierungsproblem im Endlichdimensionalen:

  Sei $T_0 := 0 < T_1 < \ldots < T_n := T$. Mit X_n bezeichnen wir den linearen Raum aller Funktionen $x : [0,T] \to \mathbb{R}$ mit folgenden Eigenschaften:

  (i) $\forall i = 0, \ldots, n-1 : \forall s,t \in [T_i, T_{i+1}) : x(t) = x(s)$

  (ii) $x(T_{n-1}) = x(T_n)$

  Wir setzen voraus, dass $v \in X_n$. Weiterhin seien $\sigma \in X_n$ (Energietarif) und eine (pumpenabhängige) stetige Funktion $\varrho : [0,\infty) \to [0,\infty)$ gegeben und $w : [0,T] \times [0, x_{\max}] \to [0,\infty)$ definiert durch

  $$w(t,r) := \sigma(t) \cdot \varrho(r)$$

  Da jede Treppenfunktion $x \in X_n$ durch die Werte $x(T_0), \ldots, x(T_{n-1})$ eindeutig charakterisiert ist, werden wir unter x den Vektor

  $$\begin{pmatrix} x(T_0) \\ \vdots \\ x(T_{n-1}) \end{pmatrix} =: \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

  verstehen. Entsprechendes gilt für σ, v. Als Spezialfall der obigen Optimierungsaufgabe die Funktion f unter den angegebenen Nebenbedingungen zu minimieren erhalten wir damit

  $$\sum_{i=1}^{n} \sigma_i \cdot \varrho(x_i) \cdot (T_i - T_{i-1}) \to \min$$

  bei

  (i) $\forall i = 1, \ldots, n : 0 \leq x_i \leq x_{\max}$

  (ii) Für $i = 1, \ldots, n$:

  $$H_{\min} \leq H_0 + \sum_{l=1}^{i} (x_l - v_l) \cdot (T_l - T_{l-1}) \leq H_{\max}$$

1.1.2 Aufgabenstellung und Grundbegriffe

- Seien $G \subseteq \mathbb{R}^n$, $f : G \to \mathbb{R}$ gegeben. Dann betrachten wir die Aufgabe, einen Punkt $x^* \in G$ zu finden, sodass

  $$\forall x \in G : f(x^*) \leq f(x) \tag{1.1}$$

- Ein solcher Punkt $x^* \in G$ wird dementsprechend *Lösung* der Aufgabe (1.2) genannt. Formal schreiben wir diese Aufgabe in der Form

  $$f(x) \to \min \text{ bei } x \in G \tag{1.2}$$

  Dabei heißt f *Zielfunktion*, G *zulässiger Bereich*. Entsprechend heißt x *zulässiger Punkt*, wenn $x \in G$. Der Funktionswert $f(x^*) = f_{\min}$ heißt *Optimalwert*.

- Falls $G = \mathbb{R}^n$, so spricht man von einem *unrestringierten* oder *freien Optimierungsproblem*.

- Häufig ist es sinnvoll die Bedingung (1.1) nur lokal zu stellen. Man verlangt dann, dass eine Umgebung $U(x^*)$ des Punktes x^* existiert, sodass

  $$\forall x \in G \cap U(x^*) : f(x^*) \leq f(x) \tag{1.3}$$

  gilt. Ein Punkt $x^* \in G$ der dieser Bedingung (1.3) genügt, wird *lokale Lösung* der Aufgabe (1.2) genannt.

- Erfüllt x^* die Bedingung (1.1), dann nennt man x^* im Gegensatz dazu *globale Lösung*.
- Von einer *isolierten lokalen Lösung* spricht man, wenn in einer Umgebung von x^* keine weitere lokale Lösung existiert. Dann folgt $f(x^*) < f(x)$ für alle $x \in G$ in einer Umgebung von x^*. Ist die letzte Eigenschaft erfüllt, dann nennt man x^* eine *strenge lokale Lösung*.
- Zur globalen Optimierung ist Satz von Weierstraß nützlich: Ist $f : G \to \mathbb{R}$ stetig, $G \subseteq \mathbb{R}^n$ mit $G \neq \emptyset$ kompakt, dann gibt es ein $x^* \in G$ mit

$$f(x^*) = \inf_{x \in G} f(x)$$

- **Beispiel 1.1** Sei $G \subset \mathbb{R}^2$ gegeben durch

$$G := \{x \in \mathbb{R}^2; g_1(x) := x_1^2 + x_2^2 \leq 1, g_2(x) := -x_2 \leq 0\}$$

Für die Aufgabe $f(x) := -x_2 \to \min$ ist einzige globale Lösung $x^* = (0,1)^T$. Für $f(x) := x_2 \to \min$ sind alle Elemente $(x_1, x_2) \in [-1,1] \times \{0\}$ globale Lösungen.

- Der zulässige Bereich G wird häufig mit Hilfe von Restriktionsfunktionen $g : \mathbb{R}^n \to \mathbb{R}^m$ und $h : \mathbb{R}^n \to \mathbb{R}^p$ definiert:

$$x \in G :\Leftrightarrow \forall i \in I := \{1, \ldots, m\} : g_i(x), \forall j \in J := \{1, \ldots, p\} : h_j(x) = 0 \tag{1.4}$$

Kurz schreibt man $g(x) \leq 0$ bzw. $h(x) = 0$ anstelle von $g_i(x) \leq 0$ für $i \in I$ und $h_j(x) = 0$ für $j \in J$. Das Optimierungsproblem (1.2) erhält damit die Gestalt

$$f(x) \to \min \quad \text{bei } g(x) \leq 0, h(x) = 0 \tag{1.5}$$

- Spezialfälle:
  1. lineares Optimierungsproblem: f, g, h sind affin-linear, z.B.

$$c^T \cdot x + c_0 \to \min \quad \text{bei } A \cdot x = b, x \geq 0$$

  2. quadratisches Optimierungsproblem: f quadratisch, d.h.

$$f(x) = \frac{1}{2} \cdot x^T \cdot A \cdot x + q^T \cdot x + q_0$$

  und g, h affin-linear.

1.1.3 Konvexitätsbegriffe

Definition 1.1

- Eine Menge $G \subseteq \mathbb{R}^n$ heißt *konvex*, wenn zu je zwei Punkten $x, y \in G$ auch deren Verbindungsstrecke vollständig zu G gehört, d.h. falls für alle $x, y \in G$:

$$\forall \lambda \in (0,1) : \lambda \cdot x + (1-\lambda) \cdot y \in G$$

- Sei $G \subseteq \mathbb{R}^n$ konvex. Eine Funktion $f : G \to \mathbb{R}$ heißt
  1. *konvex auf G*

$$:\Leftrightarrow \forall (x, y, \lambda) \in G \times G \times (0,1) : f(\lambda \cdot x + (1-\lambda) \cdot y) \leq \lambda \cdot f(x) + (1-\lambda) \cdot f(y)$$

  2. *streng konvex auf G*

$$:\Leftrightarrow \forall (x, y, \lambda) \in G \times G \times (0,1), x \neq y : f(\lambda \cdot x + (1-\lambda) \cdot y) < \lambda \cdot f(x) + (1-\lambda) \cdot f(y)$$

  3. *gleichmäßig konvex auf G* $:\Leftrightarrow \exists \gamma > 0 \forall (x, y, \lambda) \in G \times G \times (0,1)$:

$$f(\lambda \cdot x + (1-\lambda) \cdot y) \leq \lambda \cdot f(x) + (1-\lambda) \cdot f(y) - \gamma \cdot \lambda \cdot (1-\lambda) \cdot \|x - y\|^2$$

  4. *konkav auf G* $:\Leftrightarrow -f$ konvex auf G.

  Entsprechend werden die anderen Bezeichnungen übertragen.

Lemma 1.1 Seien $G \subseteq \mathbb{R}^n$ konvex und B eine offene Menge mit $G \subseteq B \subseteq \mathbb{R}^n$. Die Funktion $f : B \to \mathbb{R}$ sei differenzierbar auf G. Dann gilt:

1. f ist genau dann konvex auf G, wenn für alle $x, y \in G$:
$$f(y) - f(x) \geq \nabla f(x)^T \cdot (y - x)$$
2. f ist genau dann streng konvex, wenn für alle $x, y \in G$ mit $x \neq y$:
$$f(y) - f(x) > \nabla f(x)^T \cdot (y - x)$$
3. f ist genau dann gleichmäßig konvex auf G, wenn es $\gamma > 0$ gibt, sodass für alle $x, y \in G$:
$$f(y) - f(x) \geq \nabla f(x)^T \cdot (y - x) + \gamma \cdot \|x - y\|^2$$

Beweis: Übung

Definition 1.2 Eine Matrix $M \in \mathbb{R}^{n \times n}$ heißt

1. *positiv semidefinit* $:\Leftrightarrow \forall s \in \mathbb{R}^n : s^T \cdot M \cdot s \geq 0$
2. *positiv definit* $:\Leftrightarrow \forall s \in \mathbb{R}^n \backslash \{0\} : s^T \cdot M \cdot s > 0$

Lemma 1.2 Sei $G \subseteq \mathbb{R}^n$ eine offene, konvexe Menge, $f : G \to \mathbb{R}$ zweimal stetig differenzierbar. Dann gilt:

1. f konvex auf $G \Leftrightarrow \nabla^2 f(x)$ positiv semidefinit für alle $x \in G$
2. f streng konvex auf G, wenn $\nabla^2 f(x)$ positiv definit für alle $x \in G$. Die Umkehrung gilt im Allgemeinen nicht.
3. f gleichmäßig konvex auf $G \Leftrightarrow \exists \gamma > 0 \, \forall (s, x) \in \mathbb{R}^n \times G$:
$$s^T \cdot \nabla^2 f(x) \cdot s \geq \gamma \cdot \|s\|^2$$

Beweis:

3. „$\Rightarrow$“
   Für $x \in G, s \in \mathbb{R}^n \backslash \{0\}$ beliebig, aber fest, gilt wegen der Offenheit von G: $x + \alpha \cdot s \in G$ für alle $\alpha > 0$ hinreichend klein. Mit der Taylorformel und Lemma 1.1(3) erhält man:
$$\begin{aligned} f(x + \alpha \cdot s) - f(x) - \alpha \cdot \nabla f(x)^T \cdot s &= \frac{1}{2} \cdot \alpha^2 \cdot s^T \cdot \nabla^2 f(x) \cdot s + o(\alpha^2) \\ \overset{1.1(3)}{\Rightarrow} \gamma \cdot \alpha^2 \cdot \|s\|^2 &\leq \frac{1}{2} \cdot \alpha^2 \cdot s^T \cdot \nabla^2 f(x) \cdot s + o(\alpha^2) \\ \overset{\alpha > 0}{\Rightarrow} \gamma \cdot \|s\|^2 &\leq \frac{1}{2} s^T \cdot \nabla^2 f(x) \cdot s + \underbrace{\frac{o(\alpha^2)}{\alpha^2}}_{\to 0 \, (\alpha \to 0)} \\ \Rightarrow \gamma \cdot \|s\|^2 &\leq \frac{1}{2} \cdot s^T \cdot \nabla^2 f(x) \cdot s \end{aligned}$$
   „$\Leftarrow$“
   Mit der Taylorformel folgt aus Voraussetzung für $x, y \in G$:
$$\begin{aligned} f(y) - f(x) - \nabla f(x)^T \cdot (y - x) &= \frac{1}{2} (y - x)^T \cdot \nabla^2 f(x + \xi \cdot (y - x)) \cdot (y - x) \\ &\geq \frac{\gamma}{2} \cdot \|y - x\|^2 \end{aligned}$$
   wobei $\xi \in (0, 1)$. Wegen Lemma 1.1(3) folgt Behauptung.

Theorem 1.1 Sei $G \subseteq \mathbb{R}^n$ konvex und $f : G \to \mathbb{R}$ konvex. Dann gilt:

1. Jede lokale Lösung von (1.2) ist auch eine globale Lösung von (1.2).
2. Ist f sogar streng konvex, dann gibt es höchstens eine Lösung.
3. Falls G auch abgeschlossen, $G \neq \emptyset$, f gleichmäßig konvex und differenzierbar ist, so besitzt das Problem (1.2) genau eine Lösung.

Beweis:

1. Sei $x^* \in G$ eine lokale Lösung von (1.2). Angenommen x^* ist keine globale Lösung, dann existiert ein $y \in G$ mit $f(y) < f(x^*)$. Da G konvex ist, gilt $\lambda \cdot y + (1-\lambda) \cdot x^* \in G$ für alle $\lambda \in (0,1)$. Konvexität von f liefert für $\lambda \in (0,1)$:

$$\begin{aligned} f(\lambda \cdot y + (1-\lambda) \cdot x^*) &\leq \lambda \cdot f(y) + (1-\lambda) \cdot f(x^*) \\ &< f(x^*) \\ \Rightarrow f(x^* + \lambda \cdot (y - x^*)) &< f(x^*) \end{aligned}$$

   also ist x^* keine lokale Lösung. Widerspruch!
2. Übung 1, Aufgabe 4
3. Übung 1, Aufgabe 4

Definition 1.3 Seien $G \subseteq \mathbb{R}^n$ konvex, $B \supseteq G$ offen, $f : G \to \mathbb{R}$ heißt *quasikonvex* auf G

$$:\Leftrightarrow \forall (x, y, \lambda) \in G \times G \times (0,1) : f(\lambda \cdot x + (1-\lambda) \cdot y) \leq \max\{f(x), f(y)\}$$

Eine auf G differenzierbare Funktion $f : B \to \mathbb{R}$ heißt *pseudokonvex* auf G

$$:\Leftrightarrow \forall (x, y) \in G \times G : \{(y-x)^T \cdot \nabla f(x) \geq 0 \Rightarrow f(y) \geq f(x)\}$$

Beispiele:

1. $f(x) = x^3$ ist quasikonvex (monoton wachsend), aber nicht pseudokonvex (Wähle $x = 0$ in der Definition).
2. $f(x) := -\frac{1}{1+x^2}$ ist quasi- und pseudokonvex. Pseudokonvexität:

$$\begin{aligned} f'(x) &= \frac{2x}{(1+x^2)^2} \\ \Rightarrow (y-x)^T \cdot \nabla f(x) \geq 0 &\Leftrightarrow y \cdot x \geq x^2 \\ &\Leftrightarrow \begin{cases} y \geq x & x > 0 \\ y \leq x & x < 0 \end{cases} \end{aligned}$$

Lemma 1.3 Sei $G \subseteq \mathbb{R}^n$ konvex, $B \supseteq G$ offen. Eine auf G konvexe Funktion ist dort auch quasikonvex. Eine auf G differenzierbare konvexe Funktion $f : B \to \mathbb{R}$ ist dort auch pseudokonvex. Ist $f : B \to \mathbb{R}$ differenzierbar und pseudokonvex auf G, dann ist f auch quasikonvex.

Beweis: Übungsaufgabe

1.2 Optimalitäts- und Regularitätsbedingungen

1.2.1 Optimalitätsbedingungen mit zulässigen Richtungen

Sei $G \subseteq \mathbb{R}^n$, dann *Kegel der zulässigen Richtungen*:

$$Z(x) := \operatorname{cone}\{d \in \mathbb{R}^n; \forall \alpha \in [0,1] : x + \alpha \cdot d \in G\}$$

wobei die *Kegelhülle* $\operatorname{cone} S$ einer Menge $S \subseteq \mathbb{R}^n$ definiert ist durch

$$\operatorname{cone}(S) := \{\lambda \cdot s; \lambda \in [0, \infty)\}$$

Theorem 1.2 Es seien $G \subseteq \mathbb{R}^n$ und $B \supseteq G$ offen. $f : B \to \mathbb{R}$ sei differenzierbar auf G.

1. Falls x^* eine lokale Lösung von (1.2) ist, dann gilt:
$$\forall d \in Z(x^*) : \nabla f(x^*)^T \cdot d \geq 0 \tag{1.6}$$
Falls G zusätzlich konvex ist, gilt
$$\forall x \in G : \nabla f(x^*)^T \cdot (x - x^*) \geq 0 \tag{1.7}$$
2. Ist G konvex und f pseudokonvex auf G, so ist $x^* \in G$ zusammen mit der Bedingung (1.7) hinreichend dafür, dass x^* Lösung von (1.2) ist.

Beweis:

1. Sei $d \in Z(x^*)$ beliebig, aber fest. Da x^* lokale Lösung von (1.2) ist, muss es $\tilde{\alpha} > 0$ geben, sodass
$$x^* + \alpha \cdot d \in G \qquad f(x^*) \leq f(x^* + \alpha \cdot d)$$
für alle $\alpha \in [0, \tilde{\alpha}]$. Aus der Differenzierbarkeit von f folgt:
$$\nabla f(x^*)^T \cdot d = \lim_{\alpha \to 0} \frac{1}{\alpha} \cdot \underbrace{(f(x^* + \alpha \cdot d) - f(x^*))}_{\geq 0} \geq 0$$
Es sei nun G zusätzlich konvex, $x \in G$ beliebig, aber fest. Dann $d := x - x^* \in Z(x^*)$. Somit folgt (1.7) aus (1.6).
2. Setzen wir nun voraus, dass G konvex, f pseudokonvex auf G. Dann erhält man aus (1.7) und Definition 1.3 (Setze $(x, y) := (x^*, x)$) die Behauptung:
$$\forall x \in G : f(x) \geq f(x^*)$$

1.2.2 Optimalitätsbedingung mit Tangentialkegel

Tangentialkegel
$$T(x) := \left\{ d = \lim \frac{x^\nu - x}{t_\nu}; (x^\nu) \subseteq G, (t_\nu) \in (0, \infty), \lim t_\nu = 0 \right\}$$

Theorem 1.3 Es sei $G \subseteq \mathbb{R}^n$ und B offen mit $G \subseteq B$. Die Funktion $f : B \to \mathbb{R}$ sei differenzierbar auf G. Falls x^* lokale Lösung von (1.2) ist, dann gilt:
$$\forall d \in T(x^*) : \nabla f(x^*)^T \cdot d \geq 0 \tag{1.8}$$

Beweis:

- Angenommen es existiert ein $d \in T(x^*)$ mit $\nabla f(x^*)^T \cdot d < 0$. Dann muss es Folgen $(x^\nu) \subset G$, $(t_\nu) \in (0, \infty)$ geben, sodass
$$\lim x^\nu = x^*; \lim t_\nu = 0, \lim \frac{x^\nu - x^*}{t_\nu} = d \tag{1.9}$$
Wegen Differenzierbarkeit von f gibt es zu jedem $\varepsilon > 0$ ein $\nu(\varepsilon)$ mit $\nu(\varepsilon) > \varepsilon^{-1}$, sodass
$$\left| f(x^{\nu(\varepsilon)}) - f(x^*) - \nabla f(x^*)^T \cdot (x^{\nu(\varepsilon)} - x^*) \right| \leq \varepsilon \cdot \|x^{\nu(\varepsilon)} - x^*\|$$
Weiter ergibt sich damit:
$$\left| \frac{f(x^{\nu(\varepsilon)}) - f(x^*)}{t_{\nu(\varepsilon)}} - \frac{\nabla f(x^*)^T \cdot (x^{\nu(\varepsilon)} - x^*)}{t_{\nu(\varepsilon)}} \right| \leq \varepsilon \cdot \frac{\|x^{\nu(\varepsilon)} - x^*\|}{t_{\nu(\varepsilon)}}$$
Für $\varepsilon \to 0$ folgt mit (1.9) und $\nabla f(x^*) \cdot d < 0$, dass
$$f(x^{\nu(\varepsilon)}) - f(x^*) < 0$$
für alle hinreichend kleinen $\varepsilon > 0$. Da (x^ν) gegen x^* konvergiert, ergibt sich somit ein Widerspruch zur Voraussetzung, dass x^* ein lokales Minimum ist.

1.2.3 Optimalitätsbedingung mit Linearisierungskegel

Linearisierungskegel:

$$L(x) := \left\{ d \in \mathbb{R}^n; \forall i \in I_0(x) : \nabla g_i(x)^T \cdot d \leq 0, \nabla h(x)^T \cdot d = 0 \right\}$$

für Probleme vom Typ (1.5), wobei

$$I_0(x) := \{ i \in I; g_i(x) = 0 \}$$

(Indexmenge der in x aktiven Restriktionen)

Lemma 1.4 Die Funktionen $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$ seien differenzierbar. Mit G entsprechend (1.4) gilt dann $T(x) \subseteq L(x)$ für alle $x \in G$.

Beweis:

- Sei $x \in G$ beliebig, aber fest, und $d \in T(x)$. Dann existieren Folgen $(x^\nu) \subseteq G$, $(t_\nu) \subseteq (0, \infty)$ mit (1.9). Für $i \in I_0(x)$ ergibt sich analog zum Beweis von Theorem 1.3, dass zu jedem $\varepsilon > 0$ ein $\nu(\varepsilon) \in \mathbb{N}$ existiert mit $\nu(\varepsilon) > \varepsilon^{-1}$ und

$$\left| \frac{g_i(x^{\nu(\varepsilon)}) - g_i(x)}{t_{\nu(\varepsilon)}} - \frac{\nabla g_i(x)^T \cdot (x^{\nu(\varepsilon)} - x)}{t_{\nu(\varepsilon)}} \right| \leq \varepsilon \cdot \frac{\|x^{\nu(\varepsilon)} - x\|}{t_{\nu(\varepsilon)}}$$

  Da $g_i(x^{\nu(\varepsilon)}) \leq 0, g_i(x) = 0$ und wegen (1.9) liefert der Grenzübergang $\varepsilon \to 0$:

$$\nabla g_i(x)^T \cdot d \leq 0$$

- Analoge Argumentation führt wegen $h_j(x^{\nu(\varepsilon)}) = 0 = h_j(x)$ zu $\nabla h_j(x)^T \cdot d = 0$ für $j \in J$.

Bemerkung:

- Die Umkehrung der Inklusion in Lemma 1.4 gilt im Allgemeinen nicht, siehe folgendes Beispiel.

Beispiel 1.3 Sei

$$G := \{ x \in \mathbb{R}^2; -x_1^3 + x_2 \leq 0, -x_2 \leq 0 \}$$

dann ist $x := (0,0)^T \in G$. Man erhält:

$$\begin{aligned} Z(x) &= \{ \lambda \cdot (1,0)^T; \lambda \geq 0 \} = T(x) \\ L(x) &= \left\{ d \in \mathbb{R}^2; \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \leq 0, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \cdot \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \leq 0 \right\} \\ &= \{ d \in \mathbb{R}^2; d_2 = 0 \} = \{ \lambda \cdot (1,0)^T; \lambda \in \mathbb{R} \} \\ &\supset T(x) \end{aligned}$$

Definition 1.4 Die Funktionen $g : \mathbb{R}^n \to \mathbb{R}^m$ und $h : \mathbb{R}^n \to \mathbb{R}^p$ seien differenzierbar, G durch (1.4) gegeben. Man sagt, dass im Punkt $x \in G$ die *Abadie Constraint Qualification* (ACQ) erfüllt ist, wenn $T(x) = L(x)$.

Theorem 1.4 Es seien $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$ differenzierbar. Wenn x^* eine lokale Lösung von (1.5) ist und in x^* die ACQ erfüllt ist, gilt

$$\forall d \in L(x) : \nabla f(x^*)^T \cdot d \geq 0 \tag{1.10}$$

Beweis: Folgt direkt aus Theorem 1.3 und Definition 1.4.

Bemerkung:

- Bedingung (1.10) ist also bei der Gültigkeit der ACQ eine notwendige Optimalitätsbedingung und wird *Karush-Kuhn-Tucker-Bedingung* genannt. Eine häufig benutzte Form dieser Bedingung wird in 1.2.5 hergeleitet.

1.2.4 Regularitätsbedingungen

Eine Bedingung, die ACQ impliziert, heißt Regularitätsbedingung.

Theorem 1.5 Es seien $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$ stetig differenzierbar. G sei durch (1.4) gegeben. Jede der folgenden Bedingungen ist hinreichend dafür, dass in $x \in G$ die ACQ erfüllt ist.

1. Mangasarian-Fromorik-Constraint-Qualification (MFCQ):

   Die Vektoren $\nabla h_1(x), \ldots, \nabla h_p(x)$ sind linear unabhängig und es gibt $s \in \mathbb{R}^n$, sodass $\nabla g_i(x)^T \cdot s > 0$ für alle $i \in I_0(x)$ und $\nabla h_j(x)^T \cdot s = 0$ für alle $j \in J$.

2. Slater-Bedingung: Seien $g_1, \ldots, g_m$ konvex, $J = \emptyset$. Außerdem gibt es $\bar{x} \in \mathbb{R}^n : g_i(\bar{x}) < 0$ für $i \in I$.

3. Linear independence constraint qualification (LICQ): Die Vektoren in der Familie

   $$\{\nabla g_i(x); i \in I_0(x)\} \cup \{\nabla h_j(x); j \in J\}$$

   sind linear unabhängig.

4. Die Funktionen g_i für $i \in I_0(x)$ und h_j für $j \in J$ sind affin-linear.

Beweis:

1. Ohne Beweis
2. Übungsaufgabe
3. Folgt wegen LICQ $\Rightarrow$ MFCQ
4. Übungsaufgabe

1.2.5 Karush-Kuhn-Tacker-Bedingungen

Lemma 1.5: (Farkas) Seien $A \in \mathbb{R}^{n \times m_0}$, $B \in \mathbb{R}^{n \times p}$, $c \in \mathbb{R}^n$. Dann ist von den beiden Systemen

$$A^T \cdot z \leq 0 \qquad B^T \cdot z = 0 \qquad c^T \cdot z > 0 \tag{1.11}$$

und

$$A \cdot u + B \cdot v = c \qquad u \geq 0 \tag{1.12}$$

genau eines lösbar.

Beweis:

- Angenommen, beide Systeme besitzen gleichzeitig eine Lösung z bzw. u, v. Dann erhält man aus (1.11) und (1.12) durch Multiplikation mit u^T, v^T bzw. z^T den folgenden Widerspruch:

  $$\begin{aligned} 0 \geq u^T \cdot A^T \cdot z &= z^T \cdot A \cdot u + z^T \cdot B \cdot v \\ &= z^T \cdot c > 0 \end{aligned}$$

  Also ist wenigstens eines der beiden Systeme nicht lösbar.

- Angenommen, beide Systeme (1.11) und (1.12) sind nicht lösbar. Dann gilt wegen der Nichtlösbarkeit von (1.12):

  $$c \notin \Gamma := \{x = A \cdot u + B \cdot v; u \geq 0\}$$

  Die Minimierungsaufgabe

  $$f(x) := (x - c)^T \cdot (x - c) \to \min_{x \in \Gamma}$$

hat einen abgeschlossenen, nichtleeren, konvexen zulässigen Bereich, eine gleichmäßige konvexe Zielfunktion und besitzt daher eine eindeutige Lösung $x^* \in \Gamma$ (nach Theorem 1.2). Wegen $c \notin \Gamma$ gilt $z := c - x^* \neq 0$, also $z^T \cdot z > 0$. Die Optimalitätsbedingung (1.7) liefert:

$$\forall x \in \Gamma : \nabla f(x^*) \cdot (x - x^*) = 2 \cdot (x - c)^T \cdot (x - x^*) = -2z^T \cdot (x - x^*) \geq 0 \tag{1.13}$$

Offenbar ist Γ ein Kegel, sodass mit x^* auch $2x^*$ und $\frac{1}{2}x^*$ in Γ liegen. Setzt man diese Punkte in (1.13) für x ein, so folgt:

$$-2z^T \cdot x^* \geq 0 \qquad z^T \cdot x^* \geq 0$$

also $z^T \cdot x^* = 0$. Damit liefert (1.13):

$$\forall (u, v) \in \mathbb{R}_+^{m_0} \times \mathbb{R}^p : z^T \cdot (A \cdot u + B \cdot v) = z^T \cdot x \leq 0$$

Setzt man nacheinander speziell $v := 0$ und $u := 0$, so ergibt sich hieraus $z^T \cdot A \cdot u \leq 0$ (für $u \geq 0$) bzw. $z^T \cdot B \cdot v \leq 0$ (für $v \in \mathbb{R}^n$), also muss gelten

$$A^T \cdot z \leq 0 \qquad B^T \cdot z = 0$$

Aus der Definition von z erhält man mit $z^T \cdot z > 0$ und $z^T \cdot x^* = 0$:

$$0 < z^T \cdot z = z \cdot (c - x^*)^T = z^T \cdot c$$

Also ist (1.11) lösbar. Widerspruch!

Im Folgenden sei

$$\mathcal{L} : \mathbb{R}^{n+m+p} \to \mathbb{R} : (x, u, v) \mapsto f(x) + u^T \cdot g(x) + v^T \cdot h(x)$$

die *Lagrange-Funktion* zum Problem (1.5). Die Variablen u, v heißen auch *Lagrange-Multiplikatoren.*

Theorem 1.6 Es seien $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{R}^m, h : \mathbb{R}^n \to \mathbb{R}^p$ differenzierbar. Wenn x^* eine lokale Lösung von (1.5) ist und in x^* die ACQ erfüllt ist, dann existieren Vektoren $u^* \in \mathbb{R}^m, v^* \in \mathbb{R}^p$, sodass (x^*, u^*, v^*) das folgende System erfüllt:

$$\begin{aligned} \nabla_x \mathcal{L}(x, u, v) &= 0 \\ g(x) &\leq 0 \\ h(x) &= 0 \\ u &\geq 0 \\ u^T \cdot g(x) &= 0 \end{aligned}$$

Beweis:

- Um Lemma (1.5) anzuwenden, seien m_0, A, B, c wie folgt erklärt:

$$\begin{aligned} m_0 &:= |I_0(x^*)| & c &:= -\nabla f(x^*) \\ A &:= (\nabla g_i(x^*))_{i \in I_0(x^*)} & B &:= \nabla h(x^*) \end{aligned}$$

  Da x^* Lösung von (1.5) ist und in x^* die ACQ gilt, liefert Theorem 1.4:

$$A^T \cdot d \leq 0, B^T \cdot d = 0 \Rightarrow c^T \cdot d \leq 0$$

  Folglich ist

$$A^T \cdot d \leq 0, B^T \cdot d = 0, c^T \cdot d > 0$$

  nicht lösbar. Nach Lemma 1.5 folgt die Lösbarkeit von

$$A \cdot u + B \cdot v = c \qquad u \geq 0$$

d.h. es gibt $u^0 = (u_i^0)_{i \in I_0(x^*)} \in \mathbb{R}^{m_0}, v^* \in \mathbb{R}^p$, sodass

$$A \cdot u^0 + B \cdot v^* = c \qquad u^0 \geq 0$$

Definiert man nun $u^* \in \mathbb{R}^m$ durch

$$u_i^0 := \begin{cases} u_i^0 & i \in I_0(x^*) \\ 0 & i \in I \backslash I_0(x^*) \end{cases}$$

dann erfüllt (x^*, u^*, v^*) die Bedingungen

$$\begin{aligned} \nabla_x \mathcal{L}(x, u, v) &= 0 \\ u^* &\geq 0 \\ u^T \cdot g(x) &= 0 \end{aligned}$$

Da x^* als lokale Lösung auch zulässig ist, genügt x^* den Restriktionen $g(x) \leq 0, h(x) = 0$.

Bemerkungen:

- Äquivalenz zu KKT-Bedingungen:

$$\left.\begin{aligned} g(x) &\leq 0 \\ u &\geq 0 \\ u^T \cdot g(x) &= 0 \end{aligned}\right\} \Leftrightarrow \forall i \in I : \left\{\begin{aligned} g_i(x) &\leq 0 \\ u_i &\geq 0 \\ u_i \cdot g_i(x) &= 0 \end{aligned}\right. \tag{1.14}$$

  (*Komplementaritätsbedingung*)

- Strenge Komplementaritätsbedingung gilt in (x, u, v) genau dann wenn $u_i > 0$ für alle $i \in I_0(x)$.

- (x^*, u^*, v^*) als Lösung des KKT-Systems wird auch als *KKT-Punkt* bezeichnet und die Komponente x^* als *stationärer Punkt*.

Theorem 1.7 Es seien $f : \mathbb{R}^n \to \mathbb{R}$ differenzierbar und pseudo-konvex, $g_1, \ldots, g_m$ differenzierbar und quasi-konvex sowie $h : \mathbb{R}^n \to \mathbb{R}^p$ affin-linear. Erfüllt (x^*, u^*, v^*) die zu (1.5) gehörenden KKT-Bedingungen, dann ist x^* eine globale Lösung.

Beweis: Übung

1.2.6 Sattelpunktsbedingungen

Definition 1.5 Ein Punkt $(x^*, u^*, v^*) \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}^p$ heißt *Sattelpunkt der Lagrangefunktion* $\mathcal{L}$, wenn

$$\forall (x, u, v) \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}^p : \mathcal{L}(x^*, u, v) \leq \mathcal{L}(x^*, u^*, v^*) \leq \mathcal{L}(x, u^*, v^*)$$

Theorem 1.8 Es seien $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ und $h : \mathbb{R}^n \to \mathbb{R}^p$ beliebige Funktionen. Wenn $(x^*, u^*, v^*) \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}^p$ ein Sattelpunkt der Lagrange-Funktion $\mathcal{L}$ ist, dann ist x^* eine globale Lösung des Problems (1.5).

Beweis:

- Da (x^*, u^*, v^*) Sattelpunkt ist, muss (u^*, v^*) des Optimierungsproblems

$$-\mathcal{L}(x^*, u, v) \to \min \text{ bei } (u, v) \in \mathbb{R}_+^m \times \mathbb{R}^p$$

  sein. Die notwendige Optimalitätsbedingung aus Theorem 1.2 liefert deshalb: $\forall (u, v) \in \mathbb{R}_+^m \times \mathbb{R}^p$:

$$\begin{aligned} -\nabla_u \mathcal{L}(x^*, u^*, v^*)^T \cdot (u - u^*) - \nabla_v \mathcal{L}(x^*, u^*, v^*)^T \cdot (v - v^*) &\geq 0 \\ g(x^*)^T \cdot (u - u^*) + h(x^*) \cdot (v - v^*) &\leq 0 \end{aligned} \tag{1.15}$$

Setzt man $v := v^*$ und $u := 2u^*$ bzw. $u := \frac{1}{2}u^*$, so folgt:

$$\begin{aligned} g(x^*)^T \cdot x^* &\leq 0 \\ -\frac{1}{2} \cdot g(x^*)^T \cdot x^* &\leq 0 \\ \Rightarrow g(x^*)^T \cdot u^* &= 0 \end{aligned} \tag{1.16}$$

Dies und (1.15) ergibt

$$\forall u \in \mathbb{R}_+^m : g(x^*)^T \cdot u \leq 0$$

Daraus hat man unmittelbar $g(x^*) \leq 0$. Setzt man in (1.15) $u := u^*$, so folgt analog $h(x^*) = 0$. Also ist x^* zulässiger Punkt von (1.5).

- Aus (1.16), $u^* \geq 0$, $h(x^*) = 0$ und der rechten Ungleichung der Sattelpunktsbedingung folgt:

$$\begin{aligned} f(x^*) &= f(x^*) + g(x^*)^T \cdot u^* + h(x^*)^T \cdot v^* \\ &= \mathcal{L}(x^*, u^*, v^*) \leq \mathcal{L}(x, u^*, v^*) \\ &= f(x) + \underbrace{g(x)^T}_{\leq 0} \cdot u^* + \underbrace{h(x)^T}_{=0} \cdot v^* \leq f(x) \end{aligned}$$

für jeden zulässigen Punkt x von (1.5). Somit ist x^* globale Lösung von (1.5).

Theorem 1.9 Es seien $f : \mathbb{R}^n \to \mathbb{R}$ und $g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ differenzierbar und konvex, $h : \mathbb{R}^n \to \mathbb{R}^p$ affin-linear. Außerdem gebe es $\bar{x} \in \mathbb{R}^n$ mit $h(\bar{x}) = 0$, $g_i(\bar{x}) < 0$ für $i \in I$ (erweiterte Slater-Bedingung). Ist x^* Lösung von (1.5), dann existieren $(u^*, v^*) \in \mathbb{R}_+^m \times \mathbb{R}^p$, sodass (x^*, u^*, v^*) ein Sattelpunkt der Lagrange-Funktion ist.

Beweis:

- Die Existenz eines Vektors $\bar{x} \in \mathbb{R}^n$ mit den genannten Eigenschaften, garantiert, dass in jedem zulässigen Punkt von (1.5) die ACQ erfüllt ist (Übung). Nach Theorem 1.6 gibt es daher Vektoren (u^*, v^*), sodass (x^*, u^*, v^*) das KKT-System befriedigt. Insbesondere gilt also

$$\nabla_x \mathcal{L}(x^*, u^*, v^*) = 0$$

Dies ist jedoch die zu

$$\mathcal{L}(x, u^*, v^*) \to \min \tag{1.17}$$

gehörende KKT-Bedingung. Da $f, g_1, \ldots, g_m$ konvex, h affin-linear und $u^* \geq 0$ ist, ist auch $\mathcal{L}(., u^*, v^*)$ eine konvexe Funktion bzgl. x. Nach Theorem 1.7 ist daher x^* eine globale Lösung von (1.17). Also gilt:

$$\forall x \in \mathbb{R}^n : \mathcal{L}(x^*, u^*, v^*) \leq \mathcal{L}(x, u^*, v^*) \tag{1.18}$$

- Da (x^*, u^*, v^*) ein KKT-Punkt von (1.5) ist, folgt weiterhin

$$g(x^*) \leq 0 \qquad h(x^*) = 0 \qquad g(x^*)^T \cdot u^* = 0$$

Daraus erhält man für $\lambda^* := -g(x^*)$:

$$\begin{aligned} -\nabla_u \mathcal{L}(x^*, u^*, v^*) - \lambda^* &= 0 \\ -\nabla_v \mathcal{L}(x^*, u^*, v^*) &= 0 \\ -u^* \leq 0 \qquad \lambda^* \geq 0 \qquad (-u^*)^T \cdot \lambda^* &= 0 \end{aligned}$$

Folglich stellt (u^*, v^*, λ^*) eine Lösung des KKT-Systems zum Optimierungsproblem

$$-\mathcal{L}(x^*, u, v) \to \min \text{ bei } (u, v) \in \mathbb{R}_+^m \times \mathbb{R}^p$$

dar. Ferner sind die Zielfunktion dieses Problems und die Nebenbedingungen $-u \leq 0$ affin-linear (also konvex). Nach Theorem 1.7 löst damit (u^*, v^*) dieses Optimierungsproblem, also gilt

$$\forall (u, v) \in \mathbb{R}_+^m \times \mathbb{R}^p : \mathcal{L}(x^*, u, v) \leq \mathcal{L}(x^*, u^*, v^*)$$

1.2.7 Bedingungen mit Ableitungen zweiter Ordnung

Sei (x^*, u^*, v^*) ein KKT-Punkt. Dann für alle $d \in L(x^*)$:

$$\begin{aligned} d^T \cdot \nabla f(x^*) &= -d^T \cdot (\nabla g(x^*) \cdot u^* + \nabla h(x^*) \cdot v^*) \geq 0 \\ \Rightarrow 0 \leq d^T \cdot \nabla f(x^*) &\overset{d \in L^+}{=} d^T \cdot \nabla f(x^*) + d^T \cdot \nabla g(x^*) \cdot u^* + d^T \cdot \nabla h(x^*) \cdot v^* \\ &= d^T \cdot \nabla_x \mathcal{L}(x^*, u^*, v^*) = 0 \end{aligned}$$

Letzter Term $= 0$, falls $\nabla g_i(x^*)^T \cdot d = 0$, falls $u_i^* > 0$. Definiere

$$L^+(x, u) := \left\{ d \in L(x); \forall u_i > 0 : \nabla g_i(x)^T \cdot d = 0 \right\}$$

Man erhält somit für einen KKT-Punkt (x^*, u^*, v^*):

1. Ist $d \in L(x^*)$ und $u_i^* \cdot \nabla g_i(x^*) \cdot d = 0$ für $u_i^* > 0$, dann gilt auch $d^T \cdot f(x^*) = 0$ für $d \in L(x^*)$.
2. Ist $d^T \cdot \nabla f(x^*) = 0$ für $d \in L(x^*) \backslash \{0\}$, dann folgt nicht zwingend $\nabla g_i(x^*)^T \cdot d = 0$ für $i \in I_0(x^*)$. (Nur falls z.B. MFCQ erfüllt ist.)

Theorem 1.10 Es seien $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ und $h : \mathbb{R}^n \to \mathbb{R}^p$ zweimal stetig differenzierbar. Weiter sei x^* eine lokale Lösung von (1.5), in der die ACQ erfüllt ist und (u^*, v^*) bezeichne ein Paar zu x^* gehörender Lagrange-Multiplikatoren. Dann gilt:

$$\forall d \in L^+(x^*, u^*) : d^T \cdot \nabla_{xx} \mathcal{L}(x^*, u^*, v^*) \cdot d \geq 0$$

Ohne Beweis

Beispiel

$$f(x) := -x_1^2 + x_2^2 \to \min \qquad \text{bei } \|x\|_\infty \leq 1$$

Es gilt

$$\|x\|_\infty \leq 1 \Leftrightarrow \begin{aligned} g_1(x) &:= x_1 - 1 \leq 0 \\ g_2(x) &:= -x_1 - 1 \leq 0 \\ g_3(x) &:= x_2 - 1 \leq 0 \\ g_4(x) &:= -x_2 - 1 \leq 0 \end{aligned}$$

Aus der Linearität der Nebenbedingung folgt, dass ACQ erfüllt ist, also sind die KKT-Bedingungen notwendige Bedingungen:

$$\nabla_x \mathcal{L}(x, u) = \begin{pmatrix} -2x_1 \\ 2x_2 \end{pmatrix} + u_1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + u_2 \cdot \begin{pmatrix} -1 \\ 0 \end{pmatrix} + u_3 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} + u_4 \cdot \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

$$\begin{array}{lll} x_1 - 1 \leq 0 & u_1 \geq 0 & u_1 \cdot (x_1 - 1) = 0 \\ -x_1 - 1 \leq 0 & u_2 \geq 0 & u_2 \cdot (-x_1 - 1) = 0 \\ x_2 - 1 \leq 0 & u_3 \geq 0 & u_3 \cdot (x_2 - 1) = 0 \\ -x_2 - 1 \leq 0 & u_4 \geq 0 & u_4 \cdot (-x_2 - 1) = 0 \end{array}$$

Dieses System zerfällt in zwei unabhängige Systeme für x_1, u_1, u_2 und für x_2, u_3, u_4, die man separat lösen kann. Dies ergibt folgende drei KKT-Punkte:

$$(x_1^*, x_2^*, u_1^*, u_2^*, u_3^*, u_4^*) \in \{(0,0,0,0,0,0), (1,0,2,0,0,0), (-1,0,0,2,0,0)\}$$

Man erhält die Hesse-Matrix:

$$\nabla_{xx} \mathcal{L}(x^*, u^*) = \nabla^2 f(x^*) = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}$$

Damit

$$\begin{aligned} L^+((0,0),(0,0,0,0)) &= \mathbb{R}^2 \\ L^+((1,0),(2,0,0,0)) &= \{d \in \mathbb{R}^2; d_1 = 0\} \\ L^+((-1,0),(0,2,0,0)) &= \{d \in \mathbb{R}^2; d_1 = 0\} \end{aligned}$$

Deshalb erfüllt der KKT-Punkt $(0,0,0,0,0,0)^T$ die notwendige Optimalitätsbedingung aus Theorem 1.10 nicht, also ist $(0,0)^T$ keine lokale Lösung. Weiterhin gilt für die anderen beiden KKT-Punkte:

$$\begin{aligned} \begin{pmatrix} 0 & d_2 \end{pmatrix} \cdot \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ d_2 \end{pmatrix} = 2d_2^2 \geq 0 \qquad & (\forall d \in L^+(x^*, u^*)) \\ > 0 \qquad & (\forall d \in L^+(x^*, u^*) \backslash \{0\}) \end{aligned}$$

Theorem 1.11 Es seien $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ und $h : \mathbb{R}^n \to \mathbb{R}^p$ zweimal stetig differenzierbar. Das Tripel (x^*, u^*, v^*) genüge den KKT-Bedingungen. Falls

$$\forall d \in L^+(x^*, u^*) \backslash \{0\} : d^T \cdot \nabla_{xx} \mathcal{L}(x^*, u^*, v^*) \cdot d > 0$$

Dann ist x^* eine strenge lokale Lösung von (1.5).

Beweis:

- Angenommen x^* ist keine strenge lokale Lösung. Dann gibt es eine Folge $\{x^k\} \subset G$, sodass

  $$\lim_{k \to \infty} x^k = x \qquad x^k \neq x^* \qquad f(x^k) \leq f(x^*) \qquad (k \in \mathbb{N})$$

  Sei

  $$d^k := \frac{x^k - x^*}{\|x^k - x^*\|} \qquad \alpha_k := \|x^k - x^*\|$$

  Damit liegt (d^k) in einer kompakten Menge (der Einheitskugel B). Ohne Beschränkung der Allgemeinheit können wir annehmen, dass $d^k \to d^* \in \mathbb{R}^n$ für $k \to \infty$.

- Aus der stetigen Differenzierbarkeit von f folgt:

  $$f(x^k) = f(x^*) + \alpha_k \cdot \nabla f(x^*)^T \cdot d^k + o(\alpha_k) \qquad (k \in \mathbb{N})$$

  Wegen $f(x^k) \leq f(x^*)$ erhält man hieraus für $k \to \infty$:

  $$\begin{aligned} 0 \geq \frac{f(x^k) - f(x^*)}{\alpha_k} &= \nabla f(x^*)^T \cdot d^k + \frac{o(\alpha_k)}{\alpha_k} \\ \overset{k \to \infty}{\Rightarrow} \nabla f(x^*)^T \cdot d^* &\leq 0 \end{aligned} \tag{1.19}$$

  Ersetzt man in dieser Argumentation f durch g_i mit $i \in I_0(x^*)$ bzw. durch h_j mit $j \in J$, dann folgt unter der Berücksichtigung von $(x^k) \subset G$:

  $$\nabla g_i(x^*)^T \cdot d^* \leq 0 \qquad \nabla h_j(x^*)^T \cdot d^* = 0 \qquad (i \in I_0(x^*), j \in J) \tag{1.20}$$

- Da (x^*, u^*, v^*) den KKT-Bedingungen genügt, ergibt sich mit (1.19) und (1.20):

  $$\begin{aligned} 0 &= \nabla_x \mathcal{L}(x^*, u^*, v^*)^T \cdot d^* \\ &= \underbrace{\nabla f(x^*)^T \cdot d^*}_{\leq 0} + \sum_i \underbrace{u_i^* \cdot \nabla g_i(x^*)^T \cdot d^*}_{\leq 0} + \underbrace{\sum_j v_j^* \cdot \nabla h_j(x^*)^T \cdot d^*}_{=0} \\ &\leq u_i^* \cdot \nabla g_i(x^*)^T \cdot d^* \leq 0 \qquad (i \in I_0(x^*)) \end{aligned}$$

  Für $i \in I_0(x^*)$ mit $u_i^* > 0$ hat man also $\nabla g_i(x^*)^T \cdot d^* = 0$. Dies und (1.20) ziehen $d^* \in L^+(x^*, u^*)$ nach sich.

- Nach Voraussetzung gilt daher

$$(d^*)^T \cdot \nabla_{xx}\mathcal{L}(x^*, u^*, v^*) \cdot d^* > 0 \tag{1.21}$$

Wegen $f(x^k) \leq f(x^*)$,

$$g(x^k)^T \cdot u^* \leq 0 = g(x^*)^T \cdot u^* = h(x^k)^T \cdot v^* = h(x^*)^T \cdot v^*$$

für alle $k \in \mathbb{N}$ und $\nabla_x \mathcal{L}(x^*, u^*, v^*) = 0$ ergibt sich mit der Taylorformel:

$$\begin{aligned} 0 &\geq \mathcal{L}(x^k, u^*, v^*) - \mathcal{L}(x^*, u^*, v^*) - \alpha_k \cdot \nabla_x \mathcal{L}(x^*, u^*, v^*)^T \cdot d^k \\ &= \frac{1}{2}\alpha_k^2 \cdot (d^k)^T \cdot \nabla_{xx}\mathcal{L}(x^*, u^*, v^*) \cdot d^k + o(\alpha_k^2) \end{aligned}$$

Durch α_k^2 teilen ergibt für $k \to \infty$:

$$0 \geq (d^*)^T \cdot \nabla_{xx}\mathcal{L}(x^*, u^*, v^*) \cdot d^*$$

Widerspruch zu (1.21)!

2 Minimierung ohne Restriktionen

$$f(x) \to \min \tag{2.1}$$

mit $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar.

Abstiegsprinzip: Generiere $\{x^k\}$ mit $f(x^{k+1}) < f(x^k)$ (solange x^k noch keine lokale Lösung/stationärer Punkt)

Theorem 2.1 Falls x^* eine lokale Lösung von (2.1) ist, dann gilt

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

Beweis: Folgt aus Theorem 1.2 für $G = \mathbb{R}^n$

Bemerkung:

- Falls f pseudo-konvex ist, so ist jeder stationärer Punkt (d.h. $\nabla f(x^*) = 0$) wegen Theorem 1.2 auch eine globale Lösung von (2.1).

Theorem 2.2 Sei $f : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar, dann gilt:

1. Wenn x^* eine lokale Lösung von (2.1) ist, dann muss $\nabla^2 f(x^*)$ positiv semidefinit sein.
2. Ist $\nabla f(x^*) = 0$ und $\nabla^2 f(x^*)$ positiv definit, dann ist x^* eine strenge lokale Lösung von (2.1).

2.1 Line-Search-Verfahren

Definition 2.1 Ein Vektor $d \in \mathbb{R}^n$ heißt *Abstiegsrichtung* der Funktion f an der Stelle x, wenn $\bar{\alpha} > 0$ existiert, sodass

$$\forall \alpha \in (0, \bar{\alpha}] : f(x + \alpha \cdot d) < f(x)$$

Sei x^k bekannt und d^k eine Abstiegsrichtung. Dann kann man durch geeignete (hinreichend kleine) Wahl einer Schrittweite $\alpha_k > 0$ die Iterierte

$$x^{k+1} := x^k + \alpha_k \cdot d^k$$

ermitteln, sodass zumindest $f(x^{k+1}) < f(x^k)$ erfüllt ist.

Auftretende Fragen:

- Wie bestimmt man eine Abstiegsrichtung?
- Wie ermittelt man eine geeignete Schrittweite?
- Wie groß sollte der Abstieg $f(x^{k+1}) - f(x^k)$ mindestens sein (und wie erreicht man dies), damit ein Häufungspunkt der Folge (x^k) auch stationär ist?
- Unter welchen Voraussetzungen kann die Existenz eines Häufungspunktes der Folge (x^k) garantiert werden?

Lemma 2.1 Ein Vektor $d \in \mathbb{R}^n$ ist Abstiegsrichtung von f im Punkt x, wenn

$$\nabla f(x^*)^T \cdot d < 0$$

Falls $\nabla f(x^*)^T \cdot d > 0$ so ist d keine Abstiegsrichtung, sondern eine Anstiegsrichtung.

Beweis:

- Die Aussagen folgen direkt aus der Taylor-Formel:

$$f(x + \alpha \cdot d) = f(x) + \alpha \cdot \underbrace{\nabla f(x)^T \cdot d}_{<0} + o(\alpha)$$

$$\Rightarrow \frac{f(x + \alpha \cdot d) - f(x)}{\alpha} = \nabla f(x)^T \cdot d + \frac{o(\alpha)}{\alpha} < 0$$

  für α hinreichend klein. Außerdem:

$$\forall d \in \mathbb{R}^n \backslash \{0\} : \nabla f(x)^T \cdot \left(\frac{-\nabla f(x)}{\|\nabla f(x)\|} \right) = -\|\nabla f(x)\| \leq \nabla f(x)^T \cdot \frac{d}{\|d\|}$$

  d.h. $-\nabla f(x)$ ist die stärkste Abstiegsrichtung.

Algorithmus 2.1: Gradienten-Verfahren mit Cauchy-Schrittweitenwahl

(S1) Initialisierung: Wähle $x^0 \in \mathbb{R}^n$, $a > 0$ und setze $k := 0$.

(S2) Abbruchtest: Falls $\nabla f(x^k) = 0$, dann stoppe den Algorithmus. (x^k ist stationär.)

(S3) Abstiegsrichtung: Setze $d^k := -\nabla f(x^k)$.

(S4) Schrittweite: Berechne $\alpha_k \geq 0$, sodass

$$\forall \alpha \in [0, a] : f(x^k + \alpha_k \cdot d^k) \leq f(x^k + \alpha \cdot d^k)$$

(S5) Update: Setze $x^{k+1} := x^k + \alpha_k \cdot d^k$ und $k := k + 1$. Gehe zu (S2).

Theorem 2.3 Der Algorithmus 2.1 ist wohldefiniert und bricht entweder nach endlich vielen Schritten mit einem stationären Punkt ab oder erzeugt eine unendliche Folge (x^k). Jeder Häufungspunkt der Folge (x^k) ist ein stationärer Punkt.

Beweis:

- Wohldefiniertheit der Schritte (S1),(S2),(S3),(S5) ist offensichtlich. Schritt (S4) ist ebenfalls wohldefiniert, da $\alpha \mapsto f(x^k + \alpha \cdot d^k)$ stetig von α abhängt und auf dem kompakten Intervall $[0, a]$ nach dem Satz von Weierstraß eine Minimalstelle besitzt.

- Wir nehmen nun an, dass der Algorithmus eine unendliche Folge (x^k) erzeugt und diese wenigstens einen Häufungspunkt x^* besitzt. Die Monotonie der Folge $(f(x^k))$ zusammen mit der Stetigkeit von f liefert dann:

$$\lim_{k \to \infty} f(x^k) = f(x^*) \tag{2.2}$$

- Sei $N_1 \subseteq \mathbb{N}$, sodass

$$\lim_{k \in N_1} x^k = x^*$$

  Angenommen $\nabla f(x^*) \neq 0$. Wegen der stetigen Differenzierbarkeit von f gibt es dann $\varepsilon > 0$, sodass

$$\forall k \in N_1 : \varepsilon \leq \|\nabla f(x^k)\| \leq \frac{1}{\varepsilon}$$

Mit der Taylor-Formel folgt die Existenz von $\tilde{\alpha} \in (0,1]$:

$$\begin{aligned} f(x^k + \alpha \cdot d^k) &= f(x^k) + \alpha \cdot \nabla f(x^k)^T \cdot d^k + \dots \\ &\dots + \alpha \cdot \int_0^1 (\nabla f(x^k + t \cdot \alpha \cdot d^k) - \nabla f(x^k))^T \cdot d^k \, dt \\ &\leq f(x^k) - \alpha \cdot \varepsilon^2 + \alpha \cdot \varepsilon^{-1} \cdot \max_{t \in [0,1]} \|\nabla f(x^k + t \cdot \alpha \cdot d^k) - \nabla f(x^k)\| \\ &\leq f(x^k) - \frac{\alpha}{2} \cdot \varepsilon^2 \end{aligned}$$

für alle $k \in N_1$ und alle $\alpha \in (0, \tilde{\alpha})$. Man beachte dabei, dass die Stetigkeit von ∇f die gleichmäßige Stetigkeit von ∇f auf jeder beschränkten abgeschlossenen Menge impliziert. D.h. für eine beschränkte abgeschlossene Menge $M \subseteq \mathbb{R}^n$ gibt es zu jedem $\epsilon > 0$ ein $\delta(\epsilon) > 0$, sodass

$$\|\nabla f(x) - \nabla f(y)\| \leq \epsilon$$

für alle $x, y \in M$ mit $\|x - y\| < \delta(\epsilon)$. Für ϵ wähle man $\frac{\varepsilon^2}{2}$ und und für M eine Kugel rB mit einem so großen Radius r, dass $(x^k)_{k \in N_1} \subset rB$ und $(x^k + d^k)_{k \in N_1} \subset rB$. Offenbar gibt es dann $\tilde{\alpha} \in (0,1]$, sodass

$$\|(x^k + t \cdot \alpha \cdot d^k) - x^k\| \leq \alpha \cdot \|\nabla f(x^k)\| \leq \delta(\epsilon)$$

für alle $\alpha \in [0, \tilde{\alpha}]$ und alle $k \in N_1$. Die Art der Schrittweitenwahl in (S4) zieht damit

$$\begin{aligned} f(x^{k+1}) = f(x^k + \alpha_k \cdot d^k) &\leq f(x^k + \tilde{\alpha} \cdot d^k) \\ &\leq f(x^k) - \frac{\min\{a, \tilde{\alpha}\}}{2} \cdot \varepsilon^2 \end{aligned}$$

für alle $k \in N_1$ nach sich. Da N_1 eine unendliche Menge ist und $f(x^{k+1}) < f(x^k)$ für jedes $k \in \mathbb{N}$ gilt, hat man

$$\lim_{k \in N_1} f(x^k) = -\infty$$

was (2.2) widerspricht. Also Annahme falsch und es gilt $\nabla f(x^*) = 0$.

Bemerkungen:

1. An Stelle der negativen Gradientenrichtung $-\nabla f(x^*)$ können auch andere Richtungen $d = d(x)$ Verwendung finden, wenn sie einen hinreichend starken Abstieg gewährleisten. Dies ist insbesondere dann der Fall, wenn $d(x)$ der Forderung

$$\nabla f(x)^T \cdot \frac{d(x)}{\|d(x)\|} \leq -\varrho(\|\nabla f(x)\|) \tag{2.3}$$

genügt, wobei $\varrho : (0, \infty) \to (0, \infty)$ eine Funktion mit der Eigenschaft

$$\forall t_\nu \subset (0, \infty) : \lim_{\nu \to \infty} \varrho(t_\nu)) = 0 \Rightarrow \lim_{\nu \to \infty} t_\nu = 0$$

ist. Derartige Richtungen heißen *gradientenähnlich*.

2. Bei der Berechnung der Schrittweite wird verlangt, dass α_k ein globales Minimum der Funktion $\varphi : \mathbb{R} \to \mathbb{R}, \varphi(\alpha) = f(x + \alpha \cdot d^k)$ unter der Nebenbedingung $\alpha \in [0, a]$ ist. (Häufig auch: $a = \infty$) Diese Art der Bestimmung von α_k ist nur in Spezialfällen praktisch durchführbar. Für den allgemeinen Fall ist man auf implementierbare Schrittweitenstrategien angewiesen (vgl. Algorithmus 2.2).

Zur Beschreibung des folgenden Algorithmus definieren wir

$$\begin{aligned} \mathcal{H}(m, M) &:= \{H \in \mathbb{R}^{n \times n}; H^T = H, \forall d \in \mathbb{R}^n : m \cdot \|d\|^2 \leq d^T \cdot H \cdot d \leq M \cdot \|d\|^2\} \\ S &:= \{2^{-i}; i \in \mathbb{N}\} \end{aligned}$$

für $0 < m \leq M$.

Algorithmus 2.2: Gradientenähnliches Verfahren mit Armijo-Schrittweite

(S1) Initialisierung: Wähle $x^0 \in \mathbb{R}^n, \delta \in (0,1), 0 < m \leq M$ und setze $k := 0$.

(S2) Abbruchtest: Falls $\nabla f(x^k) = 0$, dann stoppe den Algorithmus.

(S3) Abstiegsrichtung: Wähle $H_k \in \mathcal{H}(m, M)$ und berechne d^k als Lösung von

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

(S4) Schrittweite:

$$\alpha_k := \max\{\alpha \in S; f(x^k + \alpha \cdot d^k) \leq f(x^k) + \delta \cdot \alpha \cdot \nabla f(x^k)^T \cdot d^k\}$$

(S5) Update: Setze $x^{k+1} := x^k + \alpha_k \cdot d^k$ und $k := k + 1$. Gehe zu (S2)

Theorem 2.4 Algorithmus 2.2 ist wohldefiniert und bricht entweder nach endlich vielen Schritten mit einem stationären Punkt ab oder erzeugt eine unendliche Folge (x^k). Jeder Häufungspunkt einer solchen Folge (x^k) ist ein stationärer Punkt.

Beweis: Übung

Bemerkung:

- Für $H_k := I$ ergibt sich Algorithmus 2.1. Durch spezielle Wahl von H_k im Algorithmus 2.2 kann lokal überlineare Konvergenz erzielt werden. Der nächste Algorithmus liefert eine Möglichkeit zur adaptiven Bestimmung von m, M.

Algorithmus 2.3: Gedämpftes regularisiertes Newton-Verfahren

(S1) Initialisierung: Wähle $x^0 \in \mathbb{R}^n, \delta, m_0 \in (0,1)$ und setze $k := 0$.

(S2) Abbruchtest: Falls $\nabla f(x^k) = 0$, dann stoppe den Algorithmus.

(S3) Abstiegsrichtung: Falls $\nabla^2 f(x^k) \in \mathcal{H}(m_k, m_k^{-1})$, setze $H_k := \nabla^2 f(x^k)$. Anderenfalls wähle $H_k \in \mathcal{H}(m_k, m_k^{-1})$. Berechne d^k als Lösung von

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

(S4) Schrittweite:

$$\alpha_k := \max\{\alpha \in S; f(x^k + \alpha \cdot d^k) \leq f(x^k) + \delta \cdot \alpha \cdot \nabla f(x^k)^T \cdot d^k\}$$

(S5) Update: Setze

$$m_{k+1} := \begin{cases} \frac{1}{2} \cdot m_k & \|\nabla f(x^k)\| < m_k \\ m_k & \text{sonst} \end{cases}$$
$$x^{k+1} := x^k + \alpha_k \cdot d^k$$
$$k := k + 1$$

Gehe zu (S2)

Bemerkung:

- Um die Existenz eines Häufungspunktes bei einer durch den Algorithmus 2.1, 2.2 oder 2.3 erzeugten Folge (x^k) zu sichern, wird häufig die Kompaktheit der Niveaumenge

$$W(x^0) = \{x \in \mathbb{R}^n; f(x) \leq f(x^0)\}$$

vorausgesetzt. Da die Folge offenbar in $W(x^0)$ liegt, besitzt sie unter diesen Voraussetzungen mindestens einen Häufungspunkt. Eine hinreichende Bedingung für die Kompaktheit von $W(x^0)$ ist die gleichmäßige Konvexität von f.

Definition 2.2 Es seien $(z^k) \subseteq \mathbb{R}^l$ und $z^* \in \mathbb{R}^l$ gegeben mit $z^* \notin (z^k)$. Dann heißt (z^k)

1. *Q-linear konvergent gegen* z^*, falls $\sigma \in (0,1)$ und $k_0 \in \mathbb{N}$ existieren, sodass
$$\forall k \geq k_0 : \frac{\|z^{k+1} - z^*\|}{\|z^k - z^*\|} \leq \sigma$$
2. *Q-superlinear konvergent gegen* z^*, falls
$$\lim_{k\to\infty} \frac{\|z^{k+1} - z^*\|}{\|z^k - z^*\|} = 0$$
3. mit der *Q-Ordnung* τ *konvergent gegen* z^*, falls $\lim_{k\to\infty} z^k = z^*$ und $\tau > 1$ und $\sigma > 0$ existieren, sodass
$$\lim_{k\to\infty} \frac{\|z^{k+1} - z^*\|}{\|z^k - z^*\|^\tau} \leq \sigma$$
4. *R-linear* bzw. *R-superlinear* bzw. mit der *R-Ordnung* τ *konvergent gegen* z^*, falls eine Folge $(\mu_k)_k \subset (0,\infty)$ existiert, sodass
$$\forall k \in \mathbb{N} : \|z^k - z^*\| \leq \mu_k$$
und $(\mu_k)_k$ Q-linear bzw. Q-superlinear bzw. mit der Q-Ordnung τ gegen 0 konvergiert.

Bemerkungen:

1. Die Buchstaben „R“ und „Q“ stehen für Root und Quotient.
2. Anstelle der Definition (3) kann auch folgende äquivalente Formulierung genutzt werden: (z^k) ist genau dann mit der Q-Ordnung τ gegen z^* konvergent, falls $\lim_{k\to\infty} z^k = z^*$ und $\tau > 1$ und $\sigma > 0$ existieren, sodass
$$\forall k \in \mathbb{N} : \|z^{k+1} - z^*\| \leq \sigma \cdot \|z^k - z^*\|^\tau$$

Beispiele: Es sei $z^* = 0$ und (z^k) definiert durch

1. $z^k := k^{-2}$. Dann konvergiert (z^k) gegen 0, aber wegen
$$\lim_{k\to\infty} \frac{\|z^{k+1} - z^*\|}{\|z^k - z^*\|} = \lim_{k\to\infty} \frac{k^2}{(k+1)^2} = 1$$
ist (z^k) nicht einmal Q-linear konvergent.
2. $z^k := e^{-k}$. Dann gilt
$$\lim_{k\to\infty} \frac{\|z^{k+1} - z^*\|}{\|z^k - z^*\|} = \lim_{k\to\infty} \frac{e^{-(k+1)}}{e^{-k}} = e^{-1} < 1$$
also ist (z^k) Q-linear konvergent gegen 0 mit $\sigma = e^{-1}$, aber nicht Q-superlinear.
3. $z^k = e^{-k^2}$. Dann gilt
$$\lim_{k\to\infty} \frac{\|z^{k+1} - z^*\|}{\|z^k - z^*\|} = \lim_{k\to\infty} \frac{e^{-(k+1)^2}}{e^{-k^2}} = \lim_{k\to\infty} e^{-2k-1} = 0$$
also (z^k) Q-superlinear gegen 0 konvergent, man kann jedoch leicht zeigen, dass (z^k) nicht mit einer Q-Ordnung $\tau > 1$ konvergiert.
4. $z^k := e^{-e^k}$. Dann gilt
$$\lim_{k\to\infty} \frac{\|z^{k+1} - z^*\|}{\|z^k - z^*\|^\tau} = \frac{e^{-e^{k+1}}}{e^{-\tau\cdot e^k}} = \lim_{k\to\infty} e^{e^k\cdot(\tau - e)} \leq 1$$
wenn $\tau \leq e$. Daher konvergiert (z^k) mit der Q-Ordnung $\tau = e$ gegen 0.

Theorem 2.4 Es sei f zweimal differenzierbar und $\nabla^2 f$ sei lokal lipschitz-stetig. Außerdem sei $\nabla^2 f(x^*)$ positiv definit und $\nabla f(x^*) = 0$. Für $\delta \in \left(0, \frac{1}{2}\right)$ gibt es dann ein $\varepsilon > 0$, sodass die Newton-Richtung

$$d(x) := -\nabla^2 f(x)^{-1} \cdot \nabla f(x) \tag{2.4}$$

für alle $x \in B(x^*, \varepsilon)$ definiert ist und

$$\forall x \in B(x^*, \varepsilon) : f(x + d(x)) \leq f(x) + \delta \cdot \nabla f(x)^T \cdot d(x)$$

gilt.

Beweis:

- Da $\nabla^2 f(x^*)$ positiv definit und $\nabla^2 f$ lokal lipschitz-stetig, gibt es $\varepsilon_0 > 0$ und $\gamma > 0$, sodass $\nabla^2 f(x)^{-1}$ für alle $x \in B(x^*, \varepsilon_0)$ existiert und

$$\begin{aligned} d^T \cdot \nabla^2 f(x) \cdot d &= d^T \cdot \nabla^2 f(x^*) \cdot d + d^T \cdot (\nabla^2 f(x) - \nabla^2 f(x^*)) \cdot d \\ &\geq d^T \cdot \nabla^2 f(x^*) \cdot d - \|d\|^2 \cdot \|\nabla^2 f(x) - \nabla^2 f(x^*)\| \\ &\geq \gamma \cdot \|d\|^2 \end{aligned} \tag{2.5}$$

  für alle $x \in B(x^*, \varepsilon_0)$ und alle $d \in \mathbb{R}^n$. Damit erhält man wegen (2.4) und der Taylor-Formel:

$$\begin{aligned} f(x + d(x)) &= f(x) + \delta \cdot \nabla f(x)^T \cdot d(x) + (1 - \delta) \cdot \nabla f(x)^T \cdot d(x) + \frac{1}{2} \cdot d(x)^T \cdots \\ &\quad \cdots \nabla^2 f(x) \cdot d(x) + \frac{1}{2} \cdot \int_0^1 d(x)^T \cdot (\nabla^2 f(x + t \cdot d(x)) - \nabla^2 f(x)) \cdot d(x)\, dt \\ &= f(x) + \delta \cdot \nabla f(x)^T \cdot d(x) - \left(\frac{1}{2} - \delta\right) \cdot d(x)^T \cdot \nabla^2 f(x) \cdot d(x) \\ &\quad + \frac{1}{2} \cdot \int_0^1 d(x)^T \cdot (\nabla^2 f(x + t \cdot d(x)) - \nabla^2 f(x)) \cdot d(x)\, dt \end{aligned}$$

  Nutzt man nun (2.5) aus und dass $x, \nabla^2 f(x)^{-1}$ und $d(x)$ gleichmäßig beschränkt auf $B(x^*, \varepsilon_0)$ sind, so folgt aus der lokalen Lipschitz-Stetigkeit von $\nabla^2 f(x)$ mit $L > 0$:

$$f(x + d(x)) \leq f(x) + \delta \cdot \nabla f(x)^T \cdot d(x) - \left(\frac{1}{2} - \delta\right) \cdot \gamma \cdot \|d(x)\|^2 + L \cdot \|d(x)\|^3 \tag{2.6}$$

  für alle $x \in B(x^*, \varepsilon_0)$. Aus (2.5) und (2.4) ergibt sich

$$\begin{aligned} \gamma \cdot \|d(x)\|^2 &\leq d(x)^T \cdot \nabla^2 f(x) \cdot d(x) = -\nabla f(x)^T \cdot d(x) \\ &\leq \|\nabla f(x)\| \cdot \|d(x)\| \end{aligned}$$

  und damit

$$\|d(x)\| \leq \frac{1}{\gamma} \cdot \|\nabla f(x)\|$$

  Dies, (2.6), $\nabla f(x^*) = 0$ und die Stetigkeit von ∇f liefern

$$\forall x \in B(x^*, \varepsilon) : f(x + d(x)) \leq f(x) + \delta \cdot \nabla f(x)^T \cdot d(x)$$

  für $0 < \varepsilon \leq \varepsilon_0$ hinreichend klein.

Theorem 2.5

1. Algorithmus 2.3 ist wohldefiniert und bricht entweder nach endlich vielen Schritten mit einem stationären Punkt ab oder erzeugt eine unendliche Folge (x^k). Falls diese Folge einen Häufungspunkt besitzt, so gilt

$$\liminf_{k \to \infty} \|\nabla f(x^k)\| = 0$$

   Ist die Folge (x^k) insbesondere beschränkt, so ist mindestens einer ihrer Häufungspunkte stationär.

2. Es sei f zweimal differenzierbar und $\nabla^2 f$ lokal lipschitz-stetig. Weiter sei $\delta \in \left(0, \frac{1}{2}\right)$ gewählt. Die durch Algorithmus (2.3) erzeugte Folge (x^k) habe einen Häufungspunkt x^*, der stationär ist und für den $\nabla^2 f(x^*)$ positiv definit ist. Dann konvergiert die Folge (x^k) gegen x^* mit der Q-Ordnung 2.

Beweis:

1. Die Wohldefiniertheit kann analog zu Theorem 2.4 bewiesen werden. Angenommen, die Folge (x^k) besitzt einen Häufungswert und es gibt $\varepsilon > 0$, sodass

$$\forall k \in \mathbb{N} : \|\nabla f(x^k)\| \geq \varepsilon$$

Daraus folgt

$$\min_{k \in \mathbb{N}} m_k =: \tilde{m} > 0$$

und $m_k = \tilde{m}$ für $k \in \mathbb{N}$ hinreichend groß. Setzt man $m := \tilde{m}$ und $M := \tilde{m}^{-1}$, so kann Algorithmus 2.3 für $k \geq k_0$ (k_0 hinreichend groß) als Algorithmus 2.2 angesehen werden. Für diesen ist nach Theorem 2.4 aber jeder Häufungspunkt ein stationärer Punkt. Widerspruch!

2. Nach Voraussetzung hat (x^k) einen Häufungspunkt, der stationär ist. Daher gilt nach 1. $\liminf_{k\to\infty} \|\nabla f(x^k)\| = 0$. Dies zieht

$$\lim_{k\to\infty} m_k = 0 \qquad \lim_{k\to\infty} m_k^{-1} = \infty \tag{2.7}$$

nach sich. Infolge der positiven Definitheit von $\nabla f(x^*)$ und der lokalen Lipschitz-Stetigkeit von $\nabla^2 f$ gibt es $C > 0$, sodass $\forall t \in [0,1], \forall x \in B(x^*, \varepsilon)$:

$$\|\nabla^2 f(x)^{-1}\| \leq C \tag{2.8}$$

$$\|\nabla^2 f(x + t \cdot (x^* - x)) - \nabla^2 f(x)\| \leq C \cdot \|x^* - x\| \tag{2.9}$$

mit ε aus Theorem 2.4. Wegen (2.9) für $t = 1$ und wegen (2.7) kann man $\varepsilon_1 \in (0, \varepsilon]$ finden, sodass

$$\forall k \in \mathbb{N} : x^k \in B(x^*, \varepsilon_1) \Rightarrow H_k = \nabla^2 f(x^k)$$

und

$$C^2 \cdot \varepsilon_1 \leq \frac{1}{2} \tag{2.10}$$

gilt. Aus Theorem 2.4 ergibt sich damit zunächst

$$\forall k \in \mathbb{N} : x^k \in B(x^*, \varepsilon_1) \Rightarrow \alpha_k = 1$$

Hieraus folgt für $x^k \in B(x^*, \varepsilon_1)$:

$$\begin{aligned} x^{k+1} - x^* &= x^{k+1} - x^k + x^k - x^* = \alpha_k \cdot d^k + x^k - x^* \\ &= -\nabla^2 f(x^k)^{-1} \cdot \nabla f(x^k) + x^k - x^* \end{aligned}$$

Unter Beachtung von (2.8),(2.9),(2.10) und der Taylor-Formel erhält man daraus:

$$\begin{aligned} \|x^{k+1} - x^*\| &= \| - \nabla^2 f(x^k)^{-1} \cdot (\nabla f(x^k) + \nabla^2 f(x^k) \cdot (x^* - x^k))\| \\ &\leq \|\nabla^2 f(x^k)^{-1}\| \cdot \|\nabla^2 f(x^k) \cdot (x^* - x^k) + \nabla f(x^k) - \nabla f(x^*)\| \\ &\leq C \cdot \left\| \int_0^1 (\nabla^2 f(x^k + t \cdot (x^* - x^k)) - \nabla^2 f(x^k)) \cdot (x^* - x^k)\, dt \right\| \\ &\leq C^2 \cdot \|x^k - x^*\|^2 \leq C^2 \cdot \varepsilon_1 \cdot \|x^* - x^k\| \\ &\leq \frac{1}{2} \cdot \|x^* - x^k\| \end{aligned}$$

Daraus folgt die Konvergenz der Folge (x^k) gegen x^* mit der Q-Ordnung 2.

2.2 Trust-Region-Verfahren

- Idee: quadratische Approximation von f (in Abhängigkeit von Iterationspunkten x^k):

$$f(x^k) + \nabla f(x^k)^T \cdot (x - x^k) + \frac{1}{2}(x - x^k)^T \cdot B_k \cdot (x - x_k)$$

  wobei $B_k \in \mathbb{R}^{n \times n}$ mit $B_k^T = B_k$. Zur Vereinfachung setzen wir $p := x - x^k$ und definieren die quadratische Modellfunktion

$$m_k(p) : \mathbb{R}^n \to \mathbb{R}, m_k(p) := f(x^k) + \nabla f(x^k)^T \cdot p + \frac{1}{2} p^T \cdot B_k \cdot p$$

- Der Unterschied zwischen Modell und des durch die Taylorformel

$$f(x^k + p) = f(x^k) + \nabla f(x^k)^T \cdot p + \frac{1}{2} p^T \cdot \nabla^2 f(x^k + \theta_p \cdot p) \cdot p \qquad (0 < \theta_p < 1)$$

  beschriebenen exakten lokalen Verhaltens von f ist proportional zu $\|p\|^2$, insbesondere gilt $m_k(0) = f(x^k)$.

- Für $B_k := \nabla^2 f(x^k)$ ergibt sich für den Approximationsfehler

$$\begin{aligned} |f(x^k + p) - m_k(p)| &= \frac{1}{2} \cdot \left| p^T \cdot (\nabla^2 f(x^k + \theta_p \cdot p) - \nabla^2 f(x^k)) \cdot p \right| \\ &= o(\|p\|^3) \end{aligned}$$

  falls f hinreichend glatt ist. Man spricht dann von einem *Trust-Region-Newton-Verfahren.*

- Auf Grund des von $\|p\|$-abhängenden Approximationsfehlers (und um die Lösbarkeit der Teilprobleme zu sichern) benutzt man bei Trust-Region-Verfahren die folgenden Teilprobleme:

$$m_k(p) \to \min \qquad \text{bei } \|p\| \leq \Delta_k \tag{2.11}$$

  wobei $\Delta_k > 0$ der Radius des Vertrauensbereiches ist und in Abhängigkeit von der Güte des Modells m_k verkleinert oder vergrößert werden kann. Unter $\|.\|$ verstehen wir hier immer die euklidische Norm $\|.\|_2$ (auch andere Normen können von Interesse sein).

- Die Steuerung des Parameters Δ_k muss sicherstellen, dass zumindest $f(x^k) > f(x^{k+1})$ mit $x^{k+1} := x^k + p^k$, wobei p^k eine (Näherungs-)Lösung von (2.11) bezeichnet. Dazu definiert man ein Maß für die Güte des Modells (2.11) durch

$$\varrho_k := \frac{f(x^k) - f(x^k + p^k)}{m_k(0) - m_k(p^k)}$$

  Falls $\nabla f(x^k) \neq 0$, dann gilt $m_k(p^k) < m_k(0)$, sodass der Nenner von ϱ_k positiv ist. Demzufolge bedeutet $\varrho_k \leq 0$, dass $f(x^k + p^k) \geq f(x^k)$, also kein Abstieg vorliegt und man p^k nicht zur Bestimmung einer neuen Iterierten verwenden sollte. Das heißt der Radius Δ_k ist zu groß und muss reduziert werden, um ein geeignetes p^k zu ermitteln.

  Umgekehrt, wenn ϱ_k hinreichend positiv ist, kann es zur Konvergenzbeschleunigung sinnvoll sein, den Radius Δ_k zu vergrößern.

Algorithmus 2.4: Trust-Region-Verfahren

(S1) Wähle $x^0 \in \mathbb{R}^n$, $\Delta_0 > 0$, $0 < \eta_1 < \eta_2 < 1$, $0 < \sigma_1 < 1 < \sigma_2$ und setze $k := 0$.

(S2) Falls $\nabla f(x^k) = 0$, dann stoppe Algorithmus.

(S3) Wähle eine symmetrische Matrix $B_k \in \mathbb{R}^{n \times n}$. Bestimme eine (Näherungs-)Lösung p^k von (2.11).

(S4) Berechne ϱ_k. Falls $\varrho_k \leq \eta_1$ ist, dann setze $\Delta_{k+1} := \sigma_1 \cdot \Delta_k$. Falls $\varrho_k \in (\eta_1, \eta_2)$, dann $\Delta_{k+1} := \Delta_k$. Falls $\varrho_k \geq \eta_2$, setze $\Delta_{k+1} := \sigma_2 \cdot \Delta_k$.

(S5) Falls $\varrho_k > \eta_1$, setze $x^{k+1} := x^k + p^k$, $k := k + 1$. Gehe zu (S2).

(S6) Setze $x^{k+1} := x^k$, $k := k + 1$. Gehe zu (S3).

Bemerkungen:

1. Falls $\varrho_k \leq \eta_1$, dann wird $\Delta_k, \Delta_{k+1}, \ldots$ solange verkleinert, bis ein p^k gefunden ist, dass einen hinreichenden Abstieg liefert.
2. Die exakte Lösung von (2.11) ist oft zu aufwendig, deshalb sind auch (weniger aufwendig) zu beschaffende Näherungslösungen von Interesse.

Lemma 2.3 Sei $B \in \mathbb{R}^{n\times n}$ symmetrisch und $g \in \mathbb{R}^n$, $\bar{f} \in \mathbb{R}$ und $\Delta > 0$ gegeben. Dann ist der Vektor p^* genau dann eine Lösung des Problems

$$m(p) := \bar{f} + g^T \cdot p + \frac{1}{2} p^T \cdot B \cdot p \qquad \text{bei } \|p\| \leq \Delta$$

wenn eine Zahl $\lambda \geq 0$ existiert, sodass die folgenden drei Bedingungen erfüllt sind:

1. $(B + \lambda \cdot I) \cdot p^* = -g$
2. $\lambda \cdot (\Delta - \|p^*\|) = 0$
3. $(B + \lambda \cdot I)$ ist positiv semidefinit

Beweis: Übung (KKT-Bedingungen ausnutzen!)

Theorem 2.6

1. Der Algorithmus 2.4 ist wohldefiniert.
2. Sei f stetig differenzierbar. Es werde vorausgesetzt, dass Algorithmus 2.4 eine unendliche beschränkte Folge (x^k) erzeugt. Wenn Zahlen $\beta \geq 1, \delta \in (0, 1]$ existieren, sodass

$$\begin{aligned} \forall k \in \mathbb{N} : \|p^k\| &\leq \Delta_k \\ \forall k \in \mathbb{N} : \|B_k\| &\leq \beta \end{aligned} \tag{2.12}$$

und

$$m_k(0) - m_k(p^k) \geq \delta \cdot \|\nabla f(x^k)\| \cdot \min\left\{\Delta_k, \frac{\|\nabla f(x^k)\|}{\|B_k\|}\right\} \tag{2.13}$$

für alle $k \in \mathbb{N}$ gilt, dann

(i) besitzt (x^k) mindestens einen Häufungspunkt, der stationär ist.

(ii) ist jeder Häufungspunkt von (x^k) stationär, falls ∇f auf der Niveaumenge $W(x^0)$ lipschitz-stetig ist.

Beweis:

1. Wohldefiniertheit ist offensichtlich, da alle Teilprobleme stets lösbar sind (Satz von Weierstraß).
2. (i) Zu zeigen:

$$\liminf_{k\to\infty} \|\nabla f(x^k)\| = 0$$

Aus der Definition von m_k und der Taylorformel erhält man für

$$\sigma_k := |m_k(p^k) - m_k(0) - (f(x^k + p^k) - f(x^k))|$$

die Abschätzung

$$\begin{aligned}
\sigma_k &= \left| \frac{1}{2} \cdot (p^k)^T \cdot B_k \cdot p^k + \nabla f(x^k)^T \cdot p^k - f(x^k + p^k) + f(x^k) \right| \\
&= \left| \frac{1}{2} \cdot (p^k)^T \cdot B_k \cdot p^k - \int_0^1 \nabla (f(x^k + t \cdot p^k) - \nabla f(x^k))^T \cdot p^k \, dt \right| \\
&\leq \|p^k\| \cdot \underbrace{\left(\frac{\beta}{2} \cdot \|p^k\| + \max_{t \in [0,1]} \|\nabla f(x^k + t \cdot p^k) - \nabla f(x^k)\| \right)}_{=: c(p^k)} \\
&\leq \|p^k\| \cdot c(p^k) \qquad (2.14)
\end{aligned}$$

für alle $k \in \mathbb{N}$. Dabei ist $c : \mathbb{R}^n \to [0, \infty)$ eine Funktion mit der Eigenschaft, dass zu jedem $\varepsilon > 0$ ein $\Delta(\varepsilon) \in (0, 1]$ existiert, sodass

$$c(p) \leq \varepsilon$$

für alle $p \in B(0, \Delta(\varepsilon))$. Diese Eigenschaft folgt wegen der Beschränktheit der Folge (x^k) und der Stetigkeit des Gradienten von f sofort aus der gleichmäßigen Stetigkeit von stetigen Funktionen auf kompakten Mengen.
Es wird nun angenommen, dass die Behauptung

$$\liminf_{k \to \infty} \|\nabla f(x^k)\| = 0$$

falsch ist. Dann gibt es $\mu \in (0, 1]$, sodass

$$\forall k \in \mathbb{N} : \|\nabla f(x^k)\| \geq \mu \qquad (2.15)$$

Aus (2.13) und (2.12) folgt damit

$$\begin{aligned}
m_k(0) - m_k(p^k) &\geq \delta \cdot \|\nabla f(x^k)\| \cdot \min \left\{ \Delta_k, \frac{\|\nabla f(x^k)\|}{\|B_k\|} \right\} \\
&\geq \delta \cdot \mu \cdot \min \left\{ \Delta_k, \frac{\mu}{\beta} \right\} \qquad (2.16)
\end{aligned}$$

für alle $k \in \mathbb{N}$. Mit (2.14), (2.16) und $\|p^k\| \leq \Delta_k$ ergibt sich

$$\begin{aligned}
|\varrho_k - 1| &= \left| \frac{f(x^k) - f(x^k + p^k)}{m_k(0) - m_k(p^k)} - \frac{m_k(0) - m_k(p^k)}{m_k(0) - m_k(p^k)} \right| = \frac{\sigma_k}{|m_k(0) - m_k(p^k)|} \\
&\leq \frac{\Delta_k \cdot c(p^k)}{\delta \cdot \mu \cdot \min \left\{ \Delta_k, \frac{\mu}{\beta} \right\}} \qquad (2.17)
\end{aligned}$$

Wenn

$$\Delta_k \leq \frac{\delta \cdot \mu}{\beta} \cdot \Delta((1 - \eta_2) \cdot \delta \cdot \mu) \qquad (2.18)$$

dann folgt $\Delta_k \leq \mu \cdot \beta^{-1}$ und

$$\min\{\Delta_k, \mu \cdot \beta^{-1}\} = \Delta_k$$

Deshalb erhält man mit (2.17) und (2.18):

$$\begin{aligned}
|\varrho_k - 1| &\leq \frac{\Delta_k \cdot c(p^k)}{\delta \cdot \mu \cdot \Delta_k} = \frac{c(p^k)}{\delta \cdot \mu} \\
&\leq \frac{(1 - \eta_2) \cdot \delta \cdot \mu}{\delta \cdot \mu} = 1 - \eta_2 \\
\Rightarrow -\varrho_k + 1 &\leq 1 - \eta_2
\end{aligned}$$

Also folgt $\varrho_k \geq \eta_2$ und $\Delta_{k+1} > \Delta_k$. Eine Reduktion von Δ_k (um den Faktor $\sigma_1 < 1$) kann also nur auftreten, wenn (2.18) nicht gilt, also wenn

$$\Delta_k > \frac{\delta \cdot \mu}{\beta} \cdot \Delta((1-\eta_2) \cdot \delta \cdot \mu)$$

Somit hat man

$$\Delta_{k+1} \geq \min\left\{\Delta_0, \sigma_1 \cdot \frac{\delta \cdot \mu}{\beta} \cdot \Delta((1-\eta_2) \cdot \delta \cdot \mu)\right\} \tag{2.19}$$

für alle $k \in \mathbb{N}$.

Wir können daher annehmen, dass es eine unendliche Teilmenge $N_1 \subseteq \mathbb{N}$ gibt mit $\varrho_k > \eta_1$ für alle $k \in N_1$. Unter Beachtung von (2.16) liefert dies für $k \in N_1$:

$$\begin{aligned} 0 \leq f(x^k) - f(x^{k+1}) &= f(x^k) - f(x^k + p^k) \geq \eta_1 \cdot (m_k(0) - m_k(p^k)) \\ &\overset{(2.16)}{\geq} \eta_1 \cdot \delta \cdot \mu \cdot \min\left\{\Delta_k, \mu \cdot \beta^{-1}\right\} \end{aligned}$$

für $k \in N_1$. Wegen (2.19) und da $f(x^{k+1}) \leq f(x^k)$ für alle $k \in \mathbb{N}$ gilt, folgt dass

$$\lim_{k\to\infty} f(x^k) = -\infty$$

Dies ist jedoch auf Grund der Beschränktheit von (x^k) und der Stetigkeit von f nicht möglich. Also ist die Annahme falsch und es gilt somit

$$\liminf_{k\to\infty} \|\nabla f(x^k)\| = 0$$

(ii) Ohne Beweis.

Bestimmung einer Näherungslösung der Trust-Region-Teilprobleme

- Eine einfache Möglichkeit ist die Bestimmung von p^k als *Cauchy-Punkt* :

  1. Bestimmte p_l^k als Lösung des Optimierungsproblems

     $$f(x^k) + \nabla f(x^k)^T \cdot p \to \min \qquad \text{bei } \|p\| \leq \Delta_k$$

     Dann gilt offensichtlich

     $$p_l^k = -\frac{\nabla f(x^k)}{\|\nabla f(x^k)\|} \cdot \Delta_k$$

  2. Berechne eine Cauchy-Schrittweite α_k als Lösung von

     $$m_k(\alpha \cdot p_l^k) \to \min \qquad \text{bei } \|\alpha \cdot p_l^k\| \leq \Delta_k, \alpha \geq 0$$

     Setze dann $p_c^k := \alpha_k \cdot p_l^k$.

  3. Um α_k explizit anzugeben, betrachten wir zwei Fälle:

     (i) $\nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k) \leq 0$: Dann hat man

     $$0 \leq \alpha_1 < \alpha_2 \Rightarrow m_k(\alpha_1 \cdot p_l^k) > m_k(\alpha_2 \cdot p_l^k)$$

     Folglich wird die Restriktion $\|\alpha \cdot p_l^k\| \leq \Delta_k$ aktiv, d.h. $\alpha_k = 1$.

     (ii) $\nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k) > 0$: Dann gilt entweder $\alpha_k = 1$ oder $\alpha_k \in (0,1)$ ergibt sich als freies Minimum der Funktion φ mit

     $$\varphi(\alpha) = m_k(\alpha \cdot p_l^k)$$

     In diesem zweiten Fall gilt:

     $$\begin{aligned} 0 = \varphi'(\alpha_k) &= \nabla f(x^k)^T \cdot p_l^k + \alpha_k \cdot (p_l^k)^T \cdot B_k \cdot p_l^k \\ \Rightarrow \alpha_k &= -\frac{(\nabla f(x^k)^T \cdot p_l^k}{(p_l^k)^T \cdot B_k \cdot p_l^k} = \frac{\|\nabla f(x^k)\| \cdot \Delta_k}{\frac{\Delta_k^2}{\|\nabla f(x^k)\|^2} \cdot \nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k)} \\ &= \frac{\|\nabla f(x^k)\|^3}{\Delta_k \cdot \nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k)} \end{aligned}$$

Also hat man insgesamt:

$$\alpha_k = \begin{cases} 1 & \nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k) \leq 0 \\ \min\left\{1, \dfrac{\|\nabla f(x^k)\|^3}{\Delta_k \cdot \nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k)}\right\} & \nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k) > 0 \end{cases}$$

4. Noch zu zeigen ist, dass jeder Cauchy-Punkt die Abstiegsbedingung (2.13) erfüllt.

Lemma 2.4 Es sei $\bar{\delta} \in (0, 2]$. Jeder Vektor p^k mit

$$m_k(0) - m_k(p^k) \geq \bar{\delta} \cdot (m_k(0) - m_k(p_c^k)) \tag{2.20}$$

erfüllt die Abstiegsbedingung aus (2.13) mit $\delta := \frac{\bar{\delta}}{2}$. Insbesondere erfüllen p_c^k selbst und jede exakte Lösung p^k von (2.11) die Bedingung (2.20) mit $\bar{\delta} = 1$.

Beweis:

- Wir betrachten die Erfüllbarkeit der Abstiegsbedingung (2.13) zuerst für $p^k := p_c^k$ und unterscheiden die folgenden drei Fälle:

  1. $\nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k) \leq 0$: Dann folgt

$$\begin{aligned} m_k(0) - m_k(p_c^k) &= f(x^k) - \left(f(x^k) + \nabla f(x^k)^T \cdot p_c^k + \frac{1}{2} \cdot (p_c^k)^T \cdot B_k \cdot p_c^k\right) \\ &= -\alpha_k \cdot \nabla f(x^k)^T \cdot p_l^k - \frac{\alpha_k^2}{2} \cdot (p_l^k)^T \cdot B_k \cdot p_l^k \\ &= \frac{\Delta_k \cdot \alpha_k}{\|\nabla f(x^k)\|} \cdot \nabla f(x^k)^T \cdot \nabla f(x^k) \\ &\quad \underbrace{-\frac{\alpha_k^2 \cdot \Delta_k^2}{2 \cdot \|\nabla f(x^k)\|^2} \cdot \nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k)}_{\geq 0} \\ &\overset{\alpha_k = 1}{\geq} \Delta_k \cdot \|\nabla f(x^k)\| \geq \|\nabla f(x^k)\| \cdot \min\left\{\Delta_k, \frac{\|\nabla f(x^k)\|}{\|B_k\|}\right\} \end{aligned}$$

  2. $\nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k) > 0$ und $\alpha_k < 1$: Dann folgt wie oben

$$\begin{aligned} m_k(0) - m_k(p_c^k) &= \Delta_k \cdot \alpha_k \cdot \|\nabla f(x^k)\| \underbrace{-\frac{\alpha_k^2 \cdot \Delta_k^2}{2 \cdot \|\nabla f(x^k)\|^2} \cdot \nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k)}_{<0} \\ &= \frac{\alpha_k}{2} \cdot \|\nabla f(x^k)\| \cdot \left(2\Delta_k - \frac{\alpha_k \cdot \Delta_k^2}{\|\nabla f(x^k)\|^3} \cdot \nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k)\right) \\ &\overset{\text{Def } \alpha_k}{=} \frac{\alpha_k}{2} \cdot \|\nabla f(x^k)\| \cdot (2\Delta_k - \Delta_k) = \frac{\alpha_k \cdot D_k}{2} \cdot \|\nabla f(x^k)\| \\ &\overset{\text{Def } \alpha_k}{=} \frac{\|\nabla f(x^k)\|^4}{2\nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k)} \geq \frac{\|\nabla f(x^k)\|^4}{2 \cdot \|B_k\| \cdot \|\nabla f(x^k)\|^2} \\ &= \frac{\|\nabla f(x^k)\|^2}{2 \cdot \|B_k\|} \geq \frac{1}{2} \cdot \|\nabla f(x^k)\| \cdot \min\left\{\Delta_k, \frac{\|\nabla f(x^k)\|}{\|B_k\|}\right\} \end{aligned}$$

  3. $\nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k)$ und $\alpha_k = 1$: Mit den Überlegungen aus (2) mit $\alpha_k = 1$ erhält man

$$\begin{aligned} m_k(0) - m_k(p_c^k) &= \frac{1}{2} \cdot \|\nabla f(x^k)\| \cdot \left(2\Delta_k - \frac{\Delta_k^2}{\|\nabla f(x^k)\|^3} \cdot \nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k)\right) \\ &\geq \frac{1}{2} \cdot \|\nabla f(x^k)\| \cdot \Delta_k \geq \frac{1}{2} \cdot \|\nabla f(x^k)\| \cdot \min\left\{\Delta_k, \frac{\|\nabla f(x^k)\|}{\|B_k\|}\right\} \end{aligned}$$

wegen

$$\frac{\Delta_k \cdot \nabla f(x^k)^T \cdot B_k \cdot \nabla f(x^k)}{\|\nabla f(x^k)\|^3} < 1$$

da $\alpha_k = 1$ (vgl. Wahl von α_k)

Insgesamt hat man also

$$m_k(0) - m_k(p_c^k) \geq \frac{1}{2} \cdot \|\nabla f(x^k)\| \cdot \min\left\{\Delta_k, \frac{\|\nabla f(x^k)\|}{\|B_k\|}\right\}$$

Unter Nutzung von (2.20) folgt hieraus

$$\begin{aligned} m_k(0) - m_k(p^k) &\geq \bar{\delta} \cdot (m_k(0) - m_k(p_c^k)) \\ &\geq \frac{1}{2} \cdot \bar{\delta} \cdot \|\nabla f(x^k)\| \cdot \min\left\{\Delta_k, \frac{\|\nabla f(x^k)\|}{\|B_k\|}\right\} \end{aligned}$$

Für die exakte Lösung p_*^k gilt insbesondere

$$m_k(0) - m_k(p_*^k) \geq m_k(0) - m_k(p_c^k)$$

Schnelle lokale Konvergenz eines Trust-Region-Verfahrens kann man nur erwarten, wenn lokal B_k wesentlich zur Bestimmung von p^k ausgenutzt wird und B_k eine hinreichend gute Approximation von der Hesse-Matrix von f an der Stelle x^k ist. Um trotzdem ein günstig zu lösendes Teilproblem für eine Näherungslösung p^k zu haben (p^k muss Abstiegsbedingung erfüllen) verwendet man zum Beispiel:

$$f(x^k) + \nabla f(x^k)^T \cdot p + p^T \cdot B_k \cdot p \to \min \text{ bei } \|p\| \leq \Delta_k, p \in \operatorname{span}\left\{\nabla f(x^k), B_k^{-1} \cdot \nabla f(x^k)\right\} \quad (2.21)$$

Offenbar ist der Cauchy-Punkt zulässig, sodass $m_k(p_k^c) \geq m_k(p^k)$, wobei p^k Lösung von (2.21) ist. Also ist die Abstiegsbedingung nach Lemma 2.4 auch für dieses p^k erfüllt. Wenn B_k negative Eigenwerte besitzt, wird der zweidimensionale Unterraum in (2.21) ersetzt durch

$$\operatorname{span}\left\{\nabla f(x^k), (B_k + \lambda \cdot I)^{-1} \cdot \nabla f(x^k)\right\}$$

für ein $\lambda \in (-\lambda_1, -2\lambda_1)$, wobei λ_1 der kleinste Eigenwert von B_k ist. Der Hauptaufwand solcher Techniken (two-dimensional subspace minimization) liegt in der Faktorisierung von B_k bzw. $B_k + \lambda \cdot I$.

2.3 Quasi-Newton-Verfahren

$$B_k \cdot d = -\nabla f(x^k) \quad (2.22)$$

Wie sollte B_{k+1} bei gegebenem B_k aussehen? Ziele:

1. $\operatorname{Rg}(B_{k+1} - B_k)$ klein
2. Akkumulierung der Informationen 2. Ordnung

Iterationsschritt:

$$x^{k+1} = x^k + \alpha_k \cdot d^k$$

mit Schrittweite $\alpha_k > 0$ und d^k aus (2.22). Mit Taylor-Formel ergibt sich

$$\nabla f(x^k) = \nabla f(x^{k+1}) + \nabla^2 f(x^{k+1}) \cdot (x^k - x^{k+1}) + o(\|x^k - x^{k+1}\|)$$

Vernachlässigt man das Restglied, so folgt

$$\nabla^2 f(x^{k+1}) \cdot (x^{k+1} - x^k) \approx \nabla f(x^{k+1}) - \nabla f(x^k)$$

Soll also B_{k+1} Informationen über $\nabla^2 f(x^{k+1})$ aus dem Schritt von x^k nach x^{k+1} aufnehmen, ist es sinnvoll

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

zu verlangen. Mit der Festlegung $s^k := x^{k+1} - x^k$ und $y^k := \nabla f(x^{k+1}) - \nabla f(x^k)$ geht diese Forderung über in

$$B_{k+1} \cdot s^k = y^k \tag{2.23}$$

Dies bezeichnet man in Anlehnung an das eindimensionale Sekanten-Verfahren auch als *Sekantengleichung*. Dann ist

$$(s^k)^T \cdot B_{k+1} \cdot s^k = (s^k)^T \cdot y^k > 0 \tag{2.24}$$

eine notwendige Bedingung für positive Definitheit von B_{k+1}. Dies ist für beliebige x^k, x^{k+1} im Allgemeinen nur richtig, wenn f streng konvex ist. Um bei beliebig gegebenem x^k, d^k trotzdem (2.24) für alle $x^{k+1} = x^k + \alpha_k \cdot d^k$ zu sichern, kann man Einfluss auf α_k nehmen.

Lemma 2.5 Es seien $x^k, d^k \in \mathbb{R}^n$ mit $\nabla f(x^k)^T \cdot d^k < 0$ und $0 < \delta_1 < \delta_2 < 1$ gegeben. Weiter sei f auf $\{x^k + \alpha \cdot d^k; \alpha \geq 0\}$ nach unten beschränkt.

1. Dann gibt es ein Intervall $[\alpha_1, \alpha_2]$ mit $0 < \alpha_1 < \alpha_2$ so, dass die *Wolfe-Bedingungen*

$$\begin{aligned} f(x^k + \alpha \cdot d^k) &\leq f(x^k) + \delta_1 \cdot \alpha \cdot \nabla f(x^k)^T \cdot d^k \\ \nabla f(x^k + \alpha \cdot d^k)^T \cdot d^k &\geq \delta_2 \cdot \nabla f(x^k)^T \cdot d^k \end{aligned}$$

   für alle $\alpha \in [\alpha_1, \alpha_2]$ erfüllt ist.

2. Ist für $\alpha_k > 0$ die Wolfe-Bedingung erfüllt, so gilt mit $s^k := \alpha_k \cdot d^k$:

$$(y^k)^T \cdot s^k \geq (\delta_2 - 1) \cdot \alpha_k \cdot \nabla f(x^k)^T \cdot d^k > 0$$

Beweis:

1. Übung

2. Mit der zweiten Ungleichung in der Wolfe-Bedingung und $\nabla f(x^k) \cdot d^k < 0$ erhält man

$$\begin{aligned} (y^k)^T \cdot s^k &= \alpha_k \cdot (\nabla f(x^{k+1}) - \nabla f(x^k))^T \cdot d^k \\ &\geq \alpha_k \cdot (\delta_2 - 1) \cdot \nabla f(x^k)^T \cdot d^k > 0 \end{aligned}$$

Falls B_k positiv definit und d^k entsprechend (2.22) bestimmt wird, ist $\nabla f(x^k)^T \cdot d^k < 0$ und deshalb mit Lemma 2.5(2) dann $(y^k)^T \cdot s^k > 0$, sofern α_k der Wolfe-Bedingung genügt. Zur Bestimmung von B_{k+1} betrachte das Optimierungsproblem

$$\|B_k - B\| \to \min \qquad \text{bei } B = B^T, B \cdot s^k = y^k \tag{2.25}$$

wobei s^k, y^k die Bedingung (2.24) erfüllen sowie B_k symmetrisch und positiv definit ist. Benutzt man für die Norm eine gewichtete Frobenius-Norm $\|.\|_W$ gegeben durch

$$\begin{aligned} \|A\|_W &:= \|W^{\frac{1}{2}} \cdot A \cdot W^{\frac{1}{2}}\|_F \\ \|C\|_F &:= \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}^2} \end{aligned}$$

wobei W eine positiv definite, symmetrische Matrix ist, die der Bedingung $W \cdot y^k = s^k$ genügt, so lässt sich zeigen, dass die folgende (von W unabhängige) Matrix B_{k+1} eindeutige Lösung von (2.25) ist:

$$B_{k+1} := (I - \gamma_k \cdot y^k \cdot (s^k)^T) \cdot B_k \cdot (I - \gamma_k \cdot s^k \cdot (y^k)^T) + \gamma_k \cdot y^k \cdot (y^k)^T \tag{DFP}$$

mit $\gamma_k := \frac{1}{(y^k)^T \cdot s^k}$ (DFP-Formel nach Davidon/Fletcher/Povell, 1959). Wegen

$$\begin{aligned} B_{k+1} &= B_k - \gamma_k \cdot y^k \cdot (s^k)^T - \gamma_k \cdot B_k \cdot s^k \cdot (y^k)^T + \gamma_k^2 \cdot \|s_k\|^2 \cdot y^k \cdot (y^k)^T + \gamma_k \cdot y^k \cdot (y^k)^T \\ &= B_k - \gamma_k \cdot (y^k \cdot (s^k)^T \cdot B_k + (y^k \cdot (s^k)^T \cdot B_k)^T) + \gamma_k \cdot (\gamma_k \cdot \|s_k\|^2 + 1) \cdot y^k \cdot (y^k)^T \end{aligned}$$

nennt man B_{k+1} auch eine Rang-2-Modifikation von B_k. Um Gleichung (2.22) effizient zu lösen, ist es mit Hilfe der sogenannten Sherman-Morrison-Woodbury-Formel leicht möglich aus der DFP-Formel eine Vorschrift zur Gewinnung der Inversen $H_{k+1} := B_{k+1}^{-1}$ aus $H_k := B_k^{-1}$.

$$H_{k+1} = H_k - \frac{H_k \cdot y^k \cdot (y^k)^T \cdot H_k}{(y^k)^T \cdot H^k \cdot y^k} + \frac{s^k \cdot (s^k)^T}{(y^k)^T \cdot s^k} \tag{DFP}$$

Damit folgt in (2.22):

$$\begin{aligned} B_{k+1} \cdot d^{k+1} &= -\nabla f(x^{k+1}) \\ \Rightarrow d^{k+1} &= -B_{k+1}^{-1} \cdot \nabla f(x^{k+1}) = -H_{k+1} \cdot \nabla f(x^{k+1}) \\ &= H_k \cdot \nabla f(x^{k+1}) + \left(-\frac{H_k \cdot y^k \cdot (y^k)^T \cdot H_k}{(y^k)^T \cdot H^k \cdot y^k} + \frac{s^k \cdot (s^k)^T}{(y^k)^T \cdot s^k} \right) \cdot \nabla f(x^{k+1}) \end{aligned}$$

Alternativ: BFGS-Formel (Broyden/Fletcher/Goldfarb/Shanno) Dazu beachte

$$B_{k+1} \cdot s^k = y^k \overset{\det B_{k+1} \neq 0}{\Leftrightarrow} s^k = B_{k+1}^{-1} \cdot y^k = H_{k+1} \cdot y^k$$

Betrachte das Optimierungsproblem

$$\|H_k - H\| \to \min \qquad \text{bei } H = H^T, H \cdot y^k = s^k$$

mit gewichteter Frobeniusnorm $\|.\|_W$ für $\|.\|$, wobei W positiv definit und symmetrisch mit $W \cdot s^k = y^k$. Dann erhält man als eindeutige Lösung:

$$H_{k+1} = (I - \gamma_k \cdot s^k \cdot (y^k)^T) \cdot H_k \cdot (I - \gamma_k \cdot y^k \cdot (s^k)^T) + \gamma_k \cdot s^k \cdot (s^k)^T \tag{2.26}$$

(BFGS-Formel).

Algorithmus 2.5: BFGS-Verfahren mit Wolfe-Schrittweite

(S1) Wähle $x^0 \in \mathbb{R}^n, H_0 \in \mathbb{R}^{n \times n}$ symmetrisch und positiv definit. Setze $k := 0$.

(S2) Ist $\nabla f(x^k) = 0$, dann stoppe Algorithmus.

(S3) Berechne $d^k := -H_k \cdot \nabla f(x^k)$.

(S4) Berechne $\alpha_k > 0$, sodass die Wolfe-Bedingung erfüllt ist.

(S5) $x^{k+1} := x^k + \alpha_k \cdot d^k$, $s^k := \alpha_k \cdot d^k$, $y^k := \nabla f(x^{k+1}) - \nabla f(x^k)$. Berechne H_{k+1} nach (2.26). Setze $k = k + 1$ und gehe zu Schritt 2.

Theorem 2.7 Es sei $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar. Dann ist Algorithmus 2.5 wohldefiniert und alle Matrizen der Folge $(H_k)_{k \in \mathbb{N}}$ sind symmetrisch und positiv definit. Setzt man zusätzlich voraus, dass der Algorithmus 2.5 nicht nach endlich vielen Schritten abbricht und f zweimal stetig differenzierbar ist und dass es Zahlen $m > 0, M > 0$ gibt, sodass

$$\forall x \in W(x^0) : \forall d \in \mathbb{R}^n : m \cdot \|d\|^2 \leq d^T \cdot \nabla^2 f(x) \cdot d \leq M \cdot \|d\|^2$$

und

$$W(x^0) = \{x \in \mathbb{R}^n; f(x) \leq f(x^0)\}$$

eine konvexe Menge ist, dann konvergiert die von Algorithmus 2.5 erzeugte Folge $(x^k)_{k \in \mathbb{N}}$ gegen die Lösung von (2.1).

Theorem 2.8 Es sei f zweimal stetig differenzierbar und Algorithmus (2.5) erzeuge gegen ein lokales Minimum x^* von (2.1) konvergente Folge $(x^k)_{k\in\mathbb{N}}$. Außerdem sei $\nabla^2 f(x^*)$ positiv definit und mit $\varepsilon_0 > 0, L > 0$ gelte

$$\forall x \in B(x^*, \varepsilon_0) : \|\nabla^2 f(x) - \nabla^2 f(x^*)\| \leq L \cdot \|x - x^*\|$$

sowie

$$\sum_{k=0}^{\infty} \|x^k - x^*\| < \infty$$

Dann konvergiert $(x^k)_{k\in\mathbb{N}}$ Q-superlinear gegen x^*.

3 Minimierung unter Nebenbedingungen

3.1 Lineare Optimierung

3.1.1 Grundlagen

- Standardaufgabe:
$$f(x) := c^T \cdot x \to \min \qquad \text{bei } Ax = b, x \geq 0 \tag{3.1}$$
mit $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$ gegeben. KKT-Bedingungen für (3.1) sind notwendige Optimalitätsbedingungen (da ACQ wegen Linearität der Nebenbedingungen erfüllt). Sie sind auch hinreichende Bedingungen, da f konvex und der zulässige Bereich konvex ist (vgl. Theorem 1.6 und Theorem 1.7). KKT-Bedingungen:
$$\begin{aligned} & c - u + A^T \cdot v = 0, Ax - b = 0, x \geq 0, u \geq 0, x^T \cdot u = 0 \\ \Leftrightarrow\ & -A^T \cdot v + u - c = 0, Ax - b = 0, x \geq 0, u \geq 0, \forall i : x_i \cdot u_i = 0 \\ \Leftrightarrow\ & A^T \cdot \lambda + s - c = 0, Ax - b = 0, x \geq 0, s \geq 0, x_i \cdot s_i = 0 \\ \Leftrightarrow\ & A^T \cdot \lambda + s - c = 0, Ax - b = 0, x \geq 0, s \geq 0, X \cdot S \cdot e = 0 \end{aligned} \tag{3.2}$$
mit $X := \operatorname{diag}(x_1, \ldots, x_n), S := \operatorname{diag}(s_1, \ldots, s_n), e = (1, \ldots, 1)^T \in \mathbb{R}^n$.

- Man erhält ein zu (3.2) äquivalentes KKT-System, wenn man die folgende Optimierungsaufgabe zu Grunde legt:
$$b^T \cdot \lambda \to \max \qquad \text{bei } A^T \cdot \lambda + s = c, s \geq 0 \tag{3.3}$$
Die KKT-Bedingungen zu (3.3) lauten
$$\begin{aligned} & \begin{pmatrix} -b \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -I \end{pmatrix} y + \begin{pmatrix} A \\ I \end{pmatrix} v = 0, A^T \lambda + s - c = 0, s \geq 0, y \geq 0, \forall i : y_i \cdot s_i = 0 \\ \Leftrightarrow\ & A^T \cdot \lambda + s - c = 0, Av - b = 0, v - y = 0, y \geq 0, s \geq 0, S \cdot Y \cdot e = 0 \\ \overset{x:=y}{\Leftrightarrow}\ & A^T \cdot \lambda + s - c = 0, Ax - b = 0, s \geq 0, x \geq 0, S \cdot X \cdot e = 0 \end{aligned}$$
Offenbar gilt $v = y$, also kann man v eliminieren. Setzt man außerdem $x := y$, so erhält man das System (3.2). Die Optimierungsaufgabe (3.3) wird als die zur Aufgabe (3.1) *duale Optimierungsaufgabe* bezeichnet und umgekehrt.

Theorem 3.1 Folgende Aussagen sind äquivalent:

1. Das Problem (3.1) ist lösbar.
2. Das Problem (3.3) ist lösbar.
3. Die zulässigen Bereiche der Aufgaben (3.1) und (3.3) sind nicht leer.

Beweis:

- (1) $\Leftrightarrow$ (2): Klar.

- (1) $\Rightarrow$ (3): Klar.
- (3) $\Rightarrow$ (1):

$$G_p := \{x \in \mathbb{R}^n; Ax = b, x \geq 0\}$$
$$G_D := \{(\lambda, s) \in \mathbb{R}^m \times \mathbb{R}^n; A^T \cdot \lambda + s - c = 0, s \geq 0\}$$

Für beliebig gewählte $\bar{x} \in G_P$ und $(\bar{\lambda}, \bar{s}) \in G_D$ gilt:

$$A\bar{x} = b, A^T \cdot \bar{\lambda} + \bar{s} = c, \bar{s} \geq 0, \bar{x} \geq 0 \tag{3.4}$$

Multipliziert man die erste Gleichung mit $\bar{\lambda}^T$ und die zweite Gleichung mit $\bar{x}^T$, so folgt

$$\bar{\lambda}^T \cdot A \cdot \bar{x} = \bar{\lambda}^T \cdot b, \bar{x}^T \cdot A^T \cdot \bar{\lambda} + \bar{x}^T \cdot \bar{s} = \bar{x}^T \cdot c$$

Einsetzen ergibt

$$\bar{\lambda}^T \cdot b + \bar{x}^T \cdot \bar{s} = \bar{x}^T \cdot c$$

Da $\bar{x} \geq 0, \bar{s} \geq 0$ gilt $\bar{x}^T \cdot \bar{s} \geq 0$ und daher

$$\bar{\lambda}^T \cdot b \leq \bar{x}^T \cdot c \tag{3.5}$$

Speziell hat man damit, dass die Zielfunktion von (3.1) auf G_P nach unten durch $\bar{\lambda}^T \cdot b$ beschränkt ist.

Da (3.5) für jeden beliebigen Punkt $(\bar{x}, \bar{\lambda}, \bar{s}) \in G_P \times G_D$ gilt, hat das System

$$Ax - b = 0, A^T \cdot \lambda + s - c = 0, x \geq 0, s \geq 0, \bar{\lambda}^T \cdot b > \bar{x}^T \cdot c$$

keine Lösung. Für $\alpha \neq 0$ ist deshalb auch das folgende System nicht lösbar:

$$\begin{pmatrix} A^T & 0 & -c \\ 0 & -I & 0 \\ 0 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} \lambda \\ x \\ \alpha \end{pmatrix} \leq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & A & -b \end{pmatrix} \cdot \begin{pmatrix} \lambda \\ x \\ \alpha \end{pmatrix} = 0, \begin{pmatrix} b \\ -c \\ 0 \end{pmatrix}^T \cdot \begin{pmatrix} \lambda \\ x \\ \alpha \end{pmatrix} > 0 \tag{3.6}$$

Man überlegt sich nun noch die Nichtlösbarkeit für $\alpha = 0$: Das System (3.6) hat dann die Gestalt

$$A^T \cdot \lambda \leq 0, x \geq 0, A \cdot x = 0, b^T \cdot \lambda - c^T \cdot x > 0$$

Daraus sowie aus (3.4) und aus der Existenz von Punkten $\hat{x} \in G_p \neq \emptyset$, $(\hat{\lambda}, \hat{s}) \in G_D \neq \emptyset$ erhält man den Widerspruch

$$0 = x^T \cdot A^T \cdot \hat{\lambda} \leq x^T \cdot c < b^T \cdot \lambda = \lambda^T \cdot A\hat{x} \leq 0$$

Somit ist System (3.6) nicht lösbar. Nach Lemma 1.5 (Farkas) besitzt dann aber das System

$$\begin{pmatrix} A & 0 & 0 \\ 0 & -I & 0 \\ -c^T & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} x \\ s \\ \beta \end{pmatrix} + \begin{pmatrix} 0 \\ A^T \\ -b^T \end{pmatrix} \cdot (-\lambda) = \begin{pmatrix} b \\ -c \\ 0 \end{pmatrix}, \begin{pmatrix} x \\ s \\ \beta \end{pmatrix} \geq 0$$

eine Lösung. Dies bedeutet, dass

$$Ax = b, A^T \cdot \lambda + s - c = 0, c^T \cdot x + \beta - b^T \cdot \lambda = 0, (x, s, \beta)^T \geq 0 \tag{3.7}$$

lösbar ist. Somit ergibt sich (nacheinander):

$$\lambda^T \cdot Ax = \lambda^T \cdot b, x^T \cdot A^T \cdot \lambda + x^T \cdot s = x^T \cdot c, \lambda^T \cdot b + x^T \cdot s = x^T \cdot c,$$
$$b + x^T \cdot s = 0, x^T \cdot s \leq 0$$

Wegen $(x, s) \geq 0$ ist $x^T \cdot s \geq 0$ und es folgt $x^T \cdot s = 0$. Damit ist die Lösbarkeit des KKT-Systems (3.2) gezeigt. Deshalb muss auch (3.1) (und damit (3.3)) lösbar sein.

Zur Vereinfachung der Schreibweise bezeichne S_P die Lösungsmenge des primalen Problems (3.1) und S_D die Lösungsmenge des dualen Problems (3.3). Es lässt sich zeigen, dass $S := S_P \times S_D$ die Lösungsmenge des zugehörigen KKT-Systems (3.2) ist. (Übung)

Beispiel 3.1

$$(P) : x_1 \to \min \qquad \text{bei}\, x_1 + x_2 + x_3 = 1, (x_1, x_2, x_3) \geq 0$$

Dann ist offenbar

$$S_P = \left\{ x^* = \begin{pmatrix} 0 \\ t \\ 1-t \end{pmatrix} ; t \in [0,1] \right\}$$

Die zugehörige duale Optimierungsaufgabe lautet

$$(D) : \lambda \to \max \qquad \text{bei } \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \cdot \lambda + s = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, s \geq 0$$

Man erhält

$$S_D = \left\{ (\lambda^*, s^*); \lambda^* = 0, s^* = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$$

Theorem 3.2 Es seien $\bar{x} \in G_P$ und $(\bar{\lambda}, \bar{s}) \in G_D$. Dann ist $S_P \neq \emptyset$ und $S_D \neq \emptyset$ und es gilt für beliebige $(x^*, \lambda^*, s^*) \in S_P \times S_D$:

$$b^T \cdot \bar{\lambda} \leq b^T \cdot \lambda^* = c^T \cdot x^* \leq c^T \cdot \bar{x}$$

d.h. starke Dualität (d.h. $b^T \lambda^* = c^T \cdot x^*$) und schwache Dualität (d.h. $b^T \bar{\lambda} \leq c^T \cdot \bar{x}$) ist erfüllt.

Beweis:

- Wegen Theorem 3.1 folgt $S_P \neq \emptyset$ und $S_D \neq \emptyset$. Es sei $(x^*, \lambda^*, s^*) \in S_P \times S_D$ beliebig gewählt. Dann erhält man unter Ausnutzung von (3.2):

$$\begin{aligned} b^T \cdot \lambda^* &= (Ax^*)^T \cdot \lambda^* = (x^*)^T \cdot A^T \cdot \lambda^* + (x^*)^T \cdot s^* \\ &= (x^*)^T \cdot (A^T \cdot \lambda + s^*) = c^T \cdot x^* \end{aligned}$$

  Da $x^* \in S_P$ und $(\lambda^*, s^*) \in S_D$ sowie $\bar{x} \in G_P$ und $(\bar{\lambda}, \bar{s}) \in G_D$ gilt offenbar

$$b^T \cdot \bar{\lambda} \leq b^T \cdot \lambda^* \qquad\qquad c^T \cdot x^* \leq c^T \cdot \bar{x}$$

Theorem 3.3 Es sei $S \neq \emptyset$. Dann bilden die Indexmengen

$$\begin{aligned} \mathcal{B} &:= \{i \in \{1, \ldots, n\}; \exists x^* \in S_P : x_i^* > 0\} \\ \mathcal{N} &:= \{i \in \{1, \ldots, n\}; \exists (\lambda^*, s^*) \in S_D : s_i^* > 0\} \end{aligned}$$

eine Partition von $\{1, \ldots, n\}$, d.h. es gilt

1. $\mathcal{B} \cap \mathcal{N} = \emptyset$
2. $\mathcal{B} \cup \mathcal{N} = \{1, \ldots, n\}$ (Goldman-Tacker)

Beweis:

1. Um $\mathcal{B} \cap \mathcal{N} = \emptyset$ zu zeigen, nehmen wir das Gegenteil an. Es möge also ein $j \in \mathcal{B} \cap \mathcal{N}$ existieren. Folglich gibt es $(x^*, \lambda^*, s^*) \in S$ und $(\bar{x}, \bar{\lambda}, \bar{s}) \in S$, sodass $x_j^* > 0$, $s_j^* = 0$ und $\bar{x}_j = 0, \bar{s}_j > 0$. Daraus erhält man

$$(x^* + \bar{x})^T \cdot (s^* + \bar{s}) > 0$$

   Die Konvexität von S (Übung) liefert

$$\frac{1}{2} \cdot (x^*, \lambda^*, s^*) + \frac{1}{2} \cdot (\bar{x}, \bar{\lambda}, \bar{s}) \in S$$

   also insbesondere

$$(x^* + \bar{x})^T \cdot (s^* + \bar{s}) = 0$$

   und damit einen Widerspruch.

2. Es sei $J := \{1, \ldots, n\} \backslash (\mathcal{B} \cup \mathcal{N})$. Angenommen es existiert $j \in J \neq \emptyset$. Um einen Widerspruch zu erhalten, wird das System

$$\forall i \in J \backslash \{j\} : (A_{.,i})^T \cdot w \leq 0, \forall i \in \mathcal{B} : (A_{.,i})^T \cdot w = 0, -(A_{.,j})^T \cdot w > 0 \tag{3.8}$$

betrachtet. Es werden nun zwei Fälle unterschieden:

(i) Das System (3.8) sei lösbar. Dann bezeichne w^* eine Lösung. Wegen $S \neq \emptyset$ und der Konvexität von S kann ein $(x^*, \lambda^*, s^*) \in S$ gewählt werden, sodass $s^*_{\mathcal{N}} > 0$. (Denn: Seien $i_1, i_2 \in \mathcal{N}$, dann existieren $(\bar{x}, \bar{\lambda}, \bar{s}) \in S$ mit $\bar{s}_{i_1} > 0$ und $(\hat{x}, \hat{\lambda}, \hat{s}) \in S$ mit $\hat{s}_{i_2} > 0$. Dann wegen Konvexität $\frac{1}{2} \cdot (\hat{x}, \hat{\lambda}, \hat{s}) + \frac{1}{2} \cdot (\bar{x}, \bar{\lambda}, \bar{s}) \in S$ mit $\frac{1}{2}\bar{s}_{i_j} + \frac{1}{2}\hat{s}_{i_j} > 0$ für $j \in \{1, 2\}$.) Falls $\mathcal{N} = \emptyset$ entfällt die letzte Bedingung. Nun sei $(\bar{\lambda}, \bar{s})$ definiert durch

$$\bar{\lambda} = \lambda^* + \varepsilon \cdot w^* \qquad \bar{s} := c - A^T \cdot \bar{\lambda} = s^* - \varepsilon \cdot A^T w^*$$

wobei (im Fall $\mathcal{N} \neq \emptyset$) $\varepsilon > 0$ so klein gewählt wird, sodass

$$\forall i \in \mathcal{N} : \bar{s}_i = s^*_i - \varepsilon \cdot (A^T \cdot w^*) > 0$$

Weiter gilt für jedes $\varepsilon > 0$ wegen (3.8) auch

$$\begin{aligned}
\bar{s}_j &= \underbrace{s^*_j}_{0} - \varepsilon \cdot (A_{.,j})^T \cdot w^* > 0 \\
\bar{s}_i &= \underbrace{s^*_i}_{0} - \varepsilon \cdot (A_{.,i})^T \cdot w^* \geq 0 \qquad (i \in J \backslash \{j\}) \\
\bar{s}_{\mathcal{B}} &= s^*_{\mathcal{B}} - \varepsilon \cdot \underbrace{(A_{.,\mathcal{B}})^T \cdot w^*}_{=0} = s^*_{\mathcal{B}} = 0
\end{aligned}$$

Damit genügt der Vektor $(\bar{\lambda}, \bar{s})$ offenbar den Bedingungen $\bar{s} \geq 0$, $A^T \cdot \bar{\lambda} + \bar{s} = c$, also $(\bar{\lambda}, \bar{s}) \in G_D$.
Es sei $\hat{x}$ eine beliebige Lösung von (3.1). Nach Definition der Indexmenge $\mathcal{B}$ muss dann $\hat{x}_{\mathcal{N} \cup J} = 0$ gelten, woraus mit $\bar{s}_{\mathcal{B}} = 0$ sofort $\hat{x}^T \cdot \bar{s} = 0$ folgt. Also genügt $(\hat{x}, \bar{\lambda}, \bar{s})$ dem System (3.2), d.h. $(\hat{x}, \bar{\lambda}, \bar{s}) \in S$ und damit $(\bar{\lambda}, \bar{s}) \in S_D$. Wegen $\bar{s}_j > 0$ gilt aber $j \in \mathcal{N}$. Widerspruch zu $j \in J$.

(ii) Das System (3.8) sei unlösbar. Dann liefert Lemma 1.5 (Farkas) die Lösbarkeit des Systems

$$\sum_{j \in J \backslash \{j\}} u_i \cdot A_{.,i} + \sum_{i \in \mathcal{B}} v_i \cdot A_{.,i} = -A_{.,j}, \forall i \in J \backslash \{j\} : u_i \geq 0 \tag{3.9}$$

Es sei nun ein Vektor $\bar{u} \in \mathbb{R}^{|J|}$ wie folgt definiert:

$$\bar{u}_i := \begin{cases} u_i & i \in J \backslash \{j\} \\ 1 & i = j \end{cases}$$

Damit ergibt sich aus (3.9)

$$A_{.,\mathcal{J}} \cdot \bar{u} + A_{.,\mathcal{B}} \cdot v = 0, \bar{u} \geq 0, \bar{u}_j > 0 \tag{3.10}$$

Nun sei x^* eine Lösung von (3.1) mit $x^*_{\mathcal{B}} > 0$. Der Vektor $\bar{x}$ sei definiert durch

$$\bar{x}_i := \begin{cases} x^*_i + \varepsilon \cdot v_i & i \in \mathcal{B} \\ \varepsilon \cdot \bar{u}_i & i \in \mathcal{J} \\ 0 & i \in \mathcal{N} \end{cases}$$

wobei $\varepsilon > 0$ und im Fall $\mathcal{B} \neq \emptyset$ der Teil $\bar{x}_\mathcal{B}$ entfällt. Aus (3.10) folgt damit:

$$\begin{aligned} A \cdot \bar{x} &= A_{.,J} \cdot \bar{x}_J + A_{.,\mathcal{B}} \cdot \bar{x}_\mathcal{B} + A_{.,\mathcal{N}} \cdot \underbrace{\bar{x}_\mathcal{N}}_{0} \\ &= \varepsilon \cdot A_{.,J} \cdot \bar{u} + \underbrace{A_{.,\mathcal{B}} \cdot x^*_\mathcal{B}}_{b} + \varepsilon \cdot A_{.,\mathcal{B}} \cdot v = b \end{aligned}$$

für $\varepsilon > 0$ hinreichend klein gilt $\bar{x} \geq 0$, d.h. $\bar{x} \in G_P$.
Es sei $(\hat{\lambda}, \hat{s})$ eine beliebige Lösung des dualen Problems (3.3). Nach Definition von $\mathcal{N}$ gilt $s_{\mathcal{B} \cup J} = 0$. Somit hat man $\bar{x}^T \cdot \hat{s} = 0$ (da $x_\mathcal{N} = 0$). Also ist $(\bar{x}, \hat{\lambda}, \hat{s}) \in S$ und $\bar{x} \in S_P$. Da $\bar{x}_j = \varepsilon \cdot \bar{u}_j > 0$ folgt $j \in \mathcal{B}$ im Widerspruch zu $j \in J$.

Es folgt $J = \emptyset$, also $\{1, \ldots, n\} = \mathcal{B} \cup \mathcal{N}$.

Bemerkung: Theorem 3.3 kann wie folgt interpretiert werden:

1. Falls das KKT-System (3.2) eine Lösung (x^*, λ^*, s^*) besitzt mit $x^*_i > 0$, so kann es keine Lösung $(\bar{x}, \bar{\lambda}, \bar{s})$ geben mit $\bar{s}_i > 0$.
2. Das zu den linearen Optimierungsproblem (3.1) und (3.3) gehörende KKT-System (3.2) besitzt (bei Lösbarkeit) stets eine streng komplementäre Lösung $(x^*, \lambda^*, s^*) \in S$, d.h. $x^* + s^* > 0$.

Bemerkung zum Beweis: 1. kann auch durch die Betrachtung von $(x^*, \bar{\lambda}, \bar{s}) \in S$ zum Widerspruch geführt werden (in S wegen $S = S_P \times S_D$), denn Komplementaritätsbedingung nicht erfüllt.

3.1.2 Der zentrale Pfad

Anstelle des KKT-Systems (3.2) betrachten wir das parametrisierte System

$$F_\tau(x, \lambda, s) := \begin{pmatrix} A^T \cdot \lambda + s - c \\ A \cdot x - b \\ X \cdot S \cdot e - \tau e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, x \geq 0, s \geq 0 \tag{3.11}$$

mit dem Parameter $\tau > 0$. Offenbar ist $X \cdot S \cdot e = \tau \cdot e, x > 0, s > 0$ zu $X \cdot S \cdot e = \tau \cdot e, s \geq 0, x \geq 0$ äquivalent. Es entsteht nun die Frage, unter welchen Bedingungen das System (3.11) zu jedem $\tau > 0$ eine eindeutige Lösung besitzt und was für $\tau \to 0$ passiert.
Um dies zu untersuchen, bezeichne

$$G^0 := \left\{(x, s) \in \mathbb{R}^n \times \mathbb{R}^n; \exists \lambda \in \mathbb{R}^m : A^T \cdot \lambda + s - c = 0, A \cdot x - b = 0, x > 0, s > 0\right\}$$

die Menge der streng zulässigen inneren Punkte. Im Folgenden werden nun Beziehungen zwischen der Lösungsmenge von (3.11) und der des Minimierungsproblems

$$f_\tau(x, s) := \frac{1}{\tau} \cdot x^T \cdot s - \sum_{i=1}^{n} \log(s_i \cdot x_i) \to \min \qquad \text{bei } (x, s) \in G^0 \tag{3.12}$$

hergestellt.

Lemma 3.1 Es sei $G^0 \neq \emptyset$ und $\tau > 0$. Falls $(x_\tau, \lambda_\tau, s_\tau)$ das System (3.11) löst, dann löst (x_τ, s_τ) auch das Minimierungsproblem (3.12). Umgekehrt gibt es zu jeder Lösung (x_τ, s_τ) von (3.12) ein λ_τ, sodass $(x_\tau, \lambda_\tau, s_\tau)$ das System (3.11) löst.

Beweis:

- Falls (3.12) eine Lösung (x_τ, s_τ) besitzt, so muss ein λ_τ existieren, sodass die Aufgabe

$$\tilde{f}_\tau(x, \lambda, s) \to \min \qquad \text{bei } (x, \lambda, s) \in \tilde{G} \tag{3.13}$$

mit

$$\tilde{G} := \{(x, \lambda, s) \in \mathbb{R}^{n+m+n} | A^T \cdot \lambda + s - c = 0, A \cdot x - b = 0, x > 0, s > 0\}$$
$$\tilde{f}_\tau(x, \lambda, s) := f_\tau(x, s) \qquad ((x, \lambda, s) \in \tilde{G})$$

die Lösung $(x_\tau, \lambda_\tau, s_\tau)$ besitzt und $(x_\tau, s_\tau) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$ liegt.
Die KKT-Bedingungen für (3.13) lauten

$$\begin{aligned} \frac{1}{\tau} \cdot S \cdot e - X^{-1} \cdot e + A^T \cdot v &= 0 \\ \frac{1}{\tau} \cdot X \cdot e - S^{-1} \cdot e + w &= 0 \\ A \cdot w &= 0 \\ Ax - b &= 0 \\ A^T \cdot \lambda + s - c &= 0 \\ (x, s) &\geq 0 \end{aligned}$$

(KKT-Bedingen notwendig, weil affin-lineare Nebenbedingungen. Lagrange-Multiplikatoren für Ungleichungen entfallen, sind alle 0 wegen $x > 0, s > 0$.) Wird die zweite Gleichung mit A multipliziert und berücksichtigt man die dritte, so ergibt sich

$$A \cdot \left(\frac{1}{\tau} \cdot X \cdot e - S^{-1} \cdot e \right) = 0$$

Multipliziert man $\left(\frac{1}{\tau} \cdot X \cdot e - S^{-1} \cdot e\right)^T$ mit der ersten Gleichung, so folgt

$$\left(\frac{1}{\tau} \cdot X \cdot e - S^{-1} \cdot e \right) \cdot \left(\frac{1}{\tau} \cdot S \cdot e - X^{-1} \cdot e \right) = 0$$

Dies ergibt weiter

$$\begin{aligned} 0 &= (X \cdot e - \tau \cdot S^{-1} \cdot e)^T \cdot (S \cdot e - \tau \cdot X^{-1} \cdot e) \\ &= (X \cdot e - \tau \cdot S^{-1} \cdot e)^T \cdot (X^{-\frac{1}{2}} \cdot S^{\frac{1}{2}}) \cdot (X^{\frac{1}{2}} \cdot S^{-\frac{1}{2}}) \cdot (S \cdot e - \tau \cdot X^{-1} \cdot e) \\ &= ((X \cdot S)^{\frac{1}{2}} \cdot e - \tau \cdot (X \cdot S)^{-\frac{1}{2}} \cdot e)^T \cdot ((X \cdot S)^{\frac{1}{2}} \cdot e - \tau \cdot (X \cdot S)^{-\frac{1}{2}} \cdot e) \\ &= \|((X \cdot S)^{\frac{1}{2}} \cdot e - \tau \cdot (X \cdot S)^{-\frac{1}{2}} \cdot e)\|^2 \\ &= \|(X \cdot S)^{-\frac{1}{2}} \cdot (X \cdot S \cdot e - \tau \cdot e)\|^2 \\ \Leftrightarrow 0 &= X \cdot S \cdot e - \tau \cdot e = 0 \end{aligned}$$

Jede Lösung $(x_\tau, \lambda_\tau, s_\tau)$ der KKT-Bedingungen zu (3.13) erfüllt also neben der Zulässigkeit $(x_\tau, \lambda_\tau, s_\tau) \in \tilde{G}$ die Bedingung $X_\tau \cdot S_\tau \cdot e = \tau \cdot e$ und ist somit Lösung von (3.11).

- Es sei nun umgekehrt $(x_\tau, \lambda_\tau, s_\tau)$ eine Lösung von (3.11). Dann ist $(x_\tau, s_\tau) \in G^0$, also zulässiger Punkt von (3.12). Weiterhin gilt $X_\tau \cdot S_\tau \cdot e = \tau \cdot e$, dass

$$f_\tau(x_\tau, s_\tau) = n \cdot \frac{\tau}{\tau} - n \cdot \log(\tau)$$

Mit $g(t) := t - \log(t) - 1$ ergibt sich andererseits für $(x, s) \in G^0$:

$$\begin{aligned} f_\tau(x, s) &= \sum_{i=1}^{n} \left(\frac{x_i \cdot s_i}{\tau} - \log(x_i \cdot s_i) \right) \\ &= \sum_{i=1}^{n} \left(\frac{x_i \cdot s_i}{\tau} - \log\left(\frac{x_i \cdot s_i}{\tau} \right) - \log \tau - 1 \right) + n \\ &= n - n \cdot \log \tau + \sum_{i=1}^{n} g\left(\frac{x_i \cdot s_i}{\tau} \right) \\ &= f_\tau(x_\tau, s_\tau) + \sum_{i=1}^{n} g\left(\frac{x_i \cdot s_i}{\tau} \right) \end{aligned}$$

Da $g(t) \geq 0$ für alle $t > 0$ ergibt sich die Abschätzung $f_\tau(s_\tau, s_\tau) \leq f_\tau(x, s)$ für alle $(x, s) \in G^0$. Folglich ist jede Lösung von (3.11) auch Lösung von (3.12).

Lemma 3.2 Es sei $G^0 \neq \emptyset$ und $\tau > 0$ beliebig aber fest. Dann gilt:

1. f_τ ist streng konvex auf G^0.
2. Für jedes $C > 0$ ist die Niveaumenge

$$W_\tau(C) := \{(x, s) \in G^0; f_\tau(x, s) \leq C\}$$

ist kompakt.

Beweis:

1. Übung
2. Die Abgeschlossenheit von $W_\tau(C)$ ist leicht einzusehen (Urbild abgeschlossener Menge unter stetiger Funktion). Es wird nun die Beschränktheit gezeigt. Dazu sei $(\bar{x}, \bar{s}) \in G^0$ fest gewählt. Sei nun $(x, s) \in W_\tau(C)$ beliebig aber fest gewählt. Dann gibt es $\bar{\lambda}, \lambda \in \mathbb{R}^m$, sodass

$$A^T \cdot \bar{\lambda} + \bar{s} - c = 0 \qquad A^T \cdot \lambda + s - c = 0$$

Da auch $A \cdot x = b, A \cdot \bar{x} = b$, folgt $A \cdot (x - \bar{x}) = 0$ sowie

$$A^T \cdot (\bar{\lambda} - \lambda) + (\bar{s} - s) = 0$$

Beide Gleichungen zusammen ergeben

$$(\bar{x} - x)^T \cdot (\bar{s} - s) = (\bar{x} - x)^T \cdot A^T \cdot (\bar{\lambda} - \lambda) = 0$$

Wegen $f_\tau(x, s) \leq C$ erhält man daraus

$$\begin{aligned}
\bar{x}^T \cdot s + \bar{s}^T \cdot x &= x^T \cdot s + \bar{x}^T \cdot \bar{s} \\
&= \bar{x}^T \cdot \bar{s} + \tau \cdot \underbrace{\left(\frac{1}{\tau} \cdot x^T \cdot s - \sum_{i=1}^n \log(x_i \cdot s_i) \right)}_{f_\tau(x,s)} + \tau \cdot \sum_{i=1}^n \log(x_i \cdot s_i) \\
&\leq \bar{x}^T \cdot \bar{s} + \tau \cdot \left(C + \sum_{i=1}^n \log(x_i \cdot s_i) \right) \\
&= \bar{x}^T \cdot \bar{s} + \tau \cdot \left(C + \sum_{i=1}^n \log x_i + \sum_{i=1}^n \log s_i \right) \qquad (3.14)
\end{aligned}$$

Da $(\bar{x}, \bar{s}) > 0$ gilt

$$\xi := \min\{\min\{\bar{x}_i, \bar{s}_i; i = 1, \ldots, n\}\} > 0$$

Aus (3.14) erhält man damit

$$\xi \cdot \left(\sum_{i=1}^n x_i + \sum_{i=1}^n s_i \right) \leq \bar{x}^T \cdot \bar{s} + \tau \cdot C + \tau \cdot \left(\sum_{i=1}^n \log x_i + \sum_{i=1}^n \log s_i \right)$$

woraus unter Berücksichtigung von $\lim_{a \to \infty}(a - \log a) = \infty$ die Beschränktheit von $W_\tau(C)$ folgt.

$$\underbrace{\xi \cdot \left(\sum_{i=1}^n x_i + \sum_{i=1}^n s_i \right) - \tau \cdot \left(\sum_{i=1}^n \log x_i + \sum_{i=1}^n \log s_i \right)}_{\geq M \cdot \left(\sum_{i=1}^n x_i + \sum_{i=1}^n s_i \right) \text{ für } x_i, s_i \text{ groß}} \leq K$$

$$\Rightarrow \sum_{i=1}^n x_i + \sum_{i=1}^n s_i \leq \frac{K}{M}$$

Theorem 3.4 Sei $G^0 \neq \emptyset$. Dann besitzt das System (3.11) für jedes $\tau > 0$ eine Lösung $(x_\tau, \lambda_\tau, s_\tau)$. Die Komponenten (x_τ, s_τ) sind dabei stetig eindeutig. Falls A den Rang m hat, ist auch λ_τ eindeutig.

Beweis:

- Es sei $(\hat{x}, \hat{s}) \in G^0$. Für $C := f_\tau(\hat{x}, \hat{s})$ ist die Niveaumenge $W_\tau(C)$ nicht leer und nach Lemma 3.2(2) kompakt. Falls (x_τ, s_τ) (3.12) löst, dann auch (x_τ, s_τ) Lösung des Problems

$$f_\tau(x, s) \to \min \qquad \text{bei } (x, s) \in W_\tau(C) \tag{3.15}$$

  und umgekehrt. Wegen der Kompaktheit von $W_\tau(C)$ und der Stetigkeit von f_τ auf $W_\tau(C)$ muss (3.15) nach dem Satz von Weierstraß mindestens eine Lösung besitzen. Folglich ist auch (3.12) lösbar. Für letztere Aufgabe folgert man aus Theorem 1.1(2) unter Ausnutzung der strengen Konvexität von f_τ auf G^0 (Lemma 3.2(1)) unter der Konvexität von G^0 die eindeutige Lösbarkeit von (3.12). Mit Lemma 3.1 folgt daher die Lösbarkeit von (3.11) und die Eindeutigkeit der Lösung bzgl. der Komponenten (x, s).

- Falls $\operatorname{Rg} A = m$ und $(x_\tau, \lambda_\tau, s_\tau)$ eine Lösung von (3.11) bezeichnet, dann hat die in (3.11) vorkommende Gleichung $A^T \cdot \lambda + s - c = 0$ für $s := s_\tau$ genau eine Lösung (lineare Unabhängigkeit der Spalten von A^T).

Definition 3.1 Als *zentralen Pfad* zum Paar der zueinander dualen linearen Optimierungsaufgaben (3.1) bzw. (3.3) bezeichnet man die Menge

$$\{(x_\tau, \lambda_\tau, s_\tau); \tau > 0\}$$

wobei $(x_\tau, \lambda_\tau, s_\tau)$ eine zu $\tau > 0$ gehörende Lösung von (3.11) bezeichnet.

Bemerkung:

1. Es lässt sich weiterhin zeigen, dass der Grenzwert

$$(x^*, s^*) := \lim_{\tau \downarrow 0} (x_\tau, s_\tau)$$

   existiert und $\lambda^* \in \mathbb{R}^m$ existiert, sodass $(x^*, \lambda^*, s^*) \in S$, d.h. x^* löst (3.1) und (λ^*, s^*) löst (3.3). Außerdem gilt $[x^* + s^*]_i > 0$ für $i = 1, \ldots, n$.

3.1.3 Prädiktor-Korrektor-Verfahren

Definiere

$$\mathcal{N}_2(\theta) := \left\{(x, \lambda, s) \in \tilde{G}^0; \|X \cdot S \cdot e - \mu \cdot e\|_2 \leq \theta \cdot \mu\right\}$$

mit $\mu := \frac{x^T \cdot s}{n}$ und $\theta \in [0, 1)$. Für $\theta = 0$ erhält man speziell

$$\mathcal{N}_2(0) = \{(x, \lambda, s) \in \tilde{G}^0; \forall i = 1, \ldots, n : x_i \cdot s_i = \mu\}$$

also der zentrale Pfad selbst.
Für praktische Pfadverfolungs-Algorithmen wird man daher $\theta \in (0, 1)$ wählen. Dann liefert $\|X \cdot S \cdot e - \mu \cdot e\|_2 \leq \theta \cdot \mu$ insbesondere

$$(x_i \cdot s_i - \mu)^2 \leq \sum_{i=1}^{n} (x_i \cdot s_i - \mu)^2 \leq \theta^2 \cdot \mu^2$$

also

$$\begin{aligned} -\theta \cdot \mu \leq (x_i \cdot s_i - \mu) &\leq \theta \cdot \mu \\ \Rightarrow (1 - \theta) \cdot \mu \leq x_i \cdot s_i &\leq (1 + \theta) \cdot \mu \end{aligned}$$

für $i = 1, \ldots, n$.
Entsprechend (3.11) wird der zentrale Pfad durch das System ($\tau > 0$)

$$F_\tau(x, \lambda, s) = 0 \qquad (x \geq 0, s \geq 0)$$

definiert. Sei $(x^k, \lambda^k, s^k) \in \mathcal{N}_2\left(\frac{1}{4}\right)$. Für den Prädiktorschritt führe einen Newton-Schritt für $F_0(x, \lambda, s) = 0$ mit Schrittweitenstrategie aus, sodass $(x^{k+1}, \lambda^{k+1}, s^{k+1}) \in \mathcal{N}_2\left(\frac{1}{2}\right)$. Um wieder in die kleinere Umgebung $\mathcal{N}_2\left(\frac{1}{4}\right)$ zu gelangen, wird erneut ein Newton-Schritt verwendet, jetzt allerdings für die Gleichung $F_{\mu_k}(x, \lambda, s) = 0$ mit $\mu_k = \frac{(x^{k+1})^T \cdot s^{k+1}}{n}$.

Algorithmus 3.1: Prädiktor-Korrektor-Verfahren nach Mizumo/Todd/Ye

(S1) Wähle $z^0 := (x^0, \lambda^0, s^0) \in \mathcal{N}_2\left(\frac{1}{4}\right)$, $\varepsilon \geq 0$ und setze $k := 0$.

(S2) Falls $\mu_k := \frac{(x^k)^T \cdot s^k}{n} \leq \varepsilon$, dann stoppe Algorithmus.

(S3) Falls $k \mod 2 = 0$, dann Prädiktorschritt: Bestimme $\Delta z^k := (\Delta x^k, \Delta \lambda^k, \Delta s^k)$ als Lösung von

$$\nabla F_0(z^k)^T \cdot \Delta z + F_0(z^k) = 0$$

und bestimme α_k als größtes $\alpha \in [0, 1]$, für das $z^k + \alpha \cdot \Delta z^k \in \mathcal{N}_2\left(\frac{1}{2}\right)$. Setze $z^{k+1} := z^k + \alpha_k \cdot \Delta z^k$.

Falls $k \mod 2 = 1$, dann Korrektorschritt: Bestimme Δz^k als Lösung von

$$\nabla F_{\mu_k}(z^k)^T \cdot \Delta z + F_{\mu_k}(z^k) = 0$$

und setze $z^{k+1} := z^k + \Delta z^k$.

(S4) Setze $k := k + 1$ und gehe zu (S2).

Theorem 3.5 Es sei $G^0 \neq \emptyset$. Dann gilt:

1. $\mu_{k+2} = \mu_{k+1} \leq \left(1 - \frac{0{,}4}{\sqrt{n}}\right) \cdot \mu_k$ für $k \in 2\mathbb{N}_0$.
2. Falls $\mu_0 \leq \varepsilon^{-\kappa}$ mit $\varepsilon \in (0, 1), \kappa > 0$, dann bricht der Algorithmus 3.1 in Schritt (S2) mit $\mu_k \leq \varepsilon$ spätestens ab, wenn $k \geq 2 \cdot (1 + \kappa) \cdot \frac{\sqrt{n}}{0{,}4} \cdot |\ln \varepsilon|$.
3. Es gibt $C > 0$, sodass $\mu_{k+2} = \mu_{k+1} \leq C \cdot \mu_k^2$ für alle $k \in 2\mathbb{N}_0$.

ohne Beweis

3.2 Nichtlineare Probleme

$$f(x) \to \min \qquad \text{bei } x \in G := \{x \in \mathbb{R}^n; g(x) \leq 0\} \tag{3.23}$$

Einbeziehung von Gleichungsrestriktionen meist einfach möglich, erfordert ggf. besondere Voraussetzungen (z.B. LICQ).

3.2.1 Zugang über Straf- und Barrierefunktionen

Sei $\tilde{R} := \mathbb{R} \cup \{\infty\}$ und $f : \mathbb{R}^n \to \mathbb{R}$ stetig und $g : \mathbb{R}^n \to \mathbb{R}^m$ stetig.

Definition 3.5 Es sei $(w_k)_{k\in\mathbb{N}}$ eine Folge von stetigen Funktionen $w_k : \mathbb{R}^n \to \tilde{\mathbb{R}}$, derart dass

$$\lim_{k\to\infty} x^k = x \Rightarrow \liminf_{k\to\infty} w_k(x^k) \begin{cases} \geq 0 & x \in G \\ = \infty & x \notin G \end{cases} \tag{3.24}$$

für jede Folge $(x^k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^n$ gilt. Außerdem sei für eine nichtleere Menge $B \subseteq G$ die Bedingung

$$\forall x \in B : \lim_{k\to\infty} w_k(x) = 0 \tag{3.25}$$

erfüllt. Dann heißen die Folgenglieder w_k *Straffunktionen zum Problem (3.23).* Genügt die Folge $(w_k)_{k\in\mathbb{N}}$ zusätzlich der Bedingung

$$\forall (k, x) \in \mathbb{N} \times (\mathbb{R}^n \setminus \operatorname{int} G) : w_k(x) = +\infty$$

so spricht man von *Barrierefunktionen zum Problem (3.23).*

Bemerkung:

1. Stetigkeit von $w_k : \mathbb{R}^n \to \tilde{\mathbb{R}}$ bedeutet

   $$\lim_{j\to\infty} x^j = x \Rightarrow \lim_{j\to\infty} w_k(x_j) = w_k(x) \in \tilde{\mathbb{R}}$$

   für jede Folge $(x^j)_{j\in\mathbb{N}} \subset \mathbb{R}^n$.

Anstelle des restringierten Problems (3.23) wird nun eine Folge von unrestringierten Ersatzproblemen der Form

$$T_k(x) := f(x) + w_k(x) \to \min \tag{3.26}$$

betrachtet.

Beispiel 3.3

$$2x_1^2 + x_2^2 \to \min \qquad \text{bei } g(x) := 1 - x_2 - x_2 \leq 0$$

Definiere Straffunktion zum Beispiel als $w_k(x) := r_k \cdot (\max\{0, 1-x_1-x_2\})^2$, wobei $(r_k)_{k\in\mathbb{N}} \subset (0, \infty)$ mit $\lim_{k\to\infty} r_k = +\infty$. Weiterhin erkennt man $B = G$. Ersatzzielfunktion:

$$T_k(x) = 2x_1^2 + x_2^2 + r_k \cdot (\max\{0, 1 - x_1 - x_2\})^2$$

Die notwendige (und hier wegen der Konvexität von T_k auch hinreichende) Optimalitätsbedingung für (3.26) lautet:

$$\nabla T_k(x) = \begin{pmatrix} 4x_1 - 2r_k \cdot \max\{0, 1 - x_1 - x_2\} \\ 2x_2 - 2r_k \cdot \max\{0, 1 - x_1 - x_2\} \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Daraus erhält man:

1. Falls $1 - x_1 - x_2 \leq 0$, so folgt $x_1 = x_2 = 0$. Widerspruch!
2. Falls $1 - x_1 - x_2 > 0$, so ergibt sich

   $$4x_1 + r_k \cdot (-2 + 2x_2 + x_1) = 0 \qquad 2x_2 + r_k \cdot (-2 + 2x_1 + 2x_2) = 0$$

   beziehungsweise

   $$\begin{pmatrix} 1 + \frac{2}{r_k} & 1 \\ 1 & 1 + \frac{1}{r_k} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$
   $$\Rightarrow x^k := \begin{pmatrix} x_1^k \\ x_2^k \end{pmatrix} = \frac{1}{3 + \frac{2}{r_k}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Dies liefert

$$x^* := \lim_{k\to\infty} x^k = \frac{1}{3} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Theorem 3.8 Es sei $(w_k)_{k\in\mathbb{N}}$ gegeben und die Ersatzprobleme (3.26) seien für jedes $k \in \mathbb{N}$ lösbar (mit Lösung x^k). Außerdem gelte

$$\inf_{x\in B} f(x) = \inf_{x\in G} f(x) \tag{3.27}$$

Dann ist jeder Häufungspunkt der Folge $(x^k)_{k\in\mathbb{N}}$ Lösung des Problems (3.23).

Beweis:

- Da x^k für $k \in \mathbb{N}$ Lösung des Ersatzproblems (3.26) ist, gilt

$$\forall x \in \mathbb{R}^n : f(x^k) + w_k(x^k) \leq f(x) + w_k(x) \tag{3.28}$$

Es sei nun $(x^k)_{k\in N_1} \subseteq (x^k)_{k\in\mathbb{N}}$ eine gegen $\bar{x}$ konvergente Teilfolge. Aus der Stetigkeit von f, aus $B \neq \emptyset$ sowie aus (3.24),(3.25),(3.28) folgt:

$$f(\bar{x}) \overset{(3.24),\text{stetig}}{\leq} \lim_{k\in N_1} f(x^k) + \liminf_{k\in N_1} w_k(x^k)$$
$$\overset{(3.28)}{\leq} f(x) + \lim_{k\in N_1} w_k(x) \overset{(3.25)}{=} f(x)$$

für alle $x \in B$. Somit unter Ausnutzung von (3.27):

$$f(\bar{x}) \leq \inf_{x\in B} f(x) = \inf_{x\in G} f(x) \tag{3.29}$$

Mit (3.25) und (3.28) ergibt sich außerdem

$$\liminf_{k\in N_1} w_k(x^k) \leq \liminf_{k\in N_1} (f(x) - f(x^k) + w_k(x)) = f(x) - f(\bar{x}) < \infty$$

für $x \in B$. Nach (3.24) folgt daher $\bar{x} = \liminf_{k\in N_1} x^k \in G$. Zusammen mit (3.29) erhält man, dass $\bar{x}$ Lösung von (3.23) ist.

Beispiele

1. quadratische Straffunktion:

$$w_k(x) := r_k \cdot \sum_{i=1}^{m} (\max\{0, g_i(x)\})^2 \qquad B = G$$

2. logarithmische Barrierefunktion:

$$w_k(x) := \begin{cases} -\dfrac{1}{r_k} \cdot \displaystyle\sum_{i=1}^{m} \log(-g_i(x)) & x \in G^0 \\ \infty & x \notin G^0 \end{cases}$$

mit $B = G^0 = \{x \in \mathbb{R}^n; \forall i = 1, \ldots, m : g_i(x) < 0\}$.

3. exakte nicht differenzierbare Straffunktion

$$w_k(x) := r_k \cdot \sum_{i=1}^{m} \max\{0, g_i(x)\} \qquad B = G \tag{3.30}$$

Bemerkung:

1. Mit dem Attribut „exakt“ wird eine Folge $(w_k)_{k\in\mathbb{N}}$ von Straffunktionen bezeichnet, wenn unter geeigneten Bedingungen ein $k_0 \in \mathbb{N}$ existiert, sodass

   - jede globale Minimalstelle
   - oder jede lokale Minimalstelle
   - oder jeder stationäre Punkt

   der Ersatzzielfunktion T_k für jedes $k \geq k_0$ auch globale Mimimalstelle (bzw. lokale Minimalstelle bzw. stationärer Punkt) des Ausgangsproblems (3.23) ist und umgekehrt. Die Ersatzaufgabe (3.26) spiegelt also für $k \geq k_0$ bestimmte Eigenschaften des Ausgangsproblems wieder.

Lemma 3.6 Die Lagrange-Funktion $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ zum Problem (3.23),

$$\mathcal{L}(x, u) = f(x) + \sum_{i=1}^{m} u_i \cdot g_i(x)$$

besitze einen Sattelpunkt $(x^*, u^*) \in \mathbb{R}^n \times \mathbb{R}^m_+$. Wird die Straffunktion (3.30) verwendet und gilt $r_k > \max\{u_i^*, i \in I\}$, dann ist jede Lösung x^k des Ersatzproblems (3.26) auch eine Lösung von (3.23).

ohne Beweis

Theorem 3.9 Es seien $g_1, \ldots, g_m, f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar. Weiter sei die Folge $(T_k)_{k\in\mathbb{N}}$ der Ersatzzielfunktionen gegeben durch

$$T_k(x) = f(x) + r_k \cdot \sum_{i=1}^{m} (\max\{0, g_i(x)\})^2$$

wobei $(r_k)_{k\in\mathbb{N}} \subset (0, \infty)$ mit $r_k \to \infty$ für $k \to \infty$. Die Ersatzprobleme (3.26) mögen für jedes $k \in \mathbb{N}$ eine Näherungslösung x^k derart besitzen, dass

$$\|\nabla T_k(x^k)\| \leq \varepsilon_k$$

wobei $\varepsilon_k \to 0$ $(k \to \infty)$. Definiert man nun noch Vektoren $u^k \in \mathbb{R}^m_+$ durch

$$u_i^k := 2r_k \cdot \max\{0, g_i(x^k)\}$$

für alle $(i, k) \in \{1, \ldots, m\} \times \mathbb{N}$, dann genügt jeder Häufungspunkt $(\bar{x}, \bar{u})$ der Folge $((x^k, u^k))_{k\in\mathbb{N}}$ den KKT-Bedingungen zum Problem (3.23), falls in $\bar{x}$ die MFCQ erfüllt ist.

ohne Beweis

Bemerkung:

1. Es gilt
$$\nabla T_k(x) = \nabla f(x) + 2r_k \cdot \sum_{i=1}^{m} \max\{0, g_i(x)\} \cdot \nabla g_i(x) \approx 0$$
deshalb Wahl von u^k vernünftig.

Algorithmus 3.3: Straffunktions-Verfahren

(S1) Wähle $(\varepsilon_k)_{k\in\mathbb{N}} \subset [0, \infty)$, $(r_k)_{k\in\mathbb{N}} \subset (0, \infty)$ mit $\varepsilon_k \to 0, r_k \to \infty$ $(k \to \infty)$. Setze $k := 0$.

(S2) Berechne x^k, sodass $\|\nabla T(x^k)\| \leq \varepsilon_k$. Setze
$$u_i^k := 2r_k \cdot \max\{0, g_i(x^k)\}$$
und $u^k := (u_1^k, \ldots, u_m^k)^T$.

(S3) Falls (x^k, u^k) KKT-Punkt von (3.23), dann stoppe Algorithmus. Sonst setze $k := k + 1$ und gehe zu (S2).

Bemerkung:

1. Zur Berechnung von x^{k+1} sollte x^k als Startpunkt gewählt werden.

3.2.2 Zugang über zulässige Richtungen

- Richtungssuchproblem (P2 nach Zoutendijk): Es sei $x \in G$ beliebig gegeben.

$$\lambda \to \min \text{ bei } \nabla f(x)^T \cdot d \leq \lambda, \forall i \in I_\varepsilon(x) : g_i(x) + \nabla g_i(x)^T \cdot d \leq \lambda, \|d\| \leq 1 \tag{3.31}$$

  wobei

$$I_\varepsilon(x) := \{i \in I; g_i(x) \geq -\varepsilon\}$$

  für $\varepsilon \geq 0$ als Indexmenge der ε-aktiven Restriktionen bezeichnet wird. Insbesondere gilt $\hat{\lambda} \leq 0$, da $(\lambda, d) := (0, 0)$ für jedes $x \in G$ ein zulässiger Punkt von (3.31) ist.

- Die Zielfunktion des Richtungssuchproblems (3.31) ist stetig. Da $\|d\| \leq 1$ muss es wegen $\nabla f(x)^T \cdot d \leq \lambda$ ein $\bar{\lambda} < 0$ geben, sodass $\lambda \geq \bar{\lambda}$. Ferner können wir offenbar die Nebenbedingung $\lambda \leq 0$ zum zulässigen Bereich von (3.31) hinzufügen, ohne die Lösungsmenge des Problems zu ändern. Damit ist der zulässige Bereich des so modifizierten Problems kompakt. Mit dem Satz von Weierstraß ist damit die Lösbarkeit des modifizierten Problems und damit des Richtungssuchproblems (3.31) gezeigt. Mit $\lambda(x, \varepsilon)$ werde der optimale Zielfunktionswert von (3.31) bezeichnet.

Lemma 3.7 Es sei $\varepsilon > 0$. Für $\bar{x} \in G$ sei die MFCQ erfüllt. Dann gilt:

$$\lambda(\bar{x}, \varepsilon) = 0 \Leftrightarrow \exists \bar{u} : (\bar{x}, \bar{u}) \text{ erfüllt die KKT-Bedingungen für (3.23)}$$

ohne Beweis

Algorithmus 3.4: P2-Verfahren nach Zoutendijk

(S1) Wähle $x^0 \in G, \varepsilon > 0$, $\delta \in (0, 1)$. Setze $k := 0$.

(S2) Berechne d^k als Lösung des Richtungssuchproblems (3.31), wobei $x := x^k$.

(S3) Falls $\lambda(x^k, \varepsilon) = 0$, dann stoppe Algorithmus.

(S4) Berechne Schrittweite

$$\alpha_k := \max\left\{\alpha \in S; x^k + \alpha \cdot d^k \in G, f(x^k + \alpha \cdot d^k) \leq f(x^k) + \alpha \cdot \delta \cdot \nabla f(x^k)^T \cdot d^k\right\}$$

(S5) Setze $x^{k+1} := x^k + \alpha_k \cdot d^k$, $k := k + 1$. Gehe zu (S2).

Theorem 3.10 Algorithmus 3.4 ist wohldefiniert. Besitzt eine von Algorithmus 3.4 erzeugte Folge $(x^k)_{k \in \mathbb{N}}$ einen Häufungspunkt x^*, in dem die MFCQ erfüllt ist, so gibt es $u^* \in \mathbb{R}^m_+$, sodass (x^*, u^*) den KKT-Bedingungen zu (3.23) genügt.

ohne Beweis

Bemerkungen:

1. Algorithmus 3.4 kann als „zulässiges Gradientenverfahren“ aufgefasst werden. Es gibt andere Richtungssuchprobleme, die lokal schnelle Konvergenz ermöglichen.
2. Beschränkung des Iterationsgebietes auf G kann Nachteil sein (bei spezieller Gestalt von G), jedoch Vorteil wenn beispielsweise Zielfunktion nur auf G definiert ist.
3. Die Richtungssuchprobleme (3.31) sind lineare Optimierungsaufgaben, falls $\|\cdot\| = \|\cdot\|_\infty$ gewählt wird.

3.2.3 Sequential-Quadratic-Programming Zugang

Seien f, g stetig differenzierbar. Richtungsprobleme sind quadratische Optimierungsaufgaben, wobei $x \in \mathbb{R}^n$ und $B \in \mathbb{R}^{n \times n}$ symmetrisch gegeben seien,

$$\nabla f(x)^T \cdot d + \frac{1}{2} d^T \cdot B \cdot d \to \min \qquad \text{bei } \nabla g_i(x)^T \cdot d + g_i(x) \leq 0 \qquad \forall i \in J(x) \tag{3.32}$$

In einem SQP-Verfahren werden nun Schritt für Schritt sequentiell derartige Richtungssuchprobleme verwendet.

Bemerkungen:

1. Häufig wird für B eine positiv definite Matrix verwendet (die ebenfalls von Punkt x abhängen kann) und unter Umständen Informationen 2. Ordnung enthalten (vgl. 3.24).

2. $J(x) \subseteq I$ ist eine geeignet gewählte Teilmenge von I, z.B.

   $$J_\varepsilon(x) := \{i \in I; g_i(x) \geq F(x) - \varepsilon\}$$

   wobei $F(x) := \max\{0, g_1(x), \ldots, g_m(x)\}$ mit $\varepsilon > 0$. Auch $J(x) = I$ ist denkbar. Gegebenenfalls Aufweitung des zulässigen Bereichs:

   $$\nabla g_i(x)^T \cdot d + g_i(x) \leq w_i \qquad w_i \geq 0$$

   Man hat dann jedoch (etwa durch zusätzlichen Strafterm in der Zielfunktion von (3.32)) dafür sorge zu tragen, dass w_i hinreichend klein bleibt, um die Approximationseigenschaften von (3.32) gegenüber Ausgangsproblem (3.23) nicht zu stark zu verschlechtern.

3. Falls $B = B(x)$ positiv definit ist und (3.32) einen nichtleeren zulässigen Bereich besitzt, so hat (3.32) eine eindeutige Lösung $d(x)$. Zu $d(x)$ können jedoch unter Umständen unendlich viele Vektoren von Lagrange-Multiplikatoren gehören.

4. Man kann (3.32) so modifizieren, dass in einem SQP-Verfahren in Verbindung mit einer Line-Search-Technik ausschließlich in G liegende Iterierte erzeugt werden.

5. Um ein Maß für die Güte eines Punktes x zu erhalten, etwa für die Anwendung einer Line-Search-Technik, verwendet man häufig Straffunktionsansätze, z.B. $\Phi_r : \mathbb{R}^n \to \mathbb{R}$ mit

   $$\Phi_r(x) := f(x) + r \cdot F(x)$$

Lemma 3.8 Sei B positiv definit, $J(x) := J_\varepsilon(x)$ mit $\varepsilon > 0$ und $(d(x), u(x)) \in \mathbb{R}^n \times \mathbb{R}_+^{|J(x)|}$ sei ein KKT-Punkt des Richtungssuchproblems (3.32). Falls $r > \|u(x)\|_1$ mit einem $r \geq 1$, dann gilt

$$\Phi_r(x + \alpha \cdot d(x)) \leq \Phi_r(x) - \frac{\alpha}{2} \cdot d(x)^T \cdot B \cdot d(x)$$

für alle $\alpha > 0$ hinreichend klein.

Beweis:

- Aus der Taylorformel erhält man

  $$f(x + \alpha \cdot d(x)) = f(x) + \alpha \cdot \nabla f(x)^T \cdot d(x) + o_0(\alpha \cdot d(x)) \tag{3.33}$$

  und für $i \in J_\varepsilon(x)$ ergibt sich bei Beachtung des Richtungssuchproblems (3.32):

  $$\begin{aligned} g_i(x + \alpha \cdot d(x)) &= g_i(x) + \alpha \cdot \nabla g_i(x)^T \cdot d(x) + o_i(\alpha \cdot d(x)) \\ &= \alpha \cdot \underbrace{(g_i(x) + \nabla g_i(x)^T \cdot d(x))}_{\leq 0} + (1 - \alpha) \cdot g_i(x) + o_i(\alpha \cdot d(x)) \\ &\leq (1 - \alpha) \cdot g_i(x) + o_i(\alpha \cdot d(x)) \leq (1 - \alpha) \cdot F(x) + o_i(\alpha \cdot d(x)) \end{aligned} \tag{3.34}$$

Für $i \in I \backslash J_\varepsilon(x)$ gilt:

$$\begin{aligned} g_i(x + \alpha \cdot d(x)) &= g_i(x) + (g_i(x + \alpha \cdot d(x)) - g_i(x)) < F(x) - \varepsilon + L \cdot \alpha \cdot \|\alpha \cdot d(x)\| \\ &\leq (1 - \alpha) \cdot F(x) \end{aligned} \tag{3.35}$$

wobei $L > 0$ eine (lokale) Lipschitz-Konstante ist und $\alpha \leq \varepsilon \cdot (F(x) + L \cdot \|d(x)\|)^{-1}$. Mit (3.34) folgt

$$\begin{aligned} F(x + \alpha \cdot d(x)) &= \max\{0, g_1(x + \alpha \cdot d(x)), \ldots, g_m(x + \alpha \cdot d(x)) \\ &\leq (1 - \alpha) \cdot F(x) + \sum_{i \in J_\varepsilon(x)} |O_i(\alpha \cdot d(x))| \end{aligned}$$

Dies zusammen mit (3.33) liefert

$$\begin{aligned} \Phi_r(x + \alpha \cdot d(x)) &= f(x + \alpha \cdot d(x)) + r \cdot F(x + \alpha \cdot d(x)) \\ &\leq f(x) + \alpha \cdot \nabla f(x)^T \cdot d(x) + r \cdot \left((1 - \alpha) \cdot F(x) + \sum_{i \in J_\varepsilon(x) \cup \{0\}} |\imath_i(\alpha \cdot d(x))| \right) \end{aligned} \tag{3.36}$$

Aus den KKT-Bedingungen für (3.32) ergibt sich

$$\begin{aligned} \nabla f(x) + B \cdot d(x) + \sum_{j \in J_\varepsilon(x)} u_i(x) \cdot \nabla g_i(x) &= 0 \\ \forall i \in J_\varepsilon(x) : u_i(x) \cdot (\nabla g_i(x)^T \cdot d(x) + g_i(x)) &= 0 \end{aligned}$$

und damit

$$\begin{aligned} \nabla f(x)^T \cdot d(x) &= -d(x)^T \cdot B \cdot d(x) - \sum_{i \in J_\varepsilon(x)} u_i(x) \cdot \nabla g_i(x)^T \cdot d(x) \\ &= -d(x)^T \cdot B \cdot d(x) + \sum_{i \in J_\varepsilon(x)} u_i(x) \cdot g_i(x) \\ &\leq -d(x)^T \cdot B \cdot d(x) + r \cdot F(x) \end{aligned}$$

Mit (3.36) folgt daraus

$$\begin{aligned} \Phi_r(x + \alpha \cdot d(x)) &\leq f(x) - \alpha \cdot d(x)^T \cdot B \cdot d(x) + \alpha \cdot r \cdot F(x) + r \cdot (1 - \alpha) \cdot F(x) + \\ &\quad + r \cdot \sum_{i \in J_\varepsilon(x) \cup \{0\}} |o_i(\alpha \cdot d(x))| \\ &\leq \Phi_r(x) - \frac{\alpha}{2} \cdot d(x)^T \cdot B \cdot d(x) \end{aligned}$$

für alle $\alpha > 0$ hinreichend klein.

Lemma 3.9 Es sei B positiv definit und in x sei MFCQ erfüllt, $J(x) := J_\varepsilon(x)$ mit $\varepsilon > 0$. Dann ist $d(x) = 0$ genau dann Lösung von (3.32), wenn $u \in \mathbb{R}^m_+$ existiert, sodass (x, u) KKT-Punkt von (3.23) ist.

ohne Beweis

3.2.4 Lokal superlineare Verfahren

Es seien f, g zweimal stetig differenzierbar mit lokal lipschitz-stetigen zweiten Ableitungen. Die hier beschriebenen Verfahren beruhen auf Teilproblemen, die die KKT-Bedingungen

$$\nabla_x \mathcal{L}(x, u) = 0, g(x) \leq 0, u \geq 0, u^T \cdot g(x) = 0 \tag{3.37}$$

zum Problem (3.23) hinreichend gut approximieren. Eine Möglichkeit besteht in der (nicht ganz vollständigen) Linearisierung von (3.37) in einem Punkt $(\bar{x}, \bar{u})$:

$$\begin{aligned} \nabla_x \mathcal{L}(\bar{x}, \bar{u}) + \nabla_{xx} \mathcal{L}(\bar{x}, \bar{u})^T \cdot (x - \bar{x}) + \nabla_{xu} \mathcal{L}(\bar{x}, \bar{u})^T \cdot (u - \bar{u}) &= 0 \\ g(\bar{x}) + \nabla g(\bar{x})^T \cdot (x - \bar{x}) &\leq 0 \\ u &\geq 0 \\ u^T \cdot (g(\bar{x}) + \nabla g(\bar{x})^T \cdot (x - \bar{x})) &= 0 \end{aligned}$$

Unter Beachtung der Definition von $\mathcal{L}$ und der Symmetrie von $\nabla_{xx} \mathcal{L}(\bar{x}, \bar{u})$ ergibt sich

$$\begin{aligned} \nabla_x \mathcal{L}(\bar{x}, \bar{u}) &= \nabla f(\bar{x}) + \nabla g(\bar{x})^T \cdot \bar{u} \\ \Rightarrow \nabla_{xu} \mathcal{L}(\bar{x}, \bar{u})^T \cdot (u - \bar{u}) &= \nabla g(\bar{x})^T \cdot (u - \bar{u}) \end{aligned}$$

und somit die äquivalente Formulierung

$$\begin{aligned} \nabla f(\bar{x}) + \nabla g(\bar{x}) \cdot u + \nabla_{xx} \mathcal{L}(\bar{x}, \bar{u}) \cdot (x - \bar{x}) &= 0 \\ g(\bar{x}) + \nabla g(\bar{x})^T \cdot (x - \bar{x}) &\leq 0 \\ u &\geq 0 \\ u^T \cdot (g(\bar{x}) + \nabla g(\bar{x})^T \cdot (x - \bar{x})) &= 0 \end{aligned}$$

Wie man leicht nachprüft, erweisen sich diese Bedingungen gerade als KKT-Bedingungen zur quadratischen Optimierungsaufgabe

$$\nabla f(\bar{x})^T \cdot (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T \cdot \nabla_{xx} \mathcal{L}(\bar{x}, \bar{u}) \cdot (x - \bar{x}) \to \min \text{ bei } g(\bar{x}) + \nabla g(\bar{x})^T \cdot (x - \bar{x}) \leq 0 \quad (3.38)$$

Offenbar lässt sich dieses Optimierungsproblem nun in der Form (3.32) schreiben, wenn man dort $(x, u) := (\bar{x}, \bar{u}), d := x - \bar{x}, J(x) := I$ und $B := \nabla_{xx} \mathcal{L}(\bar{x}, \bar{u})$ setzt.

Algorithmus 3.5: Wilson-Verfahren (1963)

(S1) Wähle $(x^0, u^0) \in \mathbb{R}^n \times \mathbb{R}^m$. Setze $k := 0$.

(S2) Falls (x^k, u^k) KKT-Punkt von (3.23), dann stoppe Algorithmus.

(S3) Setze $x := x^k, B := \nabla_{xx} \mathcal{L}(x^k, u^k)$ und $J(x^k) := I$. Berechne

$$(d^k, u^{k+1}) := \operatorname{argmin} \left\{ \|(\hat{d}, \hat{u} - u^k)\|; (\hat{d}, \hat{u}) \text{ ist KKT-Punkt von (3.32)} \right\}$$

(S4) Setze $x^{k+1} := x^k + d^k$ und $k := k + 1$. Gehe zu (S2).

Theorem 3.11 Es sei (x^*, u^*) ein KKT-Punkt des Problems (3.23). Weiterhin gelte

1. die strenge hinreichende Optimalitätsbedingung zweiter Ordnung in (x^*, u^*), d.h.

$$\forall d \in D(x^*, u^*) \backslash \{0\} : d^T \cdot \nabla_{xx} \mathcal{L}(x^*, u^*) \cdot d > 0 \quad (3.39)$$

   wobei

$$D(x, u) := \{d \in \mathbb{R}^n; \nabla g_i(x)^T \cdot d = 0 \text{ falls } u_i > 0\}$$

2. LICQ an der Stelle (x^*, u^*), d.h. die Vektoren in der Familie $\{\nabla g_i(x^*); i \in I_0(x^*)\}$ seien linear unabhängig.

Dann gibt es eine Umgebung U von (x^*, u^*) derart, dass Algorithmus 3.5 für jedes $(x^0, u^0) \in U$ wohldefiniert ist. Falls der Algorithmus nicht nach endlich vielen Schritten abbricht, konvergiert die erzeugte Folge (x^k, u^k) Q-quadratisch gegen (x^*, u^*).

ohne Beweis

Anderer Zugang: Es sei $\Phi : \mathbb{R}^2 \to \mathbb{R}$ mit folgender Eigenschaft gegeben:

$$\Phi(a, b) = 0 \Leftrightarrow a \geq 0, b \geq 0, a \cdot b = 0 \quad (3.40)$$

Lemma 3.10 Es sei $\Phi : \mathbb{R}^2 \to \mathbb{R}$ eine Funktion mit der Eigenschaft (3.40). Dann ist jeder KKT-Punkt von (3.23) Lösung der Gleichung

$$H_\Phi(x,u) := \begin{pmatrix} \nabla_x \mathcal{L}(x,u) \\ \Phi(-g_1(x), u_1) \\ \vdots \\ \Phi(-g_m(x), u_m) \end{pmatrix} = 0$$

und umgekehrt.

Beweis: Übung

Zwei solche Funktionen Φ, die (3.40) genügen, seien hier definiert:

$$\Phi_{\min}(a,b) := \min\{a,b\}$$
$$\varphi(a,b) := \sqrt{a^2+b^2} - (a+b)$$

$H_{\min}$ ist an solchen Punkten (x,u) nicht differenzierbar, für die $-g_i(x) = u_i$ ist für mindestens ein $i \in I$. Verwendet man φ an Stelle von $\Phi_{\min}$ so reduzieren sich die Nichtdifferenzierbarkeitsstellen auf Punkte (x,u) mit $g_i(x) = u_i = 0$ für mindestens ein $i \in I$. Die folgende Störung von φ, nämlich

$$\hat{\varphi} : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}, \hat{\varphi}(a,b,t) := \sqrt{a^2+b^2+t^2} - a - b$$

gestattet eine weitere Reduktion der nichtdifferenzierbaren Punkte auf ausschließlich diejenigen KKT-Punkte von (3.23), an denen die strenge Komplementaritätsbedingung ($g_i(x^*) = 0 \Rightarrow u_i^* > 0$ für alle $i \in I$) verletzt ist. Dazu sei $\hat{H} : \mathbb{R}^{n+m+1} \to \mathbb{R}^{n+m+1}$ gegeben durch

$$\hat{H}(x,u,t) := \begin{pmatrix} \nabla_x \mathcal{L}(x,u) \\ \hat{\varphi}(-g_1(x), u_1, t) \\ \vdots \\ \hat{\varphi}(-g_m(x), u_m, t) \\ e^t - 1 \end{pmatrix}$$

Offenbar gilt $\hat{H}(x,u,t) = 0 \Leftrightarrow (x,u)$ erfüllt KKT-Bedingung von (3.23) und $t = 0$. Da $t \neq 0$ die stetige Differenzierbarkeit von $\hat{H}$ in (x,u,t) nach sich zieht, kann man zumindest die klassische Newton-Gleichung im Punkt (x,u,t) aufstellen. Ob diese Gleichung allerdings eine Lösung besitzt, hängt von weiteren Voraussetzungen ab. Der Nachweis von lokalen Konvergenzeigenschaften eines solchen Newton-Verfahrens, vgl. Theorem 3.12, kann mit Mitteln der nichtglatten Analysis erbracht werden.

Algorithmus 3.6: Newton-Verfahren für $\hat{H}(u,x,t) = 0$

(S1) Wähle $z^0 := (x^0, u^0, t_0) \in \mathbb{R}^n \times \mathbb{R}^m \times (0,\infty)$ und setze $k := 0$.

(S2) Berechne die Newton-Richtung Δz^k, also Lösung des linearen Gleichungssystems

$$\hat{H}(z^k) + \nabla \hat{H}(z^k)^T \cdot \Delta z = 0$$

(S3) Definiere $z^{k+1} := z^k + \Delta z^k$ und $k := k+1$. Gehe zu (S2).

Bemerkung:

1. Der Newton-Schritt (falls durchführbar) sichert, dass aus $t_k > 0$ auch $t_{k+1} > 0$ folgt. Damit ist $\hat{H}(x^k, u^k, t_k) = 0$ für kein $k \in \mathbb{N}$ möglich.

Theorem 3.12 Es sei (x^*, u^*) ein KKT-Punkt des Problems (3.23), für den die strenge hinreichende Optimalitätsbedingung zweiter Ordnung (3.39) und LICQ an der Stelle x^* sind. Dann gibt es eine Umgebung U von $z^* := (x^*, u^*, 0)$ derart, dass Algorithmus 3.6 für jedes $z^0 \in U$ wohldefiniert ist und eine unendliche Folge $(z^k)_{k\in\mathbb{N}}$ erzeugt. Diese konvergiert Q-quadratisch gegen z^*.

Ohne Beweis

Bemerkung:

1. Vorteil von Algorithmus 3.6 gegenüber Wilson-Algorithmus: Nur Lösung von linearen Gleichungssystemen notwendig.
2. Die beiden vorstehenden Algorithmen besitzen neben dem Vorteil ihrer lokal schnellen Konvergenz auch verschiedene Nachteile. Insbesondere ist die Lösbarkeit der Teilprobleme nur unter ziemlich starken Voraussetzungen (und auch dann nur lokal) gesichert. Diese Voraussetzungen implizieren auch die lokale Eindeutigkeit des KKT-Punktes.

Das im folgenden beschriebene Levenberg-Marquardt-Verfahren besitzt im Unterschied dazu stets lösbare Teilprobleme und lokal schnelle Konvergenz unter Bedingungen, die nichtisolierte KKT-Punkte gestatten. Das LM-Verfahren ist ein Verfahren zur Lösung von nichtlinearen Gleichungssystemen. Zur Anwendung auf das KKT-System (3.37) benötigt man daher wieder eine geeignete Umformulierung von (3.37) als Gleichungssystem $H_\varphi(x, u) = 0$.

Bevor wir darauf näher eingehen, beschreiben wir allgemein das LM-Verfahren zur Lösung eines nichtlinearen Gleichungssystems

$$H(z) = 0 \tag{3.41}$$

wobei $H : \mathbb{R}^{\ell_1} \to \mathbb{R}^{\ell_2}$ eine differenzierbare Abbildung mit lokal lipschitz-stetiger Ableitung sei. Mit

$$Z^* := \{z \in \mathbb{R}^{\ell_1}; H(z) = 0\}$$

werde die Lösungsmenge von (3.41) bezeichnet. Im LM-Verfahren wird nun zu einer gegebenen Iterierten $s \in \mathbb{R}^{\ell_1}$ das folgende Teilproblem gelöst:

$$\psi(z) := \frac{1}{2} \cdot \|H(s) + \nabla H(s)^T \cdot (z - s)\|^2 + \frac{1}{2}\mu(s) \cdot \|z - s\|^2 \to \min \tag{3.42}$$

wobei $\mu(s) > 0$ ein von s abhängiger Parameter ist und $\|.\|$ die euklidische Norm bezeichnet. Offenbar ist die quadratische Funktion ψ (wegen $\mu(s) > 0$) gleichmäßig konvex. Somit besitzt das Teilproblem (3.42) eine eindeutige Lösung, die wir mit $z(s)$ bezeichnet werden. Außerdem folgt, dass $z(s)$ auch eindeutige Lösung der notwendigen und zugleich hinreichenden Optimalitätsbedingung

$$\nabla\psi(z) = 0$$

ist, also durch Lösen des linearen Gleichungssystems

$$\nabla H(s) \cdot H(s) + (\nabla H(s) \cdot \nabla H(s)^T + \mu(s) \cdot I) \cdot (z - s) = 0$$

bestimmt werden kann. Man beachte, dass die Systemmatrix $\nabla H(s) \cdot \nabla H(s)^T + \mu(s) \cdot I$ positiv definit ist ($\longrightarrow$ Cholesky, CG-Verfahren).

Algorithmus 3.7: Levenberg-Marquardt-Verfahren für $H(z) = 0$

(S1) Wähle $z^0 \in \mathbb{R}^{\ell_1}$. Setze $k := 0$.

(S2) Wähle $\mu(z^k) \in (0, \infty)$ und berechne $z(z^k)$.

(S3) Setze $z^{k+1} := z(z^k)$ und $k := k + 1$.

Lemma 3.11 Algorithmus 3.7 ist für beliebig gewähltes $z^0 \in \mathbb{R}^{\ell_1}$ wohldefiniert.

Beweis: Klar.

Anstelle einer beim klassischen Newton-Verfahren (für $\ell_1 = \ell_2$) üblichen Voraussetzung, dass in einer Lösung z^* des Gleichungssystems $H(z) = 0$ die Matrix $\nabla H(z^*)$ regulär ist, verwendet man eine schwächere Regularitätsbedingung (Error-Bound Condition): Dazu sei $z^* \in Z^*$. Es wird dann verlangt, dass $\omega, \delta > 0$ existieren, sodass

$$\forall z \in B(z^*, \delta) : \|H(z)\| \geq \omega \cdot \operatorname{dist}(z, Z^*) \tag{3.43}$$

Dabei bezeichne

$$\operatorname{dist}(z, M) := \inf\{\|m - z\|; m \in M\}$$

den euklidischen Abstand des Punktes $z \in \mathbb{R}^{\ell_1}$ von der nichtleeren Menge $M \subseteq \mathbb{R}^{\ell_1}$.

Lemma 3.12 Sei $\ell_1 = \ell_2$ und für $z^* \in Z^*$ sei die Matrix $\nabla H(z^*)$ regulär. Dann gibt es $\omega, \delta > 0$, sodass (3.43) erfüllt ist und $Z^* \cap B(z^*, \delta) = \{z^*\}$.

Um die Konvergenzeigenschaften von Algorithmus 3.7 näher zu untersuchen, betrachten wir zunächst das folgende Gleichungssystem

$$F(z) := \nabla H(z) \cdot H(z) = 0 \tag{3.44}$$

Das ist die notwendige Optimalitätsbedingung für das Optimierungsproblem

$$\|H(z)\|^2 \to \min \tag{3.45}$$

Offenbar ist jede Lösung von (3.41) auch globale Lösung von (3.45). Wenn (3.41) lösbar ist (wie wir hier immer annehmen), so gilt auch die Umkehrung. Somit sind (3.41) und (3.45) hier äquivalent. Das folgende Lemma zeigt, dass anstelle von $\|H(z)\|$ auch $\|F(z)\|$ als Fehlerschranke benutzt werden kann:

Lemma 3.13 Es sei die Error-Bound Condition (3.43) erfüllt. Dann gibt es $\omega_F > 0$ und $\delta_F \in (0, \delta]$, sodass

$$\forall z \in B(z^*, \delta_F) : \|F(z)\| \geq \omega_F \cdot \operatorname{dist}(z, Z^*)$$

Ohne Beweis

Lemma 3.14 Es gibt $\delta > 0, L > 0$, sodass

$$\|\nabla H(z) - \nabla H(s)\| \leq L \cdot \|z - s\| \tag{3.46}$$

$$\|H(z) - H(s) - \nabla H(s)^T \cdot (z - s)\| \leq L \cdot \|z - s\|^2 \tag{3.47}$$

$$\|H(z) - H(s)\| \leq L \cdot \|z - s\| \tag{3.48}$$

$$\|\nabla H(z)\| \leq L \tag{3.49}$$

für alle $s, z \in B(z^*, 2\delta)$ gilt.

Beweis: Folgt direkt aus der lokalen Lipschitz-Stetigkeit von ∇H.

Lemma 3.15 Die Error-Bound Condition (3.43) sei erfüllt. Außerdem sei

$$\mu(z) := \begin{cases} \|H(z)\| & z \in \mathbb{R}^{\ell_1} \backslash Z^* \\ 1 & z \in Z^* \end{cases} \tag{3.50}$$

Dann gibt es $\kappa > 0$, sodass

$$\|z(s) - s\| \leq \kappa \cdot \operatorname{dist}(s, Z^*)$$

für alle $s \in B(z^*, \delta)$.

Ohne Beweis

Lemma 3.16 Unter den Voraussetzungen von Lemma 3.15 gibt es $\hat{C} > 0$ und $\hat{\delta} > 0$, sodass

$$\forall s \in B(z^*, \hat{\delta}) : \operatorname{dist}(z(s), Z^*) \leq \hat{C} \cdot \operatorname{dist}(s, Z^*)^2$$

Bemerkung:

1. LM-Verfahren vorteilhaft gegenüber Newton-Verfahren, falls keine isolierten Lösungen vorliegen.

Beweis:

- Mit δ_F von Lemma 3.13 und κ aus Lemma 3.15 sei $\hat{\delta} := \frac{1}{\kappa+1} \cdot \delta_F$. Offenbar ist $0 < \hat{\delta} < \delta_F \leq \delta$. Sei $s \in B(z^*, \hat{\delta}) \backslash Z^*$ und $z \in B(z^*, \delta_F)$ beliebig gewählt. Unter Beachtung der Definition von ψ und F ergibt sich

$$\begin{aligned}
\nabla\psi(z) &= \nabla H(s) \cdot \nabla H(s)^T \cdot (z - s) + \nabla H(s) \cdot H(s) + \mu(s) \cdot (z - s) \\
&= \nabla H(s) \cdot (H(s) + \nabla H(s)^T \cdot (z - s)) + \mu(s) \cdot (z - s) \\
\Rightarrow F(z) - \nabla H(z) &= \nabla H(s) \cdot (H(z) - H(s) - \nabla H(s)^T \cdot (z - s)) - \mu(s) \cdot (z - s) + \\
&\quad + (\nabla H(z) - \nabla H(s)) \cdot H(z)
\end{aligned}$$

  Mit (3.49), (3.47), (3.50) und (3.46) erhalten wir

$$\|F(z) - \nabla\psi(z)\| \leq L^2 \cdot \|z - s\|^2 + \|H(z)\| \cdot \|z - s\| + L \cdot \|z - s\| \cdot \|H(z)\| \tag{3.51}$$

  Lemma 3.15 liefert

$$\begin{aligned}
\|z(s) - z^*\| &\leq \|z(s) - s\| + \|s - z^*\| \leq \kappa \cdot \operatorname{dist}(s, Z^*) + \hat{\delta} \\
&\leq (\kappa + 1) \cdot \hat{\delta} = \delta_F
\end{aligned} \tag{3.52}$$

  also gilt $z(s) \in B(z^*, \delta_F)$ und $z \in B(z^*, \delta_F)$ kann folglich durch $z(s)$ ersetzt werden. Wegen Lemma 3.13 und $\nabla\psi(z(s)) = 0$ folgt

$$\|F(z(s)) - \nabla\psi(z(s))\| = \|F(z(s))\| \geq \omega_F \cdot \operatorname{dist}(z(s), Z^*) \tag{3.53}$$

  Da $Z^* \neq \emptyset$ und abgeschlossen ist, gibt es zu jedem $s \in \mathbb{R}^n$ ein $s^\perp \in Z^*$ mit

$$\|s - s^\perp\| = \operatorname{dist}(s, Z^*)$$

  Damit haben wir

$$\|s^\perp - z^*\| \leq \|s^\perp - s\| + \|s - z^*\| \leq \hat{\delta} + \hat{\delta} \leq 2\delta$$

  und (3.48) liefert

$$\|H(s)\| = \|H(s) - H(s^\perp)\| \leq L \cdot \|s - s^\perp\| = L \cdot \operatorname{dist}(s, Z^*) \tag{3.54}$$

  Weiterhin erhalten wir mit (3.48), (3.52) und Lemma 3.15

$$\begin{aligned}
\|H(z(s))\| &= \|H(z(s)) - H(s^\perp)\| \leq L \cdot \|z(s) - s^*\| \\
&\leq L \cdot \|z(s) - s\| + L \cdot \|s - s^\perp\| \leq L \cdot (\kappa + 1) \cdot \operatorname{dist}(s, Z^*)
\end{aligned} \tag{3.55}$$

  Unter Beachtung von (3.54), (3.55) und Lemma 3.15 folgt aus (3.51):

$$\begin{aligned}
\|F(z(s)) - \nabla\psi(z(s))\| = \|F(z(s))\| &\leq \kappa^2 \cdot L^2 \cdot \operatorname{dist}(s, Z^*)^2 + \kappa \cdot L \cdot \operatorname{dist}(s, Z^*)^2 + \\
&\quad + L^2 \cdot \kappa \cdot (\kappa + 1) \cdot \operatorname{dist}(s, Z^*)^2 \\
&= \operatorname{dist}(s, Z^*) \cdot (\kappa \cdot (\kappa \cdot L + 1 + (\kappa + 1) \cdot L)) =: \operatorname{dist}(s, Z^*) \cdot c
\end{aligned}$$

  Dies zusammen mit (3.53) liefert

$$\begin{aligned}
&\omega_F \cdot \operatorname{dist}(z(s), Z^*) \leq c \cdot \operatorname{dist}(s, Z^*) \\
&\Rightarrow \operatorname{dist}(z(s), Z^*) \leq \frac{c}{\omega_F} \cdot \operatorname{dist}(s, Z^*)
\end{aligned}$$

  für alle $s \in B(z^*, \hat{\delta}) \backslash Z^*$. Also gilt die Behauptung mit $\hat{c} := \frac{c}{\omega_F}$, wobei für $s \in Z^*$ $z(s) = s$ gilt.

Theorem 3.13 Die Voraussetzungen von Lemma 3.15 seien erfüllt. Weiter sei $(z^k)_{k\in\mathbb{N}}$ durch Algorithmus 3.7 erzeugt. Dann gibt es ein $\varepsilon > 0$, sodass für $z^0 \in B(z^*, \varepsilon)$ die Folge $(z^k)_{k\in\mathbb{N}}$ Q-quadratisch gegen ein $\hat{z} \in Z^*$ konvergiert.

Ohne Beweis

Anwendung auf KKT-Systeme mittels $H_\Phi(z) = 0$ Falls $z^* = (x^*, u^*)$ die strenge Komplementaritätsbedingung erfüllt, d.h.

$$\forall i \in I : \{g_i(x^*) = 0 \Rightarrow u_i^* > 0\}, \tag{3.56}$$

dann ist H_Φ für $\Phi = \Phi_{\min}$ oder $\Phi = \varphi$ in einer Umgebung von z^* differenzierbar und besitzt dort eine lokal lipschitz-stetige Ableitung.

Korollar 3.2 Es sei $z^* = (x^*, u^*)$ ein KKT-Punkt von (3.23), an dem die strenge Komplementaritätsbedingung (3.56) erfüllt ist. Die Funktion Φ in H_Φ ist entweder $\Phi_{\min}$ oder φ. Außerdem gelte die Error-Bound-Condition, d.h. es gibt $\omega, \delta > 0$, sodass

$$\forall z \in B(z^*, \delta) : \|H_\Phi(z)\| \geq \omega \cdot \operatorname{dist}(z, Z^*)$$

Schließlich sei die Folge $(z^k)_{k\in\mathbb{N}}$ durch Algorithmus 3.7 (mit H_Φ an Stelle von H) erzeugt. Dann gibt es $\varepsilon > 0$, sodass für $z^0 \in B(z^*, \varepsilon)$ die Folge $(z^k)_{k\in\mathbb{N}}$ Q-quadratisch gegen ein $\hat{z} \in Z^*$ konvergiert.

3.2.5 Globalisierung lokal superlinear konvergenter Verfahren

Prinzip: Sei (x^0, u^0) ein beliebiger Vektor. Führe Verfahren 1 durch. Teste neue Iterierte (Test 1); falls Test scheitert, dann zurück zu Verfahren 1, ansonsten gebe Iterierte an Verfahren 2 weiter. Teste die neue Iterierte (Test 2); falls Test okay, dann wieder zu Verfahren 2, ansonsten zu Verfahren 1.

Algorithmus 3.8: Hybrid-Prinzip

(S1) Wähle $(x^0, u^0) \in \mathbb{R}^{n+m}$ und $\varepsilon > 0$. Setze $k := 0$. Wähle Verfahren 1 (mit globalen Konvergenzeigenschaften) und Verfahren 2 (mit lokal superlinearen Konvergenzeigenschaften).

(S2) Berechne (x^{k+1}, u^{k+1}), indem ein Schritt von Verfahren 1 mit Startwert (x^k, u^k) ausgeführt wird. Setze $k := k + 1$.

(S3) Test 1: Falls $\|H_{\min}(x^k, u^k)\| \leq \varepsilon$, dann gehe zu (S4). Anderenfalls gehe zu (S2).

(S4) Berechne (x^{k+1}, u^{k+1}), indem ein Schritt des Verfahrens 2 mit Startwert (x^k, u^k) ausgeführt wird. Setze $k := k + 1$.

(S5) Test 2: Falls $\|H_{\min}(x^k, u^k)\| \leq \frac{1}{2} \cdot \|H_{\min}(x^{k-1}, u^{k-1})\|$, dann gehe zu (S4). Anderenfalls setze $\varepsilon := \frac{\varepsilon}{2}$ und gehe zu (S2).

3.2.6 Das Filterprinzip zur Globalisierung

- Sei $G : \mathbb{R}^n \to [0, \infty)$ mit

$$G(x) := \sum_{i=1}^{m} \max\{0, g_i(x)\}$$

Zur Lösung der Aufgabe (3.23) müssen G und f in geeigneter Weise minimiert werden. Einige der erzeugten Paare $(G(x^k), f(x^k))$ werden dazu in einer Menge, dem *Filter*,

$$\mathcal{F}_k := \{(G_\ell, f_\ell); \ell \in L_k\}$$

gesammelt, wobei $L_k \subseteq \{0, \ldots, k\}$ bestimmte Iterationsindizes enthält.

- Man sagt x bzw. $(G(x), f(x))$ wird von einem Element des Filters (G_ℓ, f_ℓ) *dominiert*, wenn sowohl $G(x) \geq G_\ell$ als auch $f(x) \geq f_\ell$ gilt.
- Eine von einem Verfahren erzeugte Iterierte x wird vom Filter $\mathcal{F}_k$ höchstens dann akzeptiert, wenn sie von keinem Element des Filters dominiert wird. Verschärfend wird zur Sicherung globaler Konvergenzeigenschaften x nur dann vom Filter akzeptiert, wenn

$$g(x) \leq \beta \cdot G_\ell \qquad \text{oder} f(x) + \gamma \cdot G(x) \leq f_\ell \tag{3.57}$$

  für jedes $\ell \in L_k$ gilt, wobei $0 < \gamma < \beta < 1$ (etwa $\gamma = 0.01$, $\beta = 0.99$) vorzugebende Konstanten sind.
- Zur Erzeugung der Iterierten wird hier ein sogenanntes Trust-Region SQP-Verfahren benutzt. Dabei werden folgende Teilprobleme $\mathcal{P}(x, \Delta)$ betrachtet:

$$\nabla f(x)^T \cdot p + \frac{1}{2} p^T \cdot B(x) \cdot p \to \min \text{ bei } g(x) + \nabla g(x)^T \cdot p \leq 0, \|p\|_\infty \leq \Delta$$

Algorithmus 3.9: Trust-Region SQP-Filteralgorithmus

(S1) Wähle $\bar{G} > 0, \sigma \in (0,1)$ und $\bar{\Delta} > 0$. Setze $\mathcal{F}_0 := \{(\bar{G}, -\infty)\}$ und $k := 0$.

(S2) Restaurierungsschritt: Berechne x^k und $\Delta \geq \bar{\Delta}$ so, dass $(G_k, f_k) := (G(x^k), f(x^k))$ vom Filter akzeptiert wird (d.h. (3.57) muss gelten) und $\mathcal{P}(x^k, \Delta)$ einen nichtleeren zulässigen Bereich besitzt.

(S3) SQP-Schritt: Falls $\mathcal{P}(x^k, \Delta)$ nicht lösbar ist, setze $\mathcal{F}_{k+1} := \mathcal{F}_k \cup \{(G_k, f_k)\}$ und $k := k + 1$. Gehe zu (S2). Anderenfalls bestimmte $\hat{p}$ als Lösung von $\mathcal{P}(x^k, \Delta)$.

(S4) Abbruchtest: Setze $\Delta q := -\nabla f(x^k)^T \cdot \hat{p} - \frac{1}{2}\hat{p}^T \cdot B(x^k) \cdot \hat{p}$. Falls $\Delta q = 0$ und $p = 0$ zulässig für $\mathcal{P}(x^k, \Delta)$, dann stoppe Algorithmus (x^k ist stationärer Punkt von (3.23)).

(S5) Falls $x^k + \hat{p}$ vom Filter akzeptiert wird, gehe zu (S6). Anderenfalls setze $\Delta := \frac{1}{2}\Delta$ und gehe zu (S3).

(S6) Setze $\Delta f := f(x^k) - f(x^k + \hat{p})$. Falls $\Delta f \leq \sigma \cdot \Delta q$ und $\Delta q > 0$, setze $\Delta := \frac{1}{2}\Delta$ und gehe zu (S3).

(S7) Falls $\Delta q \leq 0$, dann setze $\mathcal{F}_{k+1} := \mathcal{F}_k \cup \{(G_k, f_k)\}$.

(S8) Setze $x^{k+1} := x^k + \hat{p}$ und $k := k + 1$. Wähle $\Delta \geq \bar{\Delta}$ und gehe zu (S3).

Theorem 3.14 Es seien f, g zweimal stetig differenzierbar. Der zulässige Bereich von (3.23) sei nichtleer. Mit einem $M > 0$ gelte $\|B(x)\| \leq M$ für alle $x \in \mathbb{R}^n$. Dann ist Algorithmus 3.9 wohldefiniert. Entweder der Algorithmus bricht nach endlich vielen Schritten ab mit einem stationären Punkt ab oder er erzeugt eine unendliche Folge $(x^k)_{k\in\mathbb{N}}$. Jeder Häufungspunkt x^* von $(x^k)_{k\in\mathbb{N}}$ ist ein stationärer Punkt, falls in x^* die MFCQ gilt.

Bemerkungen:

1. Alle Einträge (G_ℓ, f_ℓ) in jedem der auftretenden Filter haben die Eigenschaft $G_\ell > 0$. In (S1) ist dies klar. In (S3) folgt aus der Nichtlösbarkeit von $\mathcal{P}(x^k, \Delta)$ folgt $G(x^k) > 0$. In (S7) kann kein (G_k, f_k) mit $G_k = 0$ in den Filter aufgenommen werden, Beweis durch Widerspruch: Angenommen $G_k = 0$, dann $g(x^k) \leq 0$ und somit folgt für

$$\nabla f(x^k)^T \cdot p + \frac{1}{2} p^T \cdot B(x^k) \cdot p \to \min \text{ bei } g(x^k) + \nabla g(x^k)^T \cdot p, \|p\| \leq \Delta$$

   dass $p = 0$ zulässig ist, also ist der optimale Zielfunktionswert nichtpositiv. Damit $\Delta q := -\nabla f(x^k)^T \cdot \hat{p} - \frac{1}{2}\hat{p}^T \cdot B(x^k) \cdot \hat{p} \geq 0$. Wegen (S4) folgt $\Delta q > 0$ im Widerspruch zur Annahme $\Delta q \leq 0$.

2. Für Gleichungsnebenbedingungen kann G als

$$G(x) := \sum_{i=1}^{m} \max\{0, g_i(x)\} + \sum_{j=1}^{n} |h_j(x)|$$

definiert werden. Entsprechende Linearisierung der Gleichungsnebenbedingungen in $\mathcal{P}(x, \Delta)$ notwendig.

3. Im Restaurationsschritt: Berechne x^k und $\Delta \geq \bar{\Delta}$ so, dass $(G_k, f_k) := (G(x^k), f(x^k))$ vom Filter akzeptiert wird (d.h. (3.57) muss gelten) und $\mathcal{P}(x^k, \Delta)$ einen nichtleeren zulässigen Bereich besitzt. Dafür kann ein beliebiges Verfahren verwendet werden. Insbesondere kann jedes (x, Δ) mit $g(x) \leq 0$ und $\Delta \geq \bar{\Delta}$ als (x^k, Δ) in (S2) verwendet werden. Ein solches (x^k, Δ) wird vom Filter akzeptiert, falls dieser nur Einträge (G_ℓ, f_ℓ) mit $G_\ell > 0$ enthält (siehe Bemerkung (i)). Also ist (S2) (zumindest theoretisch) lösbar, wenn (3.23) einen zulässigen Punkt besitzt.

4 Heuristische Ansätze

4.1 Das Verfahren von Nelder-Mead

Verfahren für die Behandlung unrestringierter Optimierungsaufgaben

$$f(x) \to \min$$

Definition 4.1 Seien $x^1, \ldots, x^{n+1} \in \mathbb{R}^n$ affin unabhängig (d.h. $\{x^1 - x^2, \ldots, x^1 - x^{n+1}\}$ sind linear unabhängig), dann wird die konvexe Hülle

$$S := \text{conv}\{x^1, \ldots, x^{n+1}\}$$

als *n-dimensionales Simplex* mit den Ecken $x^1, \ldots, x^{n+1}$ bezeichnet. Man kann die Punkte $x^1, \ldots, x^{n+1}$ so ordnen, dass

$$f(x^1) \leq f(x^2) \leq \ldots \leq f(x^{n+1}) \tag{4.1}$$

gilt.

In jedem Durchlauf des folgenden Verfahrens wird ein Simplex im $\mathbb{R}^n$ ermittelt und zwar mit dem Ziel, dass der Vektor der Zielfunktionswerte (also $(f(x^1), \ldots, f(x^{n+1}))^T$) wenigstens in einer Komponente verkleinert wird. Dieses Ziel kann unter Umständen nicht erreicht werden. Immer wird jedoch der bisher beste Punkt x^1 entweder beibehalten oder verbessert.

Algorithmus 4.1: Nelder-Mead

(S1) Wähle eine nach (4.1) geordnete affin unabhängige Menge $\{x^1, \ldots, x^{n+1}\} \subset \mathbb{R}^n$. Wähle $\alpha > 0, \beta > \max\{1, \alpha\}, \gamma \in (0, 1)$.

(S2) Berechne Schwerpunkt der n besten Punkte $\hat{x} = \frac{1}{n} \sum_{i=1}^{n} x^i$ und die Reflexion des schlechtesten Punktes x^{n+1} am Schwerpunkt $\hat{x}$:

$$x^R := \hat{x} + \alpha \cdot (\hat{x} - x^{n+1})$$

(S3) Reflexionsschritt: Falls $f(x^1) \leq f(x^R) \leq f(x^n)$, dann $x^{n+1} := x^R$. Stelle (4.1) durch Umordnen her.

(S4) Expansionsschritt: Falls $f(x^R) < f(x^1)$, dann setze

$$x^E := \hat{x} + \beta \cdot (\hat{x} - x^{n+1})$$

Falls $f(x^E) < f(x^R)$, dann $x^R := x^E$. Setze $x^{n+1} := x^R$. Stelle (4.1) durch Umordnen her.

(S5) Kontraktionsschritt: Falls $f(x^R) > f(x^n)$, dann

$$x^C := \begin{cases} \hat{x} + \gamma \cdot (x^{n+1} - \hat{x}) & f(x^R) \geq f(x^{n+1}) \\ \hat{x} + \gamma \cdot (x^R - \hat{x}) & f(x^R) < f(x^{n+1}) \end{cases}$$

Falls $f(x^C) < \min\{f(x^{n+1}), f(x^R)\}$, dann $x^{n+1} := x^c$. Falls $f(x^c) \geq \min\{f(x^{n+1}), f(x^R)\}$, dann $x^i := \frac{1}{2}(x^1 + x^i)$ für $i = 2, \ldots, n+1$. Stelle (4.1) durch Umordnung her.

(S6) Gehe zu (S2).

Bemerkungen:

1. Für die Wahl der Konstanten in (S1) wird $\alpha = 1$, $\beta \in [2, 3]$ und $\gamma \in [0.4, 0.6]$ empfohlen.
2. Abbruchkriterium für Algorithmus 4.1 ist zum Beispiel

$$\frac{1}{n+1} \sum_{i=1}^{n+1} (f(x^i) - \hat{f})^2 \leq \varepsilon$$

mit $\hat{f} := \frac{1}{n} \sum_{i=1}^{n+1} f(x^i)$. Oder: $\|x^1 - x^{n+1}\| \leq \varepsilon$.

4.2 Optimierungsmechanismen aus der Natur

4.2.1 Evolutionäre Algorithmen

- $\mathcal{A}$ sei die Menge aller möglichen Individuen.
- $\mathcal{G}$ sei eine Menge, z.B. $\mathcal{G} = \mathbb{R}^\ell$, $\mathcal{G} = \mathbb{Z}^\ell$, $\mathcal{G} = \{0, 1\}^\ell$ oder Kombinationen davon.
- $c : \mathcal{A} \to \mathcal{G}$ bezeichnet die *Kodierungsabbildung*, die jedem $a \in \mathcal{A}$ ein $c(a) \in \mathcal{G}$ zuordnet. $c(a)$ wird manchmal auch als *Genotyp* des *Phänotyps* a bezeichnet. Es gilt hier jedoch $c(\mathcal{A}) \subseteq \mathcal{G}$, d.h. nicht jedem Element aus $\mathcal{G}$ muss ein Individuum entsprechen.
- $A \subseteq \mathcal{A}$ bzw. $c(A) = \{c(a) \in \mathcal{G}; a \in A\}$ bezeichnet eine Population.
- $d : \mathcal{G} \to \mathcal{A}$ bezeichne die *Dekodierungsabbildung*, wobei $d(g) \in \{a \in \mathcal{A}; c(a) = g\}$.
- $f : \mathcal{G} \to \mathbb{R}$ bezeichnet *Bewertungsfunktion* (auch Fitnessfunktion).
- $M : \mathcal{G}^m \to \mathcal{G}^m$ bezeichnet *Mutationsoperator*.
- $R : \mathcal{G}^r \to \mathcal{G}^s$ sei *Rekombinationsoperator* (zur Erzeugung von Nachkommen).
- $S : \mathcal{G}^p \to \mathcal{G}^q$ bezeichnet *Selektionsoperator* (p bzw. q die Anzahl der Individuen eine Population vor bzw. nach der Selektion).

Algorithmus 4.2: Evolutionärer Algorithmus

(S1) Wähle $A \subseteq \mathcal{A}$ mit $|A| < \infty$. Setze $P_0 := c(A)$ und $k := 0$. Wähle $r_1, r_2 \in \mathbb{N}$ mit $r_1, r_2 \geq 2$.

(S2) Abbruchtest: Wenn Abbruchbedingung erfüllt, dann stoppe Algorithmus.

(S3) Elternselektion: Setze $p := |P_k|$ und $q := r_1$ und $P_E := S(P_k)$.

(S4) Nachkommenerzeugung: $P_N := R(P_E)$.

(S5) Mutation: $m := |P_N|$ und $P_M := M(P_N)$

(S6) Umweltselektion: $p := |P_M|, q := r_2, P_U := S(P_M)$

(S7) Update: $P_{k+1} := P_U$, $k := k + 1$. Gehe zu (S2).

Bemerkung:

1. Die Bewertungsfunktion f taucht nicht explizit im Algorithmus 4.1 auf. Im Allgemeinen wird sie für die Gestaltung der Selektionsoperatoren verwendet.

4.2.2 Ant Colony Optimization

- Für eine Ameise k sei $A(x)$ die Menge der zulässigen (direkten) Folgezustände am Zustand x. Die Funktion $p(x, y)$ gebe die Wahrscheinlichkeit für Übergang der Ameise k vom Zustand x zum Zustand y an. Definiere

$$p(x, y) := \frac{\tau^\alpha(x, y) \cdot \eta^\beta(x, y)}{\sum_y \tau^\alpha(x, y) \cdot \eta^\beta(x, y)}$$

  wobei $\alpha \geq 0, \beta \leq 1$ Parameter, $\tau(x, y)$ Pheromonkonzentration auf (x, y) und $\eta(x, y)$ die Urteilhaftigkeit des Übergangs von x nach y (zum Beispiel Weglänge^{-1}) bezeichne.

- Nachdem alle Ameisen ihren Lauf beendet haben, wird Pheromonkonzentration neu berechnet:

$$\tau(x, y) = (1 - \varrho) \cdot \tau(x, y) + \Delta\tau(x, y)$$

  für $\varrho \in [0, 1)$ (Verdunstungskoeffizient), wobei $\Delta\tau(x, y)$ die neuen Pheromonablagerungen auf (x, y) bezeichnet, zum Beispiel

$$\Delta\tau(x, y) = \frac{1}{\text{Länge}} \cdot (\text{Anzahl der Ameisen, die (x,y) beschritten haben})$$