Nichtlineare Optimierung
Prof. Dr. E. Kostina
SS 2007 / WS 2011

Thorsten Ante, überarbeitet von T. Binder

Stand vom 20. Februar 2014

Inhaltsverzeichnis

Kapitel 1

Einleitung

1.1 Problemformulierungen

Nichtlineare Optimierungsverfahren (auch klassische Optimierungsverfahren genannt) zielen darauf ab, das Minimum oder Maximum einer nichtlinearen (Ziel-)Funktion zu finden. Dabei können auch Nebenbedingungen berücksichtigt werden, die den Lösungsbereich weiter einschränken. Neben der klassischen Differentialrechnung werden für die Lösung praktischer Probleme oft Suchverfahren eingesetzt.
Betrachtet werden dabei Probleme folgender Art :

$$\begin{array}{lll} \min & f(x) & f : D \subset \mathbb{R}^n \to \mathbb{R} \\ & g(x) = 0 & g : D \subset \mathbb{R}^n \to \mathbb{R}^l \quad \text{(i.A. l < n)} \\ & h(x) \geq 0 & h : D \subset \mathbb{R}^n \to \mathbb{R}^k \end{array}$$

x	:	Optimierungsvariable	
f	:	Zielfunktion	cost function, objective function
g	:	Gleichungsnebenbedingungen	equality constraints
h	:	Ungleichungsnebenbedingungen	inequality constraints
		Spezialfall von h:	
		$a \leq x \leq b$ Schranken	bounds, box constraints

Bei der Optimierung treten weiterhin zwei Fälle auf. Einmal gibt es die *unbeschränkte Optimierung*, dann ist lediglich das Problem $\min f(x)$ zu lösen, außerdem gibt es die *beschränkte Optimierung*, wobei das Problem $\min f(x)$ unter den Nebenbedingungen $g(x) = 0$ und $h(x) \geq 0$ zu lösen ist.
In dieser Vorlesung werden einige Annahmen getroffen. Zum einen befindet sich alles im Endlichdimensionalen, zum anderen sind $f, g, h \in C^3(D)$. Nicht in dieser Vorlesung behandelt werden:

- ganzzahlige und kombinatorische Probleme,
- nicht-differenzierbare Optimierung,
- infinite Optimierungsprobleme,
- semi-infinite Probleme,
- Probleme der optimalen Steuerung (Optimierung von Differentialgleichungen).

Definition 1.1 *Die* zulässige Menge S *ist definiert durch* $S = \{x \in D \mid g(x) = 0,\ h(x) \geq 0\}$. *Ein Punkt* $x \in S$ *heißt* zulässiger Punkt.

Voraussetzung für Theorie und Algorithmen ist $S \neq \emptyset$. Dies ist in der Praxis jedoch oft verletzt, da die Restriktionen oder Nebenbedingungen zu scharf sind. In diesen Fällen sind spezielle algorithmische Vorkehrungen erforderlich.

Definition 1.2 *Ein Punkt* $x^\star \in D$ *ist ein* globales Minimum, *wenn* $f(x^\star) \leq f(x)\ \forall x \in S$. *Ein Punkt* $x^\star \in D$ *ist ein* lokales Minimum, *wenn es eine Umgebung* $U \subseteq S$ *von* $x^\star$ *gibt mit* $f(x^\star) \leq f(x)$ $\forall x \in U$. *Die Minima heißen* strikt, *falls* $f(x^\star) < f(x)\ \forall x \in S \backslash \{x^\star\}$ *bzw.* $\forall\ x \in U \backslash \{x^\star\}$. *Ein Punkt* $x^\star \in D$ *heißt* isoliertes lokales Minimum *falls* $x^\star \in S, f(x) > f(x^\star) \forall x \in \{S \cap B_\varepsilon(x^\star)\} \setminus \{x^\star\}$ *Ein Punkt* $x^\star \in D$ *heißt* isoliertes globales Minimum *falls* $x^\star$ *ein globales Minimum und zudem ein isliertes lokales Minimum ist.*

Definition 1.3 *Die Mengen* $\{x \in D |\ f(x) = const\}$ *heißen* Höhenlinien. *Der Gradient von* f *steht senkrecht auf den Höhenlinien und gibt die Richtung des steilsten Anstiegs an. Die Niveaumenge ist definiert als* $N_f(x_0) = \{x \in S : f(x) \leq f(x_0)\|,\ x_0 \in S$.

Satz 1.4 *Es sei ein unbeschränktes Problem der Form* $\min\limits_{x \in S} f(x)$ *gegeben. Es sei weiter* $S \subset \mathbb{R}^n,\ f : S \to \mathbb{R}$ *stetig. Falls* $\exists\ x_0 \in S$, *so dass* $N_f(x_0)$ *kompakt ist, dann besitzt das NLP ein globales Minimum.*

Beweis: Falls $x^\star$ existiert, folgt dass $x^\star \in N_f(x_0)$. Aus dem Satz von Weierstra0 folgt, dass die stetige Funktion f auf dem Kompaktum $N_f(x_0)$ ihren Minimalwert in $x^\star \in N_f(x_0)$ hat. ■

1.2 Konvergenzraten

Sei x_k eine Folge, man sagt, x_k konvergiert gegen x^*

$$
\begin{aligned}
\textit{Q-sublinear} \quad &:\Longleftrightarrow \quad \lim_{k\to\infty} \|x_k - x^*\| = 0 \\
\textit{Q-linear} \quad &:\Longleftrightarrow \quad \exists\, 0 \leq M < 1\ \exists\ k_0 \in \mathbb{N} \forall\ k \geq k_0 :\ \tfrac{\|x_{k+1} - x^\star\|}{\|x_k - x^\star\|} \leq M \\
\textit{Q-superlinear} \quad &:\Longleftrightarrow \quad \lim_{k\to\infty} \tfrac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0 \\
&\qquad\quad (\Rightarrow\ \exists\ \text{Nullfolge}\{e_k\}, e_k \geq 0 : \|x_{k+1} - x^\star\| \leq e_k \|x_k - x^\star\|\ \forall k) \\
\textit{Q-quadratisch} \quad &:\Longleftrightarrow \quad \exists\ M < \infty \exists\ k_0 \in \mathbb{N}\ \forall\ k \geq k_0 :\ \tfrac{\|x_{k+1} - x^\star\|}{\|x_k - x^*\|^2} \leq M \\
\textit{R-linear} \quad &:\Longleftrightarrow \quad \exists \{a_k\}, \text{sodass} \|x_k - x^\star\| \leq a_k \forall k \text{ und } a_k \to 0\ \text{Q} - \text{linear}. \\
\textit{R-superlinear} \quad &:\Longleftrightarrow \quad \exists \{a_k\}, \text{sodass} \|x_k - x^\star\| \leq a_k \forall k \text{ und } a_k \to 0\ \text{Q} - \text{superlinear}. \\
&:\Longleftrightarrow \quad lim\ \sqrt[k]{\|x_k - x^\star\|} = 0. \\
\textit{R-quadratisch} \quad &:\Longleftrightarrow \quad \exists \{a_k\}, \text{sodass} \|x_k - x^\star\| \leq a_k \forall k \text{ und } a_k \to 0\ \text{Q} - \text{quadratisch}.
\end{aligned}
$$

Das Q steht hierbei für Quotient und das R für die Wurzel(root).

Beispiele:

- Eine Folge x_k mit $\|x_k - x^*\| = \frac{1}{k}$ konvergiert Q-sublinear gegen x^*.
- Eine Folge x_k mit $\|x_k - x^*\| = \left(\frac{1}{2}\right)^k$ konvergiert Q-linear gegen x^*.
- Eine Folge x_k mit $\|x_k - x^*\| = \left(\frac{1}{k}\right)^k$ konvergiert Q-superlinear gegen x^*.
- Eine Folge x_k mit $\|x_k - x^*\| = \left(\frac{1}{2}\right)^{2^k}$ konvergiert Q-quadratisch gegen x^*.

1.3 Notationen

- $x \in \mathbb{R}^\times$ Spaltenvektor
  $\|x\|_2 = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}$
- Für Matrizen $M \in \mathbb{R}^{m\times n}$ gilt die induzierte Norm: $\|M\| = \max_{m\times n \leq 1} \|Mx\|$, dann gilt: $\|M\| \cdot \|x\|$
- f differenzierbar, $\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix} \in \mathbb{R}^n$ Gradient
  $\nabla f(x) = \left(\frac{\partial f}{\partial x}(x)\right)^T$ ist ein Zeilenvektor.
- $D^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix} \in \mathbb{R}^{n\times n}$ Hesse-Matrix
  Falls f zweimal stetig differenzierbar ist
  $\Rightarrow \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \Rightarrow H$ ist symmetrisch.
- Falls $g : \mathbb{R}^n \to \mathbb{R}^l$ differenzierbar, dann gilt
  $\nabla g(x) = (\nabla g_1(x), \nabla g_2(x), ..., \nabla g_l(x)) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_l}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial g_1}{\partial x_n} & \cdots & \frac{\partial g_l}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial x} \\ \vdots \\ \frac{\partial g_l}{\partial x} \end{pmatrix}^T \in \mathbb{R}^{n\times l}$ Jacobi-Matrix von g.
- Landau-Symbole
  $O(h^k),\ o(h^k),\ k \in \mathbb{N}$
  $g(h) = O(h^k)$ für $h \to 0, \limsup_{k\to 0} \frac{\|g(h)\|}{|h|^k} < \infty$
  $g(h) = o(h^k)$ für $h \to 0, \limsup_{k\to 0} \frac{\|g(h)\|}{|h|^k} = 0$
  $g(h) = a \cdot h,\ \frac{g(h)}{h} = 1 \cdot a,\ g(h) = O(h)$
  $g(h) = a \cdot h^2,\ \frac{g(h)}{h^1} = a \cdot h \overset{h\to 0}{\to} 0,\ \frac{g(h)}{h^2} = a < \infty$
  $a \cdot h^2 = o(h^1),\ a \cdot h^2 = O(h^2)$ Vielfaches von h^2
- Taylor-Formel
  Sei $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar, dann gilt $\forall x, s \in \mathbb{R}^n$:
  $f(x+s) = f(x) + \nabla^T f(x+ts) \cdot s = f(x) + \nabla^T f(x) \cdot s + o(\|s\|),\ t \in (0,1)$.

Sei $f : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar, dann gilt:
$f(x+s) = f(x+\nabla^T f(x)\cdot s + \frac{1}{2}s^T\nabla^2 f(x+\tilde{t}s)s) = f(x) + \nabla^T f(x)s + \frac{1}{2}s^T\nabla^2 f(x)\cdot s + o(\|s\|^2)$
für ein $\tilde{t} \in (0,1)$
Weiter gilt:
$\nabla f(x+s) = \nabla f(x) + \int_0^1 \nabla^2 f(x+ts)\cdot s\ dt)$ Integral wird komponentenweise berechnet.
Beweis:
$\varphi : t \to f(x+ts)$ zweimal stetig differenzierbar.
$\varphi^{'}(t) = \nabla^T f(x+ts)\cdot s$
$\varphi^{''}(t) = s^T\nabla^2 f(x+ts)$
Taylorentwicklung in $t = 0$:
$\varphi(1) = \varphi(0) + \varphi^{'}(0) + \frac{1}{2}\varphi^{''}(\hat{t}),\ \hat{t} \in (0,1)$
$\frac{d}{dt}\nabla f(x+ts) = \nabla^2 f(x+ts)\cdot s$
$\Rightarrow \nabla f(x+ts) - \nabla f(x) = \int_0^1 \frac{d}{dt}\nabla f(x+ts)\ dt = \int_0^1 D^2 f(x+ts)\cdot s\ dt$ ■

Lemma 1.5

- *Der Gradient $\nabla f(x)$ steht senkrecht auf Höhenlinie durch x. $\{y|f(y) = f(x)\}$*
- *$-\nabla f(x)$ ist Richtung des steilsten Abstiegs (steepest descent),*
  d.h. $\min\limits_{\|d\|_2=1} \frac{\partial f}{\partial t}(x+td)|_{t=0} = \min\limits_{\|d\|=1} \nabla^T f(x)\cdot d$ *hat eine Lösung* $d^\star = \frac{-\nabla f(x)}{\|\nabla f(x)\|}$.

1.4 Spezialfälle

1.4.1 Eindimensionale Optimierung

Die eindimensionale Optimierung von $f : I \in \mathbb{R} \to \mathbb{R}$ ist im Allgemeinen ein Hilfsproblem in der mehrdimensionalen Optimierung *(Liniensuche, line search)*, aber von besonderer Bedeutung für die Theorie. Zur Lösung von

$$\min_x f(x)$$

benötigt man i.A. einen Startwert x_0 und iteriert dann $x_{k+1} = x_k + t_k d_k,\ t_k \in]0,1]$.
Im Eindimensionalen sind mehrere lokale Lösungen möglich, so dass die Lösung nicht eindeutig sein muss. Außerdem muss nicht notwendigerweise überhaupt eine Lösung existieren.

1.4.2 Lineare Optimierung

Die Lineare Optimierung beschäftigt sich mit dem Lösen linearer Optimierungsprobleme, die üblicherweise in einer der folgenden Formen gegeben sind:

$$\begin{array}{lcl} \min\limits_x\ c^T x & & \\ Ax = b, & \text{oder} & b_1 \le Ax \le b_2, \\ x \ge 0, & & c_1 \le\ x \le c_2, \end{array}$$

wobei $c, c_1, c_2, x \in \mathbb{R}^n$, $A \in M(m\times n, \mathbb{R}), m \le n$, $b, b_1, b_2 \in \mathbb{R}^m$. Alle Relationen sind dabei komponentenweise zu verstehen.
Alle Problemformulierungen der Linearen Optimierung sind durch Transformation auf die obige Form, die sogenannte *Standardform*, zurückführbar.

Beispiel

$$\begin{array}{lll} \min\limits_x & c^T x & \\ & Ax \geq b & \\ & x \geq 0 & \end{array} \quad \Longrightarrow \quad \begin{array}{lll} \min\limits_x & c^T x & \\ & Ax - z = b & \\ & x \geq 0, z \geq 0 & \end{array}$$

Die Variablen z heißen *Schlupfvariablen.*
Diese Transformationen können Strukturen zerstören! Die meisten Algorithmen der Linearen Optimierung erwarten allerdings, dass das Problem in Standardform formuliert ist.
Die Lineare Optimierung ist wichtig als eigenständiges Problem und als Hilfsproblem, zum Beispiel bei allgemeiner nichtlinearer Optimierung, bei ganzzahligen/gemischt-ganzzahligen Problemen und bei speziellen nicht-differenzierbaren Problemen. Es existieren zahlreiche Anwendungen in der Wirtschaft.

1.4.3 Lineare Ausgleichsprobleme

Bei linearen Ausgleichsproblemen wird versucht, mit einer Geraden die Punkte der Funktionswerte zu approximieren. Diese Methode hängt von der Wahl der Abstandsmessung zwischen Gerade und Punkten ab, daher wird sie anhand der verwendeten Norm in verschiedene Klassen geteilt.

Die Methode der kleinsten Quadrate *(least squares problem)*

$$\begin{aligned} \min\, f(x) &= \|A_1 x + b_1\|_2^2 \\ A_2 x + b_2 &= 0, \\ A_3 x + b_3 &\geq 0, \end{aligned}$$

ist ein spezielles quadratisches Optimierungsproblem, denn

$$f(x) = x^T (A_1^T A_1) x + 2 b_1^T A_1 x + b_1^T b_1.$$

Ohne Nebenbedingung lässt sich $\min f(x)$ eindeutig lösen, wenn $A_1^T A_1$ positiv definit ist. Aus der notwendigen Bedingung für optimale Punkte (s. Kapitel 2), nämlich dass der Gradient von f verschwindet,

$$2 A_1^T A_1 x + 2 A_1^T b_1 = 0,$$

folgt

$$x^\star = -(A_1^T A_1)^{-1} A_1^T b_1 \quad \text{(Normalengleichung)}.$$

Die Höhenlinien von $f(x)$ sind Ellipsen, und wenn $A_1^T A_1$ in $x^\star$ positiv definit ist, dann liegt ein Minimum vor.

Methoden mit anderen Normen

1.

$$\begin{aligned} \min\ & f(x) = \|A_1 x + b_1\|_\infty \\ & A_2 x + b_2 = 0, \\ & A_3 x + b_3 \geq 0. \end{aligned}$$

Die Funktion f ist nicht differenzierbar, aber man kann das Problem durch Einführung einer zusätzlichen Hilfsvariablen x_{n+1} und $2n$ weiteren Ungleichungen auf ein differenzierbares Problem zurückführen:

$$\begin{aligned} \min\ & \tilde{f}(x) = x_{n+1} \\ & -x_{n+1} \leq (A_1 x + b_1)_i \leq x_{n+1}, \quad i = 1, \ldots, n, \\ & A_2 x + b_2 = 0, \\ & A_3 x + b_3 \geq 0. \end{aligned}$$

2. (l_1)

$$\begin{aligned} \min & \|F(x)\|_1 = \sum_{i=1}^{n} |F_i(x)| \text{ nicht differenzierbar, falls } F_i(x) = 0 \\ & g(x) = 0 \\ & h(x) \geq 0. \end{aligned}$$

Transformation:

$$\begin{aligned} & \text{a)} F_i(x) = v_i - w_i,\ v_i,\ w_i \geq 0 \text{ Hilfsvariablen} \\ \Rightarrow & F_i(x) \geq 0 \ \Rightarrow\ v_i = F_i(x),\ w_i = 0 \\ & F_i(x) < 0 \ \Rightarrow\ v_i = 0,\ w_i = -F_i(x) \\ & |F_i(x)| = v_i + _i \end{aligned}$$

1.4.4 Allgemeine nicht-lineare Ausgleichsprobleme

$$\begin{aligned} \min\quad & \|f(x)\|_2, \quad f : \mathbb{R}^m \to \mathbb{R}^n \\ & g(x) = 0 \\ & h(x) \geq 0 \end{aligned}$$

1.4.5 Quadratische Optimierung

Quadratische Programme treten als Hilfsprobleme bei den meisten gängigen Optimierungsverfahren auf.

$$\min f(x) = \frac{1}{2} x^T H x + g^T x + const.$$

Hierbei ist $f(x + tp)$, mit $t \in \mathbb{R}$ und $p \in \mathbb{R}^n$ immer eine Parabel. Die Höhenlinien sind Ellipsen oder Hyperbeln:

H **positiv definit (⇒ f konvex).** Die Höhenlinien sind Ellipsen um das eindeutige Minimum $x^\star = -H^{-1}g$.

H **negativ definit (⇒ f ist konkav).** Die Höhenlinien sind Ellipsen um das eindeutige Maximum $x^\star = -H^{-1}g$.

Mit Beschränkung ist das Problem in diesen beiden Fällen nicht eindeutig lösbar. Beispielweise gilt bei linearen Beschränkungen,

$$\begin{aligned} Ax + a &= 0, \\ Bx + b &\geq 0, \end{aligned}$$

dass das Problem eventuell mehrere Minima besitzt.

H **indefinit.** Die Höhenlinien sind Hyperbeln und $x^\star = -H^{-1}g$ ist ein Sattelpunkt. Hier werden Nebenbedingungen benötigt, damit Minima existieren.

Spezialfall H semidefinit, $g = 0$. Es existiert nicht nur eine Lösung, sondern eine ganze Lösungsebene. Auch hier können Beschränkungen das Problem eindeutig lösbar machen.

Kapitel 2

Notwendige und hinreichende Optimalitätsbedingungen

Nachdem nun in der Einleitung verschiedene Optimierungprobleme exemplarisch angegeben worden sind, wollen wir uns nun damit beschäftigen, herauszufinden, ob es für jedes Optimierungsproblem eine Lösung gibt und wenn ja, wie viele dies sind und wie diese aussehen. Was natürlich zwangsläufig auf das nächste Problem führt, wie man die Lösungen berechnet. Ausgegangen wird hierbei immer von einem Optimierungsproblem der Form

$$\begin{aligned} \min_x & f(x) \\ & g(x) = 0, \\ & h(x) \geq 0. \end{aligned}$$

Im Folgenden betrachten wir Theorie und Algorithmen für lokale Minima, denn Algorithmen zum garantierten Auffinden globaler Minima existieren nur für spezielle Probleme in niedrigen Dimensionen. Festzustellen, ob ein Punkt ein globales Minimum ist, ist viel komplexer, als festzustellen, ob ein Punkt lediglich ein lokales Minimum ist.

2.1 Eindimensionale Optimierung

Die eindimensionale Optimierung beschäftigt sich mit dem Lösen des Problems min $f(x)$, wobei f eine Funktion ist von der Form

$$f : D \to \mathbb{R}, \quad f \in \mathcal{C}^3(D) \text{ 2 mal stetig diffbar}, \ D \subset \mathbb{R}. \text{ offen}$$

$$\min_{x \in \mathbb{R}} f(x)$$

Satz 2.1 (Notwendige Optimalitätsbedingungen (NOB)) *Ist $x^\star$ ein lokales Minimum, dann gilt*

$$f'(x^\star) = 0 \quad \textit{(NOB 1.Ordnung)} \qquad \textit{und} \qquad f''(x^\star) \geq 0 \quad \textit{(NOB 2.Ordnung).}$$

Beweis: x^* ist lokales Minimum, d.h. $\exists \epsilon > 0 : \ f(x^*) \leq f(x), \forall x \in B_\epsilon(x^*)$
Wähle $k > 0, k < \epsilon$, Taylorentwicklung

$$f(x^* + k) = f(x^*) + kf'(x^*) + \frac{1}{2}k^2 f''(x^*) + O(k^3) \geq f(x^*)$$

$$f(x^* - k) = f(x^*) - kf'(x^*) + \frac{1}{2}k^2 f''(x^*) + O(k^3) \geq f(x^*)$$

$$\overset{k>0}{\Rightarrow} kf'(x^*) + \frac{1}{2}k^2 f''(x*) + O(k^3) \geq 0 \text{ und } -kf'(x^*) + \frac{1}{2}k^2 f''(x^*) + O(k^3) \geq 0$$

$$\Rightarrow f'(x^*)\frac{1}{2}f''(x^*) + O(k^2) \geq 0 \text{ und } -f'(x^*) + \underbrace{\frac{1}{2}kf''(x^*) + O(k^2)}_{\to 0 \text{ für } k\to 0} \geq 0$$

$$\Rightarrow f'(x^*) \geq 0 \text{ und } -f'(x^*) \geq 0$$

$$\Rightarrow f'(x^*) = 0$$

$$\Rightarrow \frac{1}{2}kf''(x^*) + O(k^2) \geq 0$$

$$\Rightarrow \frac{1}{2}f''(x^*) + \underbrace{O(k)}_{\to 0 \text{ für } k\to 0} \geq 0$$

$$\Rightarrow f''(x^*) \geq 0$$

■

Satz 2.2 (Hinreichende Optimalitätsbedingung (HOB)) *Sei $x^\star \in D$ mit $f'(x^\star) = 0$ und $f''(x^\star) > 0$. Dann gilt: $x^\star$ ist striktes lokales Minimum.*

Beweis: Sei $f''(x^\star) > 0$ und f'' stetig. Dann existiert eine Umgebung U von $x^\star$ mit

$$U(x^*), \text{ so dass } f''(x) > 0 \quad \forall x \in U.$$

Eine Entwicklung von f nach Taylor um $x^\star$ und Anwendung des Mittelwertsatzes ergibt

$$f(x) = f(x^\star) + \underbrace{f'(x^\star)(x - x^\star)}_{=\,0} + \underbrace{\frac{1}{2}\, f''(\tilde{x})(x - x^\star)^2}_{>\,0} > f(x^\star), \quad \text{wobei } \tilde{x} \in [x, x^\star]\,.$$

$$\Rightarrow f(x) > f(x^*)$$

■

2.2 Unbeschränke Optimierung im $\mathbb{R}^n$

Die unbeschränkte Optimierung im $\mathbb{R}^n$ betrachtet das Problem $\min f(x)$, wobei $f : D \to \mathbb{R}$ und $D \subseteq \mathbb{R}^n$. Außer dem ist f 2 mal stetig differenzierbar. Ein kurzer Blick auf die Notation:

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}, \qquad \nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial {x_1}^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \cdots & \frac{\partial^2 f}{\partial {x_n}^2} \end{pmatrix}.$$

Man bezeichnet $\nabla f(x)$ als den Gradienten von $f(x)$ und die symmetrische Matrix $\nabla^2 f(x) \in \mathbb{R}^{n \times n}$ heißt Hesse-Matrix.

Satz 2.3 (Notwendige Optimalitätsbedingungen (NOB)) *Sei $x^\star$ ein lokales Minimum. Dann gilt*

1. $\nabla f(x^\star) = 0$ *(NOB 1.Ordnung),*
2. $\nabla^2 f(x^\star)$ *ist positiv semidefinit (NOB 2.Ordnung).*

Beweis: Für alle $p \in \mathbb{R}^n$ und $p \neq 0$ hat

$$\tilde{f}(t) := f(x^\star + tp), \quad t \in [-\epsilon, \epsilon] \in \mathbb{R},$$

ein lokales Minimum bezüglich t bei $t = 0$. Also erfüllt $f(t)$ bei $t = 0$ die NOB im Eindimensionalen:

$$\frac{d}{dt}\tilde{f}(t)\bigg|_{t=0} = \frac{d}{dt} f(x^\star + tp) \cdot p\bigg|_{t=0} = \nabla_x f(x^\star) p = 0 \quad \Longrightarrow \quad \nabla_x f(x^\star) = 0.$$

Damit folgt für $\nabla^2 f(x^\star)$ und alle $p \in \mathbb{R}^n$, dass

$$\frac{d^2}{dt^2}\tilde{f}(t)\bigg|_{t=0} = p^T \frac{d^2}{dt^2} f(x^\star + tp) p\bigg|_{t=0} = p^T \nabla_{xx} f(x^\star) p \geq 0.$$

Positiv-Definitheit von $\nabla^2 f(x^\star)$ bedeutet aber gerade, dass $\forall p \in \mathbb{R}^n : p^T \nabla^2 f(x^\star) p \geq 0$. ■

Definition 2.4 *Sei $\widetilde{x} \in \mathbb{R}^n$ mit $\nabla f(\widetilde{x}) = 0$ heißt stationärer Punkt. (Kann Minimum, Maximun oder Sattelpunkt sein, abhängig von $\nabla^2 f(\widetilde{x})$)*

Beispiel 1: $f(x) = -x_1^2 + x_2^2$ $\nabla f(x) = \begin{pmatrix} -2x_1 \\ 2x_1 \end{pmatrix}$

$\nabla f(x) = 0$ für $x = \widetilde{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ einziger stationärer Punkt

$\nabla^2 f(x) = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}$ indefinit

$\widetilde{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ kein Maximum, kein Minimum sondern ein Sattelpunkt.

Beispiel 2: $f(x) = x^3$

$\nabla f(x) = 3x^2$ und $\widetilde{x} = 0$ ist stationärer Punkt

$\nabla^2 f(x) = 6x$ und $\nabla^2 f(\widetilde{x}) = 0$ positiv semidefinit
$\widetilde{x}$ ist aber kein Minimum.

Satz 2.5 (Hinreichende Optimalitätsbedingungen (HOB)) *Sei $x^\star \in D$ stationär, d.h. $\nabla f(x^\star) = 0$. Falls $\nabla^2 f(x^\star)$ positiv definit, dann gilt: $x^\star$ ist striktes lokales Minimum.*

Beweis: Es existiert eine Umgebung U von $x^\star$ mit $\nabla^2 f(x)$ positiv definit für alle $x \in U$, denn die Eigenwerte einer Matrix hängen stetig von ihren Einträgen ab. Eine Entwicklung von f in eine Taylorreihe um $x^\star$ ergibt dann:

$$f(x) = f(x^\star) + \underbrace{\nabla f(x^\star)^T (x - x^\star)}_{=0} + \underbrace{\frac{1}{2}(x -^\star x)^T \nabla^2 f(\hat{x})(x - x^\star)}_{>0} > f(x^\star) \quad \text{mit } \hat{x} \in U.$$

$$\Rightarrow f(x) > f(x^*) \ , \forall x \in U(x^\star) \ , x \neq x^\star$$

■

Hinreichende Bedingung ist keine notwendige Bedingung.
Beispiel: $f(x) = x^{2k}$, $k \in \mathbb{N}\backslash\{1\}$
$x^\star = 0$ ist eindeutiges globales Minimum
$\nabla f(x^\star) = 2kx^{2k-1} = 0$
$\nabla^2 f(x^\star) = 2k(2k-1)x^{2k-2} = 0$ ist nicht positiv definit.

Wichtiger Spezialfall: f konvex

Definition 2.6 *f konvex, falls $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$.*
Wobei $\lambda \in [0,1]$, $x_1, x_2 \in \mathbb{R}^n$ $(x_1, x_2 \in D)$

Satz 2.7 *(1) Sei $x^\star$ lokales Minimum einer konexen Funktion auf einem konvexen Definitionsgebiet $D \subset \mathbb{R}^n$. Dann gilt: $x^\star$ ist globales Minimum in D.*
(2) Falls f zusätzlich differenzierbar ist. Dann ist jeder stationäre Punkt $x^\star$ mit $\nabla f(x^\star) = 0$ ein globales Minimum von f auf D.

Beweis: (1) Annahme: $x^\star$ ist ein lokales, aber kein globales Minimum. Dann existiert ein $z \in \mathbb{R}^n$ mit $f(z) < f(x^\star)$. Nun betrachte $x = \lambda z + (1-\lambda)x^\star$, $\lambda \in (0,1]$, und $x \neq x^\star$. Wegen der Konvexität von f gilt nun

$$f(x) \leq \lambda f(z) + (1-\lambda) f(x^\star) < \lambda f(x^\star) + (1-\lambda) f(x^\star) = f(x^\star).$$

Jede Umgebung U um $x^\star$ enthält ein Stück des Segments zwischen $x^\star$ und z. Somit existiert ein $x \in U(x^\star)$ mit $f(x) < f(x^\star)$. Dies ist aber ein Widerspruch dazu, dass $x^\star$ ein lokales Minimum ist. Also ist $x^\star$ schon ein globales Minimum.

(2) Annahme: $x^\star$ stationär ($\nabla f(x^\star) = 0$)), aber $x^\star$ ist kein globales Minimum. Wähle z wie oben. Dann gilt

$$\begin{aligned}\nabla f^T(x^\star)(z - x^\star) &= \frac{d}{dy} f\Big(x^\star + \lambda(z - x^\star)\Big)\Big|_{\lambda=0} \\ &= \lim_{\lambda \searrow 0} \frac{f(x^\star + \lambda(z - x^\star)) - f(x^\star)}{\lambda} \\ &= \lim_{\lambda \searrow 0} \frac{\lambda f(z) + (1-\lambda) f(x^\star) - f(x^\star)}{\lambda} = f(z) - f(x^\star) < 0\end{aligned}$$

Also ist $\nabla f(x^\star) \neq 0$ und $x^\star$ somit kein stationärer Punkt. ■

2.3 Gleichungsbeschränkte Optimierung

Wir betrachten nun das Problem

$$\min_x f(x) \quad \text{mit} \quad g(x) = 0,$$

wobei für die Gleichungsnebenbedingungen gilt

$$g : D \to \mathbb{R}^l, \quad D \subset \mathbb{R}^n, \; l \leq n.$$

Die zulässige Menge $S := \{x \mid g(x) = 0\}$ sollte mindestens die $(n-l)$-dimensionale Mannigfaltigkeit sein. Wir setzen voraus, dass der Rang von $g_x(x^\star) = l$, also maximal ist.
O.B.d.A. seien die ersten l Spalten von $g_x(x^\star)$ regulär. Dann lassen sich x und $g_x(x^\star)$ aufspalten in

$$x = \begin{pmatrix} y \\ z \end{pmatrix} \quad \text{und} \quad g_x(x^\star) = \begin{pmatrix} g_y & g_z \end{pmatrix},$$

wobei $y \in \mathbb{R}^l$, $z \in \mathbb{R}^{n-l}$ und $g_y \in \mathbb{R}^{\,l \times l}$ sowie $g_z \in \mathbb{R}^{\,l \times (n-l)}$ reguläre Matrizen sind.
Intutitiv: Höhenlinien und S berühren sich in $x^\star$ und $\nabla f(x^\star)\perp$ auf Höhenlinien und Tangentialebene (d.h. $\nabla f(x^\star)$ ist „Normalvektor“ von S.
Unser Ziel: Zu zeigen:

- Tangentialraum von S in $x^\star$ ist: $T(x^\star) = \{p \in \mathbb{R}^n | \nabla g(x^\star) p = 0\}$
- Normalvektoren von S in $x^\star$ sind: $\sum_{i=1}^{l} \lambda_i \nabla g_i(x^\star) = \nabla g(x^\star)\lambda$, $\lambda \in \mathbb{R}^l$
- $x^\star$ ist lokales Minimum: $\nabla f(x^\star) = \nabla g(x^\star)\lambda$, d.h. $\nabla f(x^\star) \perp S$ in $x^\star$

Idee: Rückführung auf den unbeschränkten Fall, dazu parametrisiere Menge S mit dem Satz über implizite Funktionen (SIF).

Satz 2.8 (Satz über implizite Funktionen (SIF)) *Seien U, V offene Mengen, $F : U \times V \to \mathbb{R}^l$ stetig differenzierbar, wobei $y, z \mapsto F(y, z)$*

$F(y_0, z_0) = 0$, $\frac{\partial F}{\partial y} |_{(y_0,z_0)}$ invertierbar, $y \in U \subset \mathbb{R}^l$, $z \in V \subset \mathbb{R}^{n-l}$, $\begin{pmatrix} y \\ z \end{pmatrix} = x \in \mathbb{R}^n$, $l < n$

Dann gilt:

- $\exists U(y_0),\ U(z_0)$
- $\exists \varphi : U(z_0) \to U(y_0)$ *eindeutig und stetig differenzierbar mit* $z \mapsto y = p(z)$
- $F(\varphi(z), z) \equiv 0 \forall z \in U(z_0),\ \varphi(z_0) = y_0$
- $\frac{\partial \varphi}{\partial z}(z = z_0) = -\frac{\partial F}{\partial y}(y_0, z_0)^{-1} \frac{\partial F}{\partial z}(y_0, z_0)$

Definition 2.9 *$x^\star$ heißt „regulärer" Punkt, wenn gilt: Rang $\nabla^T g(x^\star) = l(\leq n)$, d.h. Zeilen ($\nabla^T g_i(x^\star)$) in $\nabla^T g(x^\star)$ sind linear unabhängig.*

Mit dem Satz über implizite Funktionen (S.I.F.) folgt nun die Existenz einer Umgebung $U \subset \mathbb{R}^{n-1}$ von $z^\star$ und einer Abbildung $\varphi : U \to \mathbb{R}^l,\ z \mapsto y$, mit

1. $g(\varphi(z), z) = 0$,
2. $y^\star = \varphi(z^\star)$,
3. $\varphi \in \mathcal{C}^1(U)$,
4. $\varphi'(z^\star) = -g_y(x^\star)^{-1} g_z(x^\star)$,

denn es mit $\frac{d}{dz} g(\varphi(z), z) = 0$ folgt $g_y(x)\varphi'(z) + g_z(x) = 0$. Definiere nun Φ als

$$\Phi : U \to S, \quad z \mapsto \begin{pmatrix} \varphi(z) \\ z \end{pmatrix}, \quad S \subset \mathbb{R}^n.$$

Also gilt

1. $g(\Phi(z)) = 0$.
2. $\frac{d}{dy} g'(\Phi(z)) = g_x(\Phi(z))\Phi'(z) = 0 \quad \text{mit} \quad \Phi'(z) = \begin{pmatrix} \varphi'(z) \\ I \end{pmatrix} = \begin{pmatrix} -g_y(x^\star)^{-1} g_z(x^\star) \\ I \end{pmatrix}$.

Definition 2.10 *Die Menge $T(x^\star) = \{p \in \mathbb{R}^n \mid g_x(x^\star)p = 0\}$ heißt* Tangentialraum *von S in $x^\star$.*

Bemerkung: Der Tangentialraum von $x^\star$ ist der Kern von $g_x(x^\star)$. Außerdem folgt aus $g_x(x^\star)p = 0$ auch

$$g_y(x^\star)p_y + g_z(x^\star)p_z = 0 \quad \Longleftrightarrow \quad p_y = -g_y^{-1}(x^\star)g_z(x^\star)p_z.$$

Lemma 2.11 *Der Tangentialraum von $x^\star$ lässt sich auch schreiben als*

$$T(x^\star) = \{p \in \mathbb{R}^n \mid \exists\ p_z \in \mathbb{R}^{n-l} : \begin{pmatrix} -g_y^{-1}(x^\star)g_z(x^\star) \\ I \end{pmatrix} p_z = p\} = Im(\Phi'(z^\star))$$

Beispiel

Sei eine Kugel mit $x_1^2 + x_2^2 + x_3^2 - 1 = 0$ gegeben. Sei weiter $x^\star = (1, 0, 0)$. Dann steht $g_x(x) = (2x_1, 2x_2, 2x_3)$ senkrecht auf der Oberfläche der Kugel und es ist $g_x(x^\star) = (2, 0, 0)$. Setze $y = x_1$, $z = (x_2, x_3)^T$, dann ist

$$\varphi(z) = \varphi(x_2, x_3) = \sqrt{1 - x_2^2 - x_3^2} = x_1.$$

Also folgt

$$\Phi(z) = \begin{pmatrix} \sqrt{1-x_2^2-x_3^2} \\ x_2 \\ x_3 \end{pmatrix} \quad \text{und} \quad \Phi'(z^\star) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Es ist $p = \Phi'(z^\star)p_z = \Phi'(z^\star)\frac{\varphi_2}{p_3} = \begin{pmatrix} 0 \\ p_2 \\ p_3 \end{pmatrix}$. Für den Tangentialraum der Kugel in $x^\star$ folgt also

$$T(x^\star) = \{(0, p_2, p_3)^T \mid p_2, p_3 \in \mathbb{R}\} = \{p \in \mathbb{R}^3 \mid p_1 = 0\}.$$

Lemma 2.12 *Für „Normalen“ von S in $x^\star$ gilt: $N(x^\star) := \{q \in \mathbb{R}^n | \exists \lambda \in \mathbb{R}^l, q = \frac{\partial g^T}{\partial x}(x^\star)\lambda\}$.*
(Definition von Normalen: $N(x^\star) \perp T(x^\star)$)

Definition 2.13 *Die Funktion $L(x, \lambda) := f(x) - \lambda^T g(x)$, $\lambda \in \mathbb{R}^l$, heißt* Lagrange-Funktion*, wobei λ als* Lagrange-Multiplikator *bezeichnet wird.*

Definition 2.14 *Die Matrix $H^{red} := \frac{d^2}{dz^2} f(\Phi(z))$ heißt* reduzierte Hessematrix *oder* Krümmungstensor*.*

Satz 2.15 (Notwendige Optimalitätsbedingungen (NOB)) *Sei $x^\star \in S$ ein lokales Minimum mit Rang$(g_x(x^\star)) = l$ (constraint qualification). Dann existiert ein $\lambda \in \mathbb{R}^l$ mit*
(i) $L_x(x^\star, \lambda^\star) = 0$ *bzw.* $\nabla f(x^\star) - \nabla g(x^\star)\lambda^\star = 0$ *(NOB 1.Ordnung),*
(ii) $p^T L_{xx}(x^\star, \lambda^\star)p \geq 0 \ \forall\ p \in T(x^\star)$ *(NOB 2.Ordnung).*
(zu (ii) d.h. Hesse-Matrix der Lagrangefunktion ist positiv semidefinit auf $T(x^\star)$)

Bemerkung: NOB 1. Ordnung heißt: $\nabla f(x^\star) = \nabla g(x^\star)\lambda^\star = \sum_{i=1}^{l} \lambda_i^\star \nabla g_i(x^\star)$
(Gradient der Zielfunktion = Linearkombination der Gradienten der Restriktionen)
$\Rightarrow \nabla f(x^\star)$ ist ein Normalvektor zu S in $x^\star$

Beweis: Betrachte die Funktion $\tilde{f}(z) = f(\Phi(z))$, wobei $x = \begin{pmatrix} \varphi(z) \\ z \end{pmatrix} = \Phi(z)$ und $\varphi(z)$ aus $g(\varphi(z), z) = 0$. Da $x^\star$ ein lokales reguläres Minimum ist, ist $z^\star$ ein Minimum von $\tilde{f}(z)$. Damit folgt

$$\begin{aligned} 0 = \frac{\partial \tilde{f}(z^\star)}{\partial x} &= f_x \Phi'(z)\big|_{x=x^\star} = f_x(x^\star) \begin{pmatrix} \varphi'(z^\star) \\ I \end{pmatrix} = f_x(x^\star) \begin{pmatrix} -g_y^{-1}(x^\star) g_z(x^\star) \\ I \end{pmatrix} \\ &= f_y(x^\star)(-g_y^{-1}(x^\star) g_z(x^\star)) + f_z(x^\star) \end{aligned}$$

Definiere nun $(\lambda^\star)^T = f_y(x^\star) g_y^{-1}(x^\star)$. Dann folgt $0 = -(\lambda^\star)^T g_z(x^\star) + f_z(x^\star)$. Berechne nun $L_x(x^\star, \lambda^\star) = f_x(x^\star) - (\lambda^\star)^T g_x(x^\star) = f_y(x^\star) + f_z(x^\star) - (\lambda^\star)^T g_y(x^\star) - (\lambda^\star)^T g_z(x^\star) = 0$. Offenbar folgt aus $L_x(x^\star, \lambda^\star) = 0$ auch $f_x(x^\star)p = 0 \ \forall p \in T(x^\star)$ und somit $f_x(x^\star) \perp T(x^\star)$. Aus Satz 2.3 folgt, dass

$$p_z^T \frac{d^2}{dz^2} f(\Phi(z^\star)) p_z \geq 0 \ \forall\ p_z \in \mathbb{R}^{n-l}.$$

Das Ausrechnen der Hesse-Matrix liefert nun

$$\begin{aligned} \frac{d}{dz} f(\Phi(z^\star)) &= f_x(\Phi(z^\star))\Phi'(z^\star), \\ \frac{d^2}{dz^2} f(\Phi(z^\star)) &= \Phi'(z^\star)^T f_{xx}(\Phi(z^\star))\Phi'(z^\star) + f_x(\Phi(z^\star))\Phi''(z^\star). \end{aligned}$$

Andererseits folgt für alle $z^\star$ aus $g(\Phi(z^\star)) = 0$, dass

$$\frac{d^2}{dz^2} g(\Phi(z^\star)) = \Phi'(z)^T g_{xx}(\Phi(z^\star))\Phi'(z^\star) + g_x(\Phi(z^\star))\Phi''(z^\star) = 0.$$

Multipliziert man diese Gleichung nun mit $-\lambda^\star$ und addiert sie zu der oben berechneten Hesse-Matrix, so erhält man

$$\begin{aligned}\frac{d^2}{dz^2} f(\Phi(z^\star)) &= (\Phi'(z^\star))^T \underbrace{\Big(f_{xx}(x^\star) - \sum_i \lambda_i^\star g_{i_{xx}}(x^\star)\Big)}_{= L_{xx}(x^\star,\lambda^\star)} \Phi'(z^\star) + \underbrace{\Big(f_x(x^\star) - \sum_i \lambda_i^\star g_{i_x}(x^\star)\Big)}_{= L_x(x^\star,\lambda^\star)} \Phi''(z^\star) \\ &= \Phi'(z^\star)^T L_{xx}(\Phi(z^\star),\lambda^\star)\Phi'(z^\star) + \underbrace{L_x(\Phi(z^\star),\lambda^\star)}_{=0}\Phi''(z^\star) = \Phi'(z^\star)^T L_{xx}(\Phi(z^\star),\lambda^\star)\Phi'(z^\star).\end{aligned}$$

Dies ist positiv semidefinit, so dass nun für alle $p \in T(x^\star)$ gilt

$$p^T L_{xx}(x^\star,\lambda^\star)p \geq 0.$$

Die Hesse-Matrix der Lagrangefunktion ist somit positiv semidefinit bezüglich der Richtungen des Raumes $T(x^\star)$. ■

Beispiel:

Löse $\min f(x_1,x_2) = x_2$ unter der Nebenbedingung $g(x_1,x_2) = x_2 - x_1^2 = 0$. Somit ergibt sich für die Lagrange-Funktion $L(x,\lambda) = x_2 - \lambda(x_2 - x_1^2)$. Für die erste Ableitung ergibt sich

$$\Big(\frac{\partial L}{\partial x}\Big)^T = (2x_1\lambda,\ 1-\lambda) = 0 \quad \Longrightarrow \quad \lambda = 1,\ x_1 = 0 \Rightarrow x_2 = 0.$$

Die Hesse-Matrix ist nun $\begin{pmatrix} 2\lambda & 0 \\ 0 & 0 \end{pmatrix} \overset{\lambda=1}{=} \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$, also positiv semidefinit auf $\mathbb{R}^2$. Es ergibt sich also die Optimallösung $x^\star = (0,0)^T$ und $\lambda^\star = 1$.

Definition 2.16 *Die notwendigen Bedingungen 1. Ordnung $L_x(x^\star,\lambda^\star) = 0$ und $g(x^\star) = 0$ heißen* Karush-Kuhn-Tucker-Bedingung (KKT-Bedingung) *und das Paar $(x^\star,\lambda^\star)$ heißt* Karush-Kuhn-Tucker-Punkt (KKT-Punkt)*.*

Bemerkung: Für den Tangentialraum gilt neben der eingeführten Definition $T(x^\star) = \{p \in \mathbb{R}^n \mid g_x(x^\star)p = 0\}$ nach Lemma 2.11 außerdem noch

$$T(x^\star) = \{p \in \mathbb{R}^n \mid \exists p_z \in \mathbb{R}^{n-l} : \underbrace{\begin{pmatrix} -g_x(x^\star)^{-1}g_z(x^\star) \\ I \end{pmatrix}}_{=:Z} p_z = p\}.$$

Offenbar bilden die Spalten von Z eine Basis von $T(x^\star)$.

Korollar 2.17 *Sei $Z = \begin{pmatrix} -g_y^{-1}(x^\star)g_z(x^\star) \\ I \end{pmatrix}$. Dann sind die notwendigen Optimalitätsbedingungen aus Satz 2.15 äquivalent zu:*

(i) $f_x(x^\star)Z = 0$ (ii) $Z^T f_{xx}(x^\star)Z$ positiv semidefinit.

Der reduzierte Gradient $\gamma := f_x(x^\star)Z$ verschwindet bzw. die reduzierte Hessematrix H^{red} ist positiv semidefinit. Die Bedingung hängt von der Wahl von Z ab. Wählt man eine Orthogonalbasis, so spricht man auch vom projizierten Gradienten bzw. von der projizierten Hessematrix.

Beispiel:

Sei folgendes Optimierungsproblem gegeben:

$$\begin{aligned} \min \quad & \tfrac{1}{2}x^T Hx + c^T x \\ & g(x) = Ax + b = 0, \end{aligned}$$

wobei $H \in \mathbb{R}^{n\times n}$ positiv definit ist und Rang $A = l$. Für die Lagrange-Funktion ergibt sich $L(x,\lambda) = \frac{1}{2}x^T Hx + c^T x - \lambda^T(Ax+b)$. Es ergibt sich das KKT-System

$$\begin{cases} \frac{\partial L}{\partial x} = Hx + c - A^T\lambda = 0 & \text{NOB 1.Ordnung,} \\ Ax + b = 0 & \text{NOB 2.Ordnung,} \end{cases}$$

welches von $(x^\star, \lambda^\star)$ erfüllt wird. Es ergibt sich nun das folgenden Gleichungssystem:

$$\begin{aligned} Hx - A^T\lambda &= -c \\ Ax &= -b \end{aligned} \quad \Longleftrightarrow \quad \begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix}\begin{pmatrix} x \\ -\lambda \end{pmatrix} = \begin{pmatrix} -c \\ -b \end{pmatrix}$$

$$\begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -H^{-1}(c + A^T\lambda) \\ -(AH^{-1}A^T)^{-1}(b - AH^{-1}c) \end{pmatrix}.$$

Lemma 2.18 *Es seien A und H Matrizen und A habe den Rang l. Die Matrix H ist genau dann positiv definit auf $Ap = 0$ (d.h. $p^T Hp > 0\ \forall p : Ap = 0, p \neq 0$), wenn $\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix}$ invertierbar ist.*

2.3.1 Nullraum-Methode (null space method)

Eine mögliche Lösungsmethode ist die Nullraum-Methode. Dafür zerlegen wir das Gleichungssystem weiter,

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} & A_1^T \\ H_{21} & H_{22} & A_2^T \\ A_1 & A_2 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ -\lambda \end{pmatrix} = -\begin{pmatrix} c_1 \\ c_2 \\ b \end{pmatrix}.$$

Sei A_1 regulär. A hat vollen Rang, unter Umständen ist hier vorher Spaltentausch nötig. Damit lässt sich das System transformieren zu

$$\begin{pmatrix} H_{11} & H_{12} & I \\ H_{21} & H_{22} & (A_1^{-1}A_2)^T \\ I & A_1^{-1}A_2 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ -A_1^T\lambda \end{pmatrix} = -\begin{pmatrix} c_1 \\ c_2 \\ A_1^{-1}b \end{pmatrix}.$$

Löst man die dritte Zeile nach x_1 auf und setzt dies in die erste und zweite Zeile ein, so erhält man

$$\begin{aligned} x_1 &= -A_1^{-1}b - A_1^{-1}A_2x_2, \\ (H_{11}(-A_1^{-1}A_2) + H_{12})x_2 - H_{11}A_1^{-1}b - A_1^T\lambda &= -c_1, \\ (H_{12}^T(-A_1^{-1}A_2) + H_{22})x_2 - H_{12}^TA_1^{-1}b - A_2^T\lambda &= -c_2 \end{aligned} \quad ,$$

und somit folgt dann nach Addition der zweiten und dritten Zeile in Matrixschreibweise

$$\begin{aligned} x_1 &= -A_1^{-1}b - A_1^{-1}A_2x_2, \\ \underbrace{\begin{pmatrix} -A_1^{-1}A_2 \\ I \end{pmatrix}^T \begin{pmatrix} H_{11} & H_{12} \\ H_{12}^T & H_{22} \end{pmatrix} \begin{pmatrix} -A_1^{-1}A_2 \\ I \end{pmatrix}}_{\text{reduzierte Hessematrix}} x_2 &= \underbrace{H_{12}^TA_1^{-1}b - A_2^TA_1^{-T}H_{11}A_1^{-1}b}_{\text{reduzierter Gradient}} - c_2 + A_2^TA_1^{-T}c_1 \ . \end{aligned}$$

Lemma 2.19 *Sei $A(x)$ eine symmetrische Matrix, welche in einer Umgebung K_ε von $x^\star$ stetig ist. Dann sind auch ihre Eigenwerte reell und stetig,*

$$\begin{aligned} -\infty < \underline{\lambda}\|s\|^2 \leq \quad & s^TA(x)s \quad && \leq \overline{\lambda}\|s\| < \infty, \\ -\infty < \underline{\lambda} \leq \quad & s^TA(x)s \quad && \leq \overline{\lambda} < \infty, \\ -\infty < \underline{\lambda} \leq \quad & \textit{Eigenwerte von } A(x) \quad && \leq \overline{\lambda} < \infty \quad \forall\ x \in K_\varepsilon(x^\star). \end{aligned}$$

Satz 2.20 (Hinreichende Optimalitätsbedingung (HOB)) *Sei $x^\star \in S \subset D$, $\lambda^\star$ der Lagrangemultiplikator und*

1. $\nabla_x L(x^\star, \lambda^\star) = 0$,
2. $p^T\nabla_x^2 L(x^\star, \lambda^\star)p > 0 \quad \forall\ p \in T(x^\star), p \neq 0.$

Dann ist $x^\star$ ein striktes lokales Minimum.

Beweis: Falls $x^\star$ ein regulärer Punkt ist, können wird das Problem auf $\tilde{f} = f(\Phi(z))$ zurückführen und Satz 2.15 anwenden.
Annahme: $x^\star$ sei kein striktes lokales Minimum. Dann existiert eine Folge $(y_k)_k$ mit $y_k \in S$, $y_k \neq x^\star$ und $y_k \to x^\star$ sowie $f(y_k) \leq f(x^\star)$. Definiere nun

$$s_k := \frac{y_k - x^\star}{\|y_k - x^\star\|}, \quad \|s_k\| = 1, \quad \varepsilon_k := \|y_k - x^\star\| > 0.$$

Die Folge $(s_k)_k$ ist beschränkt, also existiert eine konvergente Teilfolge $(s_k')_k$ mit

$$s_k' \to s^\star, \quad \|s_k'\| = 1.$$

Behauptung: $s^\star \in T(x^\star)$.
Für alle $i = 1, \ldots, l$ gilt, da $y_k' \in S$ und $x^\star \in S$, nach der Taylorentwicklung

$$0 = g_i(y_k') - g_i(x^\star) = \frac{d}{dx}g_i(x^\star)\varepsilon_k's_k' + \frac{1}{2}(\varepsilon_k')^2(s_k')^T\frac{d^2}{dx^2}g_i(\hat{x}_k')s_k',$$

für einen Zwischenwert $\hat{x}_k'$ auf der Verbindungsstrecke von y_k' und $x^\star$. Die Eigenwerte von $\frac{d^2}{dx^2}g_i$ sind beschränkt in einer Umgebung von $x^\star$ nach Lemma 2.19, also ist $(s_k')^T\frac{d^2}{dx^2}g_i(\hat{x}_k')s_k'$ beschränkt. Nach dem Grenzübergang $k \to \infty$ und Division durch ε_k' erhält man

$$\frac{d}{dx}g_i(x^\star)s^\star = 0 \quad \Longrightarrow \quad s^\star \in T(x^\star).$$

Weiterhin gilt

$$0 \geq f(y_k') - f(x^\star) = \nabla f(x^\star)s_k'\varepsilon_k' + \frac{1}{2}(\varepsilon_k')^2(s_k')^T\nabla^2 f(\tilde{x}_k')s_k',$$

mit $\tilde{x}'_k$ wieder auf der Verbindungsstrecke von y'_k und $x^\star$. Addiert man alle Werte $-\lambda_i(g_i(y'_k) - g_i(x^\star)) = 0$, so erhält man

$$0 \geq \nabla_x L(x^\star) s'_k \varepsilon'_k + \frac{1}{2}(\varepsilon'_k)^2 (s'_k)^T \left(\nabla^2 f(\hat{x}_k) - \sum_{i=1}^{l} \lambda_i \nabla^2 g_i(\hat{x}'_k) \right) s'_k$$

Dividiert man durch $(\varepsilon'_k)^2$ und lässt nun wieder $k \to \infty$ gehen, so erhält man

$$0 \geq \underbrace{\nabla_x L(x^\star)}_{=0} s^\star + \frac{1}{2} \underbrace{s^\star \nabla^2 L(x^\star, \lambda^\star) s^\star}_{>0}.$$

Dies ist offensichtlich ein Widerspruch und somit muss $x^\star$ striktes Minimum sein. ■

2.4 Gleichungs- und ungleichungsbeschränkte Optimierung

In diesem Unterkapitel werden nun Optimierungsprobleme behandelt, die sowohl durch Gleichungen als auch durch Ungleichungen restringiert sind. Die Probleme sind somit von der Form

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ & g(x) = 0, \quad g : D \to \mathbb{R}^l, \\ & h(x) \geq 0, \quad h : D \to \mathbb{R}^k. \end{aligned}$$

Die zulässige Menge S ist dann von der Form $S = \{x \in D \mid g(x) = 0,\ h(x) \geq 0\}$.

Definition 2.21 *Sei $x \in S$ ein zulässiger Punkt. Die* Indexmenge der aktiven Ungleichungen *ist $I(x) := \{i \in \mathbb{N} \mid h_i(x) = 0\} = \{i_1, \ldots, i_s\}$. Die* Indexmenge der inaktiven Ungleichungen *ist $I^+(x) := \{i \in \mathbb{N} \mid h_i(x) > 0\}$. Die Funktion $\tilde{h}$ ist definiert als $\tilde{h} : D \to \mathbb{R}^s, \quad x \mapsto (h_{i_1}(x), \ldots, h_{i_s}(x))^T, \quad D \subset \mathbb{R}^n$, und die Funktion $\tilde{g}$ als $\tilde{g} : D \to \mathbb{R}^{l+s}, \quad x \mapsto \begin{pmatrix} g(x) \\ \tilde{h}(x) \end{pmatrix}, \quad D \subset \mathbb{R}^n$. Ein Punkt $x \in S$ heißt* regulär, *wenn $\tilde{g}_x(x) = \nabla \tilde{g}(x)^T$ vollen Rang $l + s$ hat.*

O.B.d.A. seien die ersten $l + s$ Spalten von $\tilde{g}_x$ linear unabhängig. Damit lässt sich dann die Menge $\tilde{S} := \{\xi \in D \mid \tilde{g}(\xi) = 0\}$ in einer Umgebung U von x durch eine Abbildung Φ bzw. φ parametrisieren. Es existieren also Φ bzw. φ mit

$$\tilde{g}(\Phi(z)) = \tilde{g} \begin{pmatrix} \varphi(z) \\ z \end{pmatrix} = 0.$$

Definition 2.22 *Seien $\lambda \in \mathbb{R}^l$, $\mu \in \mathbb{R}^k$. Die Funktion $L(x, \lambda, \mu) := f(x) - \lambda^T g(x) - \mu^T h(x)$ heißt* Lagrange-Funktion.

Bemerkung: Der Tangentialraum, der zuerst nur für gleichungsbeschränkte Funktionen eingeführt wurde (vgl. Def. 2.10), existiert hier natürlich auch. Allerdings bezieht er sich ebenfalls nur auf die als Gleichungen erfüllten Nebenbedingungen, also laut Definition 2.21 auf $\tilde{g}$, und somit ist

$$T(x^\star) = \{p \in \mathbb{R}^n \mid \tilde{g}_x(x^\star) p = 0\}.$$

Satz 2.23 (Notwendige Optimalitätsbedingung (NOB)) *Sei $x^\star$ regulär und ein lokales Minimum, dann existieren $\lambda^\star$, $\mu^\star$ mit (wobei gilt: $\lambda^\star \in \mathbb{R}^l, \mu^\star \in \mathbb{R}^m, \mu \geq 0$)*

$$\nabla_x L(x^\star, \lambda^\star, \mu^\star) = 0 \iff \nabla f(x^\star) = \nabla g(x^\star)\lambda^\star + \nabla h(x^\star)\mu^\star \quad \textit{(NOB 1.Ordnung)},$$
$$\mu^\star \geq 0 \textit{ und } (\mu^\star)^T h(x^\star) = 0 \quad \textit{(Komplementaritätsbedingung)},$$
das heißt, für alle $i = 1, \ldots, s$ ist $\mu_i^\star = 0$ oder $h_i^\star = 0$,
$$p^T \nabla_x^2 L(x^\star, \lambda^\star, \mu^\star) p \geq 0 \ \forall\ p \in T(x^\star) \quad \textit{(NOB 2.Ordnung)}.$$

Die Hessematrix ist also auf dem Tangentialraum positiv definit.

Beweis:(basiert auf NOB für gleichungsbeschränkte Probleme)
Sei $x^\star$ ein lokales Minimum. Dann ist $x^\star$ ein Minimum in $\hat{U} := U \cap S$, wobei U eine geeignete Umgebung von $x^\star$ ist. Eine weitere Einschränkung führt auf

$$\tilde{U} := \hat{U} \cap \{x \in \mathbb{R}^n \mid h_i(x) > 0 \ \forall\ i \in I^+(x^\star)\} \cap \{x \in \mathbb{R}^n \mid h_i(x) = 0 \ \forall\ i \in I(x^\star)\}.$$

Also ist $x^\star$ lokales Minimum der Aufgabe $\min f(x)$ wobei $g(x) = 0$ und $\tilde{h}(x) = 0$. Aus Satz 2.15 folgt mit

$$\tilde{L}(x, \lambda, \tilde{\mu}) := f(x) - \lambda^T g(x) - \tilde{\mu}^T \tilde{h}(x),$$

dass $\lambda^\star, \tilde{\mu}^\star$ existieren mit

$$\nabla_x \tilde{L}(x^\star, \lambda^\star, \tilde{\mu}^\star) = \nabla f(x^\star) - \lambda^\star \nabla g(x^\star) - \tilde{\mu}^\star \nabla \tilde{h}(x^\star) = 0$$

und $\nabla_x^2 \tilde{L}(x^\star, \lambda^\star, \mu^\star)$positiv semidefinit. Im Übrigen setzen wir

$$\mu_i^\star = 0 \quad \forall\ i \in I^+(x^\star).$$

Daraus folgt $L(x^\star, \lambda^\star, \mu^\star) = \tilde{L}(x^\star, \lambda^\star, \tilde{\mu}^\star)$ und

$$\nabla L(x^\star, \lambda^\star, \mu^\star) = \nabla \tilde{L}(x^\star, \lambda^\star, \tilde{\mu}^\star) \Longrightarrow \nabla^2 L(x^\star, \lambda^\star, \mu^\star) = \nabla^2 \tilde{L}(x^\star, \lambda^\star, \tilde{\mu}^\star).$$

Außerdem ist $\nabla_x^2 \tilde{L}(x^\star, \lambda^\star, \mu^\star)$ positiv semidefinit genau dann, wenn $\nabla_x^2 L(x^\star, \lambda^\star, \mu^\star)$ positiv semidefinit ist auf $T(x^\star)$.
Es bleibt noch zu zeigen, dass $\mu^\star \geq 0$ ist. Wir zeigen stattdessen: Ist ein $\mu_i^\star < 0$, dann existiert eine Kurve im zulässigen Bereich mit $\tilde{h}_i(x) > 0$, entlang der f fällt.
Sei also $\mu_{\hat{\imath}}^\star < 0$ für ein $\hat{\imath} \in I(x^\star)$. Betrachte nun

$$\hat{S} := \{x \in \mathbb{R}^n \mid g(x) = 0 \text{ und } h_i(x) = 0 \ \forall\ i \in I(x^\star) \setminus \{\hat{\imath}\}\} \supset \tilde{S},$$

$$\hat{T} := \{p \in \mathbb{R}^n \mid \frac{d}{dx} g(x^\star) p = 0 \text{ und } \frac{d}{dx} h_i(x^\star) p = 0 \ \forall\ i \in I(x^\star) \setminus \{\hat{\imath}\}\}.$$

Dann lässt $\hat{S}$ sich parametrisieren durch

$$\hat{S} := \left\{ \begin{pmatrix} \hat{y} \\ \hat{z} \end{pmatrix} \mid \hat{y} = \hat{\varphi}(\hat{z}) \right\} = \left\{ x \in \mathbb{R}^n \mid x = \hat{\Phi}(\hat{z}) := \begin{pmatrix} \hat{\varphi}(\hat{z}) \\ \hat{z} \end{pmatrix} \right\}.$$

Wähle jetzt $p_{\hat{z}} \in \hat{T}(x^\star)$, so dass gilt $h_{\hat{\imath}}\left(\hat{\Phi}(\hat{z^\star} + tp_{\hat{z}})\right) = h_{\hat{\imath}}\left(\hat{\varphi}(\hat{z^\star} + tp_{\hat{z}}), \hat{z^\star} + tp_{\hat{z}}\right) > 0$. Also gilt für $p^T = p_z^T \Phi'(x^\star) \in \hat{T}(x^\star)$ neben $g_x(x^\star)p = 0$ auch

$$i \in I(x^\star) \setminus \{\hat{\imath}\} : \begin{cases} \frac{d}{dx} h_{i_1}(x^\star)p = 0 \\ \vdots \\ \frac{d}{dx} h_{i_s}(x^\star)p = 0 \end{cases} \quad \text{mit Ausnahme von } \frac{d}{dx} h_{\hat{\imath}}(x^\star)p = \varepsilon.$$

Was passiert nun mit f? Es gilt

$$\begin{aligned} \frac{d}{dx} f(\Phi(z^\star + tp_z))\Big|_{t=0} &= f_x(x^\star)p = f_x \begin{pmatrix} \varphi'(z^\star) \\ I \end{pmatrix} p_{\hat{z}} \\ &= \lambda^T g_x(x^\star)p_z + \mu^T h_x(x^\star)p_z = \underbrace{\mu_{\hat{\imath}}}_{<0} \underbrace{\frac{d}{dx} h_{\hat{\imath}}(x^\star)p_z}_{=\varepsilon} < 0, \end{aligned}$$

Und dies ist nun ein Widerspruch dazu, das $x^\star$ ein lokales Minimum ist. ■

Bemerkung: Falls $(x^\star, \lambda^\star, \mu^\star)$ mit $\exists\ \mu_k^\star < 0$: $\nabla L(x^\star, \lambda^\star, \mu^\star) = 0$. Dann existiert immer eine Kurve in S, sodass f entlang dieser Kurve fällt.

Bemerkung: Als hinreichende Optimalitätsbedingung reicht nicht aus

$$\begin{aligned} p^T \nabla^2 L(x^\star, \lambda^\star, \mu^\star) &> 0, \quad \forall\ p \in T(x^\star) \setminus \{0\}, \\ T(x^\star) &= \{p \in \mathbb{R}^n \mid \tilde{g}_x(x^\star)p = 0\}. \end{aligned}$$

Das Problem sind die *schwach aktiven Ungleichungen* mit $\mu_i^\star = 0$ und $h_i(x^\star) = 0$. Die Ungleichungen sind *strikt aktiv*, falls $\mu_i^\star > 0$ und $h_i(x^\star) = 0$.

Definition 2.24 Tangentialraum für strikt aktive Ungleichungen *Der Tangentialraum bezogen auf die strikt aktiven Ungleichungen ist definiert als*

$$T^+(x^\star) := \{p \in \mathbb{R}^n \mid \frac{d}{dx} g(x^\star)p = 0,\ \frac{d}{dx} h_i(x^\star)p = 0\ \forall\ i \in I(x^\star)\ \textit{und}\ \mu_i^\star > 0\} \supset T(x^\star).$$

Korrekterweise müsste es allerdings $T^+(x^\star, \mu^\star)$ heißen, es wird aber darauf verzichtet.

Satz 2.25 (Hinreichende Optimalitätsbedingung (HOB)) *Sei $x^\star \in S$ und erfülle mit $\lambda^\star, \mu^\star \geq 0$ die NOB 1.Ordnung $\nabla_x L(x^\star, \lambda^\star, \mu^\star) = 0$. Ist außerdem die Komplementaritätsbedingung $(\mu^\star)^T h(x^\star) = 0$ und die weitere Bedingung*

$$p^T \nabla_x^2 L(x^\star, \lambda^\star, \mu^\star)p > 0\ \forall p \in T^+(x^\star) \setminus \{0\}$$

erfüllt, dann ist $x^\star$ ein strikt lokales Minimum.

Bemerkung: $x^\star$ muss nicht regulär sein.

Beweis: Widerspruchsannahme: $x^\star$ ist kein strikt lokales Minimum.

Dann existiert eine Folge $(y_k)_k$, $y_k \neq x^\star,\ y_k \to x^\star$ mit $f(y_k) \leq f(x^\star)$ und es gilt ferner $g(y_k) = 0$ und $h(y_k) \geq 0$. Definiere nun eine Folge $(s_k)_k$ durch

$$\varepsilon_k := \|y_k - x^\star\|, \quad s_k := \frac{y_k - x^\star}{\varepsilon_k}, \quad \|s_k\| = 1.$$

Die Folge $(s_k)_k$ ist beschränkt. Also existiert eine Teilfolge s'_k mit $s'_k \to x^\star$. Wie in Satz 2.20 folgt nun durch die Taylorentwicklung

$$\varphi'(y'_k) = \varphi'(x^\star + s'_k\varepsilon'_k) = \varphi(x^\star) + \nabla\varphi^T(x^\star)\varepsilon'_k s'_k + \mathcal{O}\left((\varepsilon'_k)^2\|\right)$$

$$\Rightarrow \quad \varphi'(y'_k) - \varphi(x^\star) = \nabla\varphi^T(x^\star)\varepsilon'_k s'_k + \mathcal{O}({\varepsilon'_k}^2) \quad \Leftrightarrow \quad \frac{\varphi'(y'_k) - \varphi(x^\star)}{\varepsilon'_k} = \nabla\varphi^T(x^\star)s'_k + \mathcal{O}(\varepsilon'_k).$$

Bei der Grenzwertbildung $k \to \infty$ läuft der rechte Ausdruck gegen $\nabla\varphi^T(x^\star)$. Insgesamt ergibt sich

$$\begin{aligned}
0 \geq f(y'_k) - f(^\star x) = \nabla_x^T f(x^\star)\varepsilon'_k s'_k + \mathcal{O}({\varepsilon'_k}^2) &\implies \nabla_x^T f(x^\star)\hat{s} \leq 0,\\
0 = g(y'_k) - g(x^\star) = \nabla_x^T g(x^\star)\varepsilon'_k s'_k + \mathcal{O}({\varepsilon'_k}^2) &\implies \nabla_x^T g(x^\star)\hat{s} = 0,\\
0 \leq h_i(y'_k) - h_i(x^\star) = \nabla_x^T h_i(x^\star)\varepsilon'_k s'_k + \mathcal{O}({\varepsilon'_k}^2) &\implies \nabla_x^T h_i(^\star)\hat{s} \geq 0 \ \forall\ i \in I(x^\star).
\end{aligned}$$

Betrachtet man nun die Lagrange-Funktion $L(x, \lambda, \mu) = f(x) - \lambda^T g(x) - \mu^T h(x)$, so gilt ebenfalls

$$0 \geq L(y'_k, \lambda^\star, \mu^\star) - L(x^\star, \lambda^\star, \mu^\star) = \underbrace{\nabla_x L(x^\star, \lambda^\star, \mu^\star)\varepsilon'_k s'_k}_{=0} + \frac{1}{2}\underbrace{(\varepsilon'_k)^2 (s'_k)^T \nabla_x^2 L(x^\star, \lambda^\star, \mu^\star) s'_k}_{\to \hat{s}^T\nabla_x^2 L(x^\star,\lambda^\star,\mu^\star)\hat{s} > 0 \ (\circledast)},$$

wobei $\circledast$ gilt für $k \to \infty$ und falls $\hat{s}$ in $T^+(x^\star)$. Nun sind zwei Fälle zu unterscheiden:

1. Gilt für alle $i \in I(x^\star)$, dass $\frac{d}{dx}h_i(x^\star)\hat{s} = 0$ und $\mu_i^\star > 0$, dann ist $\hat{s} \in T^+(x^\star)$. Dies führt zum Widerspruch wie in Satz 2.20.

2. Existiert ein i mit $\frac{d}{dx}h_i(x^\star)\hat{s} > 0$ und $\mu_i^\star > 0$, dann ist

$$0 \geq \nabla_x f(x^\star)\hat{s} = \underbrace{(\lambda^\star)^T \nabla_x g(x^\star)\hat{s}}_{=0} + (\mu^\star)^T \nabla_x h(x^\star)\hat{s} \geq \underbrace{\mu_i^\star}_{>0}\underbrace{\nabla_x h_i(x^\star)\hat{s}}_{>0} > 0.$$

   und dies ist ebenfalls ein Widerspruch. ■

Bemerkungen:

1. Falls es ein $\mu_i^\star = 0$ mit $h_i(x^\star) = 0$ gibt, ist $T(x^\star)$ eine echte Teilmenge von $T^+(x^\star)$.

2. Sowohl notwendige als auch hinreichende Bedingung lassen sich auch mit $\tilde{T}(x^\star)$ beweisen, wobei

$$\tilde{T}(x^\star) := \left\{ p \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \frac{d}{dx}g(x^\star)p = 0, & \\ \frac{d}{dx}h_i(x^\star)p = 0 & \forall\ i \in I(x^\star) \text{ und } \mu_i^\star > 0, \\ \frac{d}{dx}h_i(^\star x)p \geq 0 & \forall\ i \in I(x^\star) \text{ und } \mu_i^\star = 0. \end{array} \right\}.$$

   Der Kegel $\tilde{T}$ erfüllt $T(x^\star) \subset \tilde{T}(x^\star) \subset T^+(x^\star)$,denn in $T(x^\star)$ gilt $\nabla^T h_i(x^\star)p = 0$ und in $T^+(x^\star)$ gilt $\nabla^T h_i(x^\star)p = 0$ für $\mu_i^* > 0$.

Beispiel: Quadratische Probleme

$$\min_x \frac{1}{2} x^T H x + g^T x$$

$$Ax = b$$

$$L(x, \lambda) = \frac{1}{2} x^T H x + g^T x - \lambda^T (Ax - b)$$

$$\text{NOB 1.Ordung: } \frac{\partial L}{\partial x} x^T H + g^T - \lambda^T A = 0$$

$$\Rightarrow Hx - A^T \lambda = -g \text{ und } Ax = b$$

, wobei: H Hessematrix und A Jacobimatrix von den Restriktionen sind.

$$\Rightarrow x, \lambda \text{ erfüllen LGS:}$$

$$\begin{pmatrix} H & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -g \\ b \end{pmatrix}$$

Korollar 2.26 *Sei $x^\star \in S$ ein regulärer und zulässiger Punkt. Es existieren $\lambda^\star$, $\mu^\star \geq 0$ mit $\nabla_x L(x^\star, \lambda^\star, \mu^\star) = 0$ und $(\mu^\star)^T h(x^\star) = 0$. Dann ist*

$$p^T \nabla_x^2 L(x^\star, \lambda^\star, \mu^\star) p \geq \delta p^T p \quad \forall\ p \in \tilde{T}(x^\star) \setminus \{0\}.$$

Für $\delta = 0$ ist somit die NOB und für $\delta > 0$ die HOB für ein Minimum erfüllt.

Lemma 2.27 *Seien $H \in \mathbb{R}^{n \times n}$, $A \in \mathbb{R}^{l \times n}$ mit $l \leq n$. Sei*

$$K := \begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix}$$

Sei rg $A = l \leq n$. Falls $p^T H p > 0$ $\forall p \neq 0$, $Ap = 0$, dann gilt: K regulär (also invertierbar).

Beweis: Wir zeigen, dass

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ hat nur triviale Lösung } \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Es gilt:

1. $Hu + A^T v = 0$
2. $Au = 0$

Durch Multiplikation von u^T von links erhält man für 1:

$$u^T H u + \underbrace{u^T A v}_{=0} = 0 \Rightarrow u^T H u = 0 \Rightarrow u = 0$$

Aus 2 folgt:

$$A^T v = 0 \Rightarrow v = 0, \text{ da rg } A^T = l, \text{also linear unabhängige Spalten}$$

■

Definition 2.28 Strikte Komplementarität *gilt dann, wenn*

$$\begin{aligned} \mu_i^\star = 0 &\iff h_i(x^\star) > 0, \\ \mu_i^\star > 0 &\iff h_i(x^\star) = 0. \end{aligned}$$

Strikte Komplementarität wird für Lösungsalgorithmen meist vorausgesetzt, denn sie ist wichtig für die Stabilität dieser Algorithmen, was sich im folgenden Abschnitt noch zeigen wird.

2.5 Stabilitätsaussagen

Wie bereits oben schon erwähnt soll nun in diesem Abschnitt beschrieben werden, wie sich die Stabilität bei den Lösungsalgorithmen verhält und wie wichtig dafür besonders die strikte Komplementarität ist. Allgemein wird das folgende Problem betrachtet,

$$\min_x f(x,\varepsilon) \quad \text{unter} \quad g(x,\varepsilon) = 0,\ h(x,\varepsilon) \geq 0, \tag{NLP(ε)}$$

wobei f, g, h genügend oft stetig differenzierbar sind.
Für $\epsilon = 0$: $\min_x f(x,0) = f(x),\ g(x) = 0,\ h(x) \geq 0$, Lösung: $x^\star = x(0)$
Gesucht: Lösung $x(\epsilon)$
Wunsch:

1. $x(\epsilon)$ stetig differenzierbar in ϵ,
   d.h. $x(\epsilon)$ stetige Transforamtion von $x(0) \Rightarrow x(\epsilon) = x^\star + \epsilon \frac{\partial x}{\partial \epsilon}(0) + O(||\epsilon||^2)$
2. $||x(\epsilon) - x^\star|| \leq c||\epsilon||$

Um dies entsprechend zu illustrieren soll ein kleines Beispiel zuerst in die Problematik einführen.

Beispiel 1:

Gegeben sei das Optimierungsproblem $\min f(x) = x^4$. Durch die notwendige Bedingung an der ersten Ableitung $f'(x) = 4x^3$ erhält man, dass $x^\star = 0$ ein mögliches Minimum ist. Das hinreichende Kriterium der zweiten Ableitung ist aber nicht erfüllt, da $f''(x^\star) = 12 \cdot 0^2 = 0$ ist. Aus diesem Grund wird nun eine kleine Störung eingebaut. Es wird nun das gestörte Problem betrachtet:

$$\min f(x,\varepsilon) = x^4 - 2\varepsilon x^2, \quad \varepsilon > 0.$$

Es ergibt sich die Ableitung $f'(x,\varepsilon) = 4x^3 - 4\varepsilon x = 4x(x^2 - \varepsilon)$. Also ist $f'(x,\varepsilon) = 0$ für $x_1 = 0$ oder $x_{2/3} = \pm\sqrt{\varepsilon}$. Ferner ist $f''(x,\varepsilon) = 12x^2 - 4\varepsilon$, somit ergibt sich für die möglichen Extremstellen:

$$f''(0,\varepsilon) = -4\varepsilon < 0 \quad \Longrightarrow \quad x_1 \text{ ist Maximum}$$

$$f''(\pm\sqrt{\varepsilon},\varepsilon) = 8\varepsilon > 0 \quad \Longrightarrow \quad x_{2/3} \text{ sind Minima}$$

Also existiert ein Sprung in der Lösung und das Lösungsverhalten ist daher instabil. Die Stabilitätsausagen basieren auf dem Satz über implizite Funktionen.

Beispiel 2:

Gegeben: $\min_{x\in\mathbb{R}} f(x) = x_1^2$. Dann gilt:

- $\nabla f(x) = \begin{pmatrix} 2x_1 \\ 0 \end{pmatrix} \Rightarrow$ Lösung $x_1^\star = 0,\ x_2^\star$ beliebig
- $\nabla^2 f(x) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$
  Fall 1: positiv semi-definit $\Rightarrow$ NOB sind erfüllt
  Fall 2: nicht positiv definit $\Rightarrow$ HOB nicht erfüllt
  - Fall 1: $f(x,\epsilon) = x_1^2 + \epsilon x_2,\ \epsilon > 0$
    $\nabla f(x,\epsilon) = \begin{pmatrix} 2x_1 \\ \epsilon \end{pmatrix} \neq 0,\ \epsilon \neq 0$
    $\Rightarrow$ Es existiert kein Minimum
  - Fall 2: $f(x,\epsilon) = x_1^2 - \epsilon x_2^2,\ \epsilon > 0$
    $\nabla f(x,\epsilon) = \begin{pmatrix} 2x_1 \\ -\epsilon x_2 \end{pmatrix} = 0$ für $x = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
    $\nabla^2 f(x,\epsilon) = \begin{pmatrix} 2 & 0 \\ 0 & -\epsilon \end{pmatrix}$ indefinit
    $\Rightarrow x = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ Sattelpunkt

Satz 2.29 (Stabilitätssatz) *Seien $x^\star, \lambda^\star, \mu^\star$ gegeben, so dass die NOB und die HOB für $\varepsilon = 0$ erfüllt sind, d.h.*

(i) *$x^\star$ ist zulässig, d.h. $g(x^\star, 0) = 0,\ h(x^\star, 0) \geq 0$, Aktive Ungleichungen sind $I(x^\star)$*

(ii) $\nabla_x L(x^\star, \lambda^\star, \mu^\star, 0) = \nabla_x f(x^\star, 0) - \sum \lambda_i^\star \nabla_x g_i(x_i^\star, 0) - \sum \mu_i^\star \nabla_x h_i(x_i^\star, 0) = 0$,

(iii) *<u>strikte</u> Komplementarität (SK): $\mu^\star \geq 0$ und $(h_i(x^\star, 0) > 0 \Rightarrow\ \mu^\star = 0)$ und $(h_i(x^\star, 0) = 0 \Rightarrow \mu^\star > 0)$,*

(iv) *$x^\star$ regulär, d.h. der Rang von* $\begin{pmatrix} \nabla^T g(x^\star, 0) \\ \nabla^T \tilde{h}(x^\star, 0) \end{pmatrix} = l + |I(x^\star)|$, *wobei* $|I(x^\star)| = s$

(v) $p^T \nabla^2 L(x^\star, \lambda^\star, \mu^\star, 0) p > 0\ \forall\ p \neq 0,\ p \in T^+(x^\star) (= T(x^\star)$ *wegen strikter Komplementarität,* $T(x^\star) = \{p \mid \nabla^T g_i(x^\star, 0) p = 0,\ i = 1, \ldots, l,\ \nabla^T h_i(x^\star, 0) p = 0,\ i \in I(x^\star)\})$.

Dann existieren eine Umgebungen E von $\varepsilon = 0$ und $U \times V \times W$ von $(x^\star, \lambda^\star, \mu^\star)$ und eine stetig differenzierbare Abbildung

$$\begin{pmatrix} x \\ \lambda \\ \mu \end{pmatrix} : E \to U \times V \times W, \quad \begin{pmatrix} x(\varepsilon) \\ \lambda(\varepsilon) \\ \mu(\varepsilon) \end{pmatrix}$$

mit:

- $x(0) = x^\star,\ \lambda(0) = \lambda^\star,\ \mu(0) = \mu^\star$,
- $L(x(\varepsilon), \lambda(\varepsilon), \mu(\varepsilon)) = 0$,

- $g(x(\varepsilon),\varepsilon)=0,\ h(x(\varepsilon),\varepsilon)\geq 0,$
- *SK:* $\mu(\varepsilon)\geq 0$ *und* $h_i(x(\varepsilon),\varepsilon)=0\Leftrightarrow\mu_i(\varepsilon)>0,$
- *aktive Ungleichungen bleiben gleich:* $I(x(\varepsilon))=I(x^\star),$
- $p^T\nabla^2 L(x(\varepsilon),\lambda(\varepsilon),\mu(\varepsilon),\varepsilon)p>0\quad\forall\ p\neq 0,\ p\in T(x(\varepsilon),\varepsilon)$
  $T(x(\varepsilon),\varepsilon)=\big\{p\mid\nabla^T g_i(x(\varepsilon),\varepsilon)p=0,\ i=1,\dots,l,\ \nabla^T h_i(x(\varepsilon),\varepsilon)p=0,\ i\in I(x(\varepsilon))\big\},$

und $x(\varepsilon)$ *ist ein strikt lokales Minimum in (NLP(*ε*)).*

Beweis: Das Grundprinzip des Beweises beruht auf dem Satz über implizite Funktionen und der stetige Differenzierbarkeit. Wir definieren

$$F=\begin{pmatrix}\nabla L(x,\lambda,\tilde\mu,\varepsilon)=0\\ g(x,\varepsilon)=0\\ h_i(x,\varepsilon)=0,\ i\in I(x^\star)\end{pmatrix}.$$

Außerdem ist $\tilde\mu=(\mu_i,\ i\in I(x^\star))$ und $F(x^\star,\lambda^\star,\tilde\mu^\star,0)=0$ nach Voraussetzung. Nun bildet man den Gradienten und es folgt

$$\frac{\partial F(x(\varepsilon),\lambda(\varepsilon),\tilde\mu(\varepsilon),\varepsilon)}{\partial(x,\lambda,\mu)}\bigg|_{\varepsilon=0}=\begin{pmatrix}\nabla^2 L(x^\star,\lambda^\star,\tilde\mu^\star,0) & \nabla g(x^\star,0) & \nabla\tilde h(x^\star,0)\\ \nabla^T g(x^\star,0) & 0 & 0\\ \nabla^T\tilde h(x^\star,0) & 0 & 0\end{pmatrix}.$$

Diese Matrix ist nach einer Übungsaufgabe regulär. Aus dem Satz über implizite Funktionen folgt nun, dass Umgebungen E''' und $U\times V\times W$ existieren mit $(x,\lambda,\tilde\mu)^T\colon E'''\to U\times V\times W$, so dass $F(x(\varepsilon),\lambda(\varepsilon),\tilde\mu(\varepsilon),\varepsilon)=0$ ist. Nun schränkt man E''' auf E'' ein, so dass $h_i(x(\varepsilon),\varepsilon)>0\ \forall\,\varepsilon\in E'',\ i\notin I(x^\star)$. Dann schränkt man E'' ein auf E', so dass $\tilde\mu_i(\varepsilon)>0\forall\ \varepsilon\in E',\ i\in I(x^\star)$, und setzt nun $\mu_i(\varepsilon)=0$ für $i\notin I(x^\star)$. Zuletzt schränkt man E' so ein auf E, dass $p^T\nabla^2 L(x(\varepsilon),\lambda(\varepsilon),\mu(\varepsilon),\varepsilon)p>0$ auf $T(x(\varepsilon),\varepsilon)$. Somit folgt

$$\frac{d(x(\varepsilon),\lambda(\varepsilon),\tilde\mu(\varepsilon))^T}{d\varepsilon}=\begin{pmatrix}\nabla^2(L(x(\varepsilon),\lambda(\varepsilon),\tilde\mu(\varepsilon),\varepsilon) & g_x^T(\dots) & \tilde h_x^T(\dots)\\ g_x(x(\varepsilon),\varepsilon) & 0 & 0\\ \tilde h_x(x(\varepsilon),\varepsilon) & 0 & 0\end{pmatrix}^{-1}\begin{pmatrix}\partial L(x(\varepsilon),\lambda(\varepsilon),\tilde\mu(\varepsilon),\varepsilon)/\partial\varepsilon\\ \partial g/\partial\varepsilon\\ \partial\tilde h/\partial\varepsilon\end{pmatrix}.$$

Die Frage ist nun, da $h_j(x,\varepsilon)=h_j(x)-\varepsilon\geq 0,\varepsilon\in\mathbb{R},h_j(x^\star)=0$, was passiert mit $f(x,\varepsilon)=f(x)$?

$$\begin{aligned}\frac{df(x(\varepsilon))}{d\varepsilon}\bigg|_{\varepsilon=0}&=\nabla_x^T f(x^\star)\frac{dx(\varepsilon)}{d\varepsilon}\bigg|_{\varepsilon=0}+\frac{\partial f(x(\varepsilon))}{\partial\varepsilon}\bigg|_{\varepsilon=0}\\ &=\Big(\sum\lambda_i^\star\nabla_x g_i(x^\star,0)+\sum\tilde\mu_i^\star\nabla_x h_i(x^\star,0)\Big)\frac{dx(\varepsilon)}{d\varepsilon}\bigg|_{\varepsilon=0}=\mu_j^\star\nabla_x h_j(x^\star,0)\frac{dx(\varepsilon)}{d\varepsilon}\bigg|_{\varepsilon=0}\end{aligned}$$

Aus

$$\mu_j^\star\Bigg(\nabla^T h_j(x^\star,0)\frac{dx(\varepsilon)}{d\varepsilon}\bigg|_{\varepsilon=0}+\underbrace{\frac{\partial h_j(x^\star,\varepsilon)}{\partial\varepsilon}\bigg|_{\varepsilon=0}}_{=-1}\Bigg)=0$$

folgt $\mu_j^\star \nabla^T h_j(x^\star, 0) \frac{dx(\varepsilon)}{d\varepsilon}\Big|_{\varepsilon=0} = \mu_j^\star$ und somit ergibt sich

$$\frac{df(x(\varepsilon))}{d\varepsilon}\Big|_{\varepsilon=0} = \begin{pmatrix} \lambda^\star \\ \tilde{\mu}^\star \end{pmatrix}^T.$$

Hierbei sind $\lambda^\star$ und $\tilde{\mu}^\star$ die Schattenpreise. $\lambda_i^\star$ und $\tilde{\mu}_i^\star$ sind der Preis in der Zielfunktion für die Veränderung der Beschränkungen $g_i(x, \varepsilon)$ und $h_i(x, \varepsilon)$ um $\delta s = 1$. ■

Folgerung:
$\forall \epsilon \in E:\ x(\epsilon)$ striktes lokales Minimum in gestörten Problem.

2.6 Diskussion

Ausgegangen wird wieder von dem Problem $\min_x f(x)$ mit den Restriktionen $g(x) = 0,\ h(x) \geq 0$. Betrachte nun die Funktion $f(x + \varepsilon d) \doteq f(x) + \varepsilon \nabla^T f(x) d$. Die Gleichung lässt sich noch fortsetzen für Terme höherer Ordnung. Analog gilt für die Nebenbedingungen

$$\begin{aligned} g(x + \varepsilon d) &\doteq g(x) + \varepsilon \nabla^T g(x) d = 0, \\ h(x + \varepsilon d) &\doteq h(x) + \varepsilon \nabla^T h(x) d \geq 0. \end{aligned}$$

Eine zulässige Richtung d erfüllt also $\nabla^T g(x) d = 0$ und $\nabla^T h(x) d \geq -\frac{h(x)}{\varepsilon}$. Insbesondere ist $\nabla^T h_i(x) d \geq 0$ für $i \in I(x)$. Intuitiv wird nun deutlich: Falls x ein lokales Minimum in einem nichtlinearen Programm ist, dann ist für alle zulässigen Richtungen d die Änderung der Zielfunktion entlang d nicht negativ, $\nabla^T f(x) d \geq 0$. In anderen Worten bedeutet dies, dass dann die folgende Menge leer ist,

$$Z(x) = \left\{\ d \mid \nabla^T g(x) d = 0,\ \nabla^T h_i(x) d \geq 0\ i \in I(x),\ \nabla^T f(x) d < 0\ \right\} = \emptyset.$$

Satz 2.30 *Ist die Menge $Z(x)$ leer, dann existieren λ, μ, so dass $\nabla f(x) = \sum \lambda_i \nabla g_i(x) + \sum \mu_i \nabla h_i(x)$.*

- *$Z(x) = \emptyset$ garantiert die Existenz von μ und λ.*
- *Aus der Linear Independence Constraint Qualification (LICQ) folgt $Z(x) = \emptyset$.*
- *Falls $g(x), h_i(x),\ i \in I(x)$, linear abhängig sind, so ist ebenfalls Z die leere Menge.*
- Mangasarian Fromowitz Constraint Qualification (MFCQ): *x erfüllt die MFCQ, falls ein s existiert mit $\nabla^T g(x) s = 0$, $\nabla h_i^T s > 0$ für $i \in I(x)$.*

Beispiel:

Gegeben seien die drei Ungleichungen

$$\begin{aligned} g_1(x) &= -x_1^2 - x_2 - x_3 \geq 0, \\ g_2(x) &= -x_1^2 + x_2 - x_3 \geq 0, \\ g_3(x) &= -x_3 \geq 0. \end{aligned}$$

Es soll nun geprüft werden, ob der Punkt $x^\star = (0, 0, 0)^T$ die MFCQ-Bedingung und die LICQ-Bedingung erfüllt. Für MFCQ muss ein $s \in \mathbb{R}^n$ gefunden werden, so dass $\nabla g_i(x^\star)^T s > 0$ ist, für

die aktiven Ungleichungen. Die Ungleichungen sind alle aktiv, denn es gilt für alle $g_i(x^\star) = 0$. Für die jeweiligen Gradienten ergibt sich

$$\nabla g_1(x)^T = (-2x_1, -1, -1), \quad \nabla g_2(x)^T = (-2x_1, 1, -1), \quad \nabla g_3(x)^T = (0, 0, -1).$$

Somit folgt für die Gradienten im Punkt $x^\star$, dass

$$\nabla g_1(x^\star)^T = (0, -1, -1), \quad \nabla g_2(x^\star)^T = (0, 1, -1), \quad \nabla g_3(x^\star)^T = (0, 0, -1).$$

Wählt man nun $s^T = (0, -1, -2)$ so sind die MFCQ erfüllt, denn für alle $\nabla g_i(x^\star)^T s > 0$. Die LICQ sind hingegen nicht erfüllt, denn

$$\begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$$

sind linear abhängig.

Kapitel 3

Grundlegende Algorithmen

3.1 Allgemeines Abstiegsverfahren

Nachdem im vorherigen Kapitel nun die notwendigen und hinreichenden Kriterien für ein Optimum erläutert wurden, sollen nun in diesem Kapitel verschiedene Lösungsverfahren diskutiert werden, die ein Optimum bestimmen. Wir betrachten zunächst den unbeschränkten Fall. Alle Algorithmen sind iterativ, sofern es sich nicht um lineare oder quadratische Probleme handelt. Grundsätzlich sind die Probleme von folgender Form:

$$\min f(x), \quad \text{wobei } f : D \to \mathbb{R}, D \subset \mathbb{R}^n.$$

Bei den Verfahren werden folgende Bezeichnungen verwendet:

$x_0 \in D$	Startwert
d^k	Suchrichtung
t^k	Schrittweite
$x_{k+1} = x^k + t^k d^k$	Iteration

Definition 3.1 *Seien $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und $x \in \mathbb{R}^n$. Ein Vektor $d \in \mathbb{R}^n$ heißt* Abstiegsrichtung *von f in x, wenn ein $\bar{t} > 0$ existiert mit $f(x + td) < f(x)\ \forall\ t \in]\,0, \bar{t}\,]$.*

Lemma 3.2 *Seien f stetig differenzierbar, $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ mit $\nabla^T f(x)d < 0$. Dann ist d eine Abstiegsrichtung von f in x.*

Beweis: Bestimme die Richtungsableitung:

$$f'(x,d) = \lim_{t \to +0} \frac{f(x+td) - f(x)}{t} \overset{MWS}{=} \lim_{t \to +0} \frac{\nabla f(\tilde{x})d}{t} = \nabla^T f(x)d \ < \ 0 \quad \text{für } \tilde{x} \in [x, x+td].$$

Somit ist $\frac{f(x+td)-f(x)}{t} < 0$ für alle hinreichend kleinen t. ■

Bemerkung: Die Bedingung $\nabla^T f(x)d < 0$ ist nur hinreichend, aber nicht notwendig. Es kann z.B. der Fall auftreten, dass x ein lokales Maximum ist. Geometrisch bedeutet das Lemma, dass der Winkel zwischen d und $(-\nabla^T f(x))$ kleiner als 90° ist. Dies ist vergleichbar mit einer Übungsaufgabe, in der gezeigt wurde, dass $\nabla f(x) \perp \{z \mid f(z) = f(x) = const\}$.
Um nun den Algorithmus für das Abstiegsverfahren explizit angeben zu können, benötigt man mit x^0 eine Startangabe für den Schritt $k = 0$. Dann sieht der Algorithmus wie folgt aus:

1. Prüfe, ob x^k den Abbruchkriterien genügt.

2. Bestimme die Abstiegsrichtung d^k.

3. Bestimme die Schrittweite t^k mit $f(x^k + t^k d^k) < f(x^k)$.

4. Setze $x^{k+1} = x^k + t^k d^k$ und $k := k + 1$ und setze den Algorithmus bei Schritt 1 fort.

Es stellen sich hierbei nun zwei zentrale Fragen: Nämlich zum einen die nach der globalen Konvergenz, d.h. ob das Verfahren für beliebige Startwerte gegen ein lokales Minimum konveriert, und zum anderen die nach der lokalen Konvergenz, d.h. wie schnell das Verfahren in der Nähe einer Lösung konvergiert. Diese beiden Fragen gilt es bei den jeweils folgenden Algorithmen zu erläutern.

3.2 Das Gradienten-Verfahren

Lemma 3.3 *Sei $\nabla f(x) \neq 0$, dann löst $\bar{d} := -\frac{\nabla f(x)}{\|\nabla f(x)\|}$ das Problem*

$$\min_{d \in \mathbb{R}^n} \nabla^T f(x) d \quad \text{mit } \|d\| = 1.$$

Beweis: $L(d, \lambda) = \nabla^T f(x) d - \lambda(d^T d - 1)$. Somit folgt

$$\frac{\partial L(d^\star, \lambda^\star)}{\partial d} = \nabla f(x) - 2\lambda^\star d^\star = 0 \quad \Rightarrow \quad d^\star = \frac{\nabla f(x)}{2\lambda^\star}.$$

Durch die Restriktion ergibt sich $1 = (d^\star)^T d^\star = \frac{\|\nabla f(x)\|_2^2}{4(\lambda^\star)^2}$ und daraus folgt unmittelbar, dass $\lambda^\star = \pm\frac{\|\nabla^T f(x)\|}{2}$. Für die zweite Ableitung ergibt sich

$$\frac{\partial^2 L(d^\star, \lambda^\star)}{\partial d^2} = \begin{pmatrix} -2\lambda^\star & & 0 \\ & \ddots & \\ 0 & & -2\lambda^\star \end{pmatrix}.$$

Diese ist positiv definit, falls $-2\lambda^\star > 0$, was bedeutet, dass $\lambda^\star < 0$ sein muss. Somit scheidet die positive Lösung aus. Für die Optimallösung ergibt sich: $d^\star = -\frac{\nabla f(x)}{\|\nabla f(x)\|}$ ■

Algorithmus des Gradienten-Verfahrens

Bei einer Dateneingabe von einem Startwert x^0 und einer Abbruchgenauigkeit $\varepsilon > 0$ sieht der Algorithmus für $k = 0, 1, \ldots$ folgendermaßen aus:

1. Berechne $f(x^k)$ und $\nabla f(x^k)$.

2. Falls $\|\nabla f(x^k)\| \leq \varepsilon$, dann beende den Algorithmus.

3. Berechne $d^k := -\frac{\nabla f(x^k)}{\|\nabla f(x^k)\|}$.

4. Liniensuche: Bestimme $t^k > 0$, so dass

$$f(x^k + t^k d^k) = \min_{t>0} f(x^k + t d^k).$$

5. Setze $x^{k+1} = x^k + t^k d^k$.

6. Setze $k := k + 1$ und setze den Algorithmus bei Schritt 1 fort.

Interpretation zur Wahl von d^k und t^k: Wie oben bereits erwähnt, löst $d^k := -\frac{\nabla f(x^k)}{\|\nabla f(x^k)\|}$ die Aufgabe $\min_d f(x^k) + \nabla f(x^k)^T d$ unter der Nebenbedingung $\|d\| = 1$. Unsere Richtung minimiert also eine lineare Approximation an die Funktion. Dies sieht man an der Taylorreihe

$$f(x) = f(x^k) + \nabla f(x^k) \underbrace{(x - x^k)}_{t^k d^k} + O(\|x - x^k\|^2).$$

Die Schrittweite t^k lässt sich als Methode zur Bestimmung des Gültigkeitsbereiches der Linearisierung interpretieren. Die Gradientenrichtung d^k minimiert die lineare Approximation, so dass die Linearisierung eine gute Approximation von $f(x)$ in $x = x^k$ ist.

3.3 Globale Konvergenz

Definition 3.4 *Die Menge*

$$N(\hat{x}) := \{x \in \mathbb{R}^n \mid f(x) \leq f(\hat{x})\} \subset D$$

heißt Niveaumenge *zum Punkt $\hat{x}$.*

Satz 3.5 (Globale Konvergenz) *Sei $x^0 \in D \subset \mathbb{R}^n$ beliebig und die Niveaumenge $N(x^0)$ sei kompakt. Dann existiert ein Häufungspunkt $x^\star$ der Folge $(x^k)_k$ mit $\nabla f(x^\star) = 0$ oder es ist $\nabla f(x^k) = 0$ für ein k.*

Beweis: *Annahme:* $\nabla f(x^k) \neq 0$ für alle k. Es gilt $f(x^{k+1}) < f(x^k)$ für alle k, also ist $\{(x^k)_k\} \subset N(x^0)$. Also existiert ein Häufungspunkt $x^\star \in N(x^0)$, und da $N(x^0)$ kompakt ist, existiert eine konvergente Teilfolge $x^{k'} \to x^\star$ mit

$$f(x^\star) \leq f(x^{k'+1}) \leq f\left(x^{k'} - t^{k'}\left(\frac{\nabla f(x^{k'})}{\|\nabla f(x^{k'})\|_2}\right)\right) < f(x^{k'}). \quad (*)$$

Annahme: $\nabla f(x^\star) \neq 0$. Dann existiert ein $t^\star > 0$ und $f(x^\star + t^\star d^\star) = f\left(x^\star - t^\star \frac{\nabla f(x^\star)}{\|\nabla f(x^\star)\|_2}\right) \leq f(x^\star) - \delta$ mit $\delta > 0$. Da aber f und ∇f stetig sind, gilt auch

$$f(x^{k'} + t^\star d^{k'}) \to f(x^\star + t^\star d^\star) \iff f\Big(\underbrace{x^{k'}}_{\to x^\star} - t^\star \underbrace{\frac{\nabla f(x^{k'})}{\|\nabla f(x^{k'})\|_2}}_{\to \frac{\nabla f(x^\star)}{\|\nabla f(x^\star)\|_2}}\Big) \leq f(x^\star) - \frac{\delta}{2} \quad \forall k' \geq \bar{k},$$

also für genügend große k. Dann kann das t^k aus der Ungleichung $(*)$ nicht optimal im Sinne der Liniensuche sein, denn $f(x^{k'} + t^\star d^{k'}) < f(x^\star)$ und dies ist ein Widerspruch dazu, dass t^k das Optimum ist. Das Gradientenverfahren konvergiert also global für einen beliebigen Startwert. ■

Lemma 3.6 *Für das Gradientenverfahren mit exakter Liniensuche gilt $d^{k+1} \perp d^k$, falls das gefundene Minimum im Inneren des Intervalls $[0, \alpha[$ liegt.*

In der Praxis konvergiert das Verfahren gut, wenn x^k noch weit von der Lösung entfernt ist. Aus der Liniensuche folgt aber $d^{k+1} \perp d^k$. Dies kann zu einer langsamen Zickzack-Konvergenz in der

Nähe des Minimums führen. Eine genauere Betrachtung in der Nähe des Minimums zeigt, dass f für den Algorithmus "quadratisch" aussieht:

$$f(x) = f(x^\star) + \frac{1}{2}(x - x^\star)^T \underbrace{\nabla^2 f(x^\star)}_{\text{pos.def}}(x - x^\star) + \dots$$

Es reicht also, für die Asymptotik folgendes Problem zu betrachten:

$$\min \quad \frac{1}{2}x^T Ax, \quad \text{mit } A \text{ positiv definit.}$$

Falls wir auf den Halbachsen starten, haben wir in einem Schritt Konvergenz.

Lemma 3.7 *Bei exakter Liniensuche gilt für $f_{QP}(x) = \frac{1}{2}x^T Ax$ und $g_k := \nabla f(x^k) = Ax^k$:*

$$x^{k+1} = x^k - \frac{g_k^T g_k}{g_k^T A g_k} g_k,$$

$$f(x^{k+1}) = \left(1 - \frac{(g_k^T g_k)^2}{(g_k^T A g_k)(g_k^T A^{-1} g_k)}\right) f(x^k).$$

Beweis: Gegeben sei $f(x) = \frac{1}{2}x^T Ax$, wobei A symmetrisch und positiv definit ist. Somit ist $\nabla f(x) = Ax$. Bei der exakten Liniensuche muss nun gelten:

$$0 = \nabla^T f(x^k + t^k d^k) d^k = (A(x^k + t^k d^k))^T d^k = (\underbrace{(x^k)^T A}_{=\nabla^T f(x^k) = g_k^T} + t^k (d^k)^T A) d^k = g_k^T d^k + t^k (d^k)^T A d^k$$

$$\iff \quad t^k = -\frac{g_k^T d^k}{(d^k)^T A d^k}.$$

Mit $x^{k+1} = x^k + t^k d^k$, Ersetzen von $d^k = -\frac{g_k}{\|g_k\|}$ und Kürzen der Normen folgt schließlich

$$x^{k+1} = x^k - \frac{g_k^T d^k}{(d^k)^T A d^k} d^k = x^k - \frac{g_k^T \left(-\frac{g_k}{\|g_k\|}\right)}{\left(-\frac{g_k}{\|g_k\|}\right)^T A \left(-\frac{g_k}{\|g_k\|}\right)} \left(-\frac{g_k}{\|g_k\|}\right) = x^k - \frac{g_k^T g_k}{g_k^T A g_k} g_k \ .$$

Setze nun $\alpha := \frac{g_k^T g_k}{g_k^T A g_k}$. Weiter lässt sich, da A invertierbar (positiv definit) ist, schreiben: $x^k = A^{-1} g_k$. Es folgt:

$$f(x^{k+1}) = \frac{1}{2}(x^{k+1})^T A x^{k+1} = \frac{1}{2}((x^k)^T - \alpha g_k^T) A (x^k - \alpha g_k)$$

$$= \frac{1}{2}((x^k)^T A x^k - \alpha (x^k)^T A g_k - \alpha g_k^T A x^k + \alpha^2 g_k^T A g_k^T) = f(x^k) - \alpha \underbrace{(x^k)^T A}_{g_k^T} g_k + \frac{1}{2}\alpha^2 g_k^T A g_k$$

$$= f(x^k) - \frac{(g_k^T g_k)^2}{g_k^T A g_k} + \frac{1}{2}\frac{(g_k^T g_k)^2}{g_k^T A g_k} = f(x^k) - \frac{1}{2}\frac{(g_k^T g_k)^2}{g_k^T A g_k} \underbrace{(x^k)^T A}_{=g_k^T} x^k = f(x^k)\left(1 - \frac{(g_k^T g_k)^2}{g_k^T A g_k \ g_k^T A^{-1} g_k}\right)$$

■

Satz 3.8 (Ungleichung von Kantorovich) *Sei A positiv definit. Dann gilt für alle $x \in \mathbb{R}^n$:*

$$\frac{(x^T x)^2}{(x^T Ax)(x^T A^{-1} x)} \geq 4\frac{\lambda_{\min}\lambda_{\max}}{(\lambda_{\min}+\lambda_{\max})^2} \geq \frac{\lambda_{\min}}{\lambda_{\max}},$$

wobei $0 < \lambda_{\min} \leq \cdots \leq \lambda_{\max}$ der kleinste und der größte Eigenwert von A sind.

Beweis: Sei $y = \mathbb{R}^n$ beliebig mit $\|y\|_2 = 1$. Nach dem Spektralsatz gilt $A = U^T \Lambda U$, wobei $\Lambda = diag(\lambda_i)$. Ferner ist $U = \{u_1, \ldots, u_n\}$ eine Orthonormalbasis von Eigenvektoren zu Eigenwerten $\lambda_1 \leq \ldots \leq \lambda_n$ von A. Es existieren α_i, so dass $y = \sum_{i=1}^n \alpha_i u_i$ für alle y. Für y gilt dann:

$$1 = y^T y = (\sum_{i=1}^n \alpha_i u_i)^T \sum_{i=1}^n \alpha_i u_i) = \sum_{i=1}^n \alpha_i^2$$

Betrachte nun die Funktionen

$$q(\lambda) := \frac{1}{\lambda}, \quad Q(\lambda) := \frac{\lambda_{\min}+\lambda_{\max}-\lambda}{\lambda_{\min}\lambda_{\max}}.$$

Für alle $\bar{\lambda} \in [\lambda_{\min}, \lambda_{\max}]$ gilt nun wegen der Konvergenz von $q(\lambda)$

$$q(\bar{\lambda}) \leq \sum_{i=1}^n \alpha_i^2 q(\lambda_i) \leq Q(\bar{\lambda}).$$

Setze nun $\bar{\lambda} := \frac{1}{2}(\lambda_{\min}+\lambda_{\max})$. Hieraus folgt:

$$y^T Ay = (\sum_{i=1}^n \alpha_i u_i)^T Ay = \sum_{i=1}^n \alpha_i u_i^T Ay = \sum_{i=1}^n \alpha_i \lambda_i u_i^T y = \sum_{i=1}^n \alpha_i \lambda_i \underbrace{u_i^T \left(\sum_{j=1}^n \alpha_j u_j\right)}_{=0 \text{ für } i \neq j} = \sum_{i=1}^n \alpha_i^2 \lambda_i.$$

Außerdem ist $y^T A^{-1} y = \sum_{i=1}^n \frac{1}{\lambda_i}\alpha_i^2$, wobei $\frac{1}{\lambda_i}$ Eigenwert der Inversen von A ist. Sei nun $\bar{\lambda} = \sum_{i=1}^n \lambda_i \alpha_i^2$, dann folgt:

$$\frac{y^T y}{(y^T Ay)(y^T A^{-1} y)} = \frac{1}{\sum_{i=1}^n \lambda_i \alpha_i^2 \cdot \sum_{i=1}^n \frac{1}{\lambda_i}\alpha_i^2} = \frac{q\left(\sum_{i=1}^n \lambda_i \alpha_i^2\right)}{\sum_{i=1}^n q(\lambda_i)\alpha_i^2} \geq \frac{q(\bar{\lambda})}{Q(\bar{\lambda})} \geq \min_{\lambda_1 \leq \lambda \leq \lambda_n} \frac{q(\lambda)}{Q(\lambda)},$$

$$\frac{q(\lambda)}{Q(\lambda)} = \frac{\lambda_{\min}\cdot\lambda_{\max}}{\lambda(\lambda_{\min}+\lambda_{\max}-\lambda)}.$$

Betrachtet man nun den Gradienten von $\frac{q(\lambda)}{Q(\lambda)}$ und setzt diesen Null, so ergibt sich:

$$\left(\frac{q(\lambda)}{Q(\lambda)}\right)' = -\lambda_{\min}\lambda_{\max}\frac{\lambda_{\min}+\lambda_{\max}-2\lambda}{(\lambda(\lambda_{\min}+\lambda_{\max})-\lambda)^2} = 0 \quad \Longrightarrow \quad \lambda_{\min}+\lambda_{\max}-2\lambda = 0$$

$$\Longrightarrow \quad \lambda^\star = \frac{\lambda_{\min}+\lambda_{\max}}{2}$$

Nun kann man die lange Ungleichungskette von oben fortsetzen:

$$\min_{\lambda_1 \leq \lambda \leq \lambda_n} \frac{q(\lambda)}{Q(\lambda)} = \frac{2}{\lambda_{\min}+\lambda_{\max}} \cdot \frac{\lambda_{\min}\lambda_{\max}}{\left(\frac{\lambda_{\min}+\lambda_{\max}}{2}\right)} = 4\frac{\lambda_{\min}\lambda_{\max}}{(\lambda_{\min}+\lambda_{\max})^2} = \frac{\lambda_{\min}\lambda_{\max}}{(\lambda^\star)^2} \geq \frac{\lambda_{\min}\lambda_{\max}}{\lambda_{\max}^2} = \frac{\lambda_{\min}}{\lambda_{\max}}$$

■

Satz 3.9 *Die Ungleichung von Kontrovich ist scharf. Das heißt es existiert ein x, s.d.*

$$\frac{(x^Tx)^2}{(x^TAx)(x^TA^{-1}x)} = \frac{f\lambda_{min}\cdot\lambda_{max}}{(\lambda_{min}+\lambda_{max})^2}$$

Beweis: $0 < \lambda_1 \leq ... \leq \lambda_n$ Eigenwerte von A.
$u_1, ..., u_n$ Eigenvektoren von A, $u_i^T u_j = \delta_{ij}$
Wähle $x = u_1 + u_n$. Dann ist $(x^Tx) = (u_1+u_n)^T(u_1+u_n) = 1 + 2\cdot 0 + 1 = 2$
$x^TAx = (u_1+u_n)^TA(u_1+u_n) = (u_1+u_n)^T(\lambda_1u_1+\lambda_nu_n) = \lambda_1+\lambda_n$
$x^TA^{-1}x = \frac{1}{\lambda_1}+\frac{1}{\lambda_n}$
Also:

$$\frac{(x^Tx)^2}{(x^TAx)(x^TA^{-1}x)} = \frac{4}{(\lambda_1+\lambda_n)(\frac{1}{\lambda_1}+\frac{1}{\lambda_n})} = \frac{4\lambda_1\cdot\lambda_n}{(\lambda_1+\lambda_n)^2}$$

■

Satz 3.10 (Konvergenz des Gradientenverfahrens) *Das Gradientenverfahren konvergiert linear,*

$$f(x^{k+1}) \leq \left(\frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}}\right)^2 f(x^k),$$

und es gilt

$$\|x^{k+1}\|_A := \|A^{\frac{1}{2}}x^{k+1}\|_2 \leq \left(\frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}}\right)\|x^k\|_A.$$

Falls also $\lambda_{\min} \approx \lambda_{\max}$ gilt, so ist die Konvergenzrate κ viel kleiner als 1 und das Verfahren konvergiert sehr schnell, ansonsten ziemlich langsam:

$$\kappa(A) = \frac{\lambda_{\max}}{\lambda_{\min}}, \quad \frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}} = \frac{\kappa(A)-1}{\kappa(A)+1} = 1 - \frac{2}{\kappa(A)+1}.$$

κ ist dabei die Konditionszahl von A.

Beweis:

$$f(x^{k+1}) \overset{\text{Satz 3.8}}{\leq} \left(1 - \frac{4\lambda_{\max}\lambda_{\min}}{(\lambda_{\max}+\lambda_{\min})^2}\right) f(x^k) = \frac{(\lambda_{\max}-\lambda_{\min})^2}{(\lambda_{\max}+\lambda_{\min})^2} f(x^k)$$

$$f(x^{k+1}) = \frac{1}{2}(x^{k+1})^TAx^{k+1} = \frac{1}{2}(x^{k+1})^TA^{\frac{1}{2}}A^{\frac{1}{2}}x^{k+1} = \frac{1}{2}\left(A^{\frac{1}{2}}x^{k+1}\right)^TA^{\frac{1}{2}}x^{k+1} = \frac{1}{2}\|x^{k+1}\|_A^2$$

■

Bemerkung: $\nabla f(x^\star) = 0 \quad \Longrightarrow \quad f(x) = f(x^\star) + \frac{1}{2}(x-x^\star)\nabla^2 f(x^\star)(x-x^\star) + \ldots.$
Die Aussage gilt asymptotisch für allgemeine nichtlineare f, wenn man für $\lambda_{\min}$ bzw. $\lambda_{\max}$ den kleinsten bzw. größten Eigenwert der Hessematrix annimmt.

Satz 3.11 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar$\{x_k\} \to x^\star$ und $\nabla^2 f(x^\star)$ positiv definit. Wir betrachten man das Gradientenverfahren mit exakter Liniensuche. Sind $\lambda_{\min}, \lambda_{\max}$ der kleinste bzw. größte Eigenwert der Hessematrix $\nabla^2 f(x^\star)$, dann ist*

$$f(x_{k+1}) - f(x^\star) \leq \left(\frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}}\right)^2 (f(x_k) - f(x^\star)).$$

Auf einen ausführlichen Beweis wurde an dieser Stelle verzichtet, die Idee ist die Taylorreihe der Funktion f. Das Gradientenverfahren ist global konvergent, falls f in einer Umgebung von $x^\star$ konvex ist (also die Hessematrix dort positiv definit ist). Es konvergiert aber im Allgemeinen sehr langsam.

3.4 Das Newton-Verfahren

Das Newtonverfahren löst das Nullstellenproblem $F(x) = 0$, wobei $F(x) : \mathbb{R}^n \to \mathbb{R}^n$. Die Motivation ist, die Funktion $f(x)$ zu minimieren. In lokalen Extrempunkten muss der Gradient der Funktion verschwinden, deshalb sucht man mit dem Newton-Algorithmus die Nullstelle des Gradienten. Findet man hierfür eine Lösung $x^\star$, so ist $F(x^\star) = \nabla f(x^\star) = 0$. Man betrachtet die Linearisierung

$$F(x^{k+1}) \doteq F(x^k) + \frac{dF}{dx}(x^k)(x^{k+1} - x^k)$$

und setzt diese gleich Null. Damit erhält man

$$x^{k+1} = x^k \underbrace{- \left(\frac{dF}{dx}(x^k)\right)^{-1} F(x^k)}_{=:\Delta x^k}.$$

Der Wert $x^{k+1} = x^k + \Delta x^k$ sollte dann eine bessere Approximation für eine Nullstelle von F sein.

Lokaler Newton-Algorithmus

Gegeben sei folgender Dateninput: $x^0 \in \mathbb{R}^n$ und $\varepsilon \geq 0$. Setzt $k = 0$.

1.) Ist $\|F(x^\star)\| \leq \varepsilon$, dann breche den Algorithmus ab.

2.) Bestimme d^k aus $J(x^k)d^k = -F(x^k)$.

3.) Setze $x^{k+1} = x^k + d^k$ und $k = k + 1$ und setze den Algorithmus bei Schritt 1 fort.

Im obigen Algorithmus bezeichnet $J(x)$ die Jacobi-Matrix von F, $J(x) = \frac{d}{dx}F(x)$.
Das Newton-Verfahren selber hat einige Varianten, die in den folgenden Unterkapiteln genauer erläutert werden sollen.

3.4.1 Newton-ähnliche Verfahren

Den Beginn macht das Newton-ähnliche Verfahren. Hierbei ist

$$x^{k+1} = x^k - M(x^k)f(x^k) = x^k + \Delta x^k.$$

Die Matrix $M(x)$ bezeichnet in diesem Abschnitt eine näherungsweise Inverse oder verallgemeinerte Inverse der Jacobi-Matrix $J(x) = \frac{df}{dx}(x)$, d.h. $M(x) \approx J(x)^{-1}$.

Satz 3.12 (Lokaler Kontraktionssatz) *Sei $F : D \to \mathbb{R}^n$, $D \subset \mathbb{R}^n, F \in \mathcal{C}^1(D)$. Für alle $x, y \in D$, $t \in]0, 1]$ und $y - x = -M(x)f(x) = \Delta x$ gelte:*

1. *Lipschitz-Bedingung an J:*

$$\frac{\|M(y)\left(J(x + t(y - x)) - J(x)\right)(y - x)\|}{t\|y - x\|^2} \leq \bar{\omega}(x) \leq \omega < \infty.$$

2. *Güte der Iterationsmatrix M:*

$$\frac{\|M(y)R(x)\|}{\|y-x\|} \leq \bar{\kappa}(x) < \kappa < 1.$$

mit Residuum $R(x) := F(x) + J(x)\left(-M(x)F(x)\right) = \left(Id - J(x)M(x)\right)F(x)$.

Ist $x^0 \in D$ und gilt

3. $\delta_0 := \kappa + \frac{\omega}{2}\|\Delta x^0\| < 1, \quad \delta_k := \kappa + \frac{\omega}{2}\|\Delta x^k\|$,

4. $D_0 := \bar{K}\left(x_0; \frac{\|\Delta x^0\|}{1-\delta_0}\right) \subset D$,

wobei $\bar{K}$ eine Kugel ist mit $\bar{K}$ =(Mittelpunkt, Radius), dann folgt:

1. *Wohldefiniertheit: (x_k) ist wohldefiniert und bleibt in D_0,*

2. *Konvergenz: Es existiert $x^\star \in D_0$ und $x^k \to x^\star$ für $k \to \infty$,*

3. *A-priori-Abschätzung:*

$$\|x^{k+j} - x^\star\| \leq \frac{\delta_k^j}{1-\delta_k}\|\Delta x^k\| \leq \frac{\delta_0^{k+j}}{1-\delta_0}\|\Delta x^0\|,$$

4. *Konvergenzrate:*

$$\|\Delta x^{k+1}\| \leq \delta_k\|\Delta x^k\| = \left(\kappa + \frac{w}{2}\|\Delta x^k\|\right)\|\Delta x^k\| = \kappa\|\Delta x^k\| + \frac{\omega}{2}\|\Delta x^k\|^2;$$

5. *Falls $M(x)$ stetig und regulär ist, dann ist der Fixpunkt $x^\star$ eine Nullstelle von F, d.h. $F(x^\star) = 0$.*

Beweis: Um die Schreibweise im folgenden Beweis einfacher zu gestalten, gelten folgende Bezeichnungen: $M^k := M(x^k), \quad J^k := J(x^k), \quad F^k := F(x^k), \quad R^k := R(x^k)$.

1. Es gilt $x^0 \in D_0, x^1 \in D_0$.

2. Seien $x^k, x^{k+1} \in D_0 \subset D$. Betrachte folgende Umformung:

$$S(t) := F(x^k + t\Delta x^k), \quad S'(t) = J(x^k + t\Delta x^k)\Delta x^k, \quad \int_0^1 S'(t)\,dt = S(1) - S(0) = F^{k+1} - F^k.$$

Dann lässt sich abschätzen:

$$\begin{aligned}
\|\Delta x^{k+1}\| = \|M^{k+1}F^{k+1}\| &= \|M^{k+1}\left(F^{k+1} - F^k - J^k\Delta x^k\right) + M^{k+1}R^k\| \\
&\leq \left\|M^{k+1}\left(\int_0^1 J\left(x^k + t\Delta x^k\right)\Delta x^k\ dt - J^k\Delta x^k\right)\right\| + \|M^{k+1}R^k\| \\
&\leq \left\|M^{k+1}\int_0^1 \|J\left(x^k + t\Delta x^k\right) - J(x^k)\|\Delta x^k\ dt\right\| + \kappa\|\Delta x^k\| \\
&\leq \int_0^1 \left\|M^{k+1}\left(J\left(x^k + t\Delta x^k\right) - J(x^k)\right)\Delta x^k\right\|\ dt + \kappa\|\Delta x^k\| \\
&\leq \int_0^1 t\|\Delta x^k\|^2\omega\ dt + \kappa\|\Delta x^k\| = \|\Delta x^k\|^2\omega\int_0^1 t\ dt + \kappa\|\Delta x^k\| \\
&= \frac{\omega}{2}\|\Delta x^k\|^2 + \kappa\|\Delta x^k\| = \delta_k\|\Delta x^k\|
\end{aligned}$$

3. Die Folge (δ_k) fällt monoton:

$$\delta_1 \;=\; \frac{1}{2}\omega\|\Delta x^1\| + \kappa \quad \leq \quad \frac{1}{2}\omega\|\Delta x^0\| + \kappa \;=\; \delta_0$$

$$\Longrightarrow \quad \delta_k < \delta_{k-1} < \ldots < \delta_1 < \delta_0$$

4. Die Folge $\|\Delta x^k\|$ fällt monoton:

$$\begin{aligned}\|\Delta x^{k+j}\| \leq\; \delta_{k+j-1}\|\Delta x^{k+j-1}\| \;\leq\; \delta_{k+j-1}\cdot\ldots\cdot\delta_k\|\Delta x^k\| \;\leq\; \delta_k^j\|\Delta x^k\| \\ \leq\; \delta_k^j\delta_0^k\|\Delta x^0\| \;\leq\; \delta_0^{k+j}\|\Delta x^0\|\end{aligned}$$

5. Es gilt $x^{k+1} \in D_0$:

$$\|x^{k+1}-x^0\| = \|\Delta x^k + \Delta x^{k-1} + \ldots + \Delta x^0\| \leq \sum_{j=0}^{k}\|\Delta x^j\| \leq \left(\delta_0^k + \delta_0^{k-1} + \ldots + 1\right)\|\Delta x^0\|$$

$$\leq \frac{1}{1-\delta_0}\|\Delta x^0\| \quad \Longrightarrow \quad x^{k+1} \in D_0$$

6. Die Folge (x^k) ist eine Cauchy-Folge:

$$\begin{aligned}\|x^{i+k+1}-x^i\| &= \|x^{i+k+1}-x^{i+k}+x^{i+k}-x^{i+k-1}\cdots+x^{i+1}-x^i\| = \|\Delta x^{i+k}+\ldots+\Delta x^i\| \\ &= \sum_{p=0}^{k}\|\Delta x^{i+p}\| \quad < \quad \sum_{p=0}^{k}\delta_i^p\|\Delta x^i\| \leq \quad \frac{1}{1-\delta_i}\|\Delta x^i\| \quad \leq \quad \frac{\delta_0^i}{1-\delta_0}\|\Delta x^0\|, \quad \forall\; k \\ &\to 0 \text{ falls } \delta_k < 1 \;\forall\; k, \text{ und dies ist der Fall.}\end{aligned}$$

7. Falls $M(x)$ stetig und regulär ist, so ist der Fixpunkt eine Nullstelle von F: Bewiesen wurde bereits $M(x^k)F(x^k) \to 0$. Ist $M(x)$ stetig, so folgt $M(x^\star)F(x^\star) = 0$. Ist M regulär, so muss danach $F(x^\star) = 0$ gelten. ■

Bemerkung: In der Nullstelle gilt $\Delta x^\star = 0$, also ist $\Delta x^\star$ sehr klein in einer Umgebung von $x^\star$. Deshalb lassen sich die Voraussetzungen 3. und 4. in einer kleinen Umgebung einer Lösung $x^\star$ leicht erfüllen. Wir erhalten also lokale Konvergenz. Außerdem lässt sich die zweite Bedingung umformen zu

$$\frac{\|M(y)R(x)\|}{\|y-x\|} = \frac{\|M(y)\left(M^{-1}(x)-J(x)\right)M(x)F(x)\|}{\|y-x\|} \leq \|M(y)\left(M^{-1}(x)-J(x)\right)\| \leq \kappa < 1.$$

Das exakte Newton-Verfahren liefert also $\kappa = 0$ wegen $M^{-1} = J$.

3.4.2 Reines Newton-Verfahren

In diesem Abschnitt wird nun das reine Newton-Verfahren beschrieben. Dieses ist exakt, denn es gilt $M(x) = J(x)^{-1}$.

Satz 3.13 (Newton-Mysovski) *Falls die beiden folgenden Bedingungen gelten,*

(i) Es existiert ein $J^{-1}(y)$ mit $\|J^{-1}(y)\| \leq \beta < \infty$ auf D, (was zum Beispiel erfüllt ist, falls J stetig differenzierbar und regulär auf einem kompakten D ist),

(ii) $\|J(z) - J(w)\| \leq \gamma\|z - w\|$ mit $\gamma < \infty$, (was zum Beispiel erfüllt ist, falls J stetig differenzierbar auf einem kompakten und konvexen D ist),

dann existiert ein $\omega < \infty$.

Beweis:

$$\begin{aligned}\omega &= \frac{\|J^{-1}(y)\,(J(x + t(y - x)) - J(x))\,(y - x)\|}{t\|y - x\|^2} \\ &\leq \frac{\|J^{-1}(y)\| \cdot \|J(x + t(y - x)) - J(x)\| \cdot \|y - x\|}{t\|y - x\|^2} \leq \frac{\beta\,\gamma\,t\|y - x\|}{t\|y - x\|} = \beta\,\gamma < \infty\end{aligned}$$

■

Tatsächlich ist $\omega \ll \beta\,\gamma$, die Abschätzung ist zu schwach. Somit sind die 1. und 2. Bedingung des Kontraktionssatzes (Satz 3.12) erfüllt, und das Newton-Verfahren konvergiert, falls x_0 nahe an der Lösung ist, d.h. falls $\delta_0 < 1$ ist.
Wie schon oben angesprochen gilt im exakten Newton-Verfahren

$$\delta_k = \frac{\omega}{2}\|\Delta x^k\| \quad \text{und damit} \quad \|\Delta x^{k+1}\| \leq \frac{\omega}{2}\|\Delta x^k\|^2.$$

Diese Ungleichung führt direkt weiter zum folgenden Lemma.

Lemma 3.14 (Quadratische Konvergenz des reinen Newton-Verfahrens) *Das reine Newton-Verfahren konvergiert quadratisch, d.h. es existiert ein $\hat{k} \in \mathbb{N}$ und $c > 0$ mit*

$$\|x^{k+1} - x^\star\| \leq c\|x^k - x^\star\|^2 \quad \forall\ k \geq \hat{k}.$$

Beweis: Es existiert ein $\hat{k} \in \mathbb{N}$ mit $\delta_k = \frac{\omega}{2}\|\Delta x^k\| \leq \frac{1}{4} \quad \forall\ k \geq \hat{k}$. Da die Folge δ_k monoton fällt, fällt die Folge $\{\frac{\delta_k}{1-\delta_k}\}_k$ ebenfalls monoton für $k \to \infty$, also folgt

$$\frac{\delta_k}{1 - \delta_k} \leq \frac{\frac{1}{4}}{1 - \frac{1}{4}} = \frac{1}{3} \qquad \forall\ k \geq \hat{k}.$$

Weiter ist

$$\begin{aligned}\|\Delta x^k\| = \|x^{k+1} - x^k\| &= \|x^{k+1} - x^\star + x^\star - x^k\| \leq \|x^{k+1} - x^\star\| + \|x^k - x^\star\| \\ &\overset{3.12}{\leq} \frac{\|\Delta x^k\|\delta_k}{1 - \delta_k} + \|x^k - x^\star\| \leq \frac{1}{3}\|\Delta x^k\| + \|x^k - x^\star\| \qquad \Longrightarrow \qquad \frac{2}{3}\|\Delta x^k\| \leq \|x^k -^\star x\| \quad (*)\end{aligned}$$

Somit gilt aber außerdem

$$\begin{aligned}\|x^{k+1} - x^\star\| &\overset{3.12}{\leq} \frac{1}{1 - \delta_{k+1}}\|\Delta x^{k+1}\| \leq \frac{4}{3}\frac{\omega}{2}\|\Delta x^k\|^2 \\ &\overset{(*)}{\leq} \frac{4}{3}\frac{\omega}{2}\frac{9}{4}\|x^k - x^\star\|^2 = \frac{3}{2}\omega\|x^k - x^\star\|^2 \qquad \Longrightarrow \quad c := \frac{3}{2}\omega\end{aligned}$$

■

Anwendung auf Optimierungsprobleme

Die zu optimierende Funktion sei f. Wie oben schon erwähnt wird das Newton-Verfahren dann auf ∇f angewandt, um Lösungen zu finden, welche die NOB 1.Ordnung erfüllen. Gesucht wird also eine Nullstelle von $\nabla f(x)$.

1. Unbeschränkter Fall: Wähle $F(x) = \nabla f(x)$ und $J(x) = \frac{dF}{dx} = \nabla^2 f(x)$. Die Iteration sieht dann wie folgt aus:

$$x^{k+1} = x^k + t^k \Delta x^k \quad \text{mit} \quad \Delta x^k = -\left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k).$$

2. Beschränkter Fall: Gegeben ist das Problem $\min_{x\in\mathbb{R}^n} f(x)$ unter $g(x) = 0$, wobei $g(x) : \mathbb{R}^n \to \mathbb{R}^m, m < n$. Nach der NOB 1. Ordnung existiert ein $\lambda \in \mathbb{R}^m$ mit $\nabla_x L(x, \lambda) = 0$ und $g(x) = 0$. Für die Lagrangefunktion ergibt sich: $L(x, \lambda) = f(x) - \lambda^T g(x)$. Die Funktion für den Newton-Algorithmus ist nun:

$$F(x) = \begin{pmatrix} \nabla_x L(x, \lambda) \\ g(x) \end{pmatrix}.$$

   Die Iteration sieht dann wie folgt aus:

$$\begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \lambda^k \end{pmatrix} + \begin{pmatrix} \Delta x^k \\ \Delta\lambda^k \end{pmatrix} \quad \text{mit} \quad \begin{pmatrix} \Delta x^k \\ \Delta\lambda^k \end{pmatrix} = \begin{pmatrix} \nabla_x^2 L(x, \lambda) & -\nabla^T g(x) \\ -\nabla g(x) & 0 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \nabla L(x, \lambda) \\ g(x) \end{pmatrix}.$$

Es bieten sich nun andere Interpretationen der Newton-Iteration an: Beim unbeschränkten Fall ist Δx^k die Lösung des Minimierungsproblems

$$\min f_{QP}(\nabla x) = f(x^k) + \nabla f(x^k)^T \Delta x + \frac{1}{2} \Delta x^T \nabla^2 f(x^k) \Delta x,$$

falls $\nabla^2 \tilde{f}(x)$ positiv definit ist. Denn die notwendige Bedingung lautet

$$\frac{df_{QP}}{d\Delta x}(\Delta x) = \nabla f(x^k)^T + \Delta x^T \nabla^2 f(x^k) = 0 \quad \Rightarrow \quad \nabla f(x^k) = -\nabla^2 f(x^k) \Delta x,$$

d.h. Δx minimiert die quadratische Approximation von f. Vergleiche dazu das Gradientenverfahren, welches mit linearer Approximation arbeitet.
Im beschränkten Fall löst Δx^k das Minimierungsproblem

$$\min_{p\in\mathbb{R}^n} \nabla^T f(x^k) p + \frac{1}{2} p^T \nabla^2 L(x^k, \lambda^k) p$$

unter $g(x^k) + \nabla g(x^k) p = 0$. Dies ist das sogenannte SQP-Verfahren.

Lemma 3.15 *Falls der Gradient $\nabla f(x^k) \neq 0$ und die Hesse-Matrix $\nabla^2 f(x^k)$ positiv definit sind, so ist Δx^k eine Abstiegsrichtung für f.*

Beweis: Betrachte die Funktion $f((x^k + t\Delta x^k)$. Dann gilt:

$$\begin{aligned}\frac{df}{dt}(x^k + t\Delta x^k)\Big|_{t=0} &= \nabla^T f(x^k + t\Delta x^k) \cdot \Delta x^k = \nabla f(x^k)^T \left(-\nabla^2 f(x^k)^{-1} f(x^k)\right) \\ &= -\nabla f(x^k)^T \underbrace{\nabla^2 f(x^k)^{-1}}_{\text{pos. def.}} \nabla f(x^k) < 0\end{aligned}$$

■

Fazit: Falls die Hessematrix in einer geeigneten Umgebung eines Minimums positiv definit ist, so ist das Newton-Verfahren lokal quadratisch konvergent. Das Newton-Verfahren ist potentiell global konvergent, falls man Liniensuche mit Schrittweite $t^k \in]0, 1]$ einsetzt.

Problem: Starten wir in einer Umgebung eines Sattelpunkts oder Maximums, wo die Hessematrix nicht positiv definit ist, so konvergiert das Verfahren, wenn überhaupt, gegen diese kritischen Punkte, also den Sattelpunkt oder das Maximum. Das lässt sich auch bei geeigneter Wahl der t^k nicht verhindern. Die geeignete Wahl der t^k ist in diesem Fall überhaupt nicht möglich; es existiert keine Abstiegsrichtung, da $\nabla_x^2 f$ nicht positiv definit ist. Abhilfe schaffen hier zwei Varianten, die jeweils nur eine unterschiedliche Formulierung derselben Methode sind.

- **Levenberg-Markquardt**:

$$\min f(x^k) + \nabla f(x^k)^T \Delta x + \frac{1}{2}\Delta x^T \underbrace{\left(\nabla^2 f(x^k) + \lambda \mathrm{I}\right)}_{=:H(\lambda)} \Delta x$$

  wobei $\lambda \in \mathbb{R}$ so groß gewählt werden muss, dass $H(\lambda)$ positiv definit ist. Also

$$\Delta x^k = -\left(\nabla^2 f(x^k) + \lambda \mathrm{I}\right)^{-1} \nabla f(x^k).$$

  Für geeignete, hinreichend große λ geht Δx^k gegen die Gradientenrichtung. Asymptotisch geht $H(\lambda)$ allerdings gegen die Nullmatrix, λ darf also auch nicht zu groß gewählt werden. Für $\lambda = 0$ erhält man wieder das echte Newton-Verfahren. Für $\lambda \to \infty$, $H(\lambda) \to \lambda I$ folgt $H(\lambda)^{-1} \to \frac{1}{\lambda} I$ und somit geht wegen $\Delta x^k \to -\frac{1}{\lambda}\nabla f(x^k)$ das Verfahren in das Gradientenverfahren über.

- **Vertrauensgebiet (trust region)**:

$$\min f_Q = f(x^k) + \nabla f(x^k)^T p + \frac{1}{2} p^T \nabla^2 f(x^k) p$$

  unter der Einschränkung $\|p\| \leq \Delta$ (üblicherweise wird hier $\|p\|_2^2 \leq \Delta^2$ genommen). Zur richtigen Wahl des Vertrauensradius Δ^2 soll hier nur auf die weiterführende Literatur verwiesen werden. Der Wert Δ^2 beschreibt die Größe des Gebiets, in welchem f_Q eine gute Approximation von f ist. Für große Δ^2 erhält man wieder das echte Newton-Verfahren.

Nun soll noch gezeigt werden, dass beide Methoden wirklich nur unterschiedliche Formulierungen desselben Verfahrens sind.Für Levenberg-Markquardt gilt:

$$\begin{aligned} L(\Delta x, \lambda) &= f(x^k) + \nabla f(x^k)^T \Delta x + \frac{1}{2}\Delta x^T(\nabla^2 f(x^k) + \lambda I)\Delta x \\ &= f(x^k) + \nabla f(x^k)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(x^k)\Delta x + \frac{1}{2}\lambda \Delta x^T \Delta x \\ &= f(x^k) + \nabla f(x^k)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(x^k)\Delta x - \lambda\left(-\frac{1}{2}\|\Delta x\|^2\right) \end{aligned}$$

Für das Vertrauensgebiet gilt:

$$L(\Delta x, \lambda) = f(x^k) + \nabla f(x^k)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(x^k)\Delta x - \lambda\left(\Delta^2 - \|\Delta x\|^2\right)$$

Beide Ausdrücke stimmen in der 1. Ableitung bis auf eine Konstante überein. ■

3.4.3 Näherungsweises Newton-Verfahren

Das näherungsweise Newton-Verfahren (approximative Newton method) ist ähnlich zu dem bereits kennengelernten Newton-ähnlichen Verfahren. Der Unterschied besteht in erster Linie in der Änderung der Reihenfolge von Approximieren und Invertieren beim Bestimmen der Jacobi-Matrix. Gesucht ist also eine Nullstelle einer Funktion F, mit $F : \mathbb{R}^n \to \mathbb{R}^n$. Für die Iteration gilt:

$$\begin{aligned} x^{k+1} &= x^k - \Delta\, x = x^k - M(x^k)F(x^k) \\ M(x) &= B^{-1}(x) \quad \text{mit} \quad B(x) \approx J(x) \end{aligned}$$

Aus der zweiten Bedingung des Kontraktionssatzes (Satz 3.12) folgt die Bedingung

$$\frac{\|M(y)R(x)\|}{\|y - x\|} \leq \kappa < 1.$$

Das näherungsweise Newton-Verfahren konvergiert:
Für das Residuum $R(x)$ gilt

$$R(x) = F(x) - J(x)B^{-1}(x)F(x),$$

also

$$\begin{aligned} B^{-1}(y)R(x) &= B^{-1}(y)\left(B(x)B^{-1}(x)F(x) - J(x)B^{-1}(x)F(x)\right) \\ &= B^{-1}(y)\left(B(x) - J(x)\right)\underbrace{B^{-1}(x)F(x)}_{=\Delta\, x} \\ &= B^{-1}(y)(B(x) - J(x))(y - x), \end{aligned}$$

da $y - x = -B^{-1}(x)F(x)$. Zur zweiten Bedingung von Satz 3.12 folgt nun

$$\begin{aligned} & B^{-1}(y)R(x) = B^{-1}(y)\left(B(x) - J(x)\right)(x - y) \\ \Longrightarrow \quad & \frac{\|B^{-1}(y)R(x)\|}{\|y - x\|} \leq \|B^{-1}(y)\left(B(x) - J(x)\right)\| \overset{!}{\leq} \kappa < 1. \end{aligned}$$

Falls nun aber $\|B^{-1}(y)\| \leq \beta$, also beschränkt ist und $\|B(x) - J(x)\| \leq \eta$ ebenfalls, dann folgt $\kappa \leq \eta\beta$ und $\kappa < 1$ für η genügend klein. Damit erhalten wir aus dem Kontraktionssatz folgendes Ergebnis:

Das Verfahren konvergiert linear, wenn $\delta_0 = \kappa + \frac{\omega}{2}\|\Delta x^0\| < 1$. ist. Die Konvergenzrate ist κ.

Varianten des näherungsweisen Newton-Verfahrens

- Vereinfachtes Newton-Verfahren: $B(x) = J(x^0)$. Es ist $\kappa < 1$, falls x^0 nahe genug an der Lösung $x^\star$ liegt.
- $B(x)$ ist konstant, bis
$$\frac{\|\Delta x^{k+1}\|}{\|\Delta x^k\|} \leq \delta_{\max} \quad \text{z.B.} \quad \delta_{\max} = \frac{1}{4}.$$
Dann ist $B(x) = J(x^k)$.

Satz 3.16 (Superlineare Konvergenz) *Es gelte neben den Voraussetzungen für den Kontraktionssatz (Satz 3.12) noch $B(x^k) \to J(x^k)$ für $k \to \infty$ und $\|B^{-1}(x^k)\| \leq \beta$, also*

$$\frac{\| \left(B(x^k) - J(x^k)\right) \Delta x^k\|}{\|\Delta x^k\|} \to 0 \quad \textit{für} \quad k \to \infty.$$

Dann konvergiert $(x^k)_k$ superlinear und es gilt für $k \to \infty$:

$$\frac{\|x^{k+1} - x^\star\|}{\|x^k - x^\star\|} \to 0.$$

Beweis: Definiere zuerst

$$\kappa^k := \frac{\|B^{-1}(x^{k+1}) \left(B(x^k) - J(x^k)\right)(x^{k+1} - x^k)\|}{\|x^{k+1} - x^k\|} \leq \beta \frac{(\|B(x^k) - J(x^k))\Delta x^k\|}{\|\Delta x^k\|} \to 0 \quad \text{für } k \to \infty$$

$$\|\Delta x^{k+1}\| \leq \underbrace{\left(\underbrace{\kappa^k}_{\to 0} + \frac{\omega}{2}\underbrace{\|\Delta x^k\|}_{\to 0}\right)}_{=\delta_k \to 0} \|\Delta x^k\| \quad \text{für } k \to \infty, \kappa \to 0.$$

Mit der Abschätzung aus dem lokalen Kontraktionssatz 3.12 folgt

$$\|x^{k+1} - x^\star\| \leq \frac{\|\Delta x^{k+1}\|}{1 - \delta_{k+1}} \leq \frac{\delta_k}{1 - \delta_{k+1}}\|\Delta x^k\|$$

$$\implies \quad \|\Delta x^k\| = \|x^{k+1} - x^k + x^\star - x^\star\| \leq \|x^k - x^\star\| + \|x^{k+1} - x^\star\| \leq \|x^k - x^\star\| + \frac{\delta_k}{1 - \delta_{k+1}}\|\Delta x^k\|$$

$$\begin{aligned} \implies \quad \left(1 - \tfrac{\delta_k}{1-\delta_{k+1}}\right)\|\Delta x^k\| &\leq \|x^k - x^\star\| \\ \implies \quad \|\Delta x^k\| &\leq \tfrac{1-\delta_{k+1}}{1-\delta_{k+1}-\delta_k}\|x^k - x^\star\| \end{aligned}$$

$$\implies \quad \|x^{k+1} - x^\star\| \leq \|x^k - x^\star\| \leq \frac{\delta_k}{1 - \delta_{k+1}} \frac{1 - \delta_{k+1}}{1 - \delta_{k+1} - \delta_k}\|x^k - x^\star\|$$

$$= \underbrace{\frac{\delta_k}{1 - \delta_{k+1} - \delta_k}}_{\to 0} \|x^k - x^\star\| \quad \text{für} \quad k \to \infty$$

und damit ist die superlineare Konvergenz gezeigt. ■

Nun soll noch kurz auf den Aspekt der Globalisierung eingegangen werden. Mit Globalisierung ist hier die Konvergenz für einen beliebigen Startwert x^0 mit der Liniensuche gemeint. Ziel ist die Minimierung einer Funktion f. Für das näherungsweise Newton-Verfahren mit der Iteration

$$x^{k+1} = x^k + t^k \Delta x^k \quad \text{wobei} \quad \Delta x^k = -H_k^{-1} \nabla f(x^k), \quad H_k \approx \nabla^2 f(x^k)$$

und t^k bestimmt aus der exakten Liniensuche

$$f(x^{k+1}) = \min_{t \in\,]0,\alpha]} f(x^k + t\Delta x^k)$$

gilt der folgende Satz:

Satz 3.17 (Allgemeiner Konvergenzsatz) *Sei $x^0 \in D$, die Niveaumenge $N(x^0) := \{x \in D : f(x) \leq f(x^0)\}$ sei kompakt und die Approximation der Hessematrix H_k sei positiv definit. Dann gilt:*
Es existiert entweder $\hat{k}$ mit $\nabla f(x^{\hat{k}}) = 0$ oder ein Häufungspunkt $x^\star \in N(x^0)$ und $\nabla f(x^\star) = 0$.

Beweis:

1. Da H_k positiv definit ist, folgt nach dem Lemma 3.15, dass $\delta_k = -H_k^{-1} \nabla f(x^k)$ eine Abstiegsrichtung ist:

$$\begin{aligned} \nabla f(x^k)^T \delta_k = -\nabla f(x^k)^T H_k^{-1} \nabla f(x^k) &< 0, \\ \|\nabla f(x^k)\| &\neq 0, \\ f(x^{k+1}) &< f(x^k), \quad x^k \in N(x^0). \end{aligned}$$

2. Aus $x^k \in N(x^0)$ folgt, dass eine Teilfolge $(x^{k_i}) \to x^\star \in N(x^0)$ existiert mit $f(x^{k_i}) \geq f(x^\star)$.

3. Es existiert eine Teilfolge (x^{k_j}) von (x^{k_i}) mit $H_{k_j} \to H^\star$ und $H^\star$ positiv definit. Da H_k positiv definit ist, folgt, dass ε und β existieren mit $0 < \varepsilon \leq \|H_k\| \leq \beta < \infty$.

4. *Annahme:* Sei $\nabla f(x^\star) \neq 0$. Dann folgt: Es existieren $t^\star > 0$ und $\delta > 0$ mit

$$f\left(x^\star - t^\star H^{\star -1} \nabla f(x^\star)\right) \leq f(x^\star) - \delta.$$

   Außerdem ist

$$\begin{aligned} x^{k_{j+1}} &= x^{k_j} - t^{k_j} H_{k_j}^{-1} \nabla f(x^{k_j}), \\ f(x^\star) &\leq f(x^{k_{j+1}}), \\ \tilde{x}^{k_{j+1}} &= \underbrace{x^{k_j}}_{\to x^\star} - t^\star \underbrace{H_{k_j}^{-1}}_{\to H^{\star -1}} \underbrace{\nabla f(x^{k_j})}_{\to \nabla f(x^\star)}. \end{aligned}$$

   Es existiert j_0 mit $f(\tilde{x}^{k_{j+1}}) < f(x^\star) - \frac{\delta}{2}$ für alle $j \geq j_0$, also folgt

$$f(x^{k_{j+1}}) = \min_{t \in\,]0,\alpha]} f\left(x^k - t H_k^{-1} \nabla f(x^{k_j})\right) \leq f(\tilde{x}^{k_{j+1}})$$

   und daraus

$$f(x^\star) \leq f(x^{k_{j+1}}) \leq f(\tilde{x}^{k_{j+1}}) < f(x^\star) - \frac{\delta}{2}$$

■

Korollar 3.18 *Ist H_k positiv definit und $\|H_k^{-1}\| \leq \beta < \infty$, so erhalten wir globale superlineare Konvergenz, denn $H_{k+1}(x^{k+1} - x^k) \to \nabla^2 f(x^{k+1})(x^{k+1} - x^\star)$.*

Korollar 3.19 (Superlineare globale Konvergenz) *Wenn*

$$\frac{\|(H_{k+1} - \nabla^2 f(x^{k+1}))(x^{k+1} - x^k)\|}{\|x^{k+1} - x^k\|} \to 0,$$

dann gilt:

(i) $t^k = 1 \quad \forall k \geq k_0 \in \mathbb{N}$

(ii) $x^k \to x^\star$ *superlinear (sobald* $t^k = 1$ *und* $k \geq k_0$*) .*

Kapitel 4

Detail-Lösungen

Wiederholung

Im vorangegangenen Kapitel haben wir Verfahren kennengelernt, um für das Problem $\min_{x\in\mathbb{R}^n} f(x)$ eine Lösung zu bestimmen. An dieser Stelle sollen zur Erinnerung noch einmal kurz die wesentlichen Aspekte wiederholt werden, bevor wir uns den Detaillösungen nähern. Für das *allgemeine Abstiegsverfahren* aus Kapitel 3.1 ist der Algorithmus von der folgenden Form:
Wähle ein $x_0 \in \mathbb{R}^n$. Nun lautet der Algorithmus für $k = 0, 1, 2, \ldots$:

1. Wenn $\|\nabla f(x^k)\| \leq \varepsilon$, dann beende den Algorithmus mit dem Ergebnis x^k.
2. Bestimme eine Abstiegsrichtung d^k mit $\nabla f(x^k)d^k < 0$.
3. Bestimme eine Schrittweite $t^k > 0$, so dass $f(x^k + t^k d^k) < f(x^k)$.
4. Setze $x^{k+1} = x^k + t^k d^k$.

Es folgen nun noch einige kurze Bemerkungen über zulässige Abstiegsrichtungen und Schrittweiten. Eine Folge von Abstiegsrichtungen $\{d^k\}$ heißt *zulässig*, falls

(i) $\nabla f(x^k)^T d^k < 0$ (also d^k Abstiegsrichtung ist),

(ii) $\frac{\|\nabla f(x^k)^T d^k\|}{\|d^k\|} \to 0 \quad \Longrightarrow \quad \nabla f(x^k) \to 0.$

Dies lässt sich auch durch die folgende einfache Bedingung darstellen:
$\{d^k\}$ ist zulässig, falls

$$\forall\ k \geq 0\ \exists\ c_0 > 0 : \quad -\frac{\nabla f(x^k)^T d^k}{\|d^k\|} \geq c_0 \|\nabla f(x^k)\|.$$

Eine Folge von Schrittweiten $\{t^k\}$ heißt *zulässig*, falls

(i) $f(x^k + t^k d^k) < f(x^k) \forall\ k \geq 0$,

(ii) $f(x^k) - f(x^k + t^k d^k) \to 0 \quad \Longrightarrow \quad \frac{\nabla^T f(x) d^k}{\|d^k\|} \to 0.$

Außerdem lässt sich zeigen:

Satz 4.1 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und $x^{k+1} := x^k + t^k d^k\ \forall\ k$, wobei t^k, d^k zulässige Schrittweiten und Abstiegsrichtungen sind. Weiter sei $\nabla f(x^k) \neq 0$ und $N(x_0)$ kompakt. Dann gilt:*

(i) $f(x^{k+1}) < f(x^k)$*;*

(ii) Jeder Häufungspunkt ist stationärer Punkt;

(iii) $\lim \nabla f(x^k) = 0$*;*

(iv) Ist $x^\star$ *ein isolierter Häufungspunkt von* $\{x^k\}$ *und gilt für jede Teilfolge* $\{x'_k\}$ *zudem* $\{x'_{k+1} - x'_k\} \to 0$*, dann folgt* $\{x^k\} \to x^\star$*.*

4.1 Liniensuche

Da exakte Liniensuche oft zu aufwendig oder sogar praktisch unmöglich ist, versuchen wir t_k so zu bestimmen, dass näherungsweise gilt

$$f(x_k + t_k d_k) \approx \min_{t \in (0,\alpha]} f(x_k + t d_k)$$

Beispiel: $f(x) = x^2,\ x_0 = 2,\ d_k = (-1)^{k+1},\ t_k = 2 + 3\left(2^{-(k+1)}\right)$
$\implies \quad x_k = (-1)^k(1 + 2^{-k}) = 2, -\frac{3}{2}, \frac{5}{4}, -\frac{9}{8}, \ldots$
Obwohl d_k eine Abstiegsrichtung ist und $f(x_k)$ monoton fällt, konvergiert das Verfahren nicht gegen das Minimum: $\lim_{k\to\infty} f(x_k) = 1$. Grund: t^k ist zu groß.

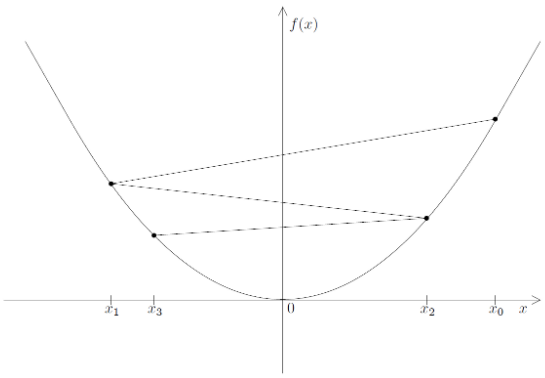

Abbildung 4.1: Schlechte Konvergenz wegen falscher Schrittweitenwahl

Beispiel: $f(x) = x^2,\ x_0 = 2,\ d_k = -1,\ t_k = 2^{-(k+1)} \implies \quad x_k = 1 + 2^{-k} = 2, \frac{3}{2}, \frac{5}{4}, \frac{9}{8}, \ldots$
Obwohl d_k eine Abstiegsrichtung ist und $f(x_k)$ monoton fällt, konvergiert das Verfahren nicht gegen das Minimum: $\lim_{k\to\infty} f(x_k) = 1$. Grund: t^k ist zu klein.

Geeignete Wahl der t_k

Wir definieren

$$\varphi(t) := f(x^k + t d^k).$$

1. Armijo-Strategie mit Backtracking:

   - Regel 1: $\varphi(t^k) \le \varphi(0) + \mu t^k \cdot \underbrace{\varphi'(0)}_{\nabla_t^T f(x^k) d^k|_{t=0}}$ mit einem $0 < \mu < 1$ (z. B. $\mu = 0.1$)

- Regel 2: Das Ziel ist sehr kleine Schritte zu vermeiden. Daher wähle t_k als größte Zahl einer Folge (σg^i), wobei $0 < \sigma \leq 1$ und $0 < g < 1$, mit $\varphi(t_k) \leq \varphi(0) + \mu t^k \varphi'(0)$ mit einem $0 < \mu \leq 1$(z. B. $\sigma = 1$, $g \in \left[\frac{1}{10}, \frac{1}{2}\right]$, $\mu = 0.1$).

Regel 1 wird auch die Armijo-Bedingung genannt.

2. Goldstein-Regeln:

- Regel 1: $\varphi(t^k) \leq \varphi(0) + \mu t^k \varphi'(0)$ mit einem $0 < \mu < 1$ (z. B. $\mu = 0.1$) $\Leftrightarrow$ $f(x^{k+1}) \leq f(x^k) + \mu t^k \nabla f(x^k)^t d^k$
- Regel 2: $\varphi(t^k) \geq \varphi(0) + (1 - \mu) t^k \varphi'(0)$

3. Powell-Wolfe-Strategie:

- Regel 1: Armijo-Bedingung
- Regel 2: $\nabla f(x_{k+1})^T d_k \geq \sigma \nabla f(x_k)^T d_k$ mit $\sigma \in [0.1, 0.9]$
  oder stabiler:
$$|\nabla f(x_{k+1})^T d_k| \ \leq \ -\sigma \nabla f(x_k)^T d_k = \sigma |\nabla^T f(x) d^k|$$

Man kann zeigen, falls $f(x^k + td^k)$ mit $t > 0$ von unten beschränkt ist, dass dann für ein $0 < \mu < \sigma < 1$ eine Schrittweite existiert, so dass die Powell-Wolfe-Bedingungen erfüllt sind.

Satz 4.2 (Schittkowski) *Sei $N(x_0) = \{x \in D \mid f(x) \leq f(x_0)\}$ kompakt und sei t_k gemäß den Armijo-Goldstein-Regeln gewählt und*

$$\nabla f(x_k)^T d_k \leq -\varepsilon \|d_k\| \|\nabla f(x_k)\| \quad \textit{mit einem } \varepsilon > 0.$$

Dann gilt: Jeder Häufungspunkt $x^\star$ ist ein stationärer Punkt, d. h. $f(x^\star) = 0$.

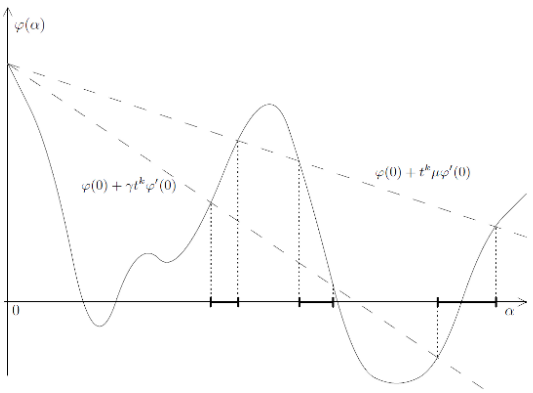

Abbildung 4.2: Goldstein-Regeln

Algorithmus zur Schrittweitenbestimmung (mit Backtracking)

1. Wähle $0 < \sigma \leq 1,\ 0 < g < 1,\ 0 < \mu \leq 1$ (z. B.$\sigma = 1,\ g = 0.5,\ C = 0.1$).
2. Setze $t_k := \sigma$.
3. Überprüfe $\varphi(t_k) \leq \varphi(0) + t_k \mu \varphi'(0)$, falls die Ungleichung gilt, ist t_k die gesuchte Schrittweite, setze $t_k := t_k \cdot g$.

Algorithmus zur Schrittweitenbestimmung (mit Interpolation)

1. Wähle $t_0 = 1,\ \beta, \mu \in [0.2, 0.5]$.

2. Für $k = 1, 2, \ldots$:

   (i) Berechne $\varphi(t_k)$.

   (ii) Falls $\varphi(t_k) < \varphi(0) + t_k C\varphi'(0)$, dann STOP.

   (iii) Interpoliere
   $$\tilde{\varphi}(t) := \varphi(0) + \varphi'(0)t + \frac{\varphi(t_k) - \varphi(0) - t_k\varphi'(0)}{t_k^2}t^2.$$
   (Stützpunkte: $\tilde{\varphi}(0) = \varphi(0),\ \tilde{\varphi}'(0) = \varphi'(0),\ \tilde{\varphi}(t_k) = \varphi(t_k)$).

   (iv) Bestimme Minimum $\tilde{t}$ der quadratischen Interpolation:
   $$\tilde{t} = -\frac{\varphi'(0)t_k^2}{2\left(\varphi(0) - \varphi'(0)t_k + \varphi(t_k)\right)}.$$

   (v) Setze $t_{k+1} := \max\{\beta t_k, \tilde{t}\}$ und gehe zu (ii).

4.2 Update-Formeln

Zusammenfassung der Konvergenzeigenschaften

Für einen Interationsschritt $x^{k+1} = x^k + t^k d^k$ gilt:

$$d^k = -H_k^{-1}(x_k)\nabla f(x_k).$$

- $\underline{H_k = I}$ (oder eine Diagonalmatrix):
  Dann ist $d^k = -\nabla f_k \Longrightarrow$ und wir erhalten das Gradientenverfahren. Mit Liniensuche konvergiert das Verfahren global, aber die lokale Konvergenz ist sehr langsam und hängt von den Eigenwerten $(\lambda_{min}, \lambda_{max})$ von $\nabla^2 f(x)$ ab.

- $\underline{H_k \text{ positiv definit:}}$
  Ist $H_k = \nabla^2 f(x^k)$, dann erhalten wir das Newtonverfahren; wenn $H_k \approx \nabla^2 f(x^k)$, dann erhalten wir ein Newton-ähnliches Verfahren. Globale Konvergenz gibt es mit der Liniensuche, wenn $0 < \varepsilon\|x^2\| \leq x^T H_k x \leq \beta\|x\|^2\ \forall\, k \in \mathbb{N}$. Die lokale Konvergenz ist abhängig von den Eigenwerten der Matrix
  $$H_k^{-\frac{1}{2}}\nabla^2 f(x_k)H_k^{-\frac{1}{2}}$$
  wegen
  $$\kappa_k = \frac{\|H_{k+1}^{-1}\left(H_k - \nabla^2 f(x_k)\right)(x_{k+1} - x_k)\,\|}{\|x_{k+1} - x_k\|} \quad \text{(siehe lokaler Kontraktionssatz 3.12)}$$
  und wir haben

  – lineare Konvergenz, wenn $\kappa_k < 1 \quad \forall\, k \in \mathbb{N}$,

  – superlineare Konvergenz, wenn $\lim_{k\to\infty} \kappa_k = 0$,

  – quadratische Konvergenz, wenn $\kappa_k = 0\ \forall\, k \in \mathbb{N}$.

- $\underline{H_k = \nabla^2 f(x_0)}$:
  Globale Konvergenz, wenn $\nabla^2 f(x_0)$ positiv definit; lokale Konvergenz mit Rate κ.
- $\underline{H_k = \nabla^2 f(x_k)}$:
  Globale Konvergenz, wenn $\varepsilon\|x^2\| \leq x^T H_k x \leq \beta\|x\|^2 \ \forall\, k \in \mathbb{N}$; lokal quadratische Konvergenz.

Das Ziel ist es daher, eine Iteration zu konstruieren, bei der

1. H_0 positiv definit ist und
2. die Updateformel für H_{k+1} aus H_k so ist, dass
   (a) H_{k+1} symmetrisch positiv definit $\Rightarrow$ globale Konvergenz;
   (b) $\lim_{k\to\infty} \frac{\|(H_{k+1}-\nabla^2 f(x_{k+1}))(x_{k+1}-x_k)\|}{\|x_{k+1}-x_k\|} = 0 \Longleftrightarrow \frac{\|H_{k+1}(x_{k+1}-x_k)-(\nabla f(x_{k+1})-\nabla f(x_k))\|}{\|x_{k+1}-x_k\|} \rightsquigarrow 0$
   $\Rightarrow$ superlineare Konvergenz.

4.2.1 Rang-1-Update-Formeln

Es wird H_{k+1} so gewählt, dass $H_{k+1}(x_{k+1}-x_k) = \nabla f(x_{k+1}) - \nabla f(x_k)$. Die allgemeine Form hierfür lautet:

$$H_{k+1} := H_k + \alpha v v^T, \qquad p_k := x_{k+1} - x_k, \qquad q_k := \nabla f(x_{k+1}) - \nabla f(x_k),$$

wobei Rang $(vv^T) = 1$ ist. Es gilt

$$\nabla f(x_{k+1}) \approx \nabla f(x_k) + \nabla^2 f(x_k) p_k \iff \nabla^2 f(x_k) p_k \approx q_k \iff H_{k+1} p_k = q_k \ \textit{(Sekantenbedingung)}.$$

Das Problem ist nun, dass hierbei n^2 Variablen vorhanden sind, die Sekantenbedingung aber nur n Gleichungen liefert.

Broyden-Bedingung

$$H_{k+1} y = H_k y \quad \text{für alle } y \in \mathbb{R}^n \text{ mit } y \perp p_k.$$

Es genügt die Gleichung für $n-1$ linear unabhängige Vektoren $y_1, \ldots, y_{n-1} \perp p_k$ zu erfüllen. Zusammen mit der Sekantenbedingung erhält man n^2 Gleichungen.

Lemma 4.3 *Es existiert genau eine Matrix H^B_{k+1}, die die Sekantenbedingung und die Broyden-Formel erfüllt. Sie hat die Form*

$$H^B_{k+1} = H_k + \frac{(q_k - H_k p_k) p_k^T}{p_k^T p_k}.$$

Beweis: Die Existenz rechnet man sofort nach. Gäbe es nun eine zweite Matrix $H^{B^*}_{k+1}$, die die Bedingungen erfüllt, dann folgt:

$$\left.\begin{array}{ll} (H^B_{k+1} - H^{B^\star}_{k+1}) p_k = 0 & \\ (H^B_{k+1} - H^{B^\star}_{k+1}) y = 0 & \forall\, y \in \mathbb{R}^n :\ y \perp p_k \end{array}\right\} \quad H^B_{k+1} = H^{B^\star}_{k+1}$$

■

Bemerkung: Der Nachteil ist, dass die Broyden-Methode keine symmetrische Update-Matrix H^B liefert, welche daher auch nicht positiv definit ist. Sie liefert bei nichtlinearen Gleichungen

bessere Lösungen als das einfache Newton-Verfahren. Für die Optimierung ist das Verfahren eher ungeeignet, da keine Abstiegsrichtung garantiert ist. Das Broyden-Update ist nur für allgemeine nicht-lineare Gleichungen $F(x) = 0$ geeignet.

4.2.2 Broyden Verfahren

1. $x_{k+1} = x_k - t_k J_k^{-1} F(x_k)$
2. $q_k = F(x_{k+1}) - F(x_k)$
3. $J_{k+1} = J_k + \frac{(q_k - J_k p_k) p_k^T}{p_k^T p_k}$

Eigenschaften: Startwert x_0 hinreichend nah an einer Lösung $x_k : J_0 := J(x_0) = F'(x_0), \forall k : t_k = 1$

$$\Rightarrow \text{ Broyden-Verfahren konvergiert superlinear, da } \frac{||(\delta_k - F'(x_k))\Delta x_k||}{||\Delta x_0||} \longrightarrow 0 \text{ für } K \to \infty$$

Symmetrisches Broyden-Update (Symmetrisches Rang-1-Update, SR1)

$$H_k^{BS} := H_k + \frac{(q_k - H_k p_k)(q_k - H_k p_k)^T}{(q_k - H_k p_k)^T p_k}$$

Lemma 4.4 *Das symmetrische Broyden-Update liefert eine eindeutige Matrix mit den Eigenschaften:*

1. H_k^{BS} *erfüllt Sekantenbedingung,*
2. H_k^{BS} *ist symmetrisch,*
3. $Rang(H_k^{BS} - H_k) = 1$.

Eigenschaften des symmetrischen Broyden-Update

Das Verfahren konvergiert superlinear, wenn

1. x_0 hinreichend nahe der Lösung $x^\star$ gewählt wird
2. $H_0 := \nabla^2 f(x_0)$
3. $t_k = 1$ (Vollschritt)

Nachteil: H_k^{BS} ist nicht notwendigerweise positiv definit. Genauer: Ist $(q_k - H_k p_k)^T p_k > 0$, dann folgt, dass H_k^{BS} positiv definit ist, falls dies schon für H_k gilt.
Fazit: Rang-1-Update Formeln sind eher unbrauchbar. Daher: Rang-2-Update

4.2.3 Rang-2-Update-Formeln

Allgemeine Form:

$$B_k = H_k + \alpha u u^T + \beta v v^T$$

Wir verwenden an hier die abkürzende Schreibweise $y_k := H_k p_k$.

- BFGS-Formel (Broyden, Fletcher, Goldfarb, Shenno):

$$H_k^{BFGS} := H_k + \frac{q_k q_k^T}{q_k^T p_k} - \frac{y_k y_k^T}{p_k^T y_k}$$

  Kann mit approximativer Liniensuche verwendet werden und ist numerisch stabil. Es gibt bislang keine befriedigende theoretische Rechtfertigung.

- DFP-Formel (Davidon, Fletcher, Powell):

$$H_k^{DFP} := H_k + \frac{(q_k - y_k)q_k^T + q_k(q_k - y_k)^T}{q_k^T p_k} - \frac{(q_k - y_k)^T p_k}{(q_k^T p_k)^2} q_k q_k^T$$

Satz 4.5 *Falls x_0 hinreichend nah an der Lösung $x^\star$; $\forall k : t_k = 1$, $H_0 = \nabla^2 f(x_0)$, dann gilt: Superlineare Konvergenz für BFGS imd DFP.*

Bemerkung: Beide Verfahren konvergieren global bei exakter Liniensuche. Bei BFGS reicht bereits die nicht exakte Liniensuche.

Lemma 4.6 *H_k positiv definit und $q_k^T p_k > 0 \Rightarrow H_{k+1}^{BFGS}$ und H_{k+1}^{DFP} sind positiv definit.*

Eigenschaften der Rang-2-Update-Formeln

Beide vorgestellten Rang-2-Update-Formeln konvergieren lokal superlinear, wenn die bereits oben genannten Bedingungen erfüllt sind. Die Verfahren konvergieren sogar global, wenn mit exakter Liniensuche gearbeitet wird; bei BFGS genügt approximative Liniensuche (z. B. Armijo-Goldstein).
Wann tritt nun globale Konvergenz auf?
Die Voraussetzungen hierfür sind, dass f 2-mal stetig differenzierbar ist, dass $N(x_0) = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$ konvex ist und dass $\nabla^2 f(x)$ positiv definit ist für alle $x \in N(x_0)$. Gilt nun zusätzlich noch für $\alpha, \beta > 0$ die Ungleichung $\alpha\|z\|^2 \leq z^T(\nabla^2 f(x))z \leq \beta\|z\|^2$, dann hat f ein eindeutiges Minimum $x^\star$.

Satz 4.7 *Sei H_0 eine beliebige positiv definite Matrix und x_0 ein Startpunkt, der die obigen Voraussetzungen erfüllt. Dann konvergiert die Folge $\{x_k\}$ gegen $x^\star$.*

Der Satz gilt für die gesamte Boyden-Klasse, mit Ausnahme der DFP-Formel.
Wann tritt nun lokale Konvergenz auf?
Sei f 2-mal stetig differenzierbar, $\{x_k\}_{BFGS} \to x^\star$, wobei $x^\star$ ein lokales Minumum mit $H(x^\star)$ Lipschitz-stetig in $U(x^\star)$ ist. Sei ferner $\sum_{k=1}^{\infty} \|x^k - x^\star\| < \infty$. Dann gilt: $t_k = 1$ erfüllt die Wolfe-Bedingung (Kapitel 3) nahe $x^\star$ und $x^k \to x^\star$ superlinear.

Der Algorithmus des DFGS-Verfahrens

Bei einer Dateneingabe von einem Startwert x^0 und einer Abbruchgenauigkeit $\varepsilon > 0$ und einer Startmatrix H_0 sieht der Algorithmus folgendermaßen aus.
Für $k = 0, 1, \ldots$:

1. Falls $\|\nabla f(x^k)\| \leq \varepsilon$, dann beende den Algorithmus.
2. Bestimme die Abstiegsrichtung $p_k := -H_k^{-1} \nabla f(x^k)$.

3. Bestimme $t^k > 0$ (z.B. mit Armijo).
4. Setze $x^{k+1} = x^k + t_k d_k$.
5. Berechne $y_k = H_k p_k$, $q_k = \nabla f_{k+1} - \nabla f_k$ und $H_{k+1} = H^{BFGS}$.
6. Setze $k := k + 1$ und danach den Algorithmus wieder bei Schritt 1 fort.

Das Broyden-Class-Verfahren

Bei einer Dateneingabe von einem Startwert x^0 und einer Abbruchgenauigkeit $\varepsilon > 0$ und einer Startmatrix M_0 sieht der Algorithmus folgendermaßen aus.
Für $k = 0, 1, \ldots$:

1. Falls $\|\nabla f(x^k)\| \leq \varepsilon$, dann beende den Algorithmus und x^k ist die Lösung.
2. Bestimme die Abstiegsrichtung $d_k := -M_k \nabla f(x^k)$ (bzw. für QP: $d_k = -M_k(Ax^k + b)$).
3. Bestimme ein geeignetes t^k (z.B. durch exakte Liniensuche, hier gilt für QP: $t_k = \frac{-(Ax^k+b)^T d_k}{d_k^T A d_k}$)
4. Setze $x^{k+1} = x^k + t_k d_k$.
5. Berechne $q_k = \nabla f_{k+1} - \nabla f_k$ (für QP: $q_k = A(x^{k+1} - x^k) = t_k A d_k$) und $p_k = x^{k+1} - x^k = t_k d_k$. Berechne M_{k+1} aus einer Update-Formel $M_{k+1}(M_k, p_k, q_k)$ (für QP gilt: $q_k^T p_k = t_k^2 d_k^T A d_k > 0$). Die Matrix M_{k+1} ist positiv definit, falls M_k dies schon war.
6. Setze $k := k + 1$ und danach den Algorithmus wieder bei Schritt 1 fort.

Lemma 4.8 *H_k positiv definit und $q_k^T p_k > 0 \implies H_k^{BFGS}$ und H_k^{DFP} sind positiv definit.*

Beweis (nur für BFGS): Sei $z \neq 0$. Wir definieren

$$z^T H_k^{BFGS} z = \underbrace{z^T H_k z}_{x_1} + \underbrace{\frac{z^T q_k q_k^T z}{p_k^T q_k}}_{x_2} - \underbrace{\frac{z^T y_k y_k^T z}{p_k^T y_k}}_{x_3}.$$

Nach Voraussetzung gilt $x_2 = \frac{\left(z^T q_k\right)^2}{p_k^T q_k} \geq 0$ und H_k ist positiv definit. Somit existiert ein L mit $H = L^T L$ (Cholesky-Zerlegung) und $a := Lz \neq 0$, $b := Lp_k$. Dann gilt nach Cauchy-Schwarz :

$$x_1 - x_3 = z^T \left(H_k - \frac{(H_k p_k)(H_k p_k)^T}{p_k^T H_k p_k} \right) z = a^T a - \frac{(a^T b)^2}{b^T b} = \begin{cases} > 0, \text{ falls } a \nparallel b \\ = 0, \text{ falls } a \parallel b, \text{ d. h. } z \parallel p. \end{cases}$$

Ist $x_1 - x_3 > 0$, dann sind wir fertig. Ansonsten gilt $z \parallel p$, also $z = \alpha p$. Damit ist $x_2 = \alpha^2 p_k^T q_k > 0$ und H_k^{BFGS} ist somit positiv definit. ■

Broyden-Familie

SR1, BFGS und DFP sind Mitglieder der Broyden-Familie

$$H_k^\delta = H_k + \frac{q_k q_k^T}{p_k^T q_k} - \frac{y_k y_k^T}{p_k^T y_k} + \delta v_k v_k^T \quad \text{mit } v_k := \left(\frac{q_k}{p_k^T q_k} - \frac{y_k}{p_k^T y_k} \right) \sqrt{p_k^T y_k}.$$

Gilt $\delta \in [0,1]$, so spricht man von der *restriktiven Broyden-Familie*. Im Falle von $\delta = 0$ bzw. $\delta = 1$ erhalten wir die BFGS- bzw. DFP-Formel. Die SR1-Formel ist ein Beispiel einer nichtrestriktiven Broyden-Formel. Gilt $p_k^T q_k > 0,\ \delta \geq 0$, so ist H_k^δ positiv definit und H_k^{DFP} ebenfalls positiv definitiv.
Andere Mitglieder der Familie erhält man durch Interpolation

$$H_k^\delta = \delta H_k^{DFP} + (1-\delta) H_k^{BFGS}.$$

Wenn H_k positiv definit ist, bleibt diese Eigenschaft erhalten:

$$\delta > \bar{\delta} = \frac{1}{1 - \frac{(p_k^T H_k p_k)(q_k^T H_k^{-1} q_k)}{(p_k^T q_k)^2}}.$$

Der Beweis benutzt die Cauchy-Schwarz 'sche Ungleichung.
Dixon (1975): Mit exakter Liniensuche erzeugt jedes H_k^δ das gleiche x^k.

Berechnung von $H_k^{-1}\nabla f(x_k)$

- 1. Variante: Update für die Inverse, kostet $\approx 3n^2$ Multiplikationen.
- 2. Variante: Update für die Hessematrix (häufiger benutzt), kostet $\approx \frac{1}{6}n^3$ Multiplikationen.

Lemma 4.9 *Für alle aufgeführten Update-Formeln gibt es eine entsprechende Update-Formel für die Inverse.*

Sei $M_k := H_k^{-1}$, dann gilt:

- Broyden:
$$M^B = M_k + \frac{(p_k - M_k q_k) M_k q_k^T}{p_k^T M_k q_k}, \text{ wenn } p_k^T M_k q_k \neq 0$$
- SR1:
$$M^{BS} = M_k + \frac{(p_k - M_k q_k)(p_k - M_k q_k)^T}{(p_k - M_k q_k)^T p}$$
(Selbstdual: $H_{k+1}^{BS} M_{k+1}^{BS} = Id$)
- BFGS und DFP: mit der Notation
$$H^{BFGS} = U^{BFGS}(H_k, q_k, p_k) \qquad H^{DFP} = U^{DFP}(H_k, q_k, p_k)$$
erhält man die entsprechenden Updates für die Inverse durch
$$M^{BFGS} = U^{DFP}(M_k, p_k, q_k) \qquad M^{DFP} = U^{BFGS}(M_k, p_k, q_k)$$
- Für allgemeine Updates aus der Broyden-Familie gilt:
$$U^\delta(H_k, q_k, p_k)\, U^\tau(M_k, p_k, q_k) = Id, \quad \text{wobei } \delta = \frac{\tau - 1}{\tau - 1 - \tau\mu} \text{ und } \mu = \frac{(p_k^T H_k p_k)(q_k^T M_k q_k)}{(p_k^T q_k)^2}$$

Satz 4.10 (Finite Termination) *Für das quadratische Problem*

$$\min f(x) = \frac{1}{2}x^T Ax + b^T x + c, \qquad \textit{mit } A \textit{ symmetrisch und positiv definit}$$

und eine beliebige positiv definite Matrix M_0 wird nach spätestens $n+1$ Schritten das Minimum $x^* = -A^{-1}b$ erreicht und es gilt:

1. $M_{i+1}q_j = p_j = t_j d_j \quad \forall\, j = 0, \ldots, i$ *(Erbeigenschaft);*
2. $d_i^T A d_j = 0 \quad \forall\, j = 0 \ldots i-1$ *(Konjugierte Eigenschaft);*
3. *Bei $i = n+1$ gilt: $M_k = A^{-1} = (\nabla^2 f(x_k))^{-1}$ und somit $x_{n+1} = x^*$.*

Beweis per Induktion: Die Induktionsvoraussetzung ($i = 0$) erfüllt 1. die Sekantenbedingung ($M_1 q_0 = p_0$) oder 2. die leere Bedingung.

$\underline{i \to i+1}$:

Sei $x_{i+1} \neq x^*$, $j \leq i-1$, dann folgt

$$\begin{aligned}\nabla f(x_{i+1}) &= Ax_{i+1} + b = Ax_{j+1} + b - A(x_{j+1} - x_{i+1}) = Ax_{j+1} + b - A(x_{j+1} - x_{j+2} + x_{j+2} - x_{j+3} \cdots - x_{i+1}) \\ &= \nabla f(x_{j+1}) + A(d_{j+1}t_{j+1} + \ldots + d_i t_i)\end{aligned}$$

und daraus

$$\nabla f(x_{i+1})^T d_j = \underbrace{\nabla f(x_{j+1})^T d_j}_{\substack{=0 \\ \text{bei exakter Liniensuche}}} + \underbrace{(t_{j+1} d_j^T A d_{j+1} + \ldots + t_i d_j^T A d_i)}_{=0} = 0.$$

Sei nun $j = i$ und $\nabla f(x^{i+1})^T d_j = 0$ (exakte Liniensuche), dann ergibt die Nebenrechnung

$$-(M_{k+1}^{-1} d_{k+1})^T d_j = -d_{k+1}^T M_{k+1}^{-1} d_j = -d_{k+1}^T \frac{1}{t_j} q_j = -d_{k+1}^T \frac{1}{t_j} t_j A d_j = -d_{k+1}^T A d_j$$

die Gleichung $p_{i+1}^T A p_j = 0$ für $j = i$ und somit $d_{k+1}^T A d_k = 0$. Aus

$$M_{i+2} = M_{i+1} + u p_{i+1}^T + v(M_{i+1} q_{i+1})^T$$

folgt direkt

$$M_{i+2} q_j = M_{i+1} q_j + u \underbrace{p_{i+1}^T q_j}_{=0} + v \underbrace{q_{i+1}^T M_{i+1} q_j}_{=t_k d_k^T A t_i d_i = 0} = M_{i+1} q_j.$$

Also bricht der Algorithmus entweder für $i < n-1$ mit $x_{i+1} = x^\star$ ab oder für $i = n-1$ gilt

$$M_n A d_j = M_n \frac{1}{t_j} q_j = \frac{1}{t_j} M_n q_j = \frac{1}{t_j} t_j d_j = d_j \quad \forall\, j = 0, \ldots, n-1.$$

Also sind die d_j Eigenvektoren zum Eigenwert 1 für $M_n A$. Da die Eigenvektoren linear unabhängig sind, folgt $M_n A = Id$, d.h. $M_n = A^{-1}$. Es folgt

$$x_{n+1} = x_n - A^{-1}\nabla f(x_n) = x_n - (x_n + A^{-1}b) = -A^{-1}b = x^\star.$$

■

Wichtige Bedeutung für praktische Umsetzung:

$$H_0 = Id \quad \text{oder} \quad H_0 = diag\left(\frac{1}{s_j^2}\right) \text{ mit Skalierungsfaktoren } s_j.$$

Kapitel 5

Sequentielle quadratische Programmierung (SQP)

5.1 Grundidee

Die Sequentielle Quadratische Programmierung ist ein direktes Verfahren zur Lösung restringierter Optimierungsprobleme. Dieses Verfahren ähnelt, abgesehen von den hier zu berücksichtigenden Ungleichungsnebenbedingungen und der speziellen Suchrichtungsbestimmung, in gewisser Weise dem Vorgehen beim Quasi-Newton-Verfahren. Zunächst werden für einen zu wählenden zulässigen Startpunkt die Funktionswerte und die jeweiligen Gradienten des Gütekriteriums und der vorhandenen Nebenbedingungen an diesem Anfangspunkt berechnet. Die Bestimmung der Suchrichtung d erfolgt durch eine zusätzlich auszuführende Minimierungsaufgabe. Die hier zu minimierende Funktion entspricht einer Approximation der Lagrange- Funktion durch eine Taylor-Reihe mit Termen inklusive der 2. Ordnung,

$$\min_{d \in \Omega_k} \nabla f(x_k)^T d + \frac{1}{2} d^T H_k d$$

$$\text{unter} \quad g(x_k) + \nabla g(x_k)^T d = 0 \quad \text{und} \quad h(x_k) + \nabla h(x_k)^T d \geq 0.$$

Die Matrix H_k entspricht einer Näherung der Hesse-Matrix der Lagrange-Funktion, die - wie auch beim Quasi- Newton-Verfahren - anfänglich als Einheitsmatrix E definiert werden kann und mit steigender Zahl von Iterationsschritten gegen die eigentliche Hesse-Matrix konvergiert. Ist die Suchrichtung d_k bekannt, folgt die Ermittlung der Schrittweite t_k in Abhängigkeit von den geltenden Restriktionen, so dass sich dadurch der nächst-bessere Parameterpunkt gemäß

$$x^{k+1} = x^k + t_k d_k$$

ergibt. Die hier vorgestellte Vorgehensweise wird dabei so lange wiederholt, bis das Abbruchkriterium erfüllt ist bzw. die restringierte Gütefunktion ausreichend minimiert ist. Der zuletzt berechnete Parameterpunkt kann dann somit als Minimierer angesehen werden. Solange das Abbruchkriterium noch nicht erfüllt ist, muss - wie auch beim Quasi-Newton-Verfahren - die Näherung der Hessematrix z.B. mittels BFGS-Update neu berechet werden, bevor innerhalb der nächsten Iteration die neue Suchrichtung bestimmt wird.

5.2 SQP-Verfahren

Ausgegangen wird bei dem SQP-Verfahren wie bisher von der Funktion $f(x)$, die über $x \in \mathbb{R}^n$ minimiert werden soll. Außerdem gelten folgende Restrinktionen:

$$g_i(x) = 0,\ i \in E, \qquad g_i(x) \geq 0,\ i \in I.$$

Der Basisalgorithmus ist dann von der folgenden Form:

Eingabe: $x^0, \lambda_0, \mu_0, \varepsilon, k$
Solange

$$\|\nabla L(x^k, \lambda_k, \mu_k)\| \geq \varepsilon, \quad \|g_E(x)\| \geq \varepsilon, \quad \min_{i \in I}\{0, g_i(x)\} < -\varepsilon,$$

verfahre wie folgt:

1. Bestimme Suchrichtung d_k aus

$$\begin{array}{ll} \min\limits_{d \in \mathbb{R}^n} & \nabla^T f(x^k)d + \frac{1}{2}d^T H_k d, \quad H_k \approx \nabla^2 L(x^k, \lambda_k, \mu_k) \\ (QP) & g_i(x_k) + \nabla g_i(x_k)^T d = 0, \quad i \in E \\ & g_i(x_k) + \nabla g_i(x_k)^T d \geq 0, \quad i \in I \end{array}$$

   Sei $\{d_k, \lambda_{QP}, \mu_{QP}\}$ ein KKT-Punkt in (QP).

2. Bestimme die Schrittweite t_k.

3. Modifikationen:

$$\begin{aligned} x^{k+1} &= x^k + t_k d_k \\ \lambda_{k+1} &= \lambda_k + t_k(\lambda_{QP} - \lambda_k) \\ \mu_{k+1} &= \mu_k + t_k(\mu_{QP} - \mu_k) \\ k &= k + 1 \end{aligned}$$

Lemma 5.1 *Für alle zulässigen x^k, die die notwendige Bedingung der 1. Ordnung erfüllen, d.h. $\exists\, \lambda^*, \mu^* :\ \nabla f(x_k) = \nabla g_E(x_k)^T \lambda^* + \nabla g_I(x_k)^T \mu^*$, und für die H_k auf $T(x_k)$ positiv definit ist, ist x^k ein Fixpunkt des SQP-Verfahrens.*

Beweis: Das Problem (QP) hat eine Lösung bei $d = 0$: Da sowohl $g_E(x_k) = 0$ als auch $g_I(x_k) \geq 0$, ist x^k somit ein zulässiger Punkt. Damit gilt

$$L^{QP}(d_k, \lambda, \mu) = \nabla f(x_k)^T d + \frac{1}{2}d^T H_k d - \lambda^T \left(g_E(x_k) + \nabla g_E(x_k)d\right) - \mu^T \left(g_I(x_k) + \nabla g_I(x_k)d\right)$$

und somit $\nabla_d L^{QP}(d_k, \lambda, \mu) = \nabla f(x_k) + H_k d - \nabla g(x_k)^T \lambda - \nabla h(x_k)^T \mu$, d.h. $\nabla_d L^{QP} = \nabla L(x, \lambda, \mu)$
■

Andere Interpretation

Betrachte die Funktion $\min_{x \in \mathbb{R}^n} f(x)$ mit $g_i(x) = 0$ für $i \in E$. Dann ergibt sich die Langrangefunktion $L(x, \lambda) = f(x) - \lambda^T g(x)$. Nun definiere die folgende Matrix:

$$A := \begin{pmatrix} \nabla g_1(x) \\ \vdots \\ \nabla g_m(x) \end{pmatrix}.$$

Aus der notwendigen Bedingung ergibt sich nun das Karush-Kuhn-Tucker-System

$$0 \;=\; F(x,\lambda) \;=\; \begin{pmatrix} \nabla L(x,\lambda) \\ g(x) \end{pmatrix} \;=\; \begin{pmatrix} \nabla f(x) - A^T\lambda \\ g(x) \end{pmatrix}.$$

Leitet man dieses System nun jeweils nach x und λ ab, so ergibt sich die Karush-Kuhn-Tucker-Matrix

$$J(x,\lambda) = \begin{pmatrix} H_k & -A^T \\ A & 0 \end{pmatrix}, \quad \text{wobei } H_k = \nabla_x^2 L(x,\lambda).$$

Die Karush-Kuhn-Tucker-Matrix (KKT-Matrix) ist nicht singulär, genau dann wenn A vollen Rang hat (CQ) und für alle $p \neq 0$ und $Ap = 0$ gilt: $p^T Hp > 0$, also H positiv definit ist auf dem Kern von A (PD).

Newton-Schritt

Der Newtonschritt ist dann von der Form: $\begin{pmatrix} x^{k+1} \\ \lambda_{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \lambda_k \end{pmatrix} + \begin{pmatrix} d_x \\ d_\lambda \end{pmatrix}$. Der Vektor $(d_x, d_\lambda)^T$ löst das lineare Gleichungssystem

$$\begin{pmatrix} H_k & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} d_x \\ d_\lambda \end{pmatrix} = - \begin{pmatrix} \nabla f(x^k) - A^T(x^k)\lambda_k \\ g(x^k) \end{pmatrix}.$$

Dasselbe Ergebnis lässt sich auch ebenfalls erzielen, wenn man für das quadratische Problem $d_x = d_k$ und $d_\lambda = \lambda_{QP} - \lambda_k$ verwendet. Aus der positiven Definitheit und der Constraint Qualification folgt dann die selbe Gleichung, denn (d_k, λ^{QP}) löst $\nabla f(x^k) + H_k d - A_k^T \lambda^{QP} = 0$ und $A_k d + g(x^k) = 0$.

Folgerung: Unter den Bedingungen der Constraint Qualification (CQ) und positiver Definitheit (PD) ergibt sich, dass das SQP-Verfahren eher eine praktische Relevanz hat, im Gegensatz zum Newton-Verfahren auf Basis des KKT-Systems. Hier handelt es sich mehr um eine theoretische Relevanz, auf Grund der globalen und lokalen Konvergenz.

5.3 Lösung des quadratischen Programms

5.3.1 Unbeschränkte Optimierung

Für das Standardproblem

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} d^T H d + g^T d,$$

wobei H positiv definit ist, existiert eine eindeutige Lösung $d^\star = -H^{-1}g$. Sie lässt sich unmittelbar aus der notwendigen Bedingung $\nabla f = Hd + g = 0$ herleiten. Falls H nicht positiv definit ist, benutzt man z.B. Trust-Region-Verfahren. Für symmetrisches H kann man das Gleichungssytem $Hd + g = 0$ mit Hilfe der *(direkten) Cholesky-Zerlegung* $H = LDL^T$ lösen:

$$Hd = LDL^T d = -g \quad \Longrightarrow \quad y := -L^{-1}g \quad \Longrightarrow \quad z := D^{-1}y \quad \Longrightarrow \quad d = L^{-T}z.$$

5.3.2 Gleichungsbeschränkte Optimierung

Bei der gleichungsbeschränkten Optimierung ist neben der vorhandenen Minimierungsfunktion noch eine Restriktion als Gleichung angegeben. Somit ergibt sich für das Problem folgende Form:

$$\min_{d\in\mathbb{R}^n} \quad \tfrac{1}{2}d^THd + g^Td \qquad s.t. \quad Ad + a = 0.$$

Unter den Voraussetzungen, dass

- $A \in \mathbb{R}^{l\times n}$ vollen Rang hat und
- $d^THd > 0 \quad \forall\, d \neq 0,\ Ad = 0 \qquad (\Longleftrightarrow H$ positiv definit auf Kern$(A))$,

existiert eine eindeutige Lösung $(d^\star, \lambda^\star)$, die das KKT-System

$$\begin{pmatrix} H & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = - \begin{pmatrix} g \\ a \end{pmatrix}$$

erfüllt. Das KKT-System ergibt sich aus der Ableitung der Lagrangefunktion $L^{QP} = \frac{1}{2}d^THd + g^Td - \lambda^T(Ad + a)$ nach d, sowie der Nebenbedingung.
Um nun die optimalen Werte für die Parameter d und λ zu bestimmen, gibt es zwei verschiedene Methoden. Die Entscheidung, welche Methode angewandt wird, hängt von der Beschaffenheit der Matrix H ab.

1. Lösungsmöglichkeit: Bildraum-Methode
Falls H leicht zu invertieren ist, kann man die KKT-Bedingungen relativ einfach umformen, so dass man direkt die Parameter bestimmen kann. Für diese ergibt sich

$$\begin{aligned} \lambda &= \left(AH^{-1}A^T\right)^{-1}\left(-a + AH^{-1}g\right), \\ d &= H^{-1}\left(A^T\lambda - g\right). \end{aligned}$$

2. Lösungsmöglichkeit: Nullraum-Methode
Dreieckszerlegung von A bzw. A^T:

$$A = (L, 0)\,Q^T \quad \Longleftrightarrow \quad A^T = Q\begin{pmatrix} L^T \\ 0 \end{pmatrix} \quad \text{mit } L \in \mathbb{R}^{l\times l},\ Q \in \mathbb{R}^{n\times n}.$$

Q ist eine orthogonale Matrix, welche sich z. B. mit Hilfe einer Householder-Transformation bestimmen lässt. Setzt man noch $z := Q^Td$ mit $z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$, wobei $z_1 \in \mathbb{R}^l,\ z_2 \in \mathbb{R}^{n-l}$, so ergibt sich damit zunächst aus

$$Ad + a = 0 \quad \Longleftrightarrow \quad AQz + a = 0 \quad \Longleftrightarrow \quad (L, 0)\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = -a \quad \Longleftrightarrow \quad Lz_1 = -a$$

die Lösung der Nebenbedingung

$$z_1^* = -L^{-1}a\,, \quad z_2^* \in \mathbb{R}^{n-l} \text{ beliebig}$$

und damit

$$d^\star = Qz^\star = (Q_1, Q_2) \begin{pmatrix} z_1^\star \\ z_2^\star \end{pmatrix} = Q_1 z_1 + Q_2 z_2.$$

Für die Zielfunktion folgt

$$\min \frac{1}{2} d^T H d + g^T d = \min_{z_2 \in \mathbb{R}^{n-l}} \frac{1}{2} \left(Q_1 z_1 + Q_2 z_2\right)^T H \left(Q_1 z_1 + Q_2 z_2\right) + g^T \left(Q_1 z_1 + Q_2 z_2\right)$$

$$\iff \quad \min_{z_2 \in \mathbb{R}^{n-l}} \frac{1}{2} z_2^T \left(Q_2^T H Q_2\right) z_2 + \left(g^T Q_2 + z_1^T Q_1^T H Q_2\right) z_2 + \text{konstante Terme mit } z_1.$$

Man berechnet nun die reduzierte Hessematrix $\widetilde{H} = Q_2^T H Q_2$ und den reduzierten Gradienten $\widetilde{g} = g^T Q_2 + z_1^T Q_1^T H Q_2$ von links nach rechts (Matrix-Vektor-Multiplikationen statt Matrix-Matrix-Multiplikationen). Dann minimiert $z_2^\star$ das unbeschränkte QP

$$\min \frac{1}{2} y^T \tilde{H} y + \tilde{g} y.$$

Es ergibt sich somit die Lösung $z_2^\star = -\tilde{H}^{-1}\tilde{g}$. Anschließend löst man $\widetilde{H} z_2 + \widetilde{g} = 0$ und erhält

$$d = Q \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

Der Multiplikator λ kann mit gleicher Zerlegung berechnet werden, denn es gibt genau ein λ mit $A^T \lambda = Hd + g$ und somit gilt

$$\begin{pmatrix} L^T \\ 0 \end{pmatrix} \lambda = Q^T (Hd + g) = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \quad \Longrightarrow \quad \lambda = L^{-T} c_1.$$

c_2 muss wegen KKT-Bedingungen 0 sein.)

5.3.3 Gleichungs- und ungleichungsbeschränkte Optimierung

Wir betrachten nun das allgemeine Problem

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \tfrac{1}{2} d^T H d + g^T d, \quad H \in \mathbb{R}^{n \times n} \text{ pos.def.}, \\ s.t. \quad & Ad + a = 0, \quad A \in \mathbb{R}^{l \times n}, \\ & Bd + b \geq 0, \quad B \in \mathbb{R}^{l \times n}. \end{aligned}$$

Die Ungleichungen werden mit der „Active-Set-Strategie“ behandelt. Man definiert sich die Menge der aktiven Indizes

$$I(d) = \{i \mid B_i d + b_i = 0\}.$$

Die Idee ist, iterativ die Menge der aktiven Indizes I^* und damit d^* zu bestimmen. Falls $I^\star$ bekannt ist, dann ist das quadratische Ungleichgungsproblem äquivalent zu dem quadratischen Gleichungsproblem

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \tfrac{1}{2} d^T H d + g^T d \\ s.t. \quad & Ad + a = 0\,, \quad B_i d + b_i = 0. \end{aligned}$$

Als Anwendungsverfahren ergibt sich nun die **Active-Set-Methode:**
Startwert: Bestimme einen zulässigen Punkt $d_0 \in \mathbb{R}^n$ und die zugehörige Menge

$$I_0 = \{i \mid B_i d_0 + b_i = 0\}.$$

Als Voraussetzung müssen hierbei gelten:

$$|I_0| + l \leq n \qquad \text{und} \qquad \begin{pmatrix} A \\ B_i \end{pmatrix} \text{ mit } i \in I_0 \text{ hat vollen Rang.}$$

Iteration: Für $k = 0, 1, 2, \ldots$:

1. Löse das $QP(I_k)$
$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \tfrac{1}{2} d^T H d + g^T d \\ s.t. \quad & Ad + a = 0, \quad B_i d + b = 0 \quad \forall\, i \in I_k, \end{aligned}$$
und erhalte die Lösung $\widehat{d}$.

2. - Falls $\widehat{d}$ ein zulässiger Punkt ist im QP mit Ungleichungen, das heißt, $B_i \widehat{d} + b \geq 0$ für $i \in \{1, \ldots, l\} \backslash I_k$, dann setze $d_{k+1} = \widehat{d}$, $I_{k+1} = I_k$ und gehe zu 3.
   - Ansonsten setze $\Delta d := \widehat{d} - d_k$, betrachte die Gerade $d_k + t\Delta d$ für $0 \leq t \leq 1$ und setze
$$t_i = \begin{cases} -\frac{B_i d_k + b_i}{B_i \Delta d}, & \text{falls } i \notin I_k \text{ oder } B_i \widehat{d} + b_i < 0, \\ \infty, & \text{sonst.} \end{cases}$$
   - Bestimme die kleinste Schrittweite $\bar{t}$, für die eine Ungleichung j verletzt wird:
$$\bar{t} := \min_i t_i = t_j,$$
und setze $d_{k+1} = d_k + \bar{t}\Delta d$, $I_{k+1} = I_k \cup \{j\}$ (dabei behält die Matrix der (aktiven Un-) Gleichungen vollen Rang).
   - Setze $k = k + 1$.
   - Falls $|I_k| + l = n$ gehe zu 3., ansonsten gehe zu 1. (Hinweis: Das I_k ist an dieser Stelle schon das I_{k+1}, da in dem Schritt vorher k bereits erhöht wurde.)

3. Wir haben eine Lösung d_k des Problems $QP(I_k)$ erreicht.
   - Berechne die zugehörigen Lagrange-Multiplikatoren λ_k, μ_k aus dem KKT-System ($\mu_i = 0,\ i \notin I_k$).
   - Falls $(\mu_k)_i \geq 0\ \forall\, i \in I_k$, ist eine Lösung des QP mit Ungleichungen gefunden $\Longrightarrow$ STOP!
   - Ansonsten gibt es ein $\xi \in I_k$ mit $(\mu)_\xi < 0$. Aus der Theorie folgt, dass die Lösung des QP für $I_k \setminus \{\xi\}$ ein besseres Zielfunktional hat. Setze also $I_{k+1} = I_k \setminus \{\xi\}$ und gehe zu 1. (Falls nur ein Index entfernt werden soll, wählt man z. B. den kleinsten Index oder den Index mit betragsgrößtem μ.)

Bemerkungen:

- Der Trick beim SQP-Verfahren besteht darin, dass das Active-Set durch die QP-Lösung automatisch identifiziert wird. Außerdem spielen Startwerte für die Multiplikatoren nur eine geringe Rolle.

- Der Algorithmus findet nach endlich vielen Schritten die Lösung: Spätestens nach $n-l$ Schritten landen wir in einer Ecke, also bei Schritt 3. Dort fällt die Zielfunktion, d. h. jede Indexmenge kann nur einmal vorkommen. Es gibt aber nur endlich viele Indexmengen (ggf. aber exponentiell viel).

- Der Algorithmus kann bei entarteten (degenerierten) Ecken versagen, d. h. falls mehr als $n-l$ Ungleichungen aktiv sind. Es können dann Zyklen entstehen, weil beim Weglassen einer Ungleichung kein Abstieg garantiert ist. *Abhilfe:* Störe die Ungleichungen (zufällig) durch

  $$B_i d + b_i + \varepsilon \geq 0 \quad \text{z. B. mit } \varepsilon \approx 10^{-8}.$$

  ε muss klein genug sein, damit die Lösung nicht zu sehr gestört wird, aber groß genug, um Rundungsfehler zu dominieren. Dabei entstehen aber viele benachbarte Ecken, also wird der Algorithmus teurer, er braucht mehr Iterationsschritte.

- Bei einem nichtlinearen Problem löst man eine Folge von QPs. Man setzt dann *Warm-Start*-Techniken ein, die das Active-Set der letzten nichtlinearen Iteration (SQP) nutzen, um die Lösung des QP „schneller“ zu starten.

Wie findet man einen zulässigen Punkt (Startwert)? Wähle ein $\bar{d}$, welches die Gleichung $A\bar{d} + a = 0$ erfüllt und setze

$$q_i := \min\left(0, B_i\bar{d} + b_i\right).$$

Löse dann folgendes relaxiertes LP:

$$\begin{aligned} \min_{q,\,d} \quad & \sum q_i \\ s.t. \quad & Ad + a = 0, \\ & Bd + b + q \geq 0, \\ & q \geq 0. \end{aligned}$$

Löse dieses Problem mit gleichem Algorithmus. Falls eine zulässige Lösung existiert, wird sie gefunden.

5.4 Konvergenzanalyse des SQP-Verfahrens

Das SQP-Verfahren bei festem Active-Set ist äquivalent zum Newton-Verfahren für das KKT-System in hinreichender Nähe der Lösung $x^\star$, welche die notwendige Bedingung 2. Ordnung und die hinreichende Bedingung erfüllt. Wenn man sich in der Nähe der Lösung befindet, in der CQ, PD und strikte Komplementarität erfüllt sind, dann bleibt das Active-Set bei dem Quadratischen Problem gleich und ist auch gleich dem Active-Set des nichtlinearen Problems.

Satz 5.2 *Startet man das Newton-Verfahren für $\nabla_x L(x,\lambda,\mu) = 0,\ g(x) = 0,\ h_i(x) \geq 0\ \forall\, i \in I^*$, in hinreichender Nähe der Lösung, so entspricht dies dem SQP-Verfahren mit exakter Hessematrix und Vollschritt.*

Beweis: Da das Active-Set in hinreichender Nähe der Lösung fest ist, nehmen wir o. B. d. A. an, dass es sich um ein gleichungsbeschränktes Problem handelt. Löst man das QP mit exakter

Hessematrix $H_k = \nabla_x^2 L(x_k, \lambda_k)$ und Vollschritt ($x_{k+1} = x_k + d_k$), dann ergibt sich folgendes KKT-System:

$$\begin{pmatrix} \nabla_x^2 L(x_k, \lambda_k) & -\nabla g(x_k)^T \\ \nabla g(x_k) & 0 \end{pmatrix} \begin{pmatrix} d_k \\ \lambda^{QP} \end{pmatrix} = - \begin{pmatrix} \nabla f(x_k) \\ g(x_k) \end{pmatrix}$$

Das Newton-Verfahren für $\nabla_x L(x, \lambda) = 0,\ g(x) = 0$, hat mit

$$F(x, \lambda) := \begin{pmatrix} \nabla_x L(x, \lambda) \\ g(x) \end{pmatrix} \quad \Longrightarrow \quad F(x_k, \lambda_k) + \nabla F(x_k, \lambda_k) \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \end{pmatrix} = 0$$

die Gestalt

$$\begin{pmatrix} \nabla_x^2 L(x_k, \lambda_k) & -\nabla g(x_k)^T \\ \nabla g(x_k) & 0 \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \end{pmatrix} = - \begin{pmatrix} \nabla L(x_k, \lambda_k) \\ g(x_k) \end{pmatrix}.$$

Subtrahiert man in der ersten Blockzeile $\nabla g(x_k)^T \lambda_k$, so erhält man

$$\begin{pmatrix} \nabla_x^2 L(x_k, \lambda_k) & -\nabla g(x_k)^T \\ \nabla g(x_k) & 0 \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k + \lambda_k \end{pmatrix} = - \begin{pmatrix} \nabla f(x_k) \\ g(x_k) \end{pmatrix}$$

und man erkennt mit $d_k = \Delta x_k$ und $\lambda^{QP} = \lambda_k + \Delta \lambda_k$ die Äquivalenz der Verfahren. ■

Korollar 5.3 *Unter den Voraussetzungen des Newton-Verfahrens konvergiert das SQP-Verfahren mit exakter Hessematrix der Lagrange-Funktion und Vollschritt lokal Q-quadratisch.*

Bemerkung: Sei

$$\widetilde{g}(x) := \begin{pmatrix} g(x) \\ h_i(x) \end{pmatrix},\ i \in I^*, \quad \Delta \widetilde{\lambda}_k := \widetilde{\lambda}_{k+1} - \widetilde{\lambda}_k.$$

Es wurde bereits gezeigt, dass das SQP-Verfahren mit exakter Hessematrix äquivalent zum Newton-Verfahren ist. Nach dem lokalen Kontraktionssatz gilt

$$\| \begin{pmatrix} \Delta x_{k+1} \\ \Delta \widetilde{\lambda}_{k+1} \end{pmatrix} \| \leq \frac{\omega}{2} \| \begin{pmatrix} \Delta x_k \\ \Delta \widetilde{\lambda}_k \end{pmatrix} \|^2, \quad \text{falls} \quad \frac{\omega}{2} \| \begin{pmatrix} \Delta x_0 \\ \Delta \widetilde{\lambda}_0 \end{pmatrix} \| < 1.$$

Satz 5.4 (Lokale Analyse) *Sei $x^\star$ ein reguläres lokales Minimum, wobei $A(x^\star)$ vollen Rang hat. Gilt außerdem*

1. $\exists\, \lambda^*,\ \mu^* \geq 0 :\ \nabla_x L(x^\star, \lambda^\star, \mu^\star) = 0$,
2. $h_i(x^\star) = 0 \iff \mu_i^* > 0\ \forall\, i \in I^\star$ *(strikte Komplementarität),*
3. $p^T H p > 0\ \forall p \in T(x^*),\ p \neq 0$ *(HOB 2. Ordnung), wobei* $T(x^\star) = \{p \mid \nabla g_i(x^\star)^T p = 0, \nabla h_i(x^\star)^T p = 0\ \forall\, i \in I^*\}$,
4. *die Näherung $\widehat{H} \approx \nabla^2 L(x^*, \lambda^*, \mu^*)$ ist positiv definit,*

dann folgt:

1. *Das $QP(x^*, \widehat{H})$*

$$\min_{d \in \Omega_k} \nabla f(x^*)^T d + \frac{1}{2} d^T \widehat{H} d$$

mit den Restriktionen

$$g(x^\star) + \nabla g(x^\star)^T d = 0 \text{ und } h(x^\star) + \nabla h(x^\star)^T d \geq 0,$$

hat genau eine Lösung $d^\star = 0$. Die Lagrange-Multiplikatoren und das Active-Set des QP sind die gleichen wie die des nichtlinearen Problems. Ferner gilt strikte Komplementarität und die Lösung des QPs erfüllt ebenfalls die HOB 2. Ordnung.

2. *Das $QP(x^*, \widehat{H})$ ist stabil gegen Störungen, d. h.: Existieren δ_1, $\delta_2 > 0$, so dass für alle x, H gilt*

$$\|x - x^\star\| < \delta_1 \ \wedge \ \|H - \widehat{H}\| < \delta_2$$

dann hat $QP(x, H)$ genau eine Lösung und das gleiche Active-Set wie das $QP(x^\star, \widehat{H})$ und $NLP(x^\star)$.

Beweis:

1. Hierfür ist der Beweis ähnlich zu dem bei der Konvergenz des Newtonverfahrens.
2. Das Ergebnis folgt direkt aus dem Störungssatz für Minimalstellen, die die HOB 2. Ordnung und strikte Komplemetarität erfüllen. ■

5.5 Update-Formeln für die Hessematrix

Dieser Abschnitt verfolgt zwei Ziele: Zum einen suchen wir eine billige Approximation H_k der Hessematrix ($H_k \approx \nabla^2 L(x, \lambda, \mu)$), zum anderen möchten wir die positive (Semi-)Definitheit von H_k erreichen.
Die Updates habe die allgemeine Form

$$H_{k+1} = U(H_k, p_k, q_k),$$

wobei U aus der Broyden-Familie (Rang-2-Updates) stammt. Alles unter der Voraussetzung, dass in $x^\star$ die Constraint Qualification, PD und die strikte Komplementarität gelten.

Satz 5.5 *Sei $x^\star$ ein lokales Minimum, $(x^\star, \lambda^\star, \mu^\star)$ das zugehörige KKT-Tripel, $H^\star := \nabla_x^2 L(x^\star, \lambda^\star, \mu^\star)$ positiv definit und seien x_0, H_0 hinreichend nahe bei $x^\star$, $H^\star$. Dann konvergiert das SQP-Verfahren mit DFP-Update Q-superlinear gegen $(x^\star, \lambda^\star, \mu^\star)$.*

Bemerkung: λ_0, μ_0 werden an dieser Stelle nicht benötigt, da das erste Update schon λ_1, μ_1 aus dem QP benutzt.

Konvergenzeigenschaften des SQP mit Vollschritt

- $H_k = \nabla_x^2 L(x_k, \lambda_k, \mu_k)$: lokal quadratische Konvergenz
- $H_k \approx \nabla_x^2 L(x_k, \lambda_k, \mu_k)$: lokal lineare Konvergenz mit Konvergenzrate κ
- H_k aus DFP-Update, H^* positiv definit: lokal superlineare Konvergenz

Powell-Modifikation

Die Powell-Modifikation wird manchmal auch *gedämpftes BFGS-Update* genannt. Ändere q ab, so dass es einer positiven Krümmungsrichtung entspricht. Wir interpolieren zwischen $H_k p_k$ und q_k, d. h. wir setzen

$$\widetilde{q} := \tau q + (1-\tau) Hp$$

mit

$$\tau := \begin{cases} 1, & \text{falls } q^T p > \frac{1}{5} p^T H p, \\ \frac{4}{5} \frac{p^T H p}{p^T H p - q^T p}, & \text{sonst.} \end{cases}$$

Der Parameter τ wird also aus dem Intervall $[0,1]$ maximal gewählt, so dass $\widetilde{q}^T p \geq \frac{1}{5} p^T H p$ gilt. Somit bekommen wir positive Definitheit in H_{k+1}, dafür geht aber die superlineare Konvergenz verloren. Trotzdem ist es sinnvoll, die Powell-Modifikation durchzuführen.

Alternativen:

- Wähle
$$\widetilde{q} := \tau q + (1-\tau) H_0 p.$$
- Benutze Hessematrix der „augmented Lagrangian“ Approximation (statt $\nabla_x^2 L$):

$$\begin{aligned} L^{aug}(x,\lambda) &= f(x) - g(x)^T \lambda + \frac{\kappa}{2} \|g(x)\|_2^2 \\ \Longrightarrow \quad \nabla_x^2 L^{aug}(x^*,\lambda^*) &= \nabla_x^2 L(x^*,\lambda^*) + \kappa \nabla g(x^*) \nabla g(x^*)^T \end{aligned}$$

  Bei hinreichend großer Wahl von κ ist $\nabla_x^2 L^{aug}(x^*,\lambda^*)$ positiv definit: Für $p \in T(x^*)$ ist $\nabla_x^2 L(x^*,\lambda^*)$ positiv definit (nach HOB 2. Ordnung) und $p^T \nabla g(x^*) \nabla g(x^*)^T p > 0$. Folglich existiert κ, so dass

$$p^T \left(\nabla_x^2 L(x^*,\lambda^*) + \kappa \nabla g(x^*) \nabla g(x^*)^T \right) p > 0 \quad \forall\, p \in T(x^*).$$

  Die Lösung des QP wird durch Addition von $\kappa \nabla g(x^*) \nabla g(x^*)^T$ nicht gestört. *Problem:* κ muss evtl. groß gewählt werden, was die Kondition stark verschlechtern kann. Außerdem ist die positive Definitheit nur für die Lösungspunkte garantiert.

Satz 5.6 *Mit beliebigem H_0, einem BFGS-Update mit Vollschritt und unter der Bedingung $x_k \to x^\star$, wobei $(x^\star, \lambda^\star, \mu^\star)$ die HOB 2. Ordnung erfüllt, ist die Konvergenz R-superlinear.*

5.6 Globale Konvergenz

Falls der Startwert weit vom nächsten lokalen Minimum entfernt ist, konvergiert das SQP-Verfahren mit Vollschritt im Allgemeinen nicht. Um zu große Schrittweiten zu vermeiden gibt es zwei Ansätze:

1. *Liniensuche:*
$$x_{k+1} = x_k + t_k \Delta x_k \quad t_k \in (0,1]$$

2. *Trust region:* Man fügt dem QP die Nebenbedingung $\|d\| \leq \delta_k$ hinzu und löst $\min \nabla^T f_k d + \frac{1}{2} d^T H d$ mit $A(x)d = -g(x)$ und $\|d\| \leq \delta_k$ für d_k und setzt dann die Berechnung mit $x^{k+1} = x^k + d^k$ fort.

Beide beruhen auf dem Konzept der Güte-Funktion (merit function), welche den Fortschritt der Iteration misst. Wir betrachten nur die Liniensuche.

Index