

Einführung in die stetige Optimierung
–Kurzsript zur Vorlesung–

P. Spellucci

SS 2007

Vorbemerkung: Dem einführenden Charakter der Veranstaltung entsprechend werden hier die allermeisten Aussagen nicht bewiesen. Wo keine Literaturstelle angegeben ist, findet der interessierte Leser die Beweise in meinem Lehrbuch “Numerische Methoden der nichtlinearen Optimierung“. Manche dieser Beweise sind recht aufwendig.

Viele der Verfahren, die in diesem Text angesprochen werden, können auf unserem Server

<http://numawww.mathematik.tu-darmstadt.de:8081>

erprobt werden, was (fast) ohne Programmierkenntnisse möglich ist. Diese sind im Text durch ein eingerücktes

NUMAWWW

gekennzeichnet. Ferner ist auch die Benutzung der optimization toolbox von MATLAB möglich. Weitere Software und viele weitere Hilfsmittel findet man im “decision tree on optimization software“

<http://plato.asu.edu/guide.html>

Wichtige Begriffe aus der Optimierung allgemein sind erklärt in dem Glossary

<http://glossary.computing.society.informs.org/index.html>

Korrekturen, Verbesserungsvorschläge und weitere Anregungen sind erwünscht.

Inhaltsverzeichnis

A Unrestringierte Minimierung	5
A.1 Einführung	5
A.2 Charakterisierung lokaler Minimalstellen, hinreichende Kriterien	11
A.2.1 Der Fall $n = 1$	11
A.2.2 Der Fall $n \geq 1$	11
A.3 Verfahren der unrestringierten Optimierung	19
A.3.1 Unrestringierte Minimierung, eindimensional	19
A.3.1.1 Einschachtelungsverfahren	19
A.3.1.2 Schrittweitenbestimmung durch Nullstellensuche	29
A.3.2 Verfahren der unrestringierten Minimierung, $n > 1$	29
A.3.2.1 Allgemeine Verfahrensstruktur: Line-Search Methoden	30
A.3.2.2 Schrittweitenverfahren	32
A.3.2.3 Richtungsbestimmung	36
A.3.2.3.1 Newtonähnliche- und Quasi-Newtonverfahren	38
A.3.2.3.2 Verfahren von cg-Typ	46
A.3.2.4 Vertrauensbereichmethoden	49
A.3.2.5 Verfahren, die nur Funktionswerte benutzen	51
A.3.3 Verfahren zur Minimierung einer Summe von Quadraten (Ausgleichsrechnung)	54
A.3.3.1 Lineare Ausgleichsrechnung	55
A.3.3.2 Nichtlineare Ausgleichsrechnung	57
A.3.3.3 Orthogonale Regression	60

B	Restringierte Optimierung	63
B.1	Einführung	63
B.2	Extremalkriterien	65
B.3	Verfahren	74
B.3.1	Klassische Penalty- und Barriereverfahren	74
B.3.2	Die Multiplikator-Methoden von Powell, Hestenes und Rockafellar	83
B.3.2.1	Gleichungsrestringierte Probleme	83
B.3.2.2	Ungleichungsrestringierte Probleme (Methode von Rockafel- lar)	88
B.3.3	Exakte differenzierbare Penalty-Funktionen (ERG)	91
B.3.3.1	Primale exakte differenzierbare Penalty-Funktionen	91
B.3.3.2	Primal-duale exakte Penalty-Funktionen	92
B.3.4	Primale Verfahren für linear restringierte Probleme	93
B.3.4.1	Das LP-Problem: Simplexverfahren von Dantzig	93
B.3.4.2	Ein Algorithmus für das definite quadratische Optimierungs- problem QP	103
B.3.4.3	Minimierung einer allgemeinen Funktion unter linearen Gleichungs- und Ungleichungsrestriktionen	108
B.3.5	SLP- und SQP-Verfahren	112
B.4	Semidefinite Optimierung	120
C	Anhang: Die Bunch-Parlett-Zerlegung	133
D	Literatur	135
E	Wichtige Quellen in Internet	137
F	Notation, Formeln	141

Kapitel A

Unrestringierte Minimierung

A.1 Einführung

Wir betrachten in diesem Kapitel den Fall ohne Restriktionen, d.h. formal $m = 0$, $p = 0$.

Dieser Fall ist schon für sich interessant, kommt aber auch als Hilfskonstruktion bei der allgemeinen restringierten Optimierung vor.

Aufgabenstellung:

$f : \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}; \quad f \in C^1(\mathcal{D}), \quad \mathcal{D}$ offen.
Gesucht: (lokale) Minimalstelle x^* von f , d.h. es gibt $\delta > 0$:
 $f(x^*) \leq f(x)$ für alle x mit $\|x - x^*\| < \delta$.

Die Existenz einer Lösung kann nur unter Zusatzvoraussetzungen garantiert werden (Gegenbeispiel: $n = 1$, $f(x) = \exp(x)$).

Eine **hinreichende Voraussetzung für die Existenz** mindestens einer lokalen Minimalstelle ist die folgende:

Es gibt $x^0 \in \mathcal{D} : \mathcal{L}_f(x^0) = \{x \in \mathcal{D} : f(x) \leq f(x^0)\}$ kompakt (beschränkt und abgeschlossen), d.h. der Rand eines Niveaubereichs ist nicht auch Rand von \mathcal{D} .

Beispiele:

1. $\mathcal{D} = \{x : x_i > 0, i = 1, \dots, n\}$

$$f(x) = \sum_{i=1}^n (x_i)^2 - \sum_{i=1}^n i x_i - \sum_{i=1}^n \ln x_i$$

hat genau ein lokales Minimum, das auch globales Minimum ist.

2. $\mathcal{D} = \mathbb{R}^2$, $f(x_1, x_2) = 2(x_1)^3 + (x_2)^2 + (x_1)^2(x_2)^2 + 4x_1x_2$
hat ein lokales, aber kein globales Minimum (die Funktion ist nach unten unbeschränkt).
3. $\mathcal{D} = \mathbb{R}^2$, $f(x_1, x_2) = x_1x_2$ hat kein lokales Minimum

Einige Höhenliniendarstellungen mögen verdeutlichen, mit welchen Schwierigkeiten man schon in diesem “einfachen“ Fall zu rechnen hat. Das erste Beispiel, die häufig in Tests benutzte Rosenbrockfunktion (ein Polynom vom Grad 4 in zwei Veränderlichen), hat nur eine Gradientennullstelle, die eine strenge globale Minimalstelle ist. Diese liegt in einem langgestreckten Tal, das entlang der Parabel $x_2 = x_1^2$ verläuft. Lokal wird die Minimalstelle umgeben von konzentrischen Ellipsen als Niveaulinien mit einem Achsenverhältnis von etwa 1:50. Wenn man also eine systematische Verkleinerung der Funktionswerte fordert, bedeutet dies, daß man sich in einer Achsenrichtung “quer zum Tal“ nur 1/50 so weit bewegen darf wie in der dazu orthogonalen. Einen solchen Fall nennt man “schlecht konditioniert“.

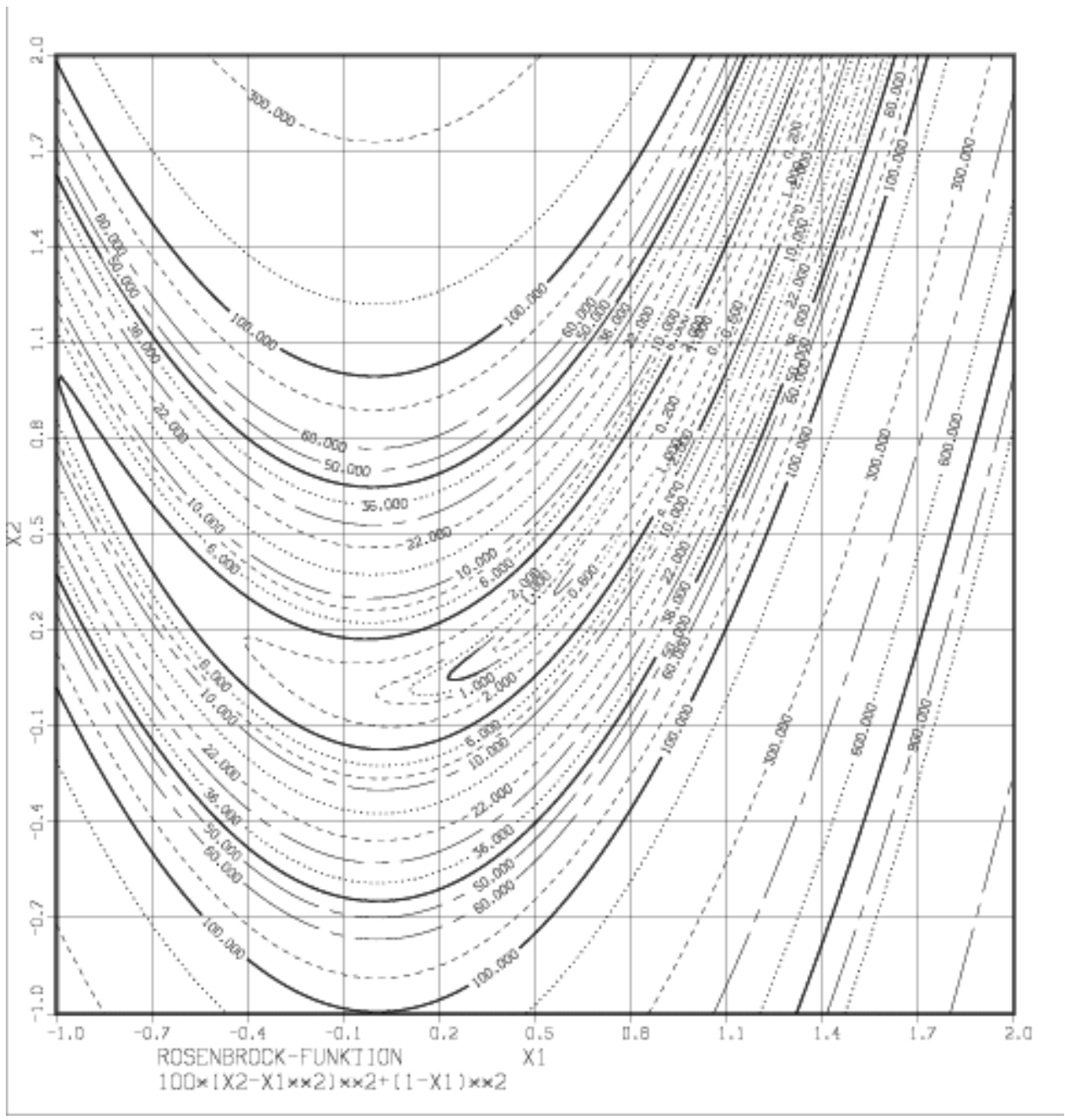
Der nächste Fall ist ebenfalls der eines Polynomes vom Grad 4 in zwei Veränderlichen (Beispiel von Himmelblau). Hier gibt es nun 8 Gradientennullstellen, davon ein lokales Maximum, 3 Sattelpunkte und vier Minimalstellen, alle mit dem globalen Optimalwert 0. Hier können kleine Änderungen in den Startwerten zu verschiedenen Lösungen führen.

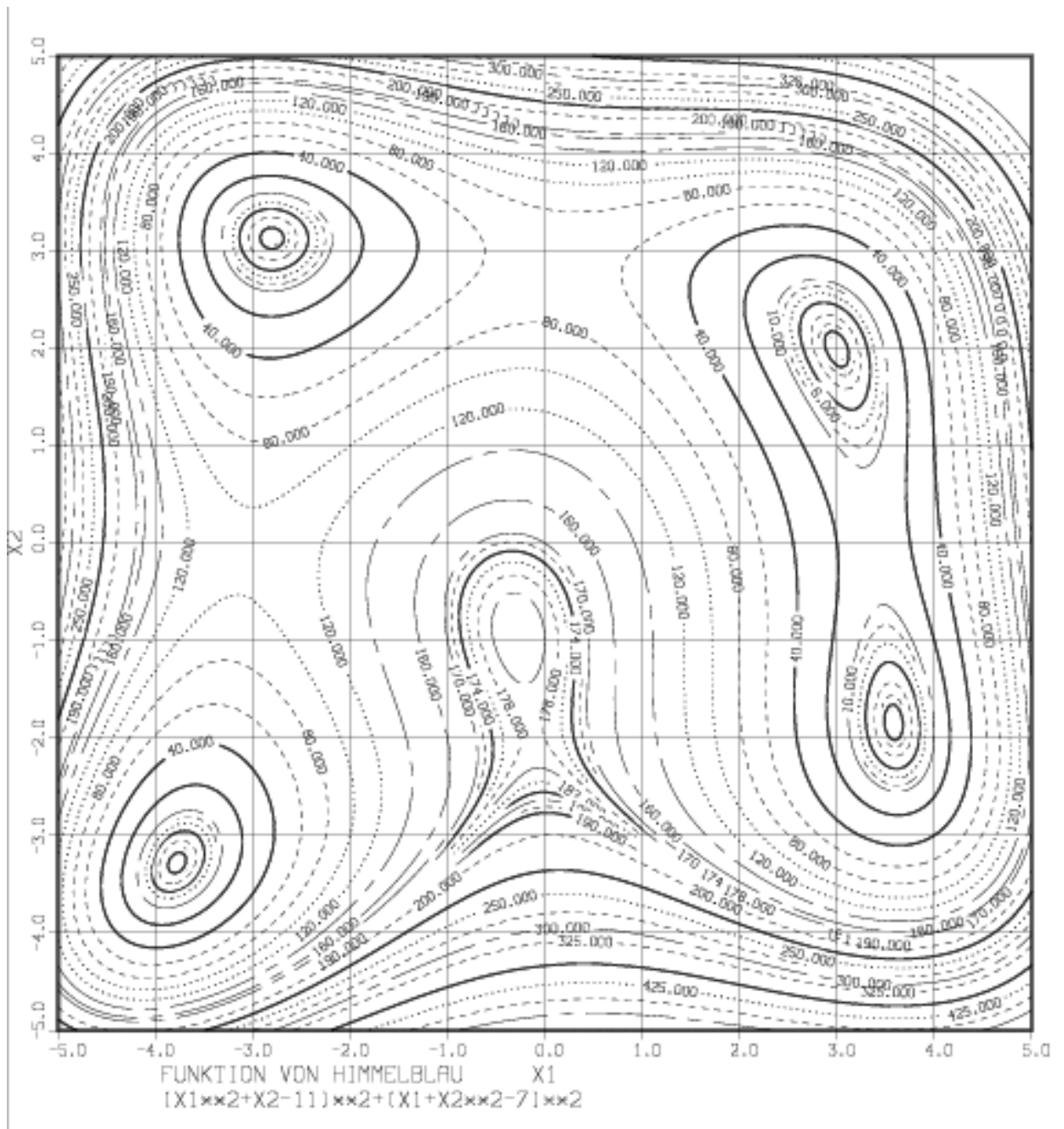
Als nächstes folgt das Beispiel von Beale, ein Polynom vom Grad 6 in zwei Veränderlichen. Hier gibt es drei Gradientennullstellen: zwei Sattelpunkte bei

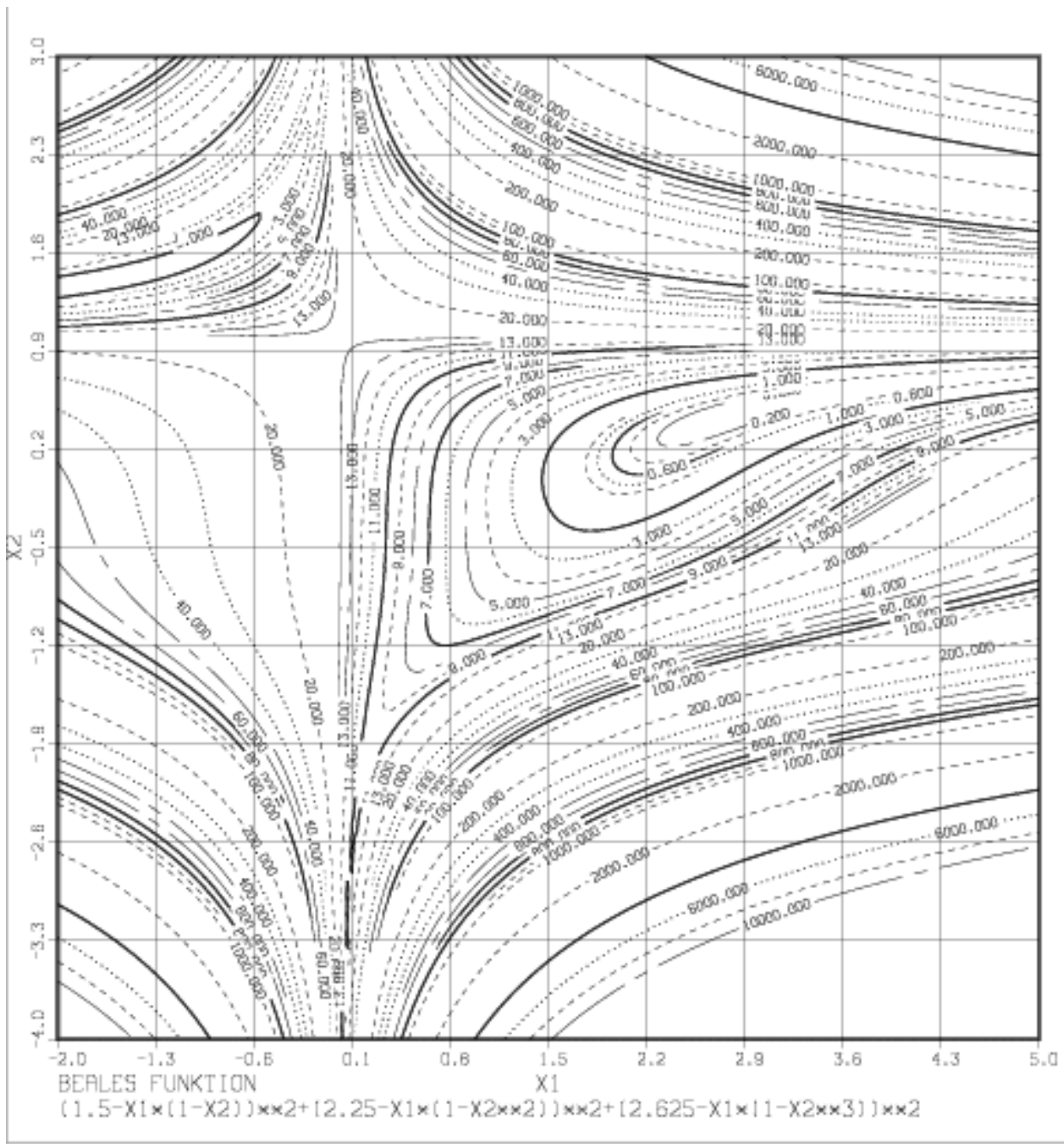
$$(0.10053793732416, -2.64451358502313) \text{ und } (0, 1)$$

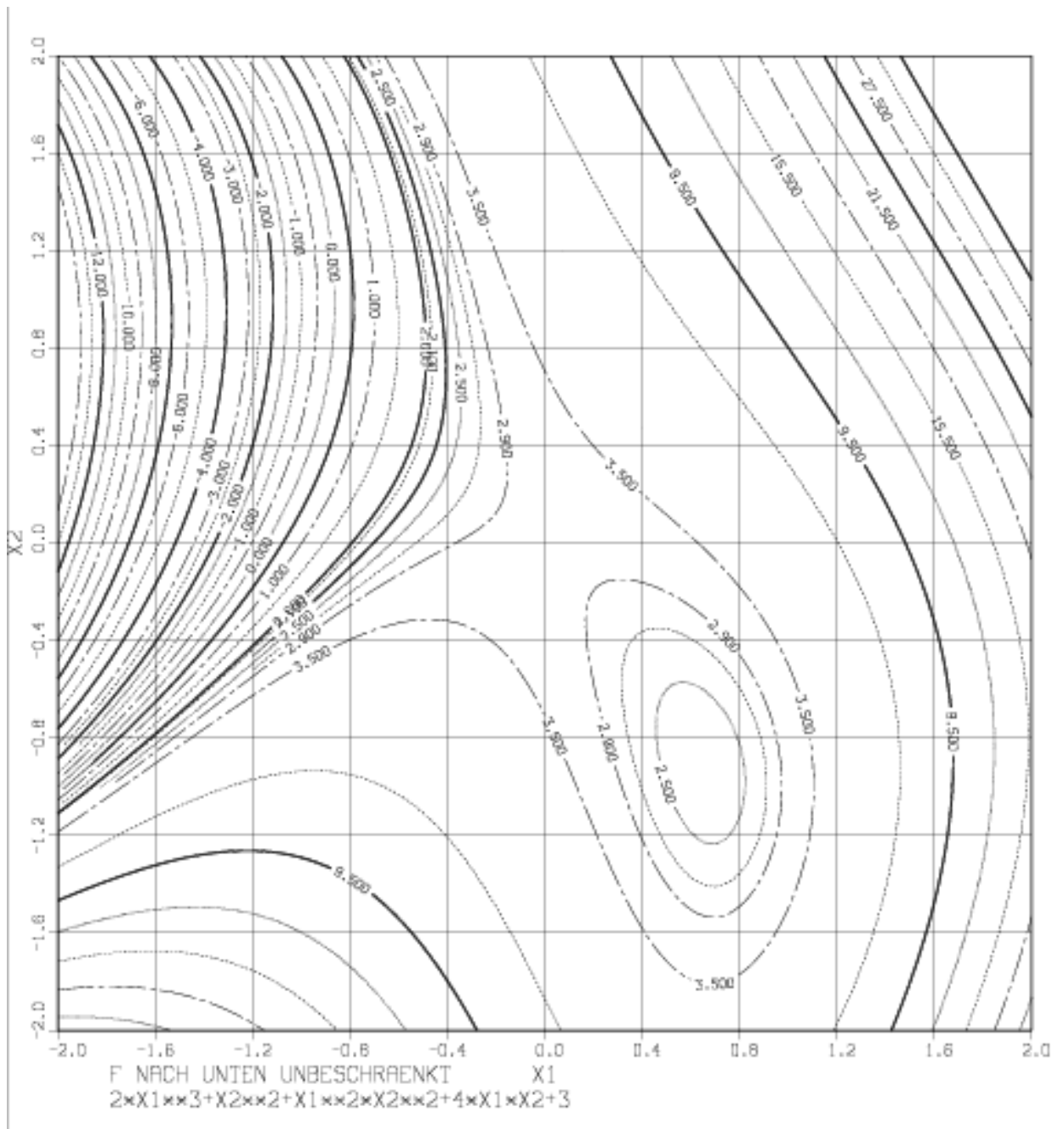
und eine Minimalstelle bei (3,0.5). Als Besonderheit treten hier unbeschränkte Niveaubereiche auf. Wegen der extrem steilen und langgestreckten “Täler“ erweist sich diese analytisch so harmlos aussehende Funktion bereits als recht schwierig für die Minimierung.

Schliesslich ein Polynom in zwei Veränderlichen von Grad 4 mit einer lokalen, aber keiner globalen Minimalstelle. Wenn man hier keine gute Startnäherung für die Minimalstelle benutzt, wird man fast immer Divergenz eines Minimierungsverfahrens beobachten.









A.2 Charakterisierung lokaler Minimalstellen, hinreichende Kriterien

A.2.1 Der Fall $n = 1$

Satz A.1. *Ist f in einer Umgebung von x^* stetig differenzierbar und x^* lokale Minimalstelle von f , dann ist $f'(x^*) = 0$. (Notwendige Bedingung erster Ordnung)*

Satz A.2. *Ist f in einer Umgebung von x^* $2k$ -mal stetig differenzierbar und gilt*

$$f'(x^*) = 0, \dots, \quad f^{(2k-1)}(x^*) = 0, \quad f^{(2k)}(x^*) > 0$$

dann ist x^ strenge lokale Minimalstelle von f , d.h. $f(x) > f(x^*)$ für alle $x \neq x^*$ mit $|x - x^*|$ hinreichend klein. □*

Beweisskizze: Benutze Taylorentwicklung in x^* bis zur Ordnung 1 bzw. $2k$.

A.2.2 Der Fall $n \geq 1$

Satz A.3. *Es sei $f: \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^1(\mathcal{D})$, \mathcal{D} offen, $x^* \in \mathcal{D}$. Falls x^* lokale Minimalstelle von f ist, gilt notwendig*

$$\nabla f(x^*) = 0 \quad (\text{notwendige Bedingung 1. Ordnung})$$

Ist $f \in C^2(\mathcal{D})$, dann gilt weiterhin:

$$\nabla^2 f(x^*) \text{ positiv semidefinit} \quad (\text{notwendige Bedingung 2. Ordnung})$$

□

Beweisskizze: Taylorentwicklung bis zur 1. bzw. 2. Ableitung. Ist $\nabla f(x^*) \neq 0$ dann betrachte $x^* - \tau \nabla f(x^*)$ für kleines τ und im Fall $\nabla f(x^*) = 0$, aber $\nabla^2 f(x^*)$ nicht positiv semidefinit betrachte man $x^* - \tau z$, wo z eine sogenannte Richtung negativer Krümmung ist, d.h. $z^T \nabla^2 f(x^*) z < 0$ (z.B. mit z als Eigenvektor zum algebraisch kleinsten Eigenwert).

Satz A.4. *Es sei $f: \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^2(\mathcal{D})$, \mathcal{D} offen, $x^* \in \mathcal{D}$. Falls gilt*

$$\nabla f(x^*) = 0, \quad \nabla^2 f(x^*) \text{ positiv definit}$$

(Hinreichende Bedingung zweiter Ordnung)

dann ist x^ strenge lokale Minimalstelle von f . □*

Definition A.5. Eine symmetrische reelle Matrix A heißt **positiv definit**, wenn gilt

$$x^T A x > 0 \quad \text{für alle } x \neq 0$$

und **positiv semidefinit**, wenn

$$x^T A x \geq 0 \quad \text{für alle } x.$$

□

Bemerkung A.6. Ist $B \in \mathbb{R}^{m \times n}$ und $\text{Rang}(B) = n$, d.h. $Bx = 0$ nur für $x = 0$, dann ist $A = B^T B \in \mathbb{R}^{n \times n}$ symmetrisch und positiv definit. □

Wir benötigen auch algorithmisch eine möglichst einfache Möglichkeit der Überprüfung einer Matrix auf positive Definitheit. Diese können wir aus dem folgenden Satz erhalten:

Satz A.7. Eine symmetrische reelle Matrix A ist positiv definit genau dann, wenn eine der folgenden gleichwertigen Aussagen gilt:

- a) Alle Eigenwerte von A sind positiv. (> 0)
- b) Alle Hauptabschnittsunterdeterminanten sind positiv, d.h.

$$a_{11} > 0, \quad \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} > 0, \quad \dots, \det(A) > 0.$$

- c) A kann zerlegt werden in ein Produkt

$$A = L \cdot D \cdot L^T$$

mit unterer Dreiecksmatrix L mit Diagonale $(1, \dots, 1)$ und Diagonalmatrix $D = \text{diag}(d_{11}, \dots, d_{nn})$ mit $d_{ii} > 0$, $i = 1, \dots, n$.

□

Hier ist es die Charakterisierung unter c), die algorithmisch am einfachsten auswertbar ist.

Ein Algorithmus für die Berechnung dieser Zerlegung einer Matrix

$$A = LDL^T$$

mit einer unteren Dreiecksmatrix L mit Diagonale $(1, \dots, 1)$ und einer Diagonalmatrix $D = \text{diag}(d_{11}, \dots, d_{nn})$ mit positiven d_{ii} ist der folgende: (formal durchführbar solange $d_{ii} \neq 0$)

A.2. CHARAKTERISIERUNG LOKALER MINIMALSTELLEN, HINREICHENDE KRITERIEN 13

$$\begin{aligned}
 & i = 1, \dots, n \\
 & d_{ii} = a_{ii} - \sum_{k=1}^{i-1} (l_{ik})^2 d_{kk}, \quad l_{ii} := 1 \\
 & j = i + 1, \dots, n : \\
 & l_{ji} = \left(a_{ji} - \sum_{k=1}^{i-1} l_{ik} l_{jk} d_{kk} \right) / d_{ii}
 \end{aligned}$$

modifizierter
Cholesky-Algorithmus

Zahlenbeispiel:

$$\begin{bmatrix} 4 & & & \\ 6 & 18 & & \\ 2 & 3 & 2 & \\ -10 & -9 & -2 & 54 \end{bmatrix} \begin{array}{l} \text{symm} \\ \\ \\ \end{array} \rightarrow \begin{array}{l} 4 = d_{11} \\ 3/2 \\ 1/2 \\ -5/2 \end{array} \begin{array}{l} \\ 9 = d_{22} \\ 0 \\ 2/3 \end{array} \begin{array}{l} \\ \\ 1 = d_{33} \\ 3 \end{array} \begin{array}{l} \\ \\ \\ 16 = d_{44} \end{array}$$

Bemerkung A.8. Aus der LDL^T -Zerlegung erhält man die Choleskyzerlegung

$$A = \tilde{L}\tilde{L}^T$$

mittels

$$\tilde{L} = L\sqrt{D} = L \operatorname{diag}(\sqrt{d_{11}}, \dots, \sqrt{d_{nn}}).$$

Hier wird nun die Positivität benötigt, um im Reellen zu bleiben.

Bemerkung A.9. Die LDL^T -Zerlegung kann auf den Fall einer Blockdiagonalmatrix mit 1×1 und 2×2 Blöcken so verallgemeinert werden, daß die sogenannte Signatur von A mit der von D übereinstimmt und die Berechnung numerisch stabil bleibt. Siehe im Anhang unter "Bunch-Parlett-Zerlegung".

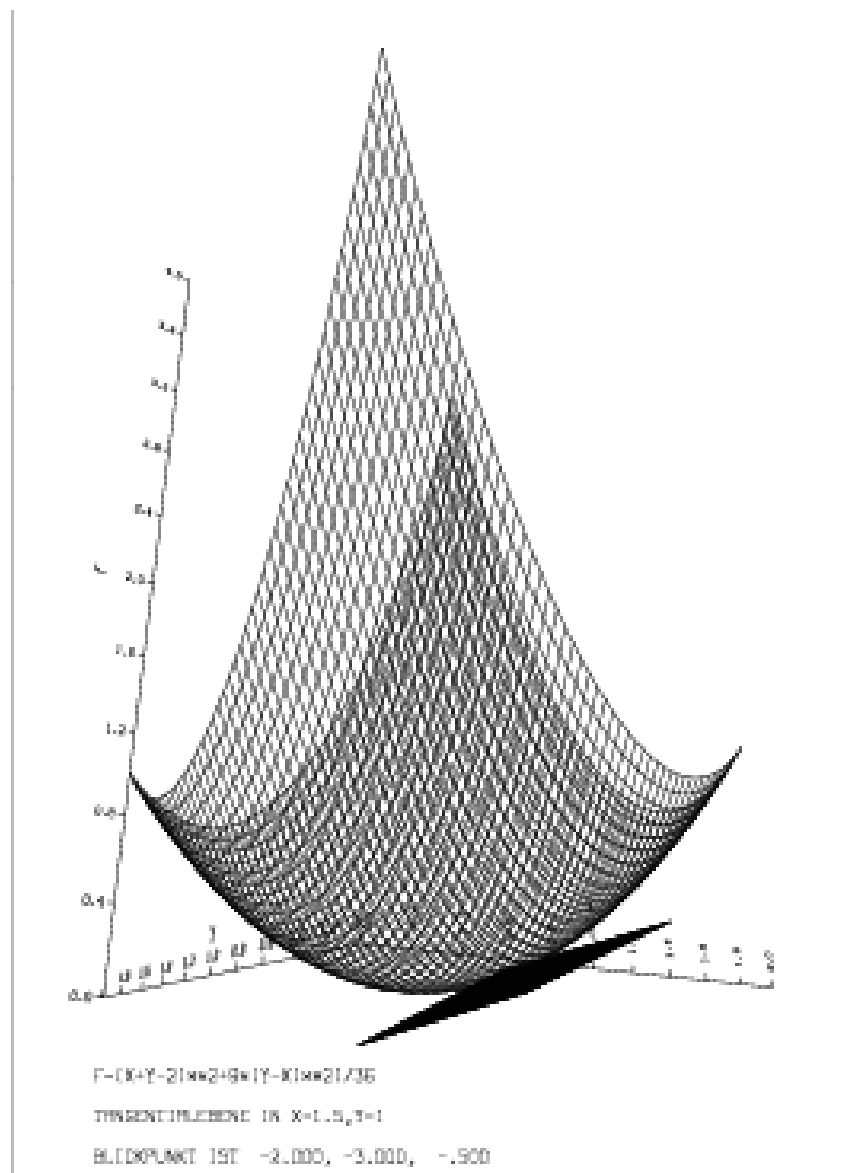
Eine hinreichende Bedingung, unter der ein lokales Minimum von f auch globales Minimum ist, ist die "Konvexität" der Funktion:

Definition A.10. $\mathcal{D} \subset \mathbb{R}^n$ heißt **konvex**, wenn mit $x \in \mathcal{D}$ $y \in \mathcal{D}$ auch stets $[x, y] \subset \mathcal{D}$.

Definition A.11. $f : \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}$, \mathcal{D} konvex, heißt **konvex** auf \mathcal{D} , wenn mit $x, y \in \mathcal{D}$ gilt

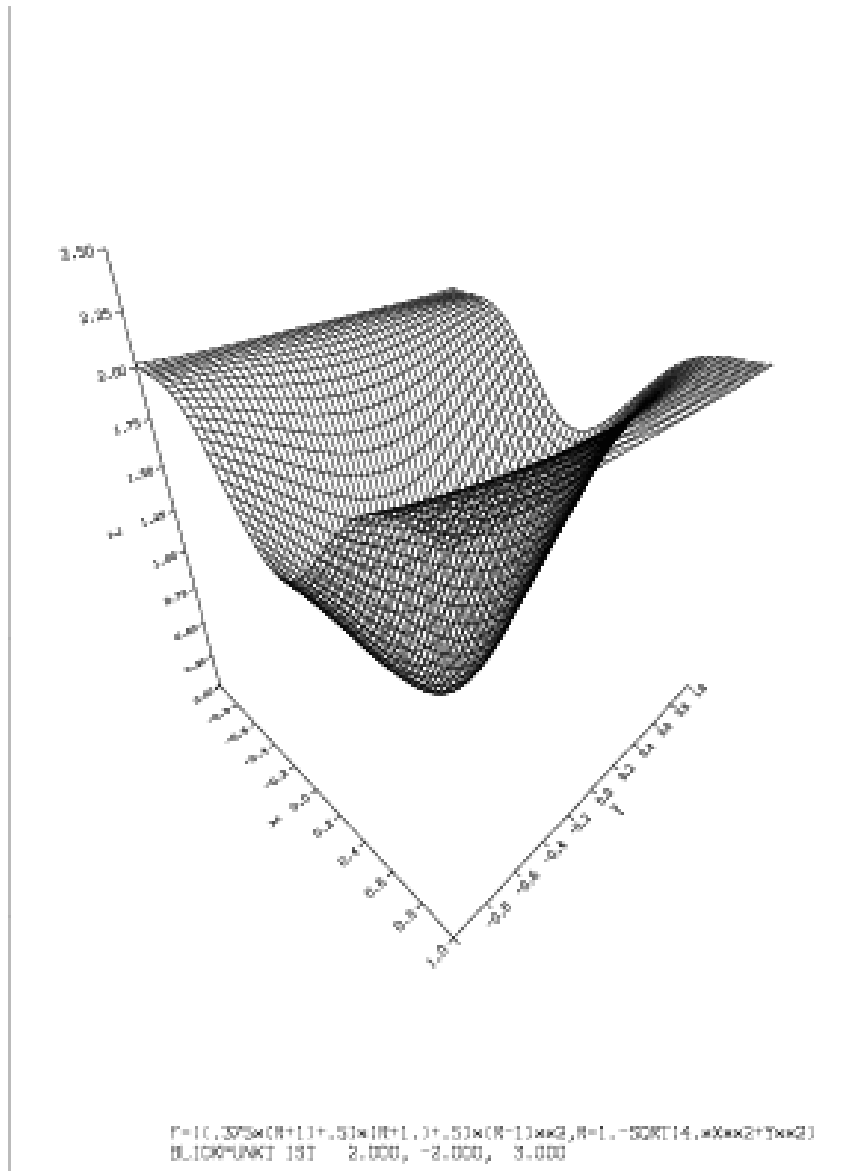
$$\lambda f(x) + (1 - \lambda)f(y) \geq f(\lambda x + (1 - \lambda)y) \quad \text{für } \lambda \in [0, 1] \tag{A.1}$$

und **streng konvex**, wenn für $0 < \lambda < 1$ nur " $>$ " gilt in (A.1). □

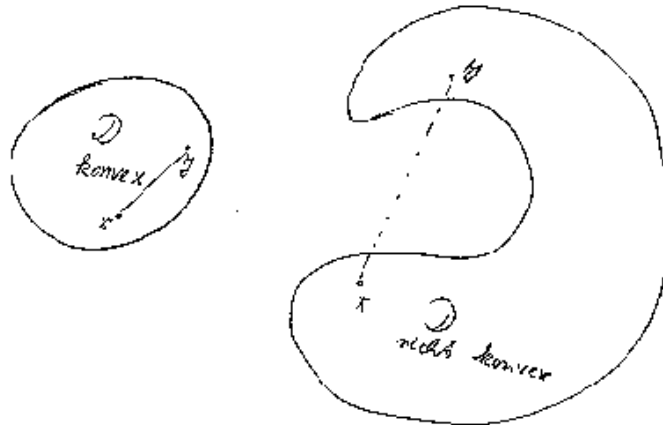


Konvexe Funktion

A.2. CHARAKTERISIERUNG LOKALER MINIMALSTELLEN, HINREICHENDE KRITERIEN 15



quasikonvexe, aber nicht konvexe Funktion



Satz A.12. Ist $\mathcal{D} \neq \emptyset \subset \mathbb{R}^n$ konvex und f konvex auf \mathcal{D} , dann ist jedes lokale Minimum auch globales Minimum. □

Beweisskizze: Sei x^* lokale Minimalstelle. Dann gibt es ein $\delta > 0$ sodaß

$$\|y - x^*\| \leq \delta \Rightarrow f(x^*) \leq f(y).$$

Sei $x \in \mathcal{D}$ beliebig. Betrachte $x^* + t(x - x^*)$. Nach Voraussetzung ist

$$f(x^* + t(x - x^*)) \leq (1 - t)f(x^*) + tf(x)$$

und

$$\|(x^* + t(x - x^*)) - x^*\| \leq \delta \text{ falls } 0 < t < \delta / (\|x\| + \|x^*\|).$$

Dies ergibt

$$f(x^*) \leq f(x^* + t(x - x^*)) \leq (1 - t)f(x^*) + tf(x)$$

für diese t . Umordnen und Division durch $t > 0$ ergibt die Behauptung.

Eine Charakterisierung differenzierbarer konvexer Funktionen liefert

Satz A.13. (Konvexitätskriterien) $\mathcal{D} \subset \mathbb{R}^n$ sei konvex, offen ($\neq \emptyset$). Ist $f \in C^1(\mathcal{D})$, dann gilt

1. f konvex auf $\mathcal{D} \Leftrightarrow f(y) \geq f(x) + \nabla f(x)^T(y - x)$ für alle $x, y \in \mathcal{D}$.

Ist $f \in C^2(\mathcal{D})$, dann gilt zusätzlich

2. f konvex auf $\mathcal{D} \iff \nabla^2 f(x)$ positiv semidefinit auf \mathcal{D} .

3. Ist $\nabla^2 f$ positiv definit auf \mathcal{D} , dann ist f streng konvex auf \mathcal{D} , d.h.

$$\lambda f(x) + (1 - \lambda)f(y) > f(\lambda x + (1 - \lambda)y) \quad \text{für } 0 < \lambda < 1$$

und beliebige $x, y \in \mathcal{D}$.

□

Korollar: $f \in C^1(\mathcal{D})$, \mathcal{D} offen konvex, $x^* \in \mathcal{D}$, f konvex auf \mathcal{D} , $\nabla f(x^*) = 0 \Rightarrow x^*$ Minimalstelle. □

In Satz A.12 wird die Existenz eines lokalen Minimums vorausgesetzt.

Das folgende Kriterium sichert die **Existenz und Eindeutigkeit** eines lokalen und zugleich globalen Minimums:

Satz A.14. Es sei $f \in C^2(\mathcal{D})$. \mathcal{D} sei offen und konvex ($\neq \emptyset$). Falls gilt:

(i) Es gibt ein $\alpha_0 \in \mathbb{R}$, sodaß $\mathcal{L}_f(\alpha_0) = \{x \in \mathcal{D} : f(x) \leq \alpha_0\}$ beschränkt und abgeschlossen ist oder es ist $\mathcal{D} = \mathbb{R}^n$.

(ii) $d^T \nabla^2 f(x) d \geq \alpha d^T d$ mit $\alpha > 0$ für alle $x \in \mathcal{D}$ und $d \in \mathbb{R}^n$, dann gibt es genau eine strenge lokale Minimalstelle von f auf \mathcal{D} , die zugleich globale Minimalstelle ist.

□

Ergänzung Definition A.11: f heißt gleichmäßig konvex auf der konvexen Menge $\mathcal{D} \subset \mathbb{R}^n$, wenn es ein $\gamma > 0$ gibt mit

$$t f(x) + (1 - t)f(y) \geq f(tx + (1 - t)y) + t(1 - t)\gamma \|x - y\|^2$$

für alle $t \in [0, 1]$ und $x, y \in \mathcal{D}$. (Dies ist genau unter den Voraussetzungen von Satz A.14 der Fall.)

Beispiele:

1. $f(x) = \exp(-x_1) + \exp(-x_2)$ ist streng konvex auf $\mathcal{D} = \mathbb{R}^2$. Es existiert **kein** lokales Minimum. (Aber: $\nabla^2 f(x) = \begin{pmatrix} \exp(-x_1) & 0 \\ 0 & \exp(-x_2) \end{pmatrix}$ ist positiv definit!)

2. $f(x) = \frac{3}{2}((x_1)^2 + (x_2)^2) + \sin x_1 \cdot \sin x_2 + 3x_1 - 4x_2$.

$$\nabla^2 f(x) = \begin{pmatrix} 3 - \sin x_1 \sin x_2 & \cos x_1 \cos x_2 \\ \cos x_1 \cos x_2 & 3 - \sin x_1 \sin x_2 \end{pmatrix}$$

$\alpha = 1$ in (ii), Satz A.14. $\mathcal{D} = \mathbb{R}^2$. Also existiert genau eine Gradientennullstelle, die das einzige lokale und zugleich globale Minimum liefert.

3. $f(x) = 1 + x_1 \ln x_1 + x_2 \ln x_2 + (x_1)^2 + (x_2)^2$

$$\mathcal{D} = \{x : x_1 > 0, x_2 > 0\}$$

$$x^0 = (0.25, 0.25); f(x^0) = 0.4318528194$$

$\mathcal{L}_f(f(x^0))$ ist kompakt, weil $f \rightarrow \infty$ für $x_1 \rightarrow \infty$ oder $x_2 \rightarrow \infty$ und auf dem Rand von \mathcal{D} $f \geq 0.7148671412$ gilt.

(Das ist der Minimalwert von $1 + z \ln z + z^2$ $z \geq 0$).

$$\nabla^2 f(x) = \begin{pmatrix} 2 + \frac{1}{x_1} & 0 \\ 0 & 2 + \frac{1}{x_2} \end{pmatrix}$$

also $\alpha = 2$ in Satz A.14.

Satz A.12 läßt sich auf andere Funktionentypen übertragen, z.B.

Definition A.15. $f : \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}$, \mathcal{D} konvex, $\neq \emptyset$, heißt **quasikonvex** auf \mathcal{D} , wenn

$$f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\} \quad \text{für alle } \lambda \in [0, 1], x, y \in \mathcal{D} \quad (\text{A.2})$$

und **streng quasikonvex**, wenn für $0 < \lambda < 1$ in (A.2) " $<$ " gilt. □

Satz A.16. $f : \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}$, \mathcal{D} konvex, $\neq \emptyset$, ist quasikonvex genau dann, wenn alle Niveaubereiche $\mathcal{L}_f(\alpha)$ von f auf \mathcal{D} konvex (oder leer) sind. □

Eine konvexe Funktion ist natürlich auch quasikonvex.

Satz A.17. Sei f auf der nichtleeren konvexen Menge \mathcal{D} quasikonvex, stetig und ein Niveaubereich von f auf \mathcal{D} sei kompakt. Dann gibt es eine lokale Minimalstelle. Jede strenge lokale Minimalstelle ist auch globale Minimalstelle. Ist zusätzlich f streng quasikonvex, dann gibt es nur eine Minimalstelle. □

Beispiel: $f(x) = 1 - \exp(-(x_1)^2 - (x_2)^2)$ ist streng quasikonvex auf \mathbb{R}^2 , aber nicht konvex.

A.3 Verfahren der unrestringierten Optimierung

Im Folgenden gelte stets

$$f : \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}; \quad f \in C^1(\mathcal{D})$$

(Meist wird zumindest implizit vorausgesetzt, daß sogar $f \in C^{2,1}(\mathcal{D})$, d.h. $f \in C^2(\mathcal{D})$ und für jede kompakte Teilmenge $\mathcal{D}_1 \subset \mathcal{D}$ existiert ein $L \geq 0$, sodaß

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\| \text{ für alle } x, y \in \mathcal{D}.$$

Es gilt $C^{2,1}(\mathcal{D}) \supset C^3(\mathcal{D})$.

Die üblichen Verfahren bestimmen Folgen $\{x^k\}$ mit

- (i) $x^k \in \mathcal{L}_f(f(x^0)) = \{x \in \mathcal{D} : f(x) \leq f(x^0)\}$
- (ii) $\nabla f(x^k) \rightarrow 0$
- (iii) $x^k - x^{k+1} \rightarrow 0$

(Es gibt auch Verfahren, die zusätzlich noch die notwendige Bedingung zweiter Ordnung “ $\nabla^2 f(x^*)$ positiv semidefinit” für jeden Häufungspunkt von $\{x^k\}$ erzwingen)

Die Bedingungen (i), (ii), (iii) werden konstruktiv erzwungen.

Die Existenz von Häufungspunkten, die Konvergenz der Gesamtfolge und die Minimalität des Grenzwertes x^* folgen nur aus Zusatzannahmen über f .

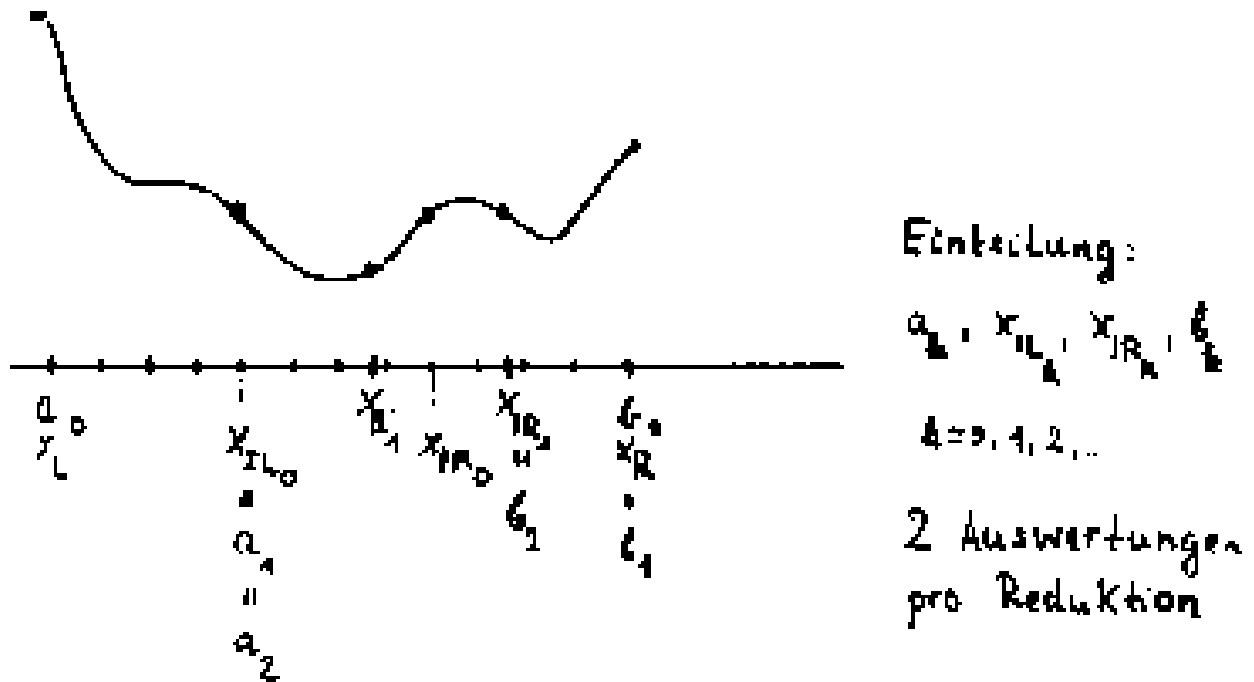
$x^0 \in \mathcal{D}$ sei bekannt.

Die Voraussetzung: $\mathcal{L}_f(f(x^0)) = \{x \in \mathcal{D} : f(x) \leq f(x^0)\}$ kompakt sichert die Existenz von Häufungswerten und die Existenz konvergenter Teilfolgen mit $\nabla f(x^*) = 0$.

A.3.1 Unrestringierte Minimierung, eindimensional

A.3.1.1 Einschachtelungsverfahren

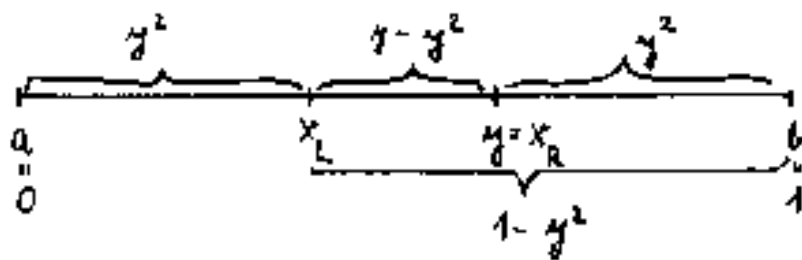
Im Folgenden nehmen wir zur Vereinfachung $\mathcal{D} = \mathbb{R}$ an. Wir wollen die (eine) Minimalstelle von f durch systematisches Suchen auf dem Graphen von f finden, nur unter Benutzung von Funktionsauswertungen. Das Analogon zur Bisektion bei der Nullstellenbestimmung ist hier die Trisektion: Dabei wird nur die Stetigkeit von f benötigt. Dies ist günstig, wenn die f -Werte nur ungenau bekannt sind (“verrauschte Daten“). Die Trisektion benutzt fortgesetzte Intervalldritteln und somit 2 Funktionsauswertungen bei einer Längenreduktion von $2/3$ pro Schritt.



Wir suchen nun eine geschicktere Einteilung des Intervalls, sodaß nur ein Testpunkt pro Intervallverkleinerung benötigt wird und die Intervallreduktion möglichst groß ist. Dies liefert die Einteilung “nach dem Prinzip des goldenen Schnitts”:

$$\frac{\text{Gesamtstrecke}}{\text{größere Teilstrecke}} = \frac{\text{größere Teilstrecke}}{\text{kleinere Teilstrecke}}$$

$$\frac{1}{y} = \frac{y}{1-y}, \quad y > 0 \Rightarrow y = \frac{1}{2}(\sqrt{5} - 1)$$



$$\Rightarrow \frac{y}{(y)^2} = \frac{(y)^2}{y - (y)^2}, \quad \frac{1 - (y)^2}{1 - y} = \frac{1 - y}{y - (y)^2}$$

Hier wird die Strecke $\overline{0, y}$ durch y^2 und die Strecke $\overline{y^2, 1}$ durch $1 - y$ nach dem Prinzip des goldenen Schnittes geteilt.

Aufgabe: $[a, b]$ gegeben. Gesucht: eine Minimalstelle x^* von f auf $[a, b]$.

Voraussetzung: $f \in C[a, b]$.

Algorithmus: Suche nach dem Prinzip des goldenen Schnitts:

$$\varrho := \frac{1}{2}(\sqrt{5} - 1); \quad a^{(0)} := a; \quad b^{(0)} := b; \quad l^{(0)} := (b^{(0)} - a^{(0)})\varrho$$

$$x_R^{(0)} := b - \varrho l^{(0)}; \quad x_L^{(0)} := a + \varrho l^{(0)};$$

Berechne $f(x_R^{(0)})$, $f(x_L^{(0)})$.

$k = 0, 1, 2, \dots$

$$l^{(k+1)} := \varrho l^{(k)}$$

Falls $f(x_R^{(k)}) > f(x_L^{(k)})$, dann

$$b^{(k+1)} := x_R^{(k)}; \quad a^{(k+1)} := a^{(k)};$$

$$x_L^{(k+1)} := a^{(k+1)} + \varrho l^{(k+1)}; \quad x_R^{(k+1)} := x_L^{(k)}$$

Berechne $f(x_L^{(k+1)})$

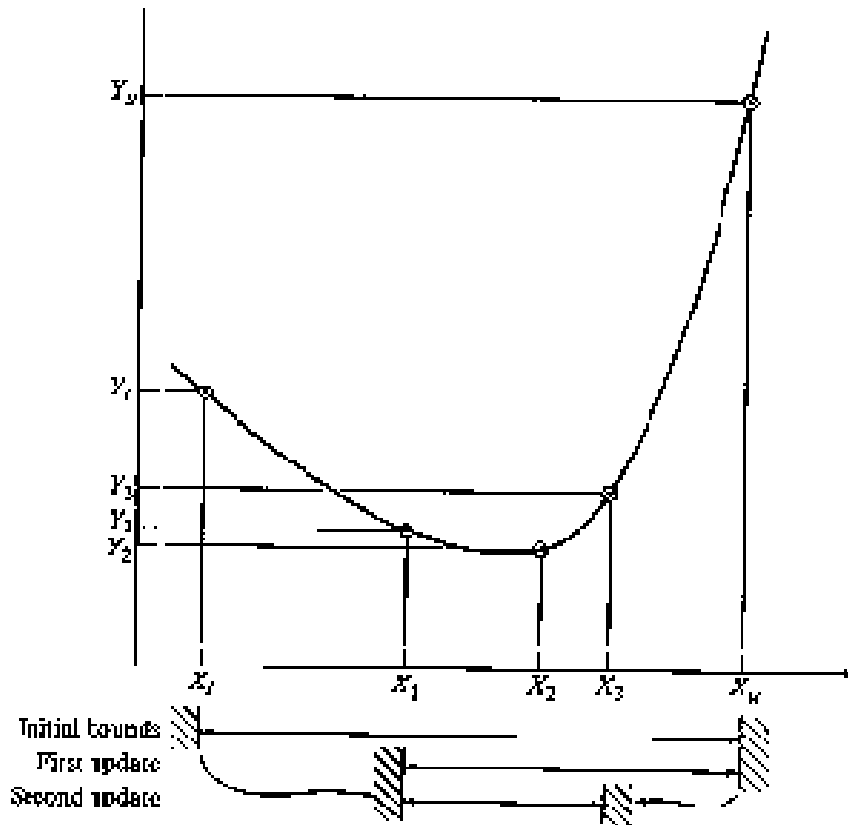
sonst

$$a^{(k+1)} := x_L^{(k)}; \quad b^{(k+1)} := b^{(k)}; \quad x_L^{(k+1)} := x_R^{(k)};$$

$$x_R^{(k+1)} := b^{(k+1)} - \varrho l^{(k+1)}$$

Berechne $f(x_R^{(k+1)})$.

NUMAWWW



Satz A.18. Sei $f : [a, b] \rightarrow \mathbb{R}$ streng quasikonvex auf $[a, b]$. Dann gilt für die Minimalstelle x^* von f $x^* \in [a^{(i)}, b^{(i)}]$ für alle i und

$$l^{(i)} = \varrho(b^{(i)} - a^{(i)}) = (\varrho)^i l^{(0)} \quad \text{für alle } i,$$

also $a^{(i)} \rightarrow x^*$, $b^{(i)} \rightarrow x^*$, d.h. die Konvergenzgeschwindigkeit ist linear. \square

Zur Orientierung

$$\frac{1}{2}(\varrho)^{10} = 0.004.$$

Kritik: Dies ist ein einfacher, zuverlässiger, aber aufwendiger Algorithmus (viele Funktionswerte bei hoher Genauigkeitsforderung)

Aber: Schnellere Verfahren erfordern bessere Differenzierbarkeitseigenschaften von f . Ein Hilfsmittel ist die Polynominterpolation:

Hier ist nur der Polynomgrad 2 oder 3 von Interesse, da wir ja das Minimum des Polynoms einfach bestimmen können müssen.

Zunächst der Fall $k = 2$.

1. Gegeben $x_0 < x_1 < x_2$

$$(x_0, f(x_0)), (x_1, f(x_1)), (x_2, f(x_2)).$$

Gesucht: $p_2 \in \Pi_2$ mit $p_2(x_j) = f(x_j)$, $j = 0, 1, 2$.

Konstruktion: (Newtonsches Interpolationspolynom)

$$\begin{aligned}\Delta_0 &:= \frac{f(x_1) - f(x_0)}{x_1 - x_0} \\ \Delta_1 &:= \frac{f(x_2) - f(x_1)}{x_2 - x_1} \\ \Delta_0^{(2)} &:= \frac{\Delta_1 - \Delta_0}{x_2 - x_0}.\end{aligned}$$

$$\begin{aligned}p_2(x) &= f(x_0) + (x - x_0)\Delta_0 + (x - x_0)(x - x_1)\Delta_0^{(2)} \\ &= f(x_2) + (x - x_2)\Delta_1 + (x - x_2)(x - x_1)\Delta_0^{(2)}.\end{aligned}$$

Fehleraussagen:

Satz A.19. a) Falls $f \in C^3([a, b])$, $x_i \in [a, b]$, dann gilt

$$f(x) - p_2(x) = \frac{f^{(3)}(\xi_x)}{3!}(x - x_0)(x - x_1)(x - x_2)$$

mit $x \in [a, b]$ und einem (unbekannten) $\xi_x \in [a, b]$.

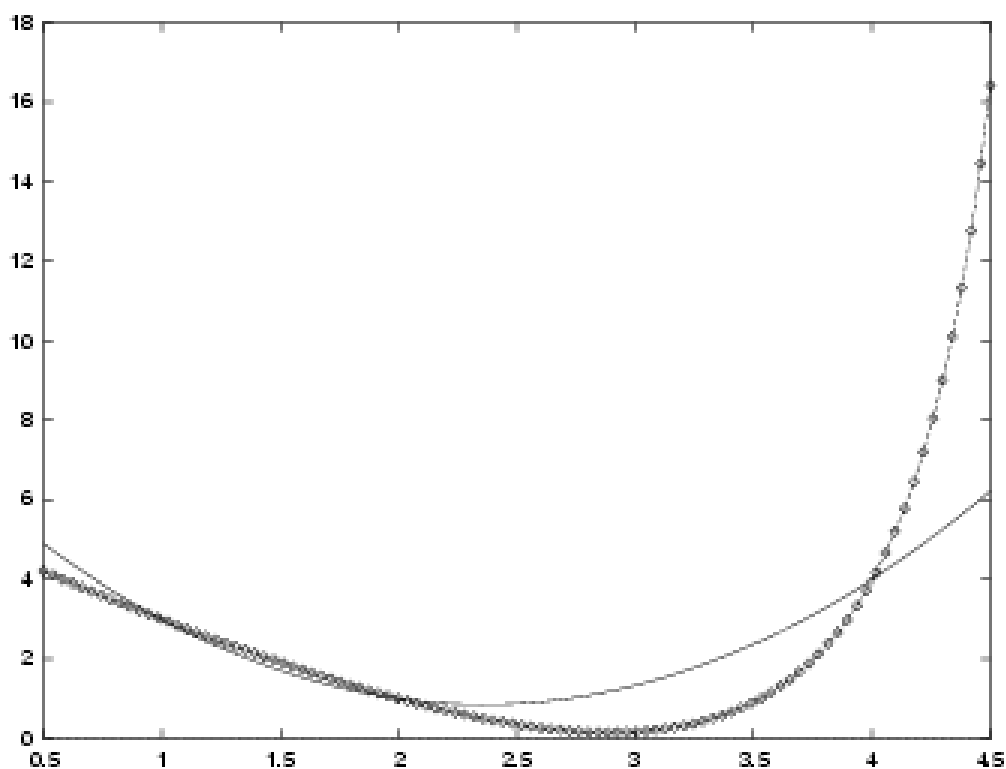
b) Falls $f(x_0) > f(x_1)$, $f(x_2) > f(x_1)$, $f''(x) > 0$ für $x \in [a, b]$ und $f \in C^4[a, b]$, dann gibt es eine eindeutige Minimalstelle x^* von f auf $[a, b]$ und x_p^* von p_2 auf $[a, b]$ und es gilt

$$|x^* - x_p^*| \leq C(x_0 - x_2)^2$$

mit einer Konstanten C , die von Werten von f'' , $f^{(3)}$, und $f^{(4)}$ auf dem Intervall $[x_0, x_2]$ abhängt.

□

Satz A.19b besagt also, daß die Minimalstelle x_p^* von p_2 eine sehr gute Näherung für x^* , die Minimalstelle von f sein wird, wenn $x_2 - x_0$ genügend klein geworden ist.



Die Abbildung zeigt diese Interpolation mit $x_0 = 1$, $x_1 = 2$ und $x_3 = 4$.

2. Der Fall $k = 2$. Gegeben x_0, x_1 , $x_1 > x_0$, sowie $f(x_0), f'(x_0), f(x_1)$.

Gesucht: $p_2 \in \Pi_2$:

$$p_2(x_0) = f(x_0), \quad p_2'(x_0) = f'(x_0), \quad p_2(x_1) = f(x_1).$$

Konstruktion: Newtonsches Interpolationspolynom mit $x_1 \rightarrow x_0$

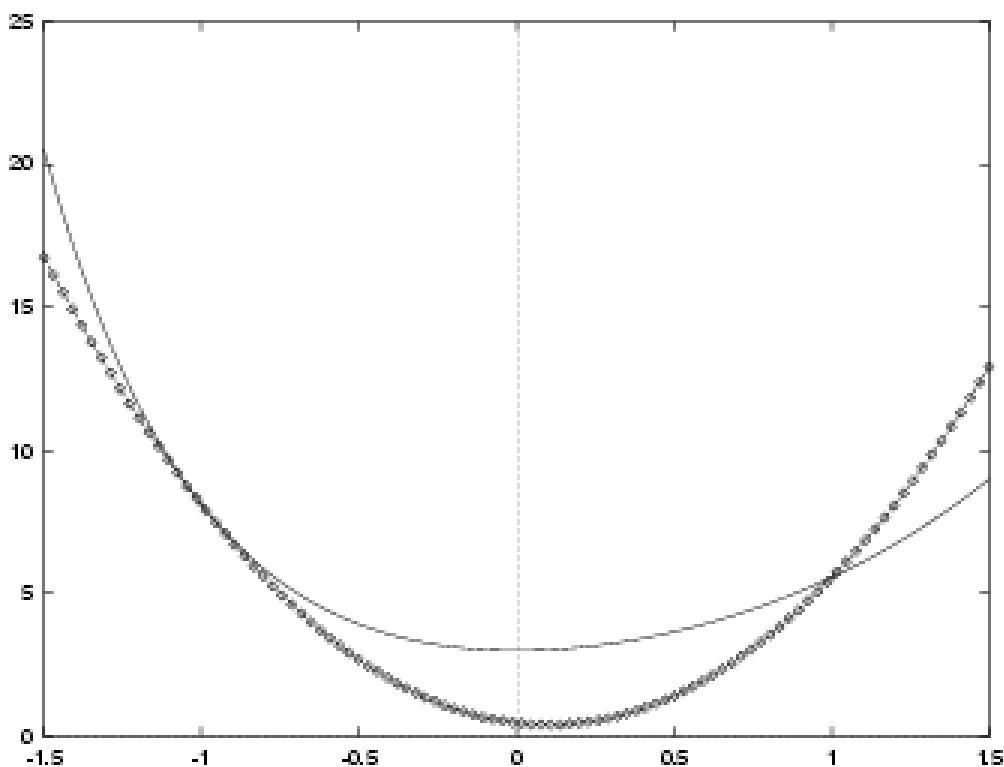
$$\begin{aligned} \Delta_0 &:= f'(x_0) \\ \Delta_1 &:= \frac{f(x_1) - f(x_0)}{x_1 - x_0} \\ \Delta_0^{(2)} &:= \frac{\Delta_1 - \Delta_0}{x_1 - x_0}. \end{aligned}$$

$$p_2(x) = f(x_0) + (x - x_0)f'(x_0) + (x - x_0)^2\Delta_0^{(2)} \quad (\text{A.3})$$

Satz A.20. Die Aussage von Satz A.19 gilt für p_2 aus (A.3) mit $x_2 := x_0$, wenn man die Bedingung $f(x_0) > f(x_1)$, $f(x_2) > f(x_1)$ ersetzt durch die Bedingungen $f'(x_0) < 0$, $f(x_1) > f(x_0) + f'(x_0)(x_1 - x_0)$,

$$x_1 > x_0 - f'(x_0)/(2\Delta_0^{(2)}) =: x_p^*.$$

An die Stelle von x_1 tritt x_0 und an die von x_2 tritt x_1 .



Die Abbildung zeigt diese Situation mit $x_0 = -1$ und $x_1 = 1$.

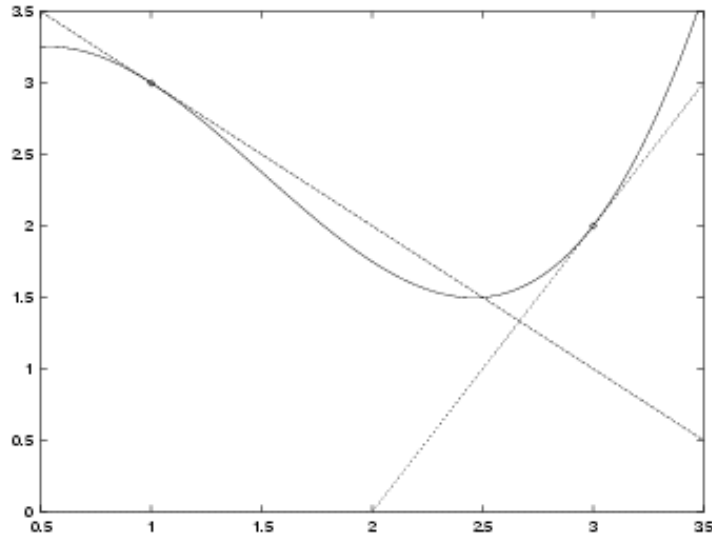
3. Der Fall $k = 3$. Gegeben: $x_0, x_1, f(x_0), f'(x_0), f(x_1), f'(x_1)$.

Gesucht: $p_3 \in \Pi_3$:

$$p_3(x_k) = f(x_k), \quad p_3'(x_k) = f'(x_k) \quad k = 0, 1$$

Konstruktion: Hermitesches Interpolationspolynom

$$\begin{aligned} \Delta_0 &:= f'(x_0) \\ \Delta_1 &:= \frac{f(x_1) - f(x_0)}{x_1 - x_0} & \Delta_0^{(2)} &:= \frac{\Delta_1 - \Delta_0}{x_1 - x_0} \\ \Delta_2 &:= f'(x_1) & \Delta_1^{(2)} &:= \frac{\Delta_2 - \Delta_1}{x_1 - x_0} & \Delta_0^{(3)} &:= \frac{\Delta_1^{(2)} - \Delta_0^{(2)}}{x_1 - x_0} \\ p_3(x) &= f(x_0) + (x - x_0)f'(x_0) + (x - x_0)^2\Delta_0^{(2)} + (x - x_0)^2(x - x_1)\Delta_0^{(3)} \end{aligned}$$



Die Abbildung zeigt ein so bestimmtes Polynom mit $x_0 = 1$ und $x_1 = 3$.

Satz A.21. Falls $f \in C^4[a, b]$, $x_0 < x_1 \in [a, b]$, dann gilt für $x \in [a, b]$

$$f(x) - p_3(x) = \frac{f^{(4)}(\xi(x))}{24} \cdot (x - x_0)^2(x - x_1)^2.$$

Falls $f \in C^5[a, b]$ und $f''(x) > 0$ auf $[a, b]$ sowie $f'(x_0)f'(x_1) < 0$, dann gilt für die Differenz zwischen der eindeutigen Minimalstelle x^* von f und der eindeutigen lokalen Minimalstelle x_p^* von p_3

$$|x^* - x_p^*| \leq C|x_1 - x_0|^3$$

mit einer Konstanten C , die von Werten von f'' , $f^{(4)}$ und $f^{(5)}$ auf dem Intervall $[x_0, x_1]$ abhängt. \square

Auf der Methode 1. beruht die Minimierung durch fortgesetzte quadratische Interpolation. Ausgehend von $x_0 < x_1 < x_2$ bestimmt man x_p^* und x_p^* ersetzt dann einen dieser drei Punkte. Kommt x_p^* zu nahe an einen dieser Punkte heran, d.h. $|x_p^* - x_j| \leq C(x_2 - x_0)^2$ mit einer Konstanten C , dann bestimmt man zwei Zusatzpunkte um x_p^* , mit denen man

die Einschachtelung neu startet: $\alpha_k \hat{=} x_0$, $\gamma_k \hat{=} x_1$, $\beta_k \hat{=} x_2$, $x_p^* \hat{=} \xi_k$. In jedem Schritt hat man eine Einschachtelung von x^* durch $[\alpha_k, \beta_k]$. Im Prinzip soll ξ_k als neue Näherung α_k oder β_k ersetzen. Wenn aber ξ_k zu nahe bei einem der drei "alten" Punkte liegt, würde die Interpolation nicht nur sehr rundungsempfindlich, auch die Intervallreduktion könnte sehr gering werden. Deshalb führt man dann einen Sonderschritt mit zwei inneren Funktionsauswertungen durch.

Sei $\gamma_0 \in]\alpha_0, \beta_0[$ so gewählt, daß $f(\alpha_0) > f(\gamma_0)$, $f(\beta_0) > f(\gamma_0)$.
 $k = 0, 1, \dots$

1.

$$\begin{aligned}\Delta_1 &:= \frac{f(\gamma_k) - f(\alpha_k)}{\gamma_k - \alpha_k}, \\ \Delta_2 &:= \frac{f(\beta_k) - f(\gamma_k)}{\beta_k - \gamma_k}, \\ \xi_k &:= \left(\alpha_k + \gamma_k - \left(\frac{\Delta_1}{\Delta_2 - \Delta_1} \right) (\beta_k - \alpha_k) \right) / 2.\end{aligned}$$

(ξ_k ist die Minimalstelle der interpolierenden Parabel zu den drei benutzten f -Werten.)

2. $\delta_k = \min\{|\xi_k - \gamma_k|, |\xi_k - \alpha_k|, |\xi_k - \beta_k|\}$.

3. Falls

$$\delta_k \leq \delta_k^* := \min\{(\beta_k - \alpha_k)^2, (\beta_k - \alpha_k)/100\},$$

(d.h. ξ_k kommt einer der drei benutzten Stellen zu nahe) dann setze

$$\begin{aligned}\tilde{\xi}_k &:= \max\{\alpha_k + \delta_k^*, \xi_k - \delta_k^*\}, \\ \tilde{\xi}_k &:= \min\{\beta_k - \delta_k^*, \tilde{\xi}_k + 2\delta_k^*\},\end{aligned}$$

$$\gamma_{k+1} := \operatorname{argmin} \{f(x), x \in \{\alpha_k, \tilde{\xi}_k, \gamma_k, \tilde{\xi}_k, \beta_k\}\}, \quad (\xi_k \text{ wird verworfen})$$

$$\alpha_{k+1} := \max\{x : x < \gamma_{k+1}, x \in \{\alpha_k, \tilde{\xi}_k, \gamma_k, \tilde{\xi}_k\}\},$$

$$\beta_{k+1} := \min\{x : x > \gamma_{k+1}, x \in \{\tilde{\xi}_k, \gamma_k, \tilde{\xi}_k, \beta_k\}\},$$

sonst

$$\begin{aligned}\gamma_{k+1} &:= \operatorname{argmin} \{f(x), x \in \{\xi_k, \gamma_k\}\} \\ \alpha_{k+1} &:= \max\{x : x < \gamma_{k+1}, x \in \{\alpha_k, \gamma_k, \xi_k\}\} \\ \beta_{k+1} &:= \min\{x : x > \gamma_{k+1}, x \in \{\gamma_k, \beta_k, \xi_k\}\}\end{aligned}$$

Satz A.22. Sei f auf $[a, b]$ streng quasikonvex, nicht monoton und viermal stetig differenzierbar. Dann gilt für die eindeutig bestimmte Minimalstelle x^* von f auf $[a, b] = [\alpha_0, \beta_0]$:

$x^* \in [\alpha_k, \beta_k]$ für alle k
und

$$l_k \leq \max\{0.99, 1 - l_{k-1}\} l_{k-1} \quad \text{d.h.} \quad \beta_k - \alpha_k \rightarrow 0.$$

Ist $f''(x) > 0$ auf $[a, b]$, dann gilt zusätzlich

$$|\beta_{k+3} - \alpha_{k+3}| \leq C |\beta_k - \alpha_k|^2 \quad \text{für} \quad k \geq 0$$

mit einer geeigneten Konstanten C . □

NUMAWWW

Die Interpolationsmethoden unter 2. und 3. werden häufig in Optimierungssoftware eingesetzt und zwar in den sogenannten Schrittweisenalgorithmen, sowohl zur Schätzung einer guten ersten Versuchsschrittweite als auch zur genauen eindimensionalen Minimierung. Die benutzten Ableitungen sind dann Richtungsableitungen einer Funktion von n Variablen in einer Korrekturrichtung d .

A.3.1.2 Schrittweitenbestimmung durch Nullstellensuche

Man kann versuchen, die eindimensionale Minimierung auch durch Lösung der Nullstellenaufgabe

$$f'(x) = 0$$

zu leisten. Diese Verfahren spielen aber im originär eindimensionalen Fall keine praktische Rolle und wir gehen deshalb nicht darauf ein. Eine zugehörige Konstruktion werden wir im Zusammenhang mit dem Prinzip des hinreichenden Abstiegs im mehrdimensionalen Fall besprechen.

A.3.2 Verfahren der unrestringierten Minimierung, $n > 1$

Bemerkung A.23. Die heute üblichen Minimierungsverfahren benutzen in der Regel den Gradienten der Zielfunktion zur Berechnung der Richtungen, in denen x^k geändert werden soll. Die Aufstellung der Formeln für den Gradienten kann aber schon ein äußerst aufwendiger Prozeß sein. Hier gibt es drei Wege, um dies zu vermeiden, und zwar

1. Benutzung der Formeln der numerischen Differentiation z.B.

$$\frac{\partial f}{\partial x_i}(y) = \frac{f(y + \tau e^i) - f(y - \tau e^i)}{2\tau} + \mathcal{O}((\tau)^2).$$

Hierbei muß man die Wahl der Diskretisierungsschrittweite τ sorgfältig vornehmen, u.a. in Abhängigkeit von der Auswertungsgenauigkeit in f , um vernünftige Resultate zu erzielen. Ist der Funktionswert von f selbst das Resultat eines Algorithmus (z.B. eines FE-Berechnungsprogramms oder eines Differentialgleichungslösers) so ist z.B. nicht einmal sicher, daß dieser Algorithmus überhaupt eine differenzierbare Funktion der Optimierungsparameter liefert. Ein Beispiel ist ein Differentialgleichungslöser mit Schrittweitensteuerung. Deshalb sollte man diese numerische Differentiation vermeiden wo immer es möglich ist. Im Prinzip kann man aber auch hier hohe Genauigkeit erzielen, wenn die Funktionswerte selbst eine hohe Genauigkeit aufweisen.

2. Benutzung automatischer Systeme (Precompiler), die die Programme dafür automatisch generieren. Dieser Weg ist gangbar, wenn die Auswertung von f als unabhängiges Unterprogramm vorliegt. Diese Vorgehensweise nennt sich "automatische Differentiation" und es liegen Programme dafür vor. (jakef, ADIFOR, ADOL-C, TAMC etc.) Hier ist die Eingabe ein Programmcode für die Funktion und die Ausgabe ein Programmcode für Funktion, Gradient und eventuell sogar die Hessematrix. Sogenannte Modellierungssysteme, die eine problemnahe Formulierung der Optimierungsaufgabe erlauben und diese dann intern in Auswertungsprogramme für die entsprechenden Funktionen umsetzen, enthalten auch die Optionen der automatischen Differentiation (AMPL, GAMS).

3. Benutzung von Formelmanipulatoren, die nach Eingabe einer Formel für f die Formeln für ∇f generieren. (MATHEMATICA, MAPLE, AXIOM, DERIVE, MUPAD, usw.)

A.3.2.1 Allgemeine Verfahrensstruktur: Line-Search Methoden

Bemerkung A.24. Allgemeiner Ansatz ist $x^{k+1} = x^k - \sigma_k d^k$ mit $-d^k$ als sogenannter Abstiegsrichtung und σ_k als einer hinreichend großen Schrittweite. Dabei wird zunächst d^k festgelegt und danach σ_k bestimmt.

Definition A.25. $-d$ heißt **Abstiegsrichtung** für f an der Stelle x , wenn $\nabla f(x)^T d > 0$. Eine Zuordnung $x \mapsto d = d(x)$ heißt **gleichmäßig gradientenbezogen** auf $\mathcal{D} \subset \mathcal{D}_f$, wenn mit Konstanten C_4, C_5 unabhängig von x, d gilt

$$C_4 \|d\| \geq \|\nabla f(x)\| \geq \frac{1}{C_4} \|d\|, \quad \nabla f(x)^T d \geq C_5 \|\nabla f(x)\| \|d\|,$$

für alle $x \in \mathcal{D}$. □

Definition A.26. σ erfüllt das **Prinzip des hinreichenden Abstiegs** in (x, d) , mit $-d$ Abstiegsrichtung für f in x , wenn

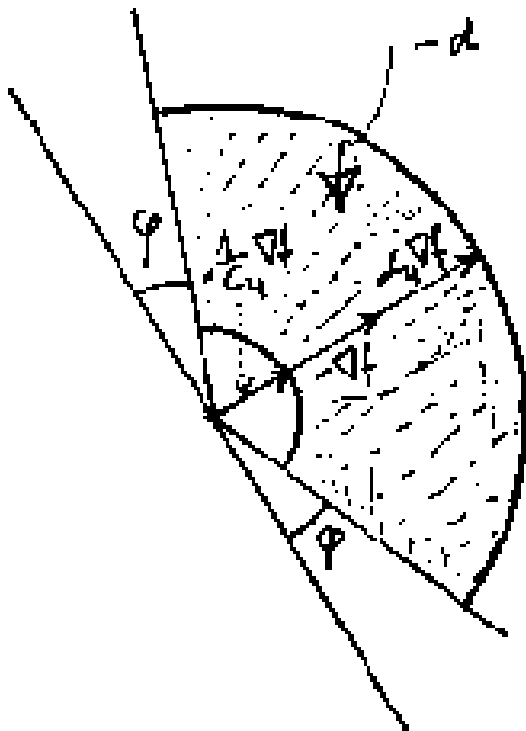
$$f(x) - f(x - \sigma d) \geq C_6 \cdot \sigma \nabla f(x)^T d$$

und

$$\sigma \geq C_7 \nabla f(x)^T d / \|d\|^2$$

gilt mit C_6, C_7 unabhängig von x, d . □

Die anschauliche Deutung von "gleichmäßig gradientenbezogen" zeigt die folgende Skizze: die Länge von d ist immer vergleichbar groß zur Länge des Gradienten und der Winkel zwischen d und dem negativen Gradienten bleibt immer von $\pi/2$ weg beschränkt.



$$\varphi = \frac{\pi}{2} - \arccos(C_5)$$

Satz A.27. Es sei $f \in C^1(\mathbb{R}^n)$, $\mathcal{L}_f(f(x^0))$ sei kompakt.
 $x^{k+1} = x^k - \sigma_k d^k$ sodaß für alle k folgendes gilt:

- (i) d^k ist gleichmäßig gradientenbezogen.
- (ii) σ_k erfüllt das Prinzip des hinreichenden Abstiegs.

Dann gilt:

1. $\{f(x^k)\}$ fällt streng monoton.
2. $\nabla f(x^k) \rightarrow 0$.
3. $x^{k+1} - x^k \rightarrow 0$, falls $\{\sigma_k\}$ beschränkt.
4. $\{x_k\}$ hat Häufungswerte und jeder Häufungswert ist eine Gradientennullstelle.
5. Hat f nur endlich viele Gradientennullstellen, dann konvergiert die Gesamtfolge gegen eine solche, falls $\{\sigma_k\}$ beschränkt ist.

Beweisskizze:

$$\begin{aligned} f(x^k) - f(x^{k+1}) &\geq C_6 \sigma_k \nabla f(x^k)^T d^k \geq C_6 C_7 (\nabla f(x^k)^T d^k / \|d^k\|)^2 \\ &\geq C_6 C_7 (C_5)^2 \|\nabla f(x^k)\|^2 \geq 0 \end{aligned}$$

$f(x^k) - f(x^{k+1}) \rightarrow 0$ da $f(x^k) \searrow$ und f nach unten beschränkt, also auch

$$\nabla f(x^k) \rightarrow 0$$

Nach Annahme unter 3. gilt damit auch

$$x^{k+1} - x^k \rightarrow 0$$

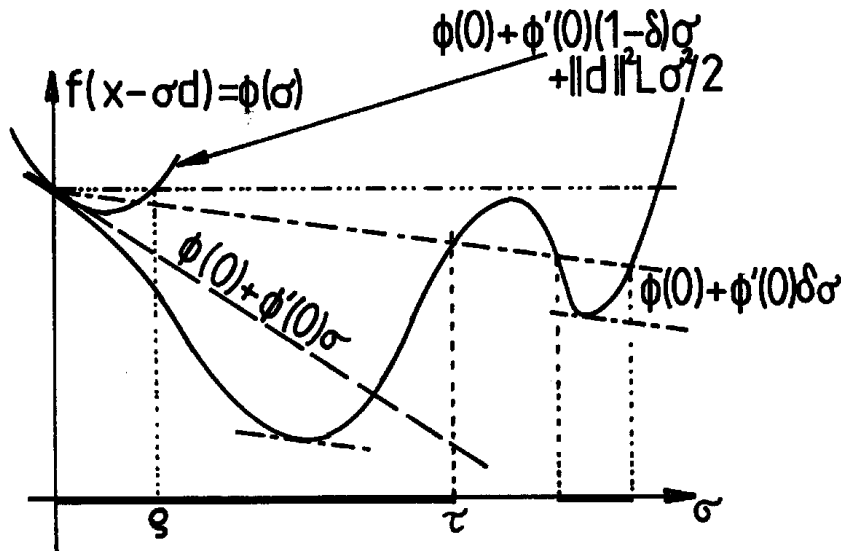
und wenn es nur endlich viele Gradientennullstellen gibt, kann dann die Folge nicht zwischen verschiedenen von diesen unendlich oft hin- und herspringen. \square

A.3.2.2 Schrittweitenverfahren

Die Konstruktion von Schrittweiten, die das Prinzip des hinreichenden Abstiegs erfüllen, wird durch folgenden Hilfssatz bereits i.w. beschrieben.

Hilfsatz A.28. Es sei $\mathcal{L}_f(f(x^0))$ kompakt (beschränkt und abgeschlossen) und $f \in C^2(\mathcal{L}_f(f(x^0)))$. Ferner sei $\nabla f(x_0)^T d > 0$. Dann existiert ein $C > 0$ unabhängig von x und d , sodaß für $0 \leq \sigma \leq C \nabla f(x)^T d / \|d\|^2$ gilt:

$$f(x_0 - \sigma d) \leq f(x_0) - (\nabla f(x_0)^T d) \sigma \delta \quad \text{falls } 0 < \delta < 1.$$



C hat die Gestalt $C = \frac{2(1-\delta)}{L}$ mit

$$L \geq n \sup \left\{ \left| \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right| : 1 \leq i, j \leq n, x \in \mathcal{D} \right\}$$

Das bedeutet, daß zumindest im Intervall $[0, C(\nabla f(x)^T d / \|d\|^2)]$ $f(x_0 - \sigma d)$ mindestens so stark abnimmt wie $f(x_0) - \delta \sigma \nabla f(x_0)^T d$, d.h. man kann zumindest einen Schritt dieser Länge ausführen.

Beweisskizze:

$$\begin{aligned} f(x_0 - \sigma d) &= f(x_0) - \delta \sigma (\nabla f(x_0)^T d) \\ &\quad + \sigma \left(-(1-\delta)(\nabla f(x_0)^T d) + \sigma \frac{1}{2} d^T \nabla^2 f(x_0 - \theta \sigma d) d \right) \end{aligned}$$

und die Klammer in der zweiten Zeile ist negativ, falls $0 \leq \sigma \leq C(\nabla f(x)^T d / \|d\|^2)$. Wenn $\{d^k\}$ gleichmäßig gradientenbezogen ist in $\{x^k\}$, gilt natürlich

$$\sigma_k \geq C_7 C_5 / C_4 > 0 \quad \text{für alle } k.$$

Man beachte für das Folgende, daß

$$f(x - d) = f(x) - \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x - \theta d) d$$

mit einem $\theta \in]0, 1[$. Dies bedeutet, daß

$$2(f(x-d) - f(x) + \nabla f(x)^T d) \approx d^T \nabla^2 f(x) d$$

Für eine streng konvexe quadratische Funktion

$$f(x) = \gamma - b^T x + \frac{1}{2} x^T A x$$

liegt das Minimum auf dem Strahl $x - \sigma d$ bei

$$\sigma = \frac{\nabla f(x)^T d}{d^T A d} = \frac{\nabla f(x)^T d}{2(f(x-d) - f(x) + \nabla f(x)^T d)}.$$

und dies bedeutet, daß die im Folgenden benutzte Anfangsschrittweite $\sigma_{k,0}$ zumindest in der Nähe eines strengen lokalen Minimums auch bei einer allgemeinen Funktion fast optimal ist.

Ein Algorithmus zur Erfüllung des Prinzips des hinreichenden Abstiegs für eine gleichmäßig gradientenbezogene Richtung d^k ist jetzt

$$\sigma_{k,0} := \begin{cases} 1 & \text{falls } f(x^k - d^k) \leq f(x^k) - \nabla f(x^k)^T d^k \\ \max \left\{ C_1, \min \left\{ C_2, \frac{\nabla f(x^k)^T d^k}{(2(f(x^k - d^k) - f(x^k) + \nabla f(x^k)^T d^k))} \right\} \right\} \end{cases}$$

$$\sigma_k = \sigma_{k,0} \beta^j \quad \text{maximal, sodaß } f(x^k) - f(x^k - \sigma_k d^k) \geq \sigma_k \delta \nabla f(x^k)^T d^k$$

Dies ist der sogenannte **Goldstein-Armijo-Abstiegstest** (“backtracking”). Dabei sind $0 < \beta < 1$ der Schrittweitenreduktionsfaktor und $0 < C_1 \ll 1 \ll C_2$ zwei geeignete “Sicherheitskonstanten”.

Ohne Zusatzinformation über die Konstruktion von d^k sollte gelten

$$C_1 < \frac{1}{\sup_{x \in \mathcal{L}_0} \|\nabla^2 f(x)\|} \inf \{ \nabla f(x)^T d / \|d\|^2 \}, \quad C_2 > \sup_{x \in \mathcal{L}_0} \|(\nabla^2 f(x))^{-1}\| \sup \{ \nabla f(x)^T d / \|d\|^2 \}^1$$

mit $\mathcal{L}_0 =$ konvexe Obermenge von $\mathcal{L}_f(f(x^0))$. Dann kann man garantieren, daß das erste lokale Minimum von f auf dem Strahl durch die Wahl von $\sigma_{k,0}$ nicht ausgeschlossen wird und für grosses k stets $\sigma_k = \sigma_{k,0}$ wird, wenn auch noch

$$0 < \delta < 1/2$$

¹Ist A eine symmetrische Matrix, so ist $\|A\|$ der Betrag des betragsgrößten Eigenwertes von A . Für allgemeines A ist $\|A\|$ definiert durch $\|A\| := \max_{\|x\|=1} \|Ax\|$, dabei ist die Vektornorm $\|\cdot\|$ vorgegeben.

gilt. Ist f streng konvex und $\nabla^2 f(x^*)$ positiv definit, dann gilt unter der Annahme

$$C_1 < 1/\|\nabla^2 f(x^*)\|, \quad C_2 > \|(\nabla^2 f(x^*))^{-1}\|$$

wieder $\sigma_k = \sigma_{k,0}$ für k hinreichend groß.

Manchmal ist es günstiger, einen etwas strengeren Abstiegstest zu verwenden, etwa den von **Powell-Wolfe**. Hier wird verlangt, daß gleichzeitig

$$f(x) - f(x - \sigma d) \geq \delta \sigma \nabla f(x)^T d \quad (\nabla f(x)^T d > 0)$$

und

$$\nabla f(x - \sigma d)^T d \leq \kappa \nabla f(x)^T d$$

oder sogar

$$|\nabla f(x - \sigma d)^T d| \leq \kappa \nabla f(x)^T d$$

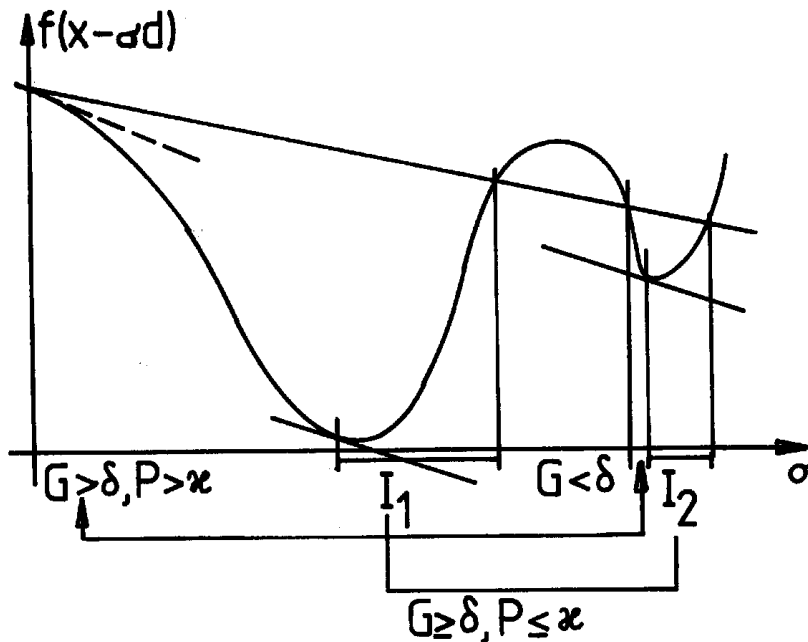
gilt mit $0 < \delta < \kappa < 1$. Letztere Bedingung wird auch als "strenge Powell-Wolfe-Bedingung" bezeichnet. Die zulässigen σ -Bereiche für den gewöhnlichen Powell-Wolfe-Test sind in der folgenden Skizze I_1 und I_2 . Es bedeutet

$$G(\sigma) = \begin{cases} 1 & \sigma = 0 \\ \frac{f(x) - f(x - \sigma d)}{\sigma \nabla f(x)^T d} & \text{sonst} \end{cases} \quad P(\sigma) = \frac{\nabla f(x - \sigma d)^T d}{\nabla f(x)^T d}$$

Algorithmus: siehe Spezialliteratur.

NUMAWWW

Der Vorteil dieses Tests: Man benötigt keine Anfangsschätzung für die Schrittweite und das Prinzip des hinreichenden Abstiegs (nicht zu kleine Schrittweiten) ist dennoch erfüllt. Ein Nachteil: Man benötigt zusätzliche Berechnungen der Richtungsableitung von f .



Die Festlegung von σ_k als $\sigma_k = \operatorname{argmin} \{f(x^k - \sigma d^k) : \sigma \in \mathbb{R}\}$ ist **in den wenigsten Fällen sinnvoll**. Der hohe Aufwand zur Bestimmung von σ_k zahlt sich nämlich nicht in verbesserten Konvergenzeigenschaften aus. Man beachte jedoch: Ist

$$f(x) = \frac{1}{2}x^T A x - b^T x, \quad A = A^T \text{ positiv definit}$$

dann ist

$$\sigma = \frac{\nabla f(x)^T d}{2(f(x-d) - f(x) + \nabla f(x)^T d)} = \operatorname{argmin} \{f(x - \sigma d) : \sigma \in \mathbb{R}\}$$

d.h. unsere obige Konstruktion ist in diesem Fall "optimal".

Allgemein gilt für σ_k aus dem angegebenen Goldstein-Armijo-Abstiegstest:

Satz A.29. *Es seien die Voraussetzungen von Satz A.27 alle erfüllt und $f \in C^{2,1}(\mathcal{U}(x^*))$, $\nabla f(x^*) = 0$, $\nabla^2 f(x^*)$ positiv definit. Ferner gelte*

$$C_5 \leq \frac{\nabla f(x^k)^T d^k}{\|d^k\|^2} \leq \frac{1}{C_5}$$

Falls im Goldstein-Armijo-Abstiegstest gilt:

1. $0 < \delta < \frac{1}{2}$,
2. $C_1 < 1/\|\nabla^2 f(x^*)\|C_5$,
3. $C_2 > \|(\nabla^2 f(x^*))^{-1}\|/C_5$,

dann ist für hinreichend großes k stets $\sigma_k = \sigma_{k,0}$. Außerdem gilt

$$|\sigma_k - \bar{\sigma}_k| \leq C_8 \|\nabla f(x^k)\|$$

mit einer geeigneten Konstanten $C_8 > 0$, wobei $\bar{\sigma}_k$ die kleinste positive Minimalstelle von $f(x^k - \sigma d^k)$ bezeichnet, d.h. σ_k ist "fast optimal" (genauer: asymptotisch exakt von erster Ordnung). □

A.3.2.3 Richtungsbestimmung

Allgemeine Vorgehensweise: 3 Typen

1. Koordinatenweise Minimierung

d^k a priori vorgegeben, in der Regel $d^k = \pm e^{(k \bmod n)+1}$ (Koordinatenrichtungen) oder $d^k = \pm v^{(k \bmod n)+1}$, $\{v^j\}$ **approximatives** Eigenvektorsystem von $\nabla^2 f(x^k)$, das u.U. erst im Lauf der Rechnung ermittelt wird. Diese Richtungen sind nicht gleichmässig gradientenbezogen. Dennoch kann man unter Zusatzvoraussetzungen die Konvergenz der zugehörigen Abstiegsverfahren beweisen.

NUMAWWW

2. Quasi-Newtonverfahren, Newtonverfahren

Hier ist d^k die Lösung (oder approximative Lösung) von

$$A_k d^k = \nabla f(x^k)$$

mit $\{A_k\}$ symmetrisch und positiv definit, wobei u.a. die Matrizenfolge die Anforderung

$$A_{k+1}(x^{k+1} - x^k) = \nabla f(x^{k+1}) - \nabla f(x^k) \quad \text{Sekantenrelation}$$

erfüllt, und A_{k+1} sich rekursiv aus A_k , $x^{k+1} - x^k$, $\nabla f(x^{k+1}) - \nabla f(x^k)$ und eventuell weiteren zurückliegenden Daten berechnet. Für einige dieser Konstruktionen kann man unter Zusatzvoraussetzungen beweisen, daß die Matrizenfolge $\{A_k\}$ gegen die Hessematrix von f im Minimum konvergiert. Beim Newtonverfahren selbst wählt man

$$A_k = \nabla^2 f(x^k)$$

Dies ist jedoch nur für gleichmässig konvexes f sinnvoll.

3. Verfahren von cg-Typ

$$d^k = \begin{cases} \nabla f(x^k) & \text{falls } 0 = k \bmod n \\ \nabla f(x^k) + \sum_{j=1}^{r(k)} \gamma_{k,j} d^{k-j} & \text{mit } (d^k, d^j) = 0 \quad \text{für } j = 0, \dots, k-1, \text{ (ergibt } \gamma_{k,j}) \end{cases}$$

Dabei ist (x, y) ein Skalarprodukt auf \mathbb{R}^n , $r(k) = k - \lfloor \frac{k}{n} \rfloor n$

Bemerkung A.30. Ein allgemeines Skalarprodukt auf \mathbb{R}^n ist eine positive Bilinearform und kann immer in der Form

$$(x, y) = x^T C y$$

mit einer symmetrischen und positiv definiten Matrix C geschrieben werden. Hier wird im Algorithmus B gewählt oder erst konstruiert (jeweils über n Schritte fest) und C ist implizit gegeben etwa als $C \approx B^{-1/2} \nabla^2 f(x^*) B^{-1/2}$.

$\lfloor x \rfloor =$ kleinste ganze Zahl $\leq x$.

Der 1. Ansatz ist nur in sehr speziellen Fällen interessant. Die damit erzielbare Konvergenzgeschwindigkeit ist sehr gering.

Der 2. Ansatz wird für Probleme kleiner bis mittlerer Dimension ($n \leq 200, \dots, 1000$) benutzt.

Der 3. Ansatz ist für Probleme sehr hoher Dimension oft die einzige Möglichkeit.

A.3.2.3.1 Newtonähnliche- und Quasi-Newtonverfahren

Konstruktion:

$$d^k \text{ ist Lösung von } A_k d^k = \nabla f(x^k) \quad (A_k \text{ symm. pos. def.})$$

Die Matrizenfolge $\{A_k\}$ wird auf verschiedenartige Weise erzeugt:

1. $A_k = \nabla^2 f(x^k)$ **Newtonverfahren**
 $\nabla^2 f(x^k)$ wird eventuell approximiert durch finite Differenzen oder berechnet durch automatische (exakte) Differentiation unter Umständen entsteht hier ein hoher Berechnungsaufwand für A_k .
 $A_k = \nabla^2 f(x^0)$: vereinfachtes Newtonverfahren.
 Die Verwendung der exakten Hessematrix ist nur für gleichmässig konvexes f sinnvoll. Andernfalls müsste man A_k modifizieren, um die positive Definitheit zu erzwingen und verliert dann den Vorteil der schnellen Konvergenz. Es gibt allerdings auch Verfahren, die für indefinite oder negativ definite Hessematrix sogenannte "Richtungen negativer Krümmung" (d.h. $(d^k)^T A_k d^k < 0$ berechnen und mit diesen einen Abstieg durchführen.
2. $\{A_k\}$ wird rekursiv erzeugt aus $\{x^k\}$ und $\{\nabla f(x^k)\}$. Dies führt zu den **Quasi-Newtonverfahren**. Mit

$$s^i := x^{i+1} - x^i, \quad y^i := \nabla f(x^{i+1}) - \nabla f(x^i)$$

hat man die Forderungen:

$$\begin{aligned} A_i d^i &= \nabla f(x^i), \quad x^{i+1} = x^i - \sigma_i d^i \\ A_{i+1} s^i &= y^i \quad \text{Sekantenrelation} \\ \text{"}A_{i+1} - A_i \text{ klein"} &\text{ im Sinne einer geeigneten Norm als zusätzliche Forderung} \\ &\text{eventuell außer Symmetrie weitere Forderungen an } A_{i+1}. \end{aligned}$$

Im Eindimensionalen ist A_{i+1} genau die Sekantensteigung und im Mehrdimensionalen wird ihre Bedeutung klar an der Taylorformel:

$$y^i = \int_0^1 \nabla^2 f(x^i + \tau s^i) d\tau s^i.$$

A_{i+1} hat also in der Richtung s^i die gleiche Abbildungseigenschaft wie ein lokaler Mittelwert der Hessematrix von f . Dies ergibt die sogenannten Quasi-Newtonverfahren minimaler Änderung. Die bekanntesten sind

DAVIDON-FLETCHER-POWELL **DFP** 1959/1963 (ältestes Verfahren dieser Art)

$$A_{i+1} = \left(I - \frac{y^i (s^i)^T}{(s^i)^T y^i} \right) A_i \left(I - \frac{s^i (y^i)^T}{(s^i)^T y^i} \right) + \frac{y^i (y^i)^T}{(y^i)^T s^i} \quad 2$$

Diese Formel erhält man, wenn man als Minimalprinzip

$$\|(H_i)^{-1/2}(A_{i+1} - A_i)(H_i)^{-1/2}\|_F = \min$$

mit den Nebenbedingungen Symmetrie und Sekantenrelation benutzt. $\|\cdot\|_F$ bezeichnet die Frobeniusnorm einer Matrix (die Wurzel aus der Quadratsumme aller Elemente). Dabei ist

$$H_i = \int_0^1 \nabla^2 f(x^i + \tau s^i) d\tau .$$

Dieses Verfahren ist nicht besonders günstig, weil es empfindlich gegen Abweichungen $\sigma_i - \bar{\sigma}_i$, $\bar{\sigma}_i =$ optimale Schrittweite, ist. Voraussetzung an σ_i ist in jedem Fall $(y^i)^T s^i > 0$, sonst geht die Definitheit verloren.

BROYDEN-FLETCHER-GOLDFARB-SHANNO 1970 BFGS

$$A_{i+1} = A_i - \frac{A_i s^i (s^i)^T A_i}{(s^i)^T A_i s^i} + \frac{y^i (y^i)^T}{(y^i)^T s^i}$$

Hier ist das Minimalprinzip

$$\|(H_i)^{1/2}(A_{i+1}^{-1} - A_i^{-1})(H_i)^{1/2}\|_F = \min$$

mit den gleichen Nebenbedingungen. Dies ergibt ein sehr gutes, robustes Verfahren. Voraussetzung ist auch hier $(y^i)^T s^i > 0$.

SR1: BROYDEN 1967

$$\begin{aligned} A_{i+1} &= A_i + \frac{(y^i - A_i s^i)(y^i - A_i s^i)^T}{(y^i - A_i s^i)^T s^i} \\ &= A_i + \frac{(H_i - A_i) s_i s_i^T (H_i - A_i)}{s_i^T (H_i - A_i) s_i} \end{aligned}$$

mit

$$H_i = \int_0^1 \nabla^2 f(x_i + \tau s_i) d\tau .$$

SR1 konvergiert oft schneller als BFGS, **aber** u.U. ist A_{i+1} nicht mehr pos.def. oder sogar nicht definiert! Dies kann jedoch durch geeignete Modifikationen der Formel umgangen werden.

Ursprünglich wurden diese Verfahren aus **dem quadratischem Modell** für f entwickelt. (in der Praxis sind sie aber für diesen Fall uninteressant). Also

$$f(x) = \frac{1}{2} x^T A x - b^T x, \quad A = A^T \text{ pos. def.}$$

² $xy^T = n \times n$ -Matrix mit den Komponenten $x_i y_j$ i =Zeile, j =Spalte

Dann gilt

$$\nabla f(x) = Ax - b \quad \text{also} \quad As^i = y^i$$

Für DFP, BFGS gilt in diesem Fall : Falls σ_j optimal gewählt ist, dann gilt

$$A_k s^j = y^j \quad j = k - 1, \dots, 0.$$

Es ist dann

$$A_k^{-1} A s^j = s^j, \quad j = k - 1, \dots, 0$$

und falls $k = n$ folgt $A_n = A$. Generell ist dann also $x^N = x^*$ mit $N \leq n$. Für das SR1-Verfahren benötigt man nicht einmal diese Voraussetzung, wohl aber die der Durchführbarkeit des Verfahrens, um die gleiche Aussage zu erzielen.

Konvergenzaussagen

Satz A.31. Sei $\{A_k\}$ eine beschränkte Folge symmetrischer, gleichmäßig positiv definiten Matrizen, d.h. es gibt $0 < \alpha^* < \alpha^{**}$:

$$\alpha^* \leq \lambda_j(A_k) \leq \alpha^{**} \quad \text{für alle Eigenwerte } \lambda_1, \dots, \lambda_n \text{ von } A_k \text{ und alle } k.$$

Dann ist

$$d^k := A_k^{-1} \nabla f(x^k) \quad k = 0, 1, \dots$$

gleichmäßig gradientenbezogen in x^k . □

Korollar: Sei $\{A_k\}$, $\{d^k\}$, $\{x^k\}$ konstruiert wie in den Sätzen A.27, A.29, A.31 beschrieben. Dann gilt $\lim_{k \rightarrow \infty} \nabla f(x^k) = 0$. □

Konvergenzaussage für das Newtonverfahren:

Satz A.32. Sei $f \in C^{2,1}(\mathbb{R}^n)$ und gleichmäßig konvex. Es sei ferner x^0 beliebig und

$$\begin{aligned} \nabla^2 f(x^k) d^k &= \nabla f(x^k) \\ \sigma_k &= \max\{\beta^j : f(x^k) - f(x^k - \beta^j d^k) \geq \delta \beta^j \nabla f(x^k)^T d^k, j = 0, 1, 2, \dots\} \end{aligned}$$

mit $0 < \beta < 1$ und $0 < \delta < 1/2$, und

$$x^{k+1} = x^k - \sigma_k d^k.$$

Dann gilt: $\{x^k\}$ konvergiert gegen die eindeutige Minimalstelle von f , $\sigma_k \equiv 1$ für k hinreichend groß und $\|x^{k+1} - x^*\| \leq C \|x^k - x^*\|^2$ mit geeignetem $C > 0$. □

NUMAWWW

Konvergenzaussagen für das BFGS-Verfahren

Satz A.33. Sei $f \in C^{2,1}(\mathbb{R}^n)$ und gleichmässig konvex. Es sei ferner x^0 beliebig. A_0 symmetrisch positiv definit und für alle k gelte

$$(i) \quad A_k d^k = \nabla f(x^k)$$

(ii) $\{\sigma_k\}$ sei asymptotisch exakt von erster Ordnung und erfülle das Prinzip des hinreichenden Abstiegs

(z.B. $\sigma_k = \max\{\beta^j : f(x^k) - f(x^k - \beta^j d^k) \geq \delta \beta^j \nabla f(x^k)^T d^k\}$ mit $0 < \beta < 1$, $0 < \delta < 1/2$ oder σ_k aus obigem Goldstein-Armijo-Abstiegstest).

$$(iii) \quad x^{k+1} = x^k - \sigma_k d^k$$

$$(iv) \quad A_{k+1} = A_k - \frac{A_k s^k (s^k)^T A_k}{(s^k)^T A_k s^k} + \frac{y^k (y^k)^T}{(y^k)^T s^k} \quad (\text{BFGS})$$

Dann gilt:

(1) $\{A_k\}$ ist beschränkt und gleichmäßig positiv definit, d.h. es gibt $0 < \alpha_0$ und α_1 mit $\alpha_0 d^T d \leq d^T A_k d \leq \alpha_1 d^T d$ für alle $d \in \mathbb{R}^n$, alle k .

(2) $\{x^k\}$ konvergiert gegen die eindeutige Minimalstelle x^* von f .

(3) $\|x^{k+1} - x^*\| \leq \varepsilon_k \|x^k - x^*\|$ mit $\varepsilon_k \rightarrow 0$ (Q -superlineare Konvergenz)

□

NUMAWWW

Analoge Sätze gelten für das DFP- und das SR1-Verfahren. Um die globale Konvergenz des DFP-Verfahrens zu zeigen, muss man die Schrittweitenwahl weiter einschränken. Beim SR1-Verfahren muss man eine Modifikation in der Matrizenkonstruktion vornehmen, um die positive Definitheit zu sichern. In der Praxis sind diese Verfahren dem BFGS-Verfahren jedoch deutlich unterlegen. Es stellt sich die Frage, warum man keine "einfachen" Verfahren, z.B. das Gradientenverfahren, d.h. $A_k \equiv I$ wählt? Es stellt sich heraus, daß das Gradientenverfahren viel zu langsam konvergiert (obwohl es natürlich alle Bedingungen des allgemeinen Konvergenzsatzes erfüllt). Diese Konvergenzaussage lautet für das gewöhnliche Gradientenverfahren:

Satz A.34. Unter den gleichen Voraussetzungen wie in Satz A.33 gilt für das Gradientenverfahren

$$\|A^{1/2}(x^{k+1} - x^*)\| \leq \left(\left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} \right)^2 + \varepsilon_k \right)^{1/2} \|A^{1/2}(x^k - x^*)\|, \quad \varepsilon_k \rightarrow 0$$

$$A := \nabla^2 f(x^*) \quad \text{pos. def.}, \quad \lambda_1 = \lambda_{\max}(A), \quad \lambda_n = \lambda_{\min}(A)$$

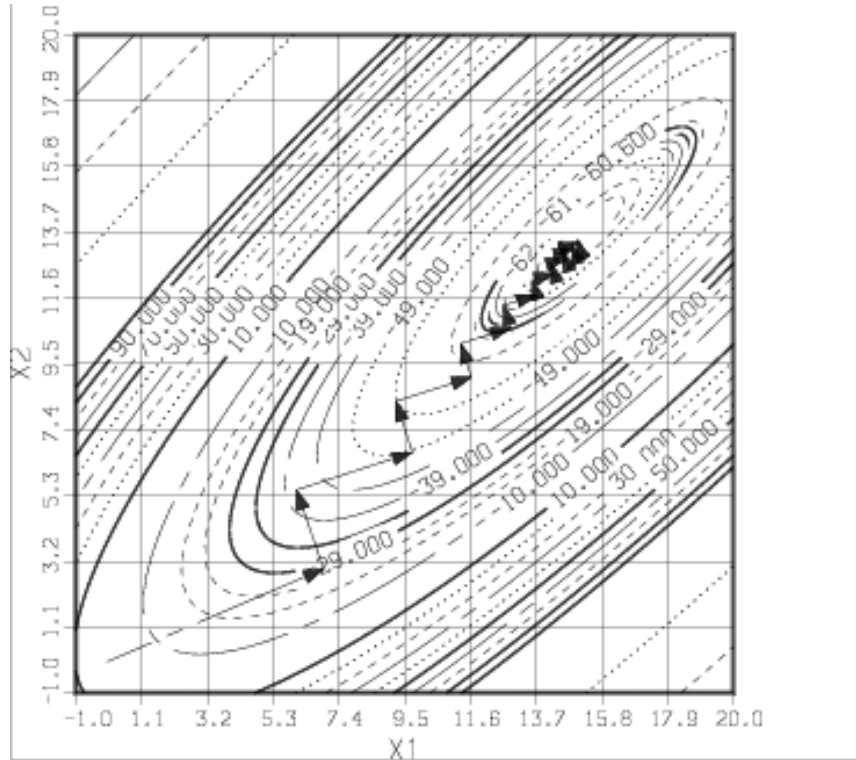
□

Ein Zahlenbeispiel mag diese Aussage verdeutlichen:

$$\lambda_1 = 100, \quad \lambda_n = 1 \quad (\text{das ist noch harmlos}) \Rightarrow \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} \approx 0.99$$

$$0.99^k = 0.1 \quad \text{für } k = 230$$

d.h. bei einer Konditionszahl von 100 benötigt man 230 Schritte (also auch 230 Gradientenauswertungen), um den Fehler um 1/10 zu verringern. Die folgende Abbildung zeigt den Ablauf des Gradientenverfahrens bei "optimal" gewählter Schrittweite.



NUMAWWW

Abstiegsverfahren für nicht gleichmäßig konvexe Funktionen erhält man durch ergänzende Maßnahmen, da man davon ausgehen kann, daß die betrachteten Funktionen zumindest in der Nähe eines (lokalen) Minimums gleichmäßig konvex sind :

1. z.B. ist $(y^i)^T s^i > 0$ gewährleistet durch Powell-Wolfe-Kriterium

$$f(x^i) - f(x^i - \sigma_i d^i) \geq \delta \sigma_i \nabla f(x^i)^T d^i$$

und

$$\nabla f(x^i - \sigma_i d^i)^T d^i \leq \kappa \nabla f(x^i)^T d^i \quad \text{mit } 0 < \delta < \kappa < 1$$

auch im nichtkonvexen Fall. Im gleichmäßig konvexen Fall gilt diese Bedingung für jede Schrittweite > 0 .

2. Die Kontrolle der gleichmässigen positiven Definitheit von A_i erfolgt am zweckmässigsten mit Hilfe der Cholesky-Zerlegung $A_i = L_i L_i^T$, die man direkt kann in $L_{i+1} L_{i+1}^T = A_{i+1}$ mit einem Aufwand von $\mathcal{O}(n^2)$.

Vorteile:

- a) Einfache Bestimmung d^i aus zwei Gleichungssystemen mit Dreiecksmatrix

$$\begin{aligned} L_i \underbrace{(L_i^T d^i)}_{=: z^i} &= \nabla f(x^i) \\ L_i z^i &= \nabla f(x^i), \quad L_i^T d^i = z^i, \quad \text{ergibt erst } z^i, \text{ dann } d^i. \end{aligned}$$

- b) Die Kontrolle von $\|A_i\|$ und $\|A_i^{-1}\|$ ist einfach.

$$\begin{aligned} \|A_i\| &\geq \max_j (l_{ii,j})^2 \\ \|A_i^{-1}\| &\geq \max_j (1/l_{ii,j})^2 \end{aligned}$$

Wenn also diese untere Schranke zu groß wird, dann wird ein **Neustart** mit $A_i := \gamma_i \cdot I$, γ_i geeignet (engl. "Restart") ausgeführt. Dies ergibt dann ein zuverlässiges, universell einsetzbares Verfahren.

Zum Aufwand z.B. beim BFGS-Verfahren: der analytische Aufwand pro Schritt beträgt durchschnittlich ca. 2 Auswertungen von f , 1 Auswertung von ∇f und der algebraische Aufwand $\mathcal{O}(n^2)$ Rechenoperationen, man hat einen Speicherbedarf von $n^2/2 + \mathcal{O}(n)$. Erfahrungsgemäss sind höchstens $30n$ Schritte für "völlige Genauigkeit" auch bei schwierigen Problemen ausreichend. Für "kleines" n ist dies alles völlig unproblematisch. Bei größerem n , etwa $n \gg 200$, sind vor allem der **Speicherbedarf** und der algebraische Aufwand kritisch zu sehen. Hochdimensionale Optimierungsprobleme besitzen in der Regel dünn besetzte Hessematrizen. Im Gegensatz zum Newtonverfahren sind aber bei den Quasi-Newton-Verfahren die Matrizen A_k **voll besetzt auch wenn $\nabla^2 f$ dünn besetzt ist**. Die direkte Forderung "konstruiere A_i mit der gleichen Besetztheitsstruktur wie $\nabla^2 f$ unter Erfüllung der Sekantenrelation" ist erfüllbar, die zugehörigen Verfahren sind jedoch in der Praxis nicht erfolgreich. Es gibt für hochdimensionale Probleme einige Alternativen:

Für den Spezialfall

$$f(x) = \sum_{i=1}^m f_i(x)$$

wobei die Bildbereiche \mathcal{R}_i von $\nabla^2 f_i(x)$, das sind die Unterräume

$$\mathcal{R}_i = \{y : y = \nabla^2 f_i(x)z \text{ mit } z \in \mathbb{R}^n \text{ bel.}\}$$

niedrigdimensionale Teilräume von \mathbb{R}^n und von x unabhängig sind (d.h. der Nichtnullteil von $\nabla^2 f_i(x)$ kann durch eine "kleine" Matrix repräsentiert werden), und alle f_i verallgemeinert konvex sind im Sinne von

$$(x - y)^T (\nabla f_i(x) - \nabla f_i(y)) \geq \varrho_i \|\mathcal{P}_i(x - y)\|^2, \quad \forall x, y$$

wobei $\mathcal{P}_i =$ die Projektion von \mathbb{R}^n auf \mathcal{R}_i , $\varrho_i \geq 0$, ist, gibt es eine Modifikation des BFGS-Verfahrens: (Toint, Griewank 1982-1989):

das "**Partitionierte BFGS-Verfahren**" Hierbei werden die Hessematrizen $\nabla^2 f_i(x)$ einzeln approximiert und A_i erst vor der Gleichungslösung zusammengesetzt.

Algorithmus:

Wähle $A_{i,0}$, $i = 1, \dots, m$ sodaß der Bildbereich von $A_{i,0}$ eine Obermenge von \mathcal{R}_i ist. (Formal ist jedes $A_{i,k}$ eine $n \times n$ -Matrix, von der aber nur der wesentliche Teil, der nicht identisch null ist, gespeichert werden muß)

Für $k = 0, 1, \dots$

$$A_k := \sum_{i=1}^m A_{i,k}$$

$$A_k d^k = \nabla f(x^k) \text{ lösen für } d^k$$

Bestimme σ_k aus einem Schrittweisenalgorithmus wie zuvor, $x^{k+1} = x^k - \sigma_k d^k$,

$$s^k = x^{k+1} - x^k$$

$$i = 1, \dots, m :$$

$$y^{i,k} := \nabla f_i(x^{k+1}) - \nabla f_i(x^k)$$

$$\nu_{i,k} = \begin{cases} 1 & \text{wenn } A_{i,k} s^k = 0 \\ (y^{i,k})^T s^k / (s^k)^T A_{i,k} s^k & \text{sonst} \end{cases}$$

$$\mu_{i,k} = \begin{cases} 0 & \text{wenn } \nu_{i,k} = 1 \\ \max\{0, \mu_i - \nu_{i,k}\} / (1 - \nu_{i,k}) & \text{sonst} \end{cases}$$

$$z^{i,k} := (1 - \mu_{i,k}) y^{i,k} + \mu_{i,k} A_{i,k} s^k$$

$$A_{i,k+1} := \begin{cases} A_{i,k} + z^{i,k} z^{i,kT} / (z^{i,kT} s^k) & \text{falls } A_{i,k} s^k = 0, y^{i,k} \neq 0 \\ A_{i,k} & \text{falls } A_{i,k} s^k = 0 \text{ und } y^{i,k} = 0 \\ A_{i,k} - (1 - \mu_i) \frac{A_{i,k} s^k (s^k)^T A_{i,k}}{(s^k)^T A_{i,k} s^k} & \text{falls } A_{i,k} s^k \neq 0 \text{ und } y^{i,k} = 0 \\ A_{i,k} - \frac{A_{i,k} s^k (s^k)^T A_{i,k}}{(s^k)^T A_{i,k} s^k} + \frac{z^{i,k} (z^{i,k})^T}{(z^{i,k})^T s^k} & \text{sonst} \end{cases}$$

Satz A.35. (Griewank 1991) *f* sei gleichmäßig konvex auf \mathbb{R}^n , die einzelnen f_i seien konvex, $\in C^2(\mathbb{R}^n)$, $\mathcal{R}_i = \mathcal{R}(\nabla^2 f_i(x))$ seien unabhängig von x und $\mathcal{R}(A_{i,0}) \supset \mathcal{R}_i$ für alle i . Ferner gelte $0 \leq \mu_i < 1/2$, $\mu_i + \varrho_i > 0$, $i = 1, \dots, m$ (bzgl. ϱ_i siehe oben) $\{\sigma_k\}$ erfülle das Prinzip des hinreichenden Abstiegs und sei asymptotisch exakt von erster Ordnung. Dann konvergiert $\{x^k\}$ Q -superlinear gegen die eindeutige Minimalstelle x^* von f , d.h.

$$\|x^{k+1} - x^*\| \leq \varepsilon_k \|x^k - x^*\| \quad \text{mit} \quad \varepsilon_k \rightarrow 0.$$

□

Limited-memory-Quasi-Newton-Verfahren (Byrd, Nocedal und Schnabel): Diese beruhen auf einer anderen Darstellung der Quasi-Newton-Matrizen. Haben z.B. s^k und y^k die gleiche Bedeutung wie beim BFGS-Verfahren, d.h.

$$s^k = x^{k+1} - x^k, \quad y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$$

und setzt man

$$\begin{aligned} S_k &= (s^0, \dots, s^{k-1}) \in \mathbb{R}^{n \times k}, \\ Y_k &= (y^0, \dots, y^{k-1}) \in \mathbb{R}^{n \times k}, \\ (R_k)_{ij} &= \begin{cases} (s^{i-1})^T y^{j-1} & i \leq j \\ 0 & i > j \end{cases} \\ D_k &= \text{diag}((s^0)^T y^0, \dots, (s^{k-1})^T y^{k-1}) \quad (\text{beachte } (s^i)^T y^i > 0) \end{aligned}$$

dann ist A_k^{-1} (BFGS) darstellbar als

$$A_k^{-1} = A_0^{-1} + (S_k, A_0^{-1} Y_k) \left(\begin{array}{c|c} (R_k)^{-T} (D_k + Y_k^T A_0^{-1} Y_k) R_k^{-1} & -R_k^{-T} \\ \hline -R_k^{-1} & 0 \end{array} \right) \left(\begin{array}{c} S_k^T \\ \hline Y_k^T A_0^{-1} \end{array} \right)$$

Man benutzt nun für höhere Schrittzahlen **nicht alle** zurückliegenden Vektorpaare (s^j, y^j) , sondern nur eine feste Anzahl m “geeignet ausgewählter” solcher Paare (z.B. 3-10 aus den zurückliegenden) Damit hat man dann oft eine gute Darstellung der wichtigsten lokalen “Krümmungsinformation” mit einem Speicheraufwand von etwa $4m^2 + 2mn$ Speicherplätzen, falls man als A_0 jeweils ein geeignetes Vielfaches der Einheitsmatrix wählt. Hier wird d^k direkt erhalten als $d^k = A_k^{-1} \nabla f(x^k)$ durch Matrix-Vektormultiplikation.

NUMAWWW

Offenbar aus Gründen mangelnder numerischer Stabilität kann man bei diesem Verfahren m nicht allzu groß wählen. Bei kleinem m und großer Dimension reicht die angesammelte Information jedoch nicht zu einer guten Approximation der Newtonrichtung aus, sodaß das Verfahren in der Praxis nicht so leistungsfähig ist wie erhofft.

A.3.2.3.2 Verfahren von cg-Typ

Die ursprüngliche Motivation zur Herleitung des cg-Verfahrens war die Lösung des linearen Gleichungssystems

$$Ax = b \quad \text{mit } A \text{ positiv definit,}$$

die äquivalent ist mit der Minimierung der streng konvexen Funktion

$$f(x) = \frac{1}{2}x^T Ax - b^T x.$$

Der Lösungsansatz besteht darin, f auf ineinandergeschachtelten linearen Mannigfaltigkeiten der Form

$$x = x^0 + V_k a$$

zu minimieren. Wenn dies einfach möglich ist, gewinnt man ein finites Verfahren mit maximal $n = \dim(x)$ Schritten. Ist V_k eine Basis ($n \times k$ -Matrix) der k -ten Mannigfaltigkeit (also kommt pro Schritt eine Spalte hinzu), dann lautet die Darstellung der Lösung explizit

$$x^{(k)} = x^{(0)} - V_k((V_k)^T A V_k)^{-1} V_k^T \nabla f(x^{(0)})$$

und man erkennt, daß dies besonders einfach zu berechnen ist, wenn gilt

$$(V_k)^T A V_k \quad \text{diagonal.}$$

Dies führt auf die Idee, f längs sogenannter A -orthogonaler Richtungen zu minimieren.

Definition A.36. Sei A positiv definit. Ein System von Vektoren $p^{(i)}$, $i = 0, \dots, n-1$ mit $p^{(i)} \neq 0$ für alle i heißt A -orthogonal (oder A -konjugiert), falls

$$p^{(i)T} A p^{(j)} = 0 \quad \text{für } i \neq j.$$

□

Zum Zusammenhang A -orthogonal und orthogonal in gewöhnlichem (euklidischen) Sinn:

Sei v^1, \dots, v^n ein Orthogonalsystem, d.h.

$$\begin{aligned} v^i &\neq 0 & i = 1, \dots, n \\ (v^i)^T v^j &= 0 & \text{für } i \neq j \end{aligned}$$

$A = LL^T$ sei eine Cholesky-Zerlegung von $A \in \mathbb{R}^{n \times n}$. Dann ist das System $\{d^i\}$ mit

$$d^i := (L^T)^{-1} v^i$$

A orthogonal.

Die Besonderheit A -orthogonaler Systeme zeigt der folgende Satz, wo ihre Kenntnis bereits angenommen ist:

Satz A.37. Sei $f(x) = \gamma - b^T x + \frac{1}{2} x^T A x$ mit $A = A^T \in \mathbb{R}^{n \times n}$ positiv definit (also $\nabla f(x^*) = 0$ genau für $x^* = A^{-1} b$, x^* strenge globale Minimalstelle). d^0, \dots, d^{n-1} ($\neq 0$) seien A -konjugiert. $x^0 \in \mathbb{R}^n$ beliebig,

$$x^{k+1} = x^k - \sigma_k d^k$$

mit

$$\sigma_k = \frac{\nabla f(x^k)^T d^k}{(d^k)^T A d^k} \quad \left(= \frac{\nabla f(x^k)^T d^k}{2(f(x^k - d^k) - f(x^k) + \nabla f(x^k)^T d^k)} \right)$$

für $k = 0, 1, \dots, n - 1$.
 Dann gilt $x^n = x^*$ □

Man erzielt also finite Konvergenz bei streng konvexen quadratischen Funktionen.

Beim cg-Verfahren (conjugate gradient) von Hestenes und Stiefel (zunächst konzipiert nur für den Fall von Satz A.37, also eigentlich für die Lösung von $Ax = b$ mit $A = A^T$ pos. def.) werden die d^j rekursiv konstruiert mit Hilfe von $\nabla f(x^j)$ und d^{j-1}, \dots, d^0 . Es stellt sich schließlich heraus, daß man dabei nur d^{j-1} benötigt, weil die übrigen Koeffizienten in der Linearkombination

$$d^j = \nabla f(x^j) + \sum_{i=0}^{j-1} \beta_{j,i} d^i$$

alle zu null werden, wenn man immer die optimale Schrittweite benutzt:

$$\left. \begin{aligned} d^0 &:= \nabla f(x^0) \\ d^j &= \nabla f(x^j) + \frac{\|\nabla f(x^j)\|^2}{\|\nabla f(x^{j-1})\|^2} d^{j-1}, \quad j = 1, 2, \dots \text{ solange } \nabla f(x^j) \neq 0. \end{aligned} \right\} \quad (\text{A.4})$$

Resultat:

Satz A.38. Sei f wie in Satz A.37, $\{x^k\}$, $\{\sigma_k\}$ wie dort konstruiert mit $\{d^k\}$ aus (A.4). Dann gibt es ein $N \leq n$ mit $x^N = x^*$. d^0, \dots, d^{N-1} sind A -konjugiert. □

Bemerkung A.39. Sind alle Eigenwerte von A verschieden und

$$\nabla f(x^0) = \sum_{i=1}^n \gamma_i v^i, \quad v^i \text{ Eigenvektoren von } A, \gamma_i \neq 0 \text{ für alle } i,$$

dann ist $N = n$. Sonst kann $N < n$ sein z.B. $A = I : N = 1!$

Das Verfahren ist interessant für $n \gg 1$ wegen des geringen Speicherbedarfs und der finiten Konvergenz.

Aber: Das Konvergenzverhalten ist irregulär. Es hängt ab von der Eigenwertverteilung

von A . Eine ungünstige Situation liegt vor, wenn viele Eigenwerte $\gg \lambda_{\min}$ sind. Die A -Orthogonalität der berechneten Richtungen geht durch Rundungsfehler schnell verloren. Diese Problematik tritt nicht auf, wenn der Quotient $\lambda_{\max}(A)/\lambda_{\min}(A)$ nicht sehr groß ist (A heißt dann "gut konditioniert"). Durch eine Transformation von A kann man versuchen, diese Bedingung zu erreichen. Die Transformation muß natürlich "einfach" sein. Man gelangt so zu den präkonditionierten cg-Verfahren. Die Präkonditionierung muss aber stets problemabhängig gewählt sein. Ein Gleichungssystem mit der Präkonditionierungsmatrix muss mit geringem Aufwand lösbar sein. Im Folgenden ist $\{A_i\}$ eine Folge "einfach strukturierter" Matrizen, die (grobe) Näherungen für $\nabla^2 f(x^*)$ sein sollten. Bei nichtquadratischen Funktionen ist die exakte eindimensionale Minimierung nicht möglich. Deshalb wird hier ein Korrekturfaktor bei $\nabla f(x^k)$ angebracht, um die strenge Gradientenbezogenheit der Richtungen sicherzustellen.

Algorithmus: modifiziertes Fletcher-Reeves-Verfahren

$\{A_i\}$ sei beschränkte Folge von positiv definiten Matrizen mit

$$d^T A_i d \geq \alpha d^T d, \quad \alpha > 0 \text{ unabhängig von } i, \text{ für alle } d \in \mathbb{R}^n.^3$$

x^0 gegeben. $k = 0, 1, \dots,$

$$\begin{aligned} B &:= A_{\lfloor k/n \rfloor} && \left\{ \begin{array}{l} B \text{ ist der sogenannte Präkonditionierer} \\ B \text{ ist über } n \text{ Schritte konstant und} \\ \text{wird dann eventuell geändert} \end{array} \right. \\ g^k &:= \nabla f(x^k) \\ B\tilde{g}^k &= g^k \text{ nach } \tilde{g}^k \text{ auflösen} \\ \gamma_k &:= (\tilde{g}^k)^T g^k \quad (> 0) \end{aligned}$$

$$d^k = \begin{cases} \tilde{g}^k & \text{falls } 0 = k \pmod{n} \\ \left(1 - \frac{(d^{k-1})^T g^k}{\gamma_{k-1}}\right) \tilde{g}^k + \frac{\gamma_k}{\gamma_{k-1}} d^{k-1} & \text{sonst} \end{cases}$$

σ_k aus dem Goldstein-Armijo-Abstiegstest, asymptotisch exakt von erster Ordnung ⁴
 $x^{k+1} = x^k - \sigma_k d^k.$

Satz A.40. Sei $f \in C^{2,1}(\mathbb{R}^n)$ und $\mathcal{L}_f(f(x^0))$ beschränkt. Dann gilt für den obigen Algorithmus: $\nabla f(x^k) \rightarrow 0$ und $\nabla f(x^*) = 0$ für jeden Häufungswert von $\{x^k\}$. Hat f nur endlich viele Gradientennullstellen auf $\mathcal{L}_f(f(x^0))$, dann gilt $\lim_{x \rightarrow \infty} x^k = x^*$ mit $\nabla f(x^*) = 0$. Falls dort $\nabla^2 f(x^*)$ pos. def. ist, gilt sogar $\|x^{(k+1)n} - x^*\| \leq c \|x^{kn} - x^*\|^2$ mit einer geeigneten Konstanten $c > 0$, d.h. die Konvergenzgeschwindigkeit ist n -Schrittquadratisch. \square

Man kann diesen Algorithmus modifizieren, indem man nach weniger als n Schritten ein neues B wählt und d^k durch \tilde{g}^k neu initialisiert.

³In der Regel wird $\{A_i\}$ erst im Laufe des Algorithmus mitkonstruiert

⁴ $|\sigma_k - \bar{\sigma}_k| \leq \text{cons} \|\nabla f(x^k)\|$ mit $\bar{\sigma}_k$ "exaktes eindimensionales Minimum", d.h. $\bar{\sigma}_k : \nabla f(x^k - \bar{\sigma}_k d^k)^T d^k = 0, \quad \bar{\sigma}_k > 0$ minimal.

NUMAWWW

A.3.2.4 Vertrauensbereichsmethoden

Beim Vertrauensbereichskonzept steht zu Beginn des k -ten Schrittes ein quadratisches (oder lineares) Approximationsmodell für f zur Verfügung,

$$f(x) \approx \varphi_k(x) = f(x^k) + \nabla f(x^k)^T(x - x^k) + \frac{1}{2}(x - x^k)^T A_k(x - x^k)$$

sowie ein (vorläufiger) "Vertrauensbereichradius" $\tilde{\Delta}_k$, d.h. wir erwarten, daß das quadratische Modell f genügend genau beschreibt in der Kugel

$$\|x - x^k\|_u \leq \tilde{\Delta}_k.$$

Dabei kann die Norm $\|\cdot\|_u$ sehr verschiedenartig gewählt sein, z.B. als euklidische Norm, als euklidische Norm mit Gewichtung oder auch als Maximumnorm. Nun löst man das restringierte Ersatzproblem

$$\varphi_k(x) \stackrel{!}{=} \min_x \text{ mit der Nebenbedingung } \|x - x^k\|_u \leq \tilde{\Delta}_k. \quad (\text{A.5})$$

Hierbei ist tatsächlich an die globale Lösung des Problems gedacht. Diese kann im Fall der euklidischen Norm wie folgt charakterisiert werden:

Satz A.41. $d^k = x - x^k$ ist genau dann eine globale Lösung des Problems A.5 (mit der euklidischen Vektornorm) wenn gilt:

- 1 $\lambda_k \geq 0$, $\|d^k\| \leq \tilde{\Delta}_k$, $\lambda_k(\tilde{\Delta}_k - \|d^k\|) = 0$
- 2 $(A_k + \lambda_k I)d^k = -\nabla f(x^k)$
- 3 $A_k + \lambda_k I$ ist positiv semidefinit.

Bew.: bei Sorensen, SIAM J. Numer. Anal. 19, 1982, 409–426. □

Ist die Norm die euklidische und ist A_k positiv semidefinit, dann läuft dies auf die Lösung einer nichtlinearen skalaren Gleichung für den Parameter λ_k und die Lösung des linearen Gleichungssystems

$$(A_k + \lambda_k I)d^k = -\nabla f(x^k)$$

hinaus, ist also recht einfach zu realisieren. Dabei ist entweder $\lambda_k = 0$, wenn nämlich A_k positiv definit ist und

$$A_k d^k = -\nabla f(x^k) \quad \text{mit} \quad \|d^k\|_u \leq \tilde{\Delta}_k,$$

oder es ist $\lambda_k > 0$ und λ_k dann bestimmt durch die Gleichung

$$\|d^k\|_u = \tilde{\Delta}_k.$$

Ist A_k nicht positiv semidefinit, muss man auch noch die Eigenwerte von $A_k + \lambda_k I$ kontrollieren.

Ist $A_k = O$ und die Norm die Maximumnorm, dann ist ein lineares Optimierungsproblem zu lösen, was auch routinemässig möglich ist. Sodann wird gesetzt

$$\tilde{x}^{k+1} = x^k + d^k$$

und es wird getestet, ob $\tilde{\Delta}_k$ angemessen gewählt war, also ob die Verkleinerung in f der von φ_k in etwa entspricht. Natürlich ist

$$\varphi_k(x^k) - \varphi_k(\tilde{x}^{k+1}) = f(x^k) - \varphi_k(\tilde{x}^{k+1}) > 0$$

Wir bilden die Testgrösse

$$\varrho_k \stackrel{\text{def}}{=} \frac{f(x^k) - f(\tilde{x}^{k+1})}{f(x^k) - \varphi_k(\tilde{x}^{k+1})}.$$

Wenn $\varrho_k \leq \varepsilon$ mit $0 < \varepsilon \ll 1$ gilt, dann wird der Schritt verworfen, wir halbieren z.B. $\tilde{\Delta}_k$ und wiederholen den Schritt. Wenn $\varepsilon \leq \varrho_k \leq 1 - \eta_1$ mit $0 < \eta_1 < 1 - \varepsilon$, dann akzeptieren wir den Schritt und setzen

$$x^{k+1} = \tilde{x}^{k+1}, \tilde{\Delta}_{k+1} = \Delta_k = \tilde{\Delta}_k.$$

War aber $\varrho_k > 1 - \eta_1$, dann akzeptieren wir den Schritt ebenfalls, erhöhen aber möglicherweise $\tilde{\Delta}_{k+1}$:

$$x^{k+1} = \tilde{x}^{k+1}, \Delta_k = \tilde{\Delta}_k, \tilde{\Delta}_{k+1} = \min\{2\Delta_k, \Delta_{\max}\},$$

wobei Δ_{\max} ebenfalls benutzerspezifiziert ist. Der wesentliche Unterschied zum ersten Konzept besteht hier darin, daß die Korrekturrichtung $(x^{k+1} - x^k)/\|x^{k+1} - x^k\|$ sich mit Δ_k ändert. Auch sind die Voraussetzungen, die man an A_k stellen muß, schwächer. Der algebraische Aufwand pro Schritt ist aber häufig höher. Es gilt folgender allgemeiner Konvergenzsatz:

Satz A.42. *f sei zweimal stetig differenzierbar auf der offenen Menge \mathcal{D} und nach unten beschränkt. Die Matrizenfolge $\{A_k\}$ sei beschränkt. Dann erfüllt jeder Häufungspunkt x^* von $\{x^k\}$ die Bedingung*

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

Ist insbesondere der Niveaubereich $\mathcal{L} = \{x : f(x) \leq f(x^0)\}$ kompakt, dann hat jede unendliche Teilfolge dieser Folge einen solchen Häufungspunkt. Ist in einem solchen Häufungspunkt $\nabla^2 f(x^)$ positiv definit, dann konvergiert die Gesamtfolge gegen diesen. Ist*

$$A_k = \nabla^2 f(x^k)$$

dann erfüllt jeder Häufungspunkt auch die notwendige Bedingung zweiter Ordnung $\nabla^2 f(x^)$ positiv semidefinit.*

Man beachte, daß hier sogenannte Richtungen negativer Krümmung nicht explizit konstruiert werden müssen. Bew.: bei Schultz, Byrd und Schnabel: A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties. SIAM J. Numer. Anal. 22, (1985), 47–67. \square

A.3.2.5 Verfahren, die nur Funktionswerte benutzen

In vielen Anwendungen sind die zu minimierenden Funktionen zwar differenzierbar, aber die Gradienten sind nicht in analytischer Form verfügbar, entweder weil sie nur implizit definiert sind oder der Algorithmus zur Berechnung der Funktionswerte selbst dem Anwender nur als “black box“ vorliegt. Die Berechnung der Gradienten durch Differenzenformeln ist sehr aufwendig und überdies fragwürdig, wenn die Funktionswerte nur mit vergleichsweise geringer Genauigkeit (etwa nur auf 3 Dezimalstellen) verfügbar sind, was häufig der Fall ist. Deshalb haben Verfahren, die nur mit Funktionswerten arbeiten und die Funktion auf einem im Vergleich zu Differenzenformeln groben Gitter abtasten, immer grosses Interesse gefunden. Diese sogenannte “ableitungsfreie Minimierung“ ist Gegenstand aktueller Forschung. Das bei den Anwendern beliebteste Verfahren dieser Art, das Verfahren von Nelder und Mead (Computer Journal 7, (1965),308-313), soll hier kurz skizziert werden. In jedem Schritt des Verfahrens liegt ein Simplex im \mathbb{R}^n vor, also die konvexe Hülle von $n+1$ affin unabhängigen Punkten x^i , $i = 0, \dots, n$, an denen die Funktionswerte $f(x^i)$ bekannt sind. Ein Schritt des Verfahrens erzeugt einen neuen Simplex so, daß der grösste Funktionswert auf den Ecken dieses neuen Simplex kleiner ist als der grösste Funktionswert auf dem zuvor berechneten. Dies geschieht durch eine Anwendung von Modifikationsregeln. Zu einem Simplex mit den Ecken x^i , $i = 0, \dots, n$, sei definiert

$$\begin{aligned} h &= \operatorname{argmax}\{i : f(x^i) = \max\{f(x^j) : j = 0, \dots, n\}\} \\ sh &= \operatorname{argmax}\{i : f(x^i) = \max\{\{f(x^0), \dots, f(x^n)\} \setminus \{f(x^h)\}\}\} \\ l &= \operatorname{min}\{i : f(x^i) = \min\{f(x^j) : j = 0, \dots, n\}\} \end{aligned}$$

d.h. h ist der grösste Index unter den (möglicherweise mehreren) Ecken, die den höchsten Funktionswert tragen, sh der grösste Index für den zweithöchsten Funktionswert und l der kleinste Index für den kleinsten Funktionswert. Ferner ist

$$\bar{x} = \frac{1}{n} \left(\sum_{i=0}^n x^i - x^h \right).$$

Man hat Verfahrensparameter $\alpha > 0$, $\gamma > 1$ und $0 < \beta < 1$ und $0 < \delta < 1$, den Reflektionskoeffizienten, den Expansionskoeffizienten, den Kontraktionskoeffizienten und den Koeffizienten der massiven Kontraktion. Typische Werte sind

$$\alpha = 1, \beta = 0.5, \gamma = 2, \delta = 0.5.$$

Folgende Operationen sind vorgesehen:

1 Der Reflektionsschritt. Definiere

$$x^r = (1 + \alpha)\bar{x} - \alpha x^h .$$

x^r liegt also auf der Geraden durch x^h und \bar{x} , den Schwerpunkt der n von x^h verschiedenen Punkte, jedoch auf der x^h gegenüberliegenden Seite zu x^h . Drei Fälle können bei der Reflektion auftreten:

- 1.1 $f(x^r) < f(x^l)$. In diesem Fall wird ein Expansionsschritt ausgeführt.
- 1.2 $f(x^l) \leq f(x^r) < f(x^{sh})$. In diesem Fall wird x^h durch x^r ersetzt und ein neuer Schritt des Verfahrens gestartet.
- 1.3 $f(x^{sh}) \leq f(x^r)$. In diesem Fall wird unterschieden: ist $f(x^r) < f(x^h)$, so wird x^h durch x^r ersetzt und anschliessend ein Kontraktionsschritt ausgeführt (s.u.). Andernfalls wird sofort ein Kontraktionsschritt ausgeführt. Sei dazu

$$f(x^{h'}) = \min\{f(x^r), f(x^h)\} .$$

2 Der Kontraktionsschritt. Der Kontraktionspunkt x^c ist definiert durch

$$x^c = \beta x^{h'} + (1 - \beta)\bar{x} .$$

β ist das Verhältnis der Abstände $\|x^c - \bar{x}\|$ und $\|x^{h'} - \bar{x}\|$. Nach Definition von h' ist also

$$x^c = \beta x^h + (1 - \beta)\bar{x} \text{ wenn } f(x^r) \geq f(x^h)$$

bzw.

$$x^c = \beta x^r + (1 - \beta)\bar{x} \text{ wenn } f(x^r) < f(x^h) .$$

Ist nun $f(x^c) < f(x^h)$, so wird x^h durch x^c ersetzt und der nächste Schritt wird begonnen, andernfalls wird eine massive Kontraktion ausgeführt.

3 Die massive Kontraktion. In diesem Fall werden für $i \neq l$ alle Punkte durch

$$\hat{x}^i = x^i + \delta(x^l - x^i)$$

ersetzt. Diese massive Kontraktion tritt vor allem in zwei Situationen auf: x^h befindet sich in der Nähe eines Sattelpunktes und die Reflektion wird in Richtung eines Anstieges ausgeführt oder die Niveaubereiche von f sind sehr stark gekrümmt und $f(x^h)$ ist sehr viel grösser als alle anderen Funktionswerte.

4 Der Expansionsschritt. Dieser Schritt wird ausgeführt, wenn $f(x^r) < f(x^l)$. Man kann dann hoffen, in der Richtung von x^h nach x^r noch kleinere Werte von f zu finden. Man setzt

$$x^e = \gamma x^r + (1 - \gamma)\bar{x} .$$

Ist $f(x^e) < f(x^r)$, dann wird x^h durch x^e , sonst wird x^h durch x^r ersetzt und der Algorithmus neu gestartet.

Bei diesem Algorithmus werden also pro Schritt (pro neuem Simplex) ein, zwei oder n neue Funktionswerte berechnet, letzteres nur im Falle der massiven Kontraktion. Als Abbruchkriterium schlagen Nelder und Mead

$$\sqrt{\frac{1}{n} \sum_{i=0}^n (f(x^i) - \bar{f})^2} < \varepsilon$$

vor mit

$$\bar{f} = \frac{1}{n+1} \sum_{i=0}^n f(x^i).$$

Dieses Abbruchkriterium für sich alleine ist aber unsinnig, da es nichts über die Grösse des Simplex aussagt. Es ist ja z.B. möglich, daß alle Funktionswerte (fast) gleich sind, ohne daß ein Minimum von f auch nur in der Nähe des Simplex liegt. Das Verfahren funktioniert in der Praxis häufig erstaunlich gut, trotz seiner primitiven und heuristischen Struktur. Es sind aber (konvexe) Fälle bekannt, in denen es beweisbar gegen einen nichtstationären Punkt konvergiert, d.h. $\nabla f(x^*) \neq 0$. (McKinnon, SIOPT 1998) Dies sind Fälle, in denen der Simplex immer (allmählich) entartet, d.h. die Ecken werden affin abhängig. Man kann den Algorithmus aber durch Modifikationen beweisbar konvergent machen. Dazu gehören eine Kontrolle der hinreichend guten affinen Unabhängigkeit der Ecken, was man durch eine QR-Zerlegung der Vektoren $x^i - x^l$ testen kann, mit eventuellem Neuaufbau eines besseren Simplex. Sei dazu

$$X = (x^1 - x^0, \dots, x^n - x^0)$$

wobei man zweckmässigerweise $x^0 = x^l$ wählt, die Richtungsmatrix des Simplex \mathcal{S} und

$$QXP = R,$$

mit einer unitären Matrix Q , einer Permutationsmatrix P und einer oberen Dreiecksmatrix R mit

$$|r_{1,1}| \geq |r_{2,2}| \geq \dots \geq |r_{n,n}|.$$

(Dies kann durch die Spaltenvertauschung erreicht werden). Dann ist bekanntlich

$$vol(\mathcal{S}) = |\det(R)|/n!$$

Entartet der Simplex, dann wird

$$\frac{|r_{n,n}|}{|r_{1,1}|} \ll 1$$

und man kann die entsprechende Richtung $x^j - x^0$ durch eine Richtung $\pm q^n$ ersetzen, wo q^n die letzte Spalte von Q^T ist. Die Länge des Schrittes wählt man zweckmässig in der Grössenordnung

$$\max\{\|x^j - x^0\| : 1 \leq j \leq n\}.$$

Ist der Simplex zusammengeschrumpft, verwendet man vor dem Abbruch noch Schritte mit "exakter" eindimensionaler Minimierung durch z.B. eines der zuvor besprochenen Verfahren zur eindimensionalen Minimierung längs Richtungen, die durch die Spalten von Q^T gegeben sind, vom laufenden Punkt x^l aus.

Man kann auch daran denken, aus der linearen Approximation an f , die man durch affin lineare Interpolation der $n + 1$ Funktionswerte erhält, eine Gradientenapproximation für f zu erhalten und diese als Suchrichtung von x^l aus zu benutzen. Man hat dann wieder eine Art Differenzenformel. Ebenso kann man aus $(n + 1)(n + 2)/2$ Funktionswerten eine quadratische Approximation an f gewinnen und diese minimieren, um so einen neuen Näherungspunkt zu gewinnen. Für $n = 2$ kann man z.B. f an den Ecken und Seitenmitten eines Dreiecks auswerten und durch eine quadratische Funktion

$$p(x_1, x_2) = f_0 + g_1(x - x_0) + g_2(y - y_0) + \frac{1}{2}(h_{11}(x - x_0)^2 + 2h_{12}(x - x_0)(y - y_0) + h_{22}(y - y_0)^2)$$

interpolieren, p unter einer trust region-Restriktion

$$(x - x_0)^2 + (y - y_0)^2 \leq \Delta$$

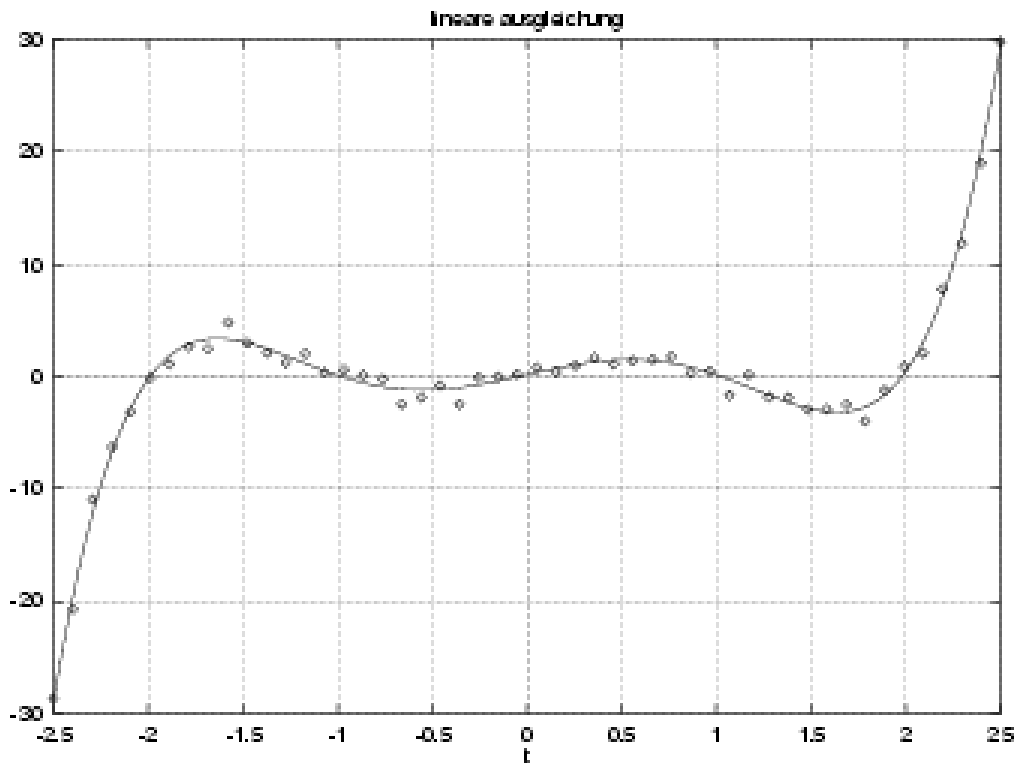
minimieren und so einen neuen verbesserten Näherungswert für das Minimum von f gewinnen.

Verfahren dieser Art sind Gegenstand aktueller Forschung. ("derivative free optimization", DFO).

A.3.3 Verfahren zur Minimierung einer Summe von Quadraten (Ausgleichsrechnung)

Ein wichtiger und häufiger Spezialfall der unrestringierten Minimierung ist der der Parameteranpassung eines mathematischen Modells für gegebene Messwerte:

$$f(x) = \frac{1}{2} \sum_{i=1}^N (y_i - g_i(x))^2, \quad g_i : \mathbb{R}^n \rightarrow \mathbb{R}, \quad y_i \text{ gegeben}, \quad N \geq n, \quad \text{oft } N \gg n$$



A.3.3.1 Lineare Ausgleichsrechnung

Hier sind die Ansatzfunktionen g_i **linear in** x , also $\nabla g_i = a^i$, $g_i(x) = (a^i)^T x$ z.B.

$$(a^i)^T = (1, t_i, t_i^2, \dots, t_i^{n-1}) \text{ Polynomannpassung}$$

Mit

$$y := \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \quad A := \begin{pmatrix} (a^1)^T \\ \vdots \\ (a^N)^T \end{pmatrix}$$

ergibt sich folgende Darstellung:

$$\begin{aligned} f(x) &= \frac{1}{2}(y - Ax)^T(y - Ax) \\ \nabla f(x) &= A^T(Ax - y) \\ \nabla^2 f(x) &= A^T A \quad \text{pos. semidefinit und positiv definit, wenn Rang}(A) = n \end{aligned}$$

Die Lösung des Problems ergibt sich theoretisch aus dem linearen Gleichungssystem

$$(A^T A)x^* = A^T y,$$

dem ‘Gauß’schen Normalgleichungssystem‘.

Dieses System ist jedoch sehr rundungsfehlerempfindlich wenn $\lambda_{\max}(A^T A)/\lambda_{\min}(A^T A) \gg 1$. Meist besser geeignet sind hier Orthogonalisierungsmethoden, siehe im Folgenden.

Die Householder-QR-Zerlegung:

Es sei $A \in \mathbb{R}^{N \times n}$, $N \geq n$.

Gesucht wird eine orthonormale $(N \times N)$ -Matrix Q , sodaß

$$QA = \begin{pmatrix} R \\ \cdots \\ 0 \end{pmatrix}$$

mit einer oberen $n \times n$ -Dreiecksmatrix R . Dies ist die sogenannte QR-Zerlegung.

(Ist A spaltenregulär, dann ist R invertierbar und umgekehrt!)

Q wird nicht explizit gebildet, sondern entsteht nur implizit als Produkt von n sogenannten "Householdermatrizen" U_i :

$$Q = U_n \cdots U_1.$$

Die U_i sind spezielle orthonormale Matrizen, nämlich Spiegelungen an Hyperebenen mit Normalenvektor u^i , also

$$U_i = I - \beta_i u^i (u^i)^T \quad \text{mit} \quad \beta_i = 2 / (u^i)^T u^i.$$

Es werden im Algorithmus nur die u^i gebildet.

u^1 berechnet sich aus der ersten Spalte von A , u^2 aus der zweiten Spalte von $U_1 A, \dots$, u^i aus der i -ten Spalte von $U_{i-1} \cdots U_1 A$, jeweils aus der Forderung in dieser Spalte die Dreiecksstruktur zu erzeugen. Setzt man

$$\begin{aligned} A_1 &:= A \\ A_i &:= U_{i-1} \cdots U_1 A \quad \text{für} \quad i \geq 2 \end{aligned}$$

und

$$A_j = (a_{ik}^{(j)})$$

dann ergibt sich mit

$$\begin{aligned} \sigma_i &:= \left(\sum_{j=i}^N (a_{ji}^{(i)})^2 \right)^{1/2} \\ \theta_i &:= \begin{cases} 1 & \text{falls } a_{ii}^{(i)} \geq 0 \\ -1 & \text{sonst} \end{cases} \end{aligned}$$

$$u^i = \left(\underbrace{0, \dots, 0}_{i-1}, \underset{\uparrow}{a_{ii}^{(i)}} + \sigma_i, a_{i+1,i}^{(i)}, \dots, a_{N,i}^{(i)} \right)^T$$

Bei der Bildung von $A_{i+1} = U_i A_i$ beachte man

$$A_{i+1} = (I - \beta_i u^i (u^i)^T) A_i = A_i - (\beta_i u^i) \underbrace{((u^i)^T A_i)}_{(A_i^T u^i)^T}$$

Man bildet also

$$v^i := \beta_i u^i, \quad w^i := A_i^T u^i$$

und subtrahiert das dyadische Produkt $(v^i)(w^i)^T$ von A_i

Danach ergibt sich die Lösung der linearen Ausgleichsaufgabe wie folgt: Q bestimmt aus A wie oben beschrieben:

$$\begin{aligned} 2f(x) &= (y - Ax)^T (y - Ax) \\ &= (y - Ax)^T Q^T Q (y - Ax) \\ &= \left(\hat{y} - \begin{pmatrix} R \\ 0 \end{pmatrix} x \right)^T \left(\hat{y} - \begin{pmatrix} R \\ 0 \end{pmatrix} x \right) \quad \text{mit} \quad Qy = \begin{pmatrix} \hat{y}^1 \\ \hat{y}^2 \end{pmatrix} \begin{matrix} \} n \\ \} N - n \end{matrix} \\ &= (\hat{y}^1 - Rx)^T (\hat{y}^1 - Rx) + \hat{y}^{2T} \hat{y}^2 \end{aligned}$$

minimal für $x = R^{-1} \hat{y}^1$, falls A spaltenregulär (R invertierbar) d.h. zu lösen ist lediglich

$$Rx = \hat{y}^1 .$$

NUMAWWW

A.3.3.2 Nichtlineare Ausgleichsrechnung

Jetzt betrachten wir den Spezialfall

$$f(x) = \frac{1}{2} \|F(x)\|_2^2 \quad \text{mit} \quad F: \mathbb{R}^n \rightarrow \mathbb{R}^N, \quad N \geq (\gg) n .$$

Z.B. kann bei einer Datenanpassung

$$F_i(x) = y_i - g(t_i, x)$$

sein mit einer gegebenen "Modellfunktion" g , die nichtlinear von x abhängt, z.B.

$$g(t, x) = x_1 + x_2 \exp(x_3 t) + x_4 \exp(x_5 t)$$

Die iterative Lösung des Problems kann durch das **Gauß-Newton-Verfahren** geleistet werden. Dies entsteht, indem man pro Schritt die Abstiegsrichtung aus einer linearen Ausgleichsaufgabe, der Linearisierung von $F_i(x) = y_i - g_i(x)$, $g_i(x) = g(t_i, x)$ bei x^k berechnet:

$$\begin{aligned} \frac{1}{2} \sum_{i=1}^N (y_i - g_i(x^k + d^k))^2 &= \frac{1}{2} \sum_{i=1}^N \left(\underbrace{y_i - g_i(x^k)}_{=: r_i^k} - \underbrace{\nabla g_i(x^k)^T}_{=: (a_i^k)^T} d^k + \mathcal{O}(\|d^k\|^2) \right)^2 \\ &\approx \frac{1}{2} (r^k - A_k d^k)^T (r^k - A_k d^k) \quad \text{mit} \quad A_k = \begin{pmatrix} (a_k^1)^T \\ \dots \\ \vdots \\ \dots \\ (a_k^N)^T \end{pmatrix} \end{aligned}$$

d^k wird also bestimmt aus

$$(r^k - A_k d^k)^T (r^k - A_k d^k) = \min_{d^k} .$$

$$x^{k+1} = x^k + \sigma_k d^k$$

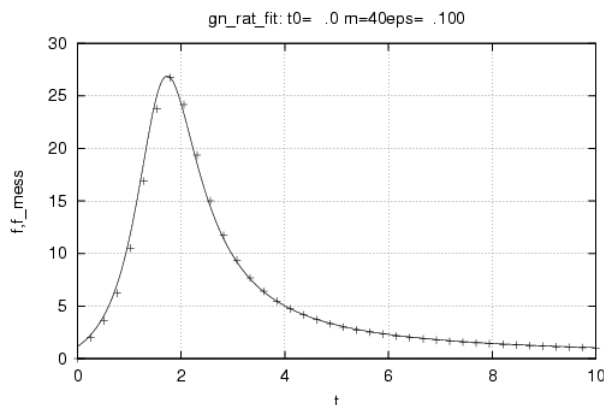
mit σ_k aus einem Abstiegstest für $f(x^k + \sigma d^k)$, $f(x) = \sum_{i=1}^N (y_i - g_i(x))^2$, also

$$f(x^k) - f(x^k + \sigma d^k) \geq \delta \sigma \nabla f(x^k)^T d^k = -\delta \sigma (A_k^T r^k)^T d^k$$

wobei $\sigma = \beta^j \sigma_{0,k}$ wie im Goldstein-Armijo-Test gebildet wird.

($0 < \delta < \frac{1}{2}$, $0 < \beta < 1$). Dies ist das sogenannte ‘‘gedämpfte‘‘ Gauß-Newton-Verfahren.

In der folgenden Abbildung ist $g_i(x) = (x_1 + x_2 t_i)/(1 + x_3 t_i + x_4 t_i^2)$ und $y_i = f_{\text{mess}}(t_i)$ ist aus g_i durch Überlagerung mit Pseudozufallszahlen entstanden.



Für das gedämpfte Gauß-Newton-Verfahren gilt folgender Konvergenzsatz:

Satz A.43. Die N Funktionen g_1, \dots, g_N seien zweimal stetig differenzierbar auf der offenen, konvexen Menge \mathcal{D} und für $x^0 \in \mathcal{D}$ sei

$$\mathcal{L}_f(f(x^0)) = \{x \in \mathcal{D} : f(x) \leq f(x^0)\} \quad \text{mit} \quad f(x) = \frac{1}{2} \sum_{i=1}^N (y_i - g_i(x))^2$$

kompakt. Ferner sei mit $g := (g_1, \dots, g_N)^T$ die Matrix

$$A(x) = \mathcal{J}_g(x) \in \mathbb{R}^{N \times n}$$

für alle $x \in \mathcal{L}_f(f(x^0))$ spaltenregulär. Dann konvergiert das gedämpfte Gauß-Newton-Verfahren im Sinne von $\nabla f(x^k) \rightarrow 0$. Falls ∇f nur endlich viele Nullstellen auf $\mathcal{L}_f(f(x^0))$ hat, konvergiert die Gesamtfolge gegen eine davon.

Die Konvergenzgeschwindigkeit ist abhängig vom Optimalwert $f(x^*)$. Wenn die benutzten Schrittweiten asymptotisch exakt sind und $\nabla^2 f(x^*)$ positiv definit ist, dann ist für $f(x^*) = 0$ die Konvergenz quadratisch, sonst nur linear und um so langsamer, je größer $f(x^*)$. Genauer gilt dann

$$\|x^{k+1} - x^*\|_* \leq \frac{|\lambda_1 - \lambda_n|}{2 + \lambda_1 + \lambda_n} \|x^k - x^*\|_* + \mathcal{O}(\|x^k - x^*\|_*^2)$$

wobei

$$\begin{aligned} \lambda_1 &= \lambda_{\max}((A^T(x^*)A(x^*))^{-1}C(x^*)) , \\ \lambda_n &= \lambda_{\min}((A^T(x^*)A(x^*))^{-1}C(x^*)) , \\ C(x) &= \sum_{i=1}^N F_i(x)\nabla^2 F_i(x) , \\ \|z\|_* &= (z^T \nabla^2 f(x^*) z)^{1/2} \end{aligned}$$

($1 + \lambda_n$ ist positiv, wenn $\nabla^2 f(x^*)$ positiv definit ist). □

NUMAWWW

Das Gauss-Newton-Verfahren versagt, wenn J_F nicht den vollen Rang besitzt. In diesem Fall ist das Levenberg-Marquardt-Verfahren von Nutzen. Dieses entspricht genau dem Schema der Vertrauensbereich-Verfahren. Das Subproblem lautet jetzt:

Bestimme d^k aus

$$\|F(x^k) + J_F(x^k)d\| = \min_d \text{ mit } \|d\| \leq \Delta_k .$$

Für d^k ergibt sich so eine Gleichung

$$(J_F(x^k)^T J_F(x^k) + \lambda_k I) d^k = -J_F(x^k)^T F(x^k)$$

worin $\lambda_k = 0$ oder durch die Bedingung $\|d^k\| = \Delta_k$ festgelegt ist und Δ_k durch das Abstiegsverhalten gesteuert wird.

NUMAWWW

A.3.3.3 Orthogonale Regression

Die gewöhnliche Ordinatenfehlerausgleichung liefert oft unzureichende Resultate, etwa wenn die in den Funktionen g_i eingearbeiteten Meßstellen t_i selbst nur ungenau bekannt sind:

$$y_i \approx F(t_i, x), \quad y_i \text{ und } t_i \text{ meßfehlerbehaftet.}$$

Dann ist die sogenannte orthogonale Regression angebracht:

$$\sum_{i=1}^N \{\delta_i^2 + (y_i - F(t_i + \delta_i; x))^2\} \stackrel{!}{=} \min_{x, \delta_1, \dots, \delta_N} .$$

Ebenso bei der Anpassung einer implizit definierten Kurve mit Parametern a_1, \dots, a_n an eine Punktmenge (x_i, y_i) , $i = 1, \dots, N$,

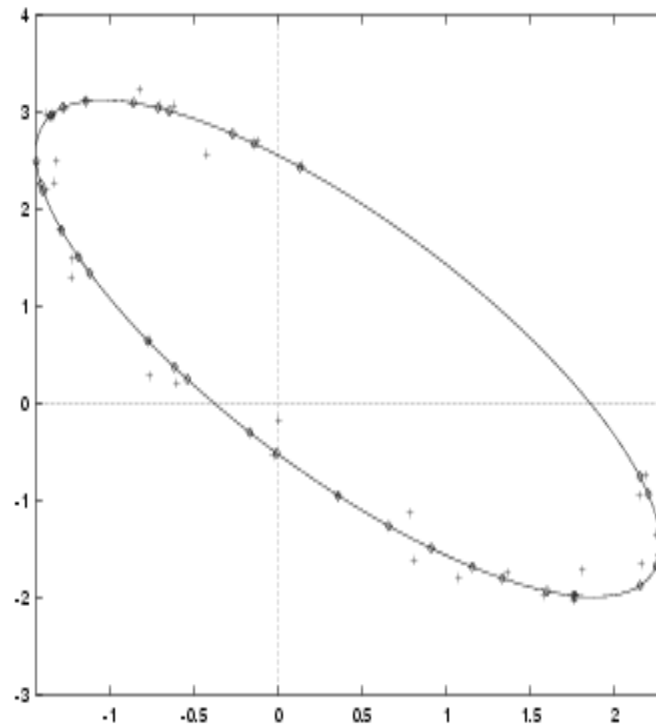
$$\sum_{i=1}^N \{(x_i - \bar{x}_i)^2 + (y_i - \bar{y}_i)^2\} = \min_{\bar{x}_1, \dots, \bar{y}_N, a}$$

mit

$$F(\bar{x}_i, \bar{y}_i, a) = 0 \quad i = 1, \dots, N$$

als Nebenbedingungen (siehe Kapitel C). Es gibt speziell an die Struktur dieser Aufgabe angepasste Versionen der nichtlinearen Ausgleichsrechnung.

NUMAWWW (z.Zt. nur Ellipsenfit)



Diese orthogonale Regression ist ausserordentlich aufwendig, weil nun ein hochdimensionales Problem zu lösen ist. Das gilt selbst dann, wenn man die spezielle Struktur des Problems ausnutzt. Häufig begnügt man sich deshalb auch in diesem Fall damit, die Quadratsumme der Einsetzfehler in der impliziten Gleichung zu minimieren. Beim Ellipsenfit an eine Punktmenge (x_i, y_i) in der Ebene bedeutet diese z.B. daß man

$$\sum_i r_i^2$$

minimiert mit

$$r_i = (l_{11}(x_i - x^{(m)}) + l_{12}(y_i - y^{(m)}))^2 + (l_{22}(y_i - y^{(m)}))^2 - 1$$

unter einer geeigneten Normierungsrestriktion, z.B.

$$l_{11} \geq 0 \quad l_{22} \geq 0 .$$

Die Optimierungsvariablen sind hier die Koeffizienten l_{11} , l_{12} , l_{22} , $x^{(m)}$, $y^{(m)}$. Dabei wurde die positiv (semi)definite Koeffizientenmatrix der Ellipsendarstellung sogleich in der Zerlegung nach Cholesky dargestellt, wodurch die Nebenbedingung der Semidefinitheit automatisch realisiert ist.

Kapitel B

Restringierte Optimierung

B.1 Einführung

Die Problemstellung lautet nun:

$$\left. \begin{aligned} f(x) &\stackrel{!}{=} \min_{x \in \mathfrak{S}} \\ \mathfrak{S} &= \{x \in \mathbb{R}^n : g(x) \geq 0, h(x) = 0\} \quad \mathfrak{S} \text{ "zulässige Menge"} \end{aligned} \right\} \text{NLO}$$

Voraussetzung: $f, g, h \in C^1(\mathbb{R}^n)$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$
 $\mathfrak{S} \neq \emptyset$.

Bemerkung B.1. *In der Literatur sind oft auch andere Formen der Problemstellung in Gebrauch, z.B. mit Ungleichungen ≤ 0 oder zweiseitigen Ungleichungen, die sich jedoch alle in das obige Modell umformen lassen.*

Bemerkung B.2. *Es gibt auch brauchbare Optimierungsverfahren für konvexe nichtdifferenzierbare Probleme, solange man sogenannte Subgradienten, d.h. Vektoren g mit*

$$f(x+y) \geq f(x) + g^T y \quad \forall y,$$

einfach berechnen kann. Diese erfordern aber spezielle Konstruktionen, auf die wir hier nicht eingehen können.

Die Existenz einer Lösung von NLO ist gesichert, wenn z.B. $\mathcal{L}_f(f(x^0)) \cap \mathfrak{S}$ beschränkt ist für ein $x^0 \in \mathfrak{S}$. Die Überprüfung der Voraussetzung $\mathfrak{S} \neq \emptyset$ ist durchaus nichttrivial.

Hier sind Schwierigkeitsgrade zu unterscheiden: (steigend)

- a) Einfachster Fall: $\mathfrak{S} = \{x \in \mathbb{R}^n : a \leq x \leq b\}$, $a < b \in \mathbb{R}^n$ gegeben
("box constraints")

- b) Ebenfalls noch effizient und zuverlässig behandelbar: h, g affin linear. Die zulässige Menge ist dann ein Polyeder und im kompakten Fall ein Polytop. Die Struktur dieser Mengen ist genau bekannt und kann algorithmisch gut ausgenutzt werden.

Dabei Spezialfall $f(x) = c^T x \quad \mapsto \quad \text{lineare Optimierung}$
 $f(x) = \frac{1}{2}x^T A x - b^T x \quad \mapsto \quad \text{quadratische Optimierung.}$

Für lineare und konvexe quadratische Optimierung sind finite Verfahren bekannt. (s.h.)

- c) h nichtlinear, aber ∇h immer spaltenregulär. Keine weiteren Restriktionen. Dann kann man unter Benutzung des Hauptsatzes über implizite Funktionen das Problem i.w. in ein unrestringiertes umwandeln. Dies geschieht natürlich numerisch, nicht analytisch.
- d) f konvex, h affin linear, g_i konkav, $i = 1, \dots, m$
 \mapsto "konvexe Optimierung"
- e) Allgemeiner obiger Fall (häufig z.B. in der Strukturoptimierung).

Eine sinnvolle Umformung kann sein:

$$g_i(x) \geq 0 \quad g_i \text{ nichtlinear} \Leftrightarrow g_i(x) - x_{n+i} = 0, \quad x_{n+i} \geq 0.$$

(d.h. die Einführung zusätzlicher vorzeichenbeschränkter Variablen, sogenannte "Schlupfvariablen") jedenfalls wenn die Anzahl m der Ungleichungen nicht $\gg n$.

In der Regel **nicht** sinnvolle Umformungen sind:

$$h_i(x) = 0 \Leftrightarrow h_i(x) \geq 0 \quad \text{und} \quad h_i(x) \leq 0 \quad i = 1, \dots, p$$

$$g_i(x) \geq 0 \Leftrightarrow g_i(x) - (x_{n+i})^2 = 0, \quad i = 1, \dots, m.$$

(Nichtlineare Gleichungen sind oft wesentlich schwieriger zu behandeln als nichtlineare Ungleichungen, Konvexität geht verloren. $\nabla h_i(x)$ und $-\nabla h_i(x)$ sind stets linear abhängig!)

Aber die Umformung nichtlinearer Gleichungen in Ungleichungen

$$\begin{aligned} 0 &\leq x_{n+i} \\ h_i &\leq x_{n+i} \\ -h_i &\leq x_{n+i}, \quad (i = 1, \dots, p) \end{aligned}$$

mit x_{n+i} als zusätzlichen Minimierungsvariablen und eine entsprechende Erweiterung der Zielfunktion, die dafür sorgt, daß sich die Lösung von NLO nicht ändert, also

$$f(x) \mapsto f(x) + \sum_{i=1}^p \alpha_{n+i} x_{n+i} \quad \alpha_{n+i} \gg 1, \quad i = 1, \dots, p \quad \text{geeignet gewählt}$$

kann sinnvoll sein, wenn das gleichungsrestringierte Originalproblem "schwierig" ist.

B.2 Extremalkriterien

Satz B.3. *Es sei $x^* \in \mathfrak{S}$ eine lokale Minimalstelle von f auf \mathfrak{S} .*

Dann existieren Multiplikatoren $\lambda_0^, \lambda_1^*, \dots, \lambda_m^* \geq 0$ und $\mu_1^*, \dots, \mu_p^* \in \mathbb{R}$ mit*

$$\lambda_0^* \nabla f(x^*) - \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^p \mu_j^* \nabla h_j(x^*) = 0$$

$$\lambda_i^* g_i(x^*) = 0 \quad i = 1, \dots, m$$

□

Dies ist die (**Multiplikatorregel von Fritz John 1948**).

Hier gibt es ein Problem: $\lambda_0^* = 0$ liefert eine nichtverwertbare Bedingung, weil f darin nicht vorkommt. Der Fall $\lambda_0^* = 0$ kann jedoch nicht immer ausgeschlossen werden, wie folgendes Beispiel zeigt:

Beispiel von Kuhn und Tucker: $n = 2, m = 2, p = 0$

$$f(x) = -x_1, \quad g_1(x) = x_2, \quad g_2(x) = (1 - x_1)^3 - x_2$$

$$x^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \lambda_0^* = 0, \quad \lambda_1^* = \alpha \in \mathbb{R}, \quad \lambda_2^* = -\alpha$$

□

Solche Zusatzvoraussetzungen an die **Restriktionsfunktionen** g und h , die $\lambda_0^* > 0$ nach sich ziehen, heißen "**Restriktionsqualifikationen**". Die wichtigsten Restriktionsqualifikationen faßt die folgende Definition zusammen.

Definition B.4. *Für $x \in \mathfrak{S}$ sei $\mathcal{A} = \mathcal{A}(x) = \{i : g_i(x) = 0\}$. $x^* \in \mathfrak{S}$ erfüllt die **Regularitätsbedingung**, wenn*

$$(\nabla h(x^*), \nabla g_{\mathcal{A}}(x^*)) \text{ spaltenregulär}$$

*ist und die **Mangasarian-Fromowitz-Bedingung**, wenn*

$$\nabla h(x^*) \text{ spaltenregulär und } \exists z \in \mathbb{R}^n : \nabla h(x^*)^T z = 0, \quad \nabla g_{\mathcal{A}}(x^*)^T z > 0$$

$\mathcal{A}(x)$: heisst die **Indexmenge** der in x "**aktiven**" Restriktionen.

□

Bemerkung B.5. *Ist $\mathcal{A} \subset \{1, \dots, m\}$ eine Teilindexmenge, etwa*

$$\mathcal{A} = \{i_1, \dots, i_s\},$$

dann bezeichnet $\nabla g_{\mathcal{A}}$ die Matrix $(\nabla g_{i_1}, \dots, \nabla g_{i_s})$, $g_{\mathcal{A}}$ den Vektor $(g_{i_1}, \dots, g_{i_s})^T$ usw.

Definition B.6. Seien die Funktionen g_i , $i = 1, \dots, m$ konkav, d.h. $-g_i$ konvex und die Gleichungsrestriktionen h_j affin linear. (In diesem Fall ist \mathfrak{S} konvex). Die **Slaterbedingung** ist erfüllt, wenn es ein x^0 gibt mit

$$h(x^0) = 0, \quad g_{\text{NL}}(x^0) > 0, \quad g_{\text{L}}(x^0) \geq 0.$$

Dabei sind g_{L} die affin-linearen und g_{NL} die nichtlinearen Ungleichungsrestriktionen. \square

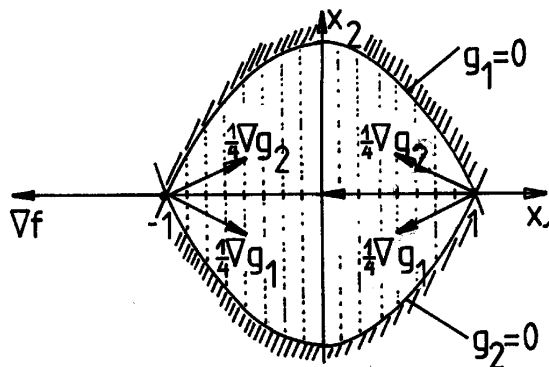
Die Slaterbedingung bezieht sich also immer auf konvexes \mathfrak{S} .

Satz B.7. Sei $x^* \in \mathfrak{S}$. Falls x^* lokale Minimalstelle von f auf \mathfrak{S} ist und in x^* die Regularitätsbedingung oder die Mangasarian-Fromowitz-Bedingung erfüllt ist, oder die $-g_i$ konvex für $i = 1, \dots, m$, und h_j affin linear, $j = 1, \dots, p$ sind und die Slaterbedingung erfüllt ist, dann gibt es Multiplikatoren $\lambda_1^*, \dots, \lambda_m^* \geq 0$, $\mu_1^*, \dots, \mu_p^* \in \mathbb{R}$ mit

$$\nabla f(x^*) - \nabla g(x^*)\lambda^* - \nabla h(x^*)\mu^* = 0, \quad \lambda_i^* g_i(x^*) = 0, \quad i = 1, \dots, m.$$

\square

Die notwendigen Bedingungen von Satz B.7 heißen “**Kuhn-Tucker-Bedingungen**” oder “**Multiplikatorregel**”, im Falle $m = 0$: “**Multiplikatorregel von Lagrange**”



Wir geben im Folgenden einen konstruktiven Beweis für diesen Satz für den Fall, daß die Regularitätsbedingung erfüllt ist. Dieser Fall ist besonders angenehm, auch algorithmisch. Der Beweis läuft so, daß wir für den Fall, daß die Aussagen des Satzes nicht gelten, eine Kurve konstruieren, die für kleinen Kurvenparameter t von x^* aus in \mathfrak{S} verläuft und längs der f abnimmt. Setze

$$N = N(x) = (\nabla h(x), \nabla g_{\mathcal{A}}(x)).$$

Wir beachten, daß nach Annahme $N^* = N(x^*)$ vollen Rang besitzt und damit auch $N(x)$ in einer Umgebung von x^* , wegen der vorausgesetzten Stetigkeit der Ableitungen.

Sei weiter für x in einer Umgebung von x^*

$$P \stackrel{\text{def}}{=} I - N(N^T N)^{-1} N^T$$

also der orthogonale Projektor auf den Nullraum von $N^T(x)$.
Man beachte, daß mit N auch P von x abhängt. Es ist

$$P(y) = y \text{ falls } N^T y = 0$$

und

$$P(y) = 0 \text{ falls } y = Na \text{ mit geeignetem } a .$$

Fall 1:

Sei

$$\nabla f(x^*) - N^* w = 0$$

nicht lösbar. Das bedeutet wegen der Rangbedingung also $|\mathcal{A}| < n$. Definiere $x(t)$ als Lösung der Anfangswertaufgabe

$$\dot{x}(t) = -P(x(t))\nabla f(x(t)) , \quad x(0) = x^* .$$

Nach dem Existenzsatz von Peano hat dieses Problem Lösungen und wenn die Gradienten Lipschitzstetig sind, dann ist die Lösung sogar eindeutig. Nach Voraussetzung ist $\dot{x}(0) \neq 0$. Setze

$$c(x) = \begin{pmatrix} h(x) \\ g_{\mathcal{A}}(x) \end{pmatrix}$$

Damit ist

$$N^T(x) = J_c(x)$$

und nach der Kettenregel

$$\frac{d}{dt} c(x(t)) = N^T(x(t))\dot{x}(t) \equiv 0$$

somit wegen $c(x^*) = 0$

$$c(x(t)) \equiv 0 , \quad \text{d.h. } x(t) \in \mathfrak{S} .$$

jedenfalls für hinreichend kleines t , weil $g_i(x(t)) > 0$ bleibt für $i \notin \mathcal{A}(x(0))$ für hinreichend kleines t . Nach der Taylorformel ist mit $x(0) = x^*$

$$f(x(t)) = f(x^*) + \nabla f(x^*)^T \dot{x}(0)t + o(t)$$

während wegen $P(x) = P^2(x) = P(x)^T$

$$\nabla f(x^*)^T \dot{x}(0) = -\|P(x(0))\nabla f(x(0))\|^2 < 0 .$$

D.h. $f(x(t))$ fällt streng monoton für kleines $t > 0$.

Fall 2:

Sei jetzt

$$\nabla f(x^*) - N^* \begin{pmatrix} v \\ w_{\mathcal{A}} \end{pmatrix} = 0$$

lösbar, somit wegen der Rangbedingung eindeutig lösbar, aber $w_{\mathcal{A}} \not\geq 0$. Das bedeutet

$$\exists i_0 \in \mathcal{A}(x(0)) : w_{i_0} < 0 .$$

Wir setzen

$$\begin{aligned} \tilde{\mathcal{A}} &= \mathcal{A} \setminus \{i_0\} , \\ \tilde{N} &= (\nabla h(x), \nabla g_{\tilde{\mathcal{A}}}(x)) , \\ \tilde{P} &= I - \tilde{N}(\tilde{N}^T \tilde{N})^{-1} \tilde{N}^T \end{aligned}$$

und betrachten nun die Anfangswertaufgabe

$$\dot{x}(t) = -\tilde{P}(x(t)) \nabla f(x(t)) , \quad x(0) = x^* .$$

Dann ist offenbar wegen $w_{i_0} \neq 0$ $\dot{x}(0) \neq 0$ und nach der gleichen Rechnung wie eben

$$\begin{aligned} h(x(t)) &\equiv 0 , \\ g_{\tilde{\mathcal{A}}}(x(t)) &\equiv 0 , \\ f(x(t)) &< f(x^*) \text{ für kleines positives } t . \end{aligned}$$

Es bleibt zu zeigen, daß

$$g_{i_0}(x(t)) \geq 0 \text{ für kleines positives } t .$$

Wegen

$$g_{i_0}(x(t)) = g_{i_0}(x^*) + \nabla g_{i_0}(x^*)^T \dot{x}(0)t + o(t)$$

und $g_{i_0}(x^*) = 0$ gilt dies, denn

$$\begin{aligned} \nabla g_{i_0}(x^*)^T \dot{x}(0) &= -\nabla g_{i_0}(x^*)^T \tilde{P}(x^*) \nabla f(x^*) \\ &= -\nabla g_{i_0}(x^*)^T \tilde{P}(x^*) (\tilde{N}(x^*) \tilde{w} + \nabla g_{i_0}(x^*) w_{i_0}) \\ &= -w_{i_0} \nabla g_{i_0}(x^*)^T \tilde{P}(x^*)^T \tilde{P}(x^*) \nabla g_{i_0}(x^*) \\ &= -w_{i_0} (\tilde{N} \tilde{a} + z)^T \tilde{P}(x^*)^T \tilde{P}(x^*) (\tilde{N} \tilde{a} + z) \\ &= -w_{i_0} z^T z \\ &> 0 . \end{aligned}$$

Hierbei haben wir benutzt, daß N vollen Rang hat, also

$$\nabla g_{i_0}(x^*) = \tilde{N} \tilde{a} + z, \quad z^T \tilde{N} = 0 ,$$

mit einem eindeutig bestimmten \tilde{a} und $z \neq 0$. Dies beendet den Beweis.

Löst man das oben beschriebene Anfangswertproblem mit dem primitivsten Integrationsverfahren, dem expliziten Eulerverfahren, dann erhält man das Gradientenprojektionsverfahren von Rosen (1960). Man schreitet dann auf einer Tangente an \mathfrak{S} von x^* aus fort (wenn \mathcal{A} bzw. $\tilde{\mathcal{A}}$ nicht leer sind, andernfalls liegt $x(t)$ im Inneren von \mathfrak{S} jedenfalls für $t > 0$), man verlässt also in der Regel \mathfrak{S} . Durch eine Zusatzmassnahme, eine sogenannte ‘‘Restoration‘‘, gelangt man dann auf den Rand von \mathfrak{S} zurück. Das ist eine Korrektur der Form

$$t^2Nr \text{ bzw. } t^2\tilde{N}\tilde{r},$$

$$r \text{ bzw. } \tilde{r} \text{ so, daß } x^* + td + t^2Nr \text{ bzw. } x^* + td + t^2\tilde{N}\tilde{r} \in \partial\mathfrak{S}$$

mit

$$d = \dot{x}(0).$$

Bei festem d und t ist dies ein nichtlineares Gleichungssystem für r bzw. \tilde{r} , das in der Regel mit dem vereinfachten Newtonverfahren gelöst wird.

Die Optimalitätsbedingungen kann man auch zweckmässig mit Hilfe der Lagrangefunktion formulieren:

Definition B.8.

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

heißt die NLO zugeordnete **Lagrange-Funktion**. □

Bemerkung B.9. In der Literatur sind auch andere Vorzeichenkonventionen für die Multiplikatoren in Gebrauch, man muß also immer erst die Vorzeichenkonvention der Problembeschreibung und die Vorzeichenkonvention für die Lagrangefunktion kennen, um Resultate zu vergleichen

Die Bedingung der Zulässigkeit und die Bedingungen der Multiplikatorregel lauten damit:

$$\begin{aligned} \nabla_x L(x^*, \lambda^*, \mu^*) &= 0^1 \\ h(x^*) &= 0 \\ \min(\lambda_i^*, g_i(x^*)) &= 0 \quad i = 1, \dots, m \end{aligned}$$

(Dies sind also $n + m + p$ nichtlineare Gleichungen für $n + m + p$ Unbekannte, aber **nicht differenzierbar** wegen $\min(\dots)$. Es gibt spezielle Lösungsansätze für dieses System, die sich jedoch in der Praxis (noch) nicht durchgesetzt haben.)

Im allgemeinen sind diese Bedingungen nur notwendig für Optimalität, nicht jedoch hinreichend. Es gilt jedoch

Satz B.10. $-g_i, i = 1, \dots, m$ seien konvex, $h_j, j = 1, \dots, p$ affin linear und f konvex auf \mathbb{R}^n . Dann ist die Multiplikatorregel **hinreichend für (globale) Optimalität** von x^* . Ist f streng konvex, dann ist x^* eindeutig. □

¹ $\nabla_x L =$ Gradient bzgl. x , $(\lambda, \mu$ fest), $\nabla_\mu L, \nabla_\lambda L$ analog
 $\nabla_{xx}^2 L =$ Hessematrix bzgl. x , $(\lambda, \mu$ fest), $\nabla_{x\mu} L$ usw. analog

Eine Aufgabe NLO mit den Bedingungen von Satz B.10 heißt **konvexe Optimierungsaufgabe**. Im nichtkonvexen Fall gibt es sowohl notwendige als auch hinreichende Bedingungen (für ein lokales Optimum), die sich auf die Hessematrix der Lagrangefunktion beziehen:

Ist $x^* \in \mathfrak{S}$ und

$$N^* := (\nabla h_1(x^*), \dots, \nabla h_p(x^*), \nabla g_{i_1}(x^*), \dots, \nabla g_{i_l}(x^*)),$$

mit $\mathcal{A} = \mathcal{A}(x^*)$, (d.h. $N^* = (\nabla h(x^*), \nabla g_{\mathcal{A}}(x^*))$), dann heißt

$$\mathcal{Z}_1^0(x^*) := \{z : (N^*)^T z = 0\}$$

der linearisierende Unterraum zu \mathfrak{S} in x^* . ($\{x^* + \tau z\}$ mit $\tau \in \mathbb{R}$ und $z \in \mathcal{Z}_1^0(x^*)$) ist die **Tangentialmannigfaltigkeit** an \mathfrak{S} in x^* . Für N^* n -spaltig und spaltenregulär ist sie ausgeartet, nämlich x^* selbst. x^* heißt dann **“Ecke”** von \mathfrak{S})

Satz B.11. Seien $f, g, h \in C^2(\mathcal{U}(x^*))$ und $x^* \in \mathfrak{S}$ sei lokale Minimalstelle von f auf \mathfrak{S} . $N^* = (\nabla h(x^*), \nabla g_{\mathcal{A}}(x^*))$ sei spaltenregulär. Dann gilt: Es gibt eindeutig bestimmte Multiplikatoren $\lambda^* \geq 0$ ($\in \mathbb{R}^m$) und $\mu^* \in \mathbb{R}^p$ mit

$$\left. \begin{aligned} \nabla_x L(x^*, \lambda^*, \mu^*) &= 0 \\ (\lambda^*)^T g(x^*) &= 0 \\ z^T \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) z &\geq \alpha z^T z \quad \text{für alle } z \in \mathcal{Z}_1^0(x^*) \text{ mit } \alpha \geq 0. \end{aligned} \right\} \quad (\text{B.1})$$

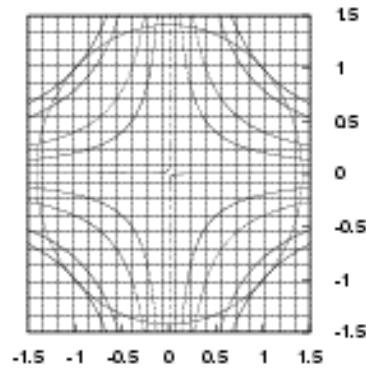
Gilt umgekehrt (B.1) und ist zusätzlich $\alpha > 0$ und $\lambda^* + g(x^*) > 0$ (d.h. nicht gleichzeitig $\lambda_i^* = g_i(x^*) = 0$ für ein i), dann ist x^* strenge Minimalstelle von f auf \mathfrak{S} . \square

Es gibt auch hinreichende Optimalitätsbedingungen, die ohne die in diesem Satz benutzte sogenannte **strikte Komplementarität** $\lambda_i + g_i(x_i) > 0 \forall i$ auskommen, diese sind jedoch komplizierter und auch in der Praxis viel schwerer überprüfbar.

Beispiele:

- $n = 2, m = 0, p = 1, f(x) = -x_1 - x_2, h_1(x) = \frac{1}{2}((x_1)^2 + (x_2)^2) - 1$.
 $x^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ erfüllt (B.1) mit $\alpha > 0$: strenges lokales (hier sogar globales) Minimum.
 $x = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$ erfüllt $\nabla_x L(x, \lambda, \mu) = 0$ mit $\mu = 1$, nicht jedoch (B.1) (Maximum).
- $n = 2, m = 2, p = 0, f(x) = -x_1, g_1(x) = 1 - (x_1)^2 - x_2, g_2(x) = 1 + x_2 - (x_1)^2$.
 $x^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \lambda_1^* = \lambda_2^* = \frac{1}{4}, \mathcal{A}(x^*) = \{1, 2\}$. x^* Ecke. (B.1) daher erfüllt mit $\alpha > 0$ ($\mathcal{Z}_1^0 = \{0\}$). Globales Minimum. (Konvexe Aufgabe, Slaterbedingung erfüllt.
 $x^{**} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$ Maximum, hier $\lambda_1^{**} = \lambda_2^{**} = -\frac{1}{4}$).
- $n = 2, m = 0, p = 1, f(x) = -10x_1x_2, h_1(x) = \frac{1}{2}((x_1)^2 + (x_2)^2) - 1$.
 $\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}$ erfüllen (B.1) mit $\alpha = 20 : 2$ strenge **lokale** Minimalstellen. \square

Das folgende Diagramm zeigt das Beispiel c. Die zulässige Menge ist der Kreis mit Radius $\sqrt{2}$ und die Niveaulinien von f tangieren diesen Kreis in $(\pm 1, \pm 1)$, zwei lokalen (und globalen) Minima und zwei Maxima.

**Ergänzungen:**

1. Gilt in einem Punkt x^* , der die Multiplikatorregel erfüllt, die Regularitätsbedingung, dann sind die Multiplikatoren eindeutig bestimmt. Man kann sie berechnen aus

$$\nabla f(x^*) - N(x^*, \mathcal{A}) \begin{pmatrix} \mu^* \\ \lambda_{\mathcal{A}}^* \end{pmatrix} = 0, \quad N(x^*, \mathcal{A}) = (\nabla h(x^*), \nabla g_{\mathcal{A}}(x^*))$$

mittels der QR-Zerlegung von N :

$$QN = \begin{pmatrix} R \\ 0 \end{pmatrix} \implies R \begin{pmatrix} \mu^* \\ \lambda_{\mathcal{A}}^* \end{pmatrix} = \left(Q \nabla f(x^*) \right)_{i=1, \dots, p+|\mathcal{A}|} \left. \vphantom{\begin{pmatrix} \mu^* \\ \lambda_{\mathcal{A}}^* \end{pmatrix}} \right\} \text{erste } p + |\mathcal{A}| \text{ Komponenten}$$

2. Genau dann ist die Menge der möglichen Multiplikatoren in der Multiplikatorregel beschränkt, wenn die Mangasarian-Fromowitz-Bedingung gilt (Gauvin 1978).
3. Genau dann ist die zulässige Menge lokal stabil unter Änderung der Restriktionen im Sinne von

$$\begin{aligned} \text{zu } x \in \mathfrak{S} \exists \tilde{x} \in \tilde{\mathfrak{S}} &= \{x : h(x) = c^1, g(x) \geq c^2\} \\ \text{mit } \|x - \tilde{x}\| &\leq C(x)(\|c^1\| + \|(c^2)^+\|) \end{aligned}$$

wenn in x die Mangasarian-Fromowitz-Bedingung gilt. (Robinson 1976).

4. Die Bedingung (B.1) in Satz B.11 mit $\alpha = 0$ wird auch als **notwendige Bedingung zweiter Ordnung** bezeichnet.
5. Die Bedingung

$$\lambda^* + g(x^*) > 0,$$

$$\text{also } g_i(x^*) = 0 \Rightarrow \lambda_i^* > 0, \quad g_i(x^*) > 0 \Rightarrow \lambda_i^* = 0$$

(wegen $\lambda_i^* g_i(x^*) = 0$) heißt Bedingung des **strikten komplementären Schlupfes** oder **strikte Komplementarität**.

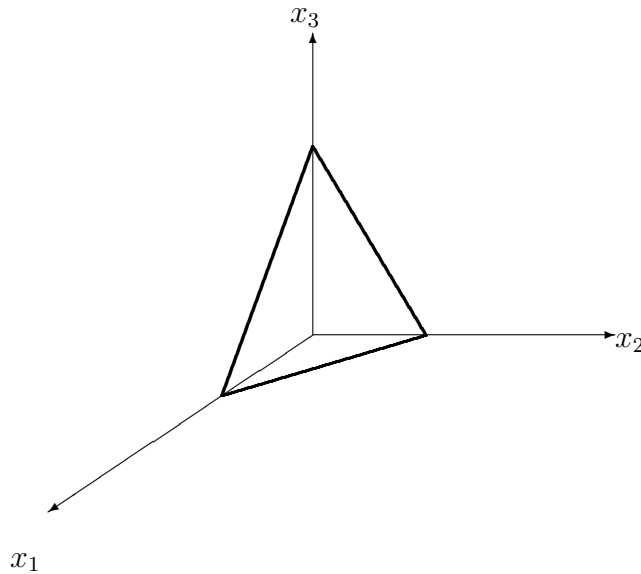
Wichtige Spezialfälle konvexer Optimierungsaufgaben:

Lineare Optimierungsaufgabe: (LP)

$$\begin{aligned} f(x) &= c^T x \\ g(x) &= G^T x + g^0 \quad (\geq 0) \\ h(x) &= H^T x + h^0 \quad (= 0) \end{aligned}$$

Standardform der linearen Optimierung:

$$\begin{aligned} f(x) &= c^T x \\ h(x) &= H^T x + h^0 \\ g(x) &= x \end{aligned}$$



Konvexe quadratische Optimierungsaufgabe: (QP)

$$\begin{aligned} f(x) &= -b^T x + \frac{1}{2} x^T A x \quad A \text{ pos. semidefinit} \\ g(x) &= G^T x + g^0 \quad (\geq 0) \\ h(x) &= H^T x + h^0 \quad (= 0) \end{aligned}$$

(Ist A indefinit oder negativ definit, ist die Aufgabe nicht konvex und ihre Lösung wesentlich schwieriger).

Für diese beiden Aufgaben gibt es spezielle finite Algorithmen.

Konvexe Optimierungsaufgaben lassen sich besonders elegant behandeln mit Hilfe der **Sattelpunkteigenschaft der Lagrangefunktion**:

x, λ, μ variieren im Folgenden unabhängig:

Gesucht $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}^p$ mit

$$L(x^*, \lambda, \mu) \leq L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*) \tag{B.2}$$

für alle $x \in \mathbb{R}^n, \lambda \geq 0 \in \mathbb{R}^m, \mu \in \mathbb{R}^p$.

Satz B.12. Falls (x^*, λ^*, μ^*) eine Lösung von (B.2) ist, ist x^* Lösung von NLO. \square

Bemerkung: Für viele Probleme NLO gibt es keinen Sattelpunkt.

Aber im Spezialfall konvexer Optimierungsaufgaben hat man eine weitreichende Aussage:

Satz B.13. $f, -g_1, \dots, -g_m : \mathbb{R}^n \rightarrow \mathbb{R}$ seien konvex, h_1, \dots, h_p seien affin linear. Die Slater-Bedingung sei erfüllt. Dann ist x^* eine Lösung von NLO genau dann, wenn es $\lambda^* \in \mathbb{R}^m, \lambda^* \geq 0$ und $\mu^* \in \mathbb{R}^p$ gibt mit

$$L(x^*, \lambda, \mu) \leq L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*)$$

für alle $x \in \mathbb{R}^n, \text{ alle } \lambda \geq 0, \text{ alle } \mu \in \mathbb{R}^p$ \square

Eine Vorgehensweise in diesem Fall, für f gleichmäßig konvex auf \mathbb{R}^n (d.h. $\lambda_{\min}(\nabla^2 f(x)) \geq \gamma > 0$ für alle $x \in \mathbb{R}^n$) ist die Folgende: Zu $\lambda \geq 0, \mu \in \mathbb{R}^p$ bestimme man

$$x = x(\lambda, \mu) \quad \text{mit} \quad \nabla_x L(x, \lambda, \mu) = 0.$$

Dieses Problem ist eindeutig lösbar, wenn f gleichmäßig konvex ist. Man setze

$$\Phi(\lambda, \mu) = L(x(\lambda, \mu), \lambda, \mu)$$

und löse das Problem

$$\Phi(\lambda, \mu) \stackrel{!}{=} \max, \quad \lambda \geq 0.$$

Dies ist ein einfach strukturiertes Problem mit leicht zu kontrollierenden Restriktionen. Für lineare und konvexe quadratische Optimierungsaufgaben führt Satz B.13 zu den sogenannten

Dualitätssätzen der linearen und quadratischen Optimierung. So lautet z.B. eine lineare Optimierungsaufgabe in Standardform

$$\begin{aligned} f(x) &= c^T x \stackrel{!}{=} \min \\ Ax &= b, \\ x &\geq 0. \end{aligned}$$

Die Kuhn-Tucker-Gleichungen dafür lauten also

$$\begin{aligned} c - A^T \mu - \lambda &= 0, \\ Ax - b &= 0, \\ x_i \lambda_i &= 0, \quad i = 1, \dots, n \end{aligned}$$

und daraus folgt mittels der Sattelpunktbedingung das duale Problem

$$\begin{aligned} \hat{f}(\mu) &= \mu^T b \stackrel{!}{=} \max \\ c - A^T \mu &\geq 0 \end{aligned}$$

und für jedes Paar (x, μ) , das für die primale und duale Aufgabe zulässig ist, gilt

$$\mu^T b \leq c^T x$$

und darüberhinaus ist im Optimum die Dualitätslücke $c^T x - \mu^T b$ null:

$$(\mu^*)^T b = c^T x^*.$$

Bezeichnungen:

λ, μ	duale Variablen
$\lambda \geq 0$	duale Zulässigkeit
$(\lambda^*)^T g(x^*) = 0$	Komplementaritätsbedingung
$\lambda^* + g(x^*) > 0$	Bedingung der strikten Komplementarität

B.3 Verfahren

B.3.1 Klassische Penalty- und Barriereverfahren

Hier wird versucht, das restringierte Problem in eine Schar unrestringierter Probleme so einzubetten, daß die Lösungsschar der unrestringierten Probleme gegen die eigentlich gesuchte Lösung konvergiert. In der Praxis führt man diesen Grenzübergang nicht aus, sondern

begnügt sich mit einer “genügend genauen“ Näherung.

Penalty-Terme bestrafen das Verlassen von \mathfrak{G} :

$$\begin{aligned} \psi : \psi(x) &= 0 & x \in \mathfrak{G} \\ \psi(x) &> 0 & x \notin \mathfrak{G} \end{aligned}$$

Differenzierbare Penalty-Terme:

für Gleichungen: $\psi(x) = \sum_{i=1}^p (h_i(x))^2, \quad \sum_{i=1}^p w_i (h_i(x))^2, \quad w_i > 0 \text{ fest}$

$$\psi(x) = \sum_{i=1}^p (h_i(x))^{2m}, \quad m = 2, 3, \dots \text{ (weniger geeignet)}$$

für Ungleichungen: $\psi(x) = \sum_{i=1}^m (\min\{0, g_i(x)\})^2$

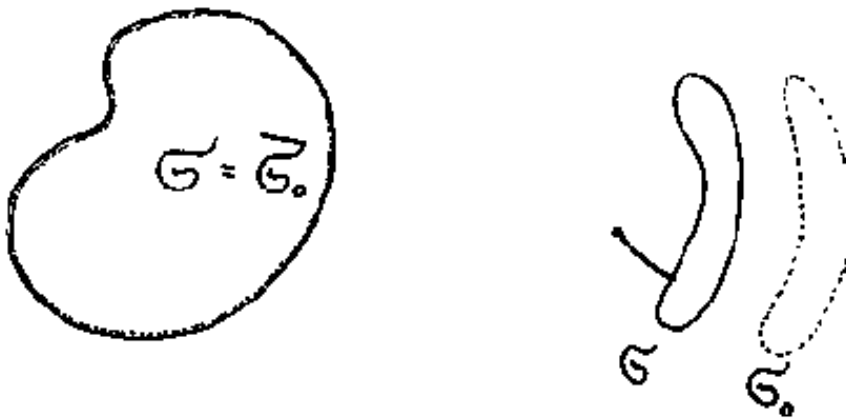
Nichtdifferenzierbare Penaltyterme sind:

$$\sum_{i=1}^p w_i |h_i(x)|, \quad - \sum_{i=1}^m w_i \min\{0, g_i(x)\}$$

Barriere-Terme verhindern das Verlassen des zulässigen Gebietes durch Errichtung einer "Barriere" auf dem Rand. Bei dieser Methode benötigt man als Voraussetzung für die Konstruierbarkeit, daß der Abschluss des offenen Kernes der Menge gleich dem Abschluss der Menge ist: ²

$$\overline{\{x : g_i(x) > 0 \text{ für } i = 1, \dots, m\}} = \{x : g_i(x) \geq 0, i = 1, \dots, m\}$$

Für Gleichungsrestriktionen kann man natürlich keine Barriereansätze benutzen.



²Ist $\mathcal{A} \subset \mathbb{R}^n$ eine Menge, dann bezeichnet $\bar{\mathcal{A}}$ die Menge \mathcal{A} vereinigt mit allen ihren Häufungspunkten (Abschluß von \mathcal{A}).

Ein für Ungleichungen $g_i(x) \geq 0$ geeigneter Barriereterm ist $\psi(x) = -\sum_{i=1}^m \ln(g_i(x))$

$$\psi(x^k) \rightarrow \infty \text{ wenn } g_i(x^k) \rightarrow 0 \text{ für ein } i, x^k \in \mathfrak{S}$$

Konstruktion einer **Penalty-Funktion**:

$$\Phi(x; \varrho) = f(x) + \frac{1}{\varrho} \left(\sum_{i=1}^p w_i (h_i(x))^2 + \sum_{i=1}^m w_{p+i} (\min(0, g_i(x)))^2 \right)$$

Dabei sind w_j **fest**e Gewichte, die geeignet gewählt wurden.

$$x \in \mathfrak{S} : \Phi(x; \varrho) \equiv f(x)$$

$$x \notin \mathfrak{S} : \Phi(x; \varrho) \rightarrow \infty \text{ mit } \varrho \rightarrow 0.$$

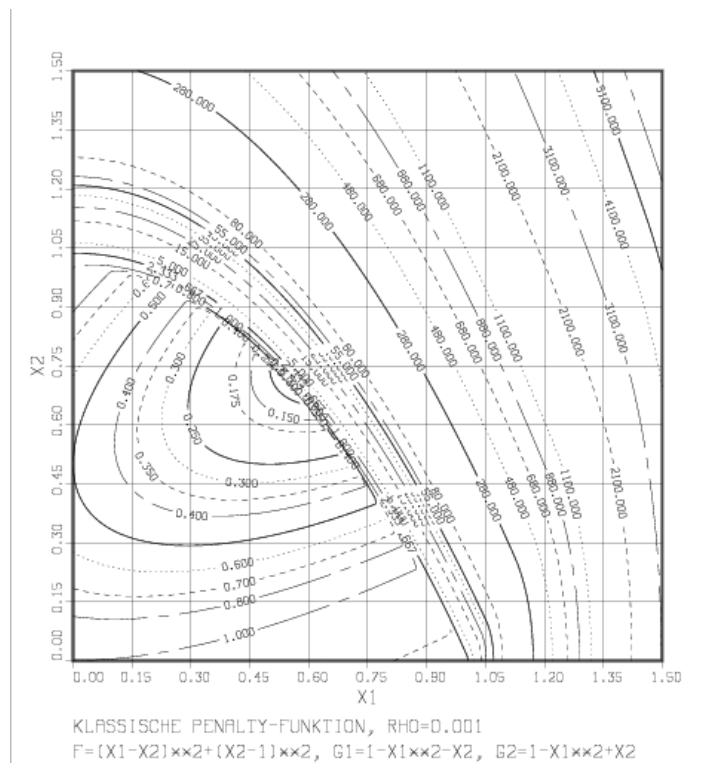
Eine (**gemischte**) **Penalty-Barriere-Funktion** ist

$$B(x; \varrho) = f(x) + \frac{1}{\varrho} \sum_{i=1}^p w_i (h_i(x))^2 - \varrho \sum_{i=1}^m w_{i+p} \ln(g_i(x))$$

Sie ist nur definiert, wenn $g_i(x) > 0$ $i = 1, \dots, m$.

Wegen $\varrho \rightarrow 0$ spielen Terme mit $g_i(x) > 1$ hier keine Rolle, d.h. i.w. gilt

$$B(x; \varrho) \geq f(x) \text{ für } x \in \mathfrak{S}.$$



Die prinzipielle **Vorgehensweise** ist in beiden Fällen die gleiche:
Wähle monotone Nullfolge $\{\varrho_k\} \searrow 0$ und definiere

$$x^*(\varrho_k) \in \operatorname{argmin}_{x \in \mathbb{R}^n} \Phi(x; \varrho_k)$$

bzw.

$$x^*(\varrho_k) \in \operatorname{argmin}_{x \in \mathfrak{S}_0} B(x; \varrho_k)$$

$$\mathfrak{S}_0 = \{x \in \mathbb{R}^n : g_1(x) > 0, \dots, g_m(x) > 0\}.$$

In beiden Fällen handelt es sich um **unrestringierte** Minima, d.h. $x^*(\varrho_k)$ wird bestimmt mit den Methoden aus Abschnitt B.3.

Zu erwarten ist : $x^*(\varrho_k) \rightarrow x^*$ Lösung von NLO.

Beispiele:

1. $f(x) = (x)^2$, $p = 0$, $m = 1$, $g(x) = x - 1$.

Lösung $x^* = 1$, $\lambda^* = 2$.

$$\Phi(x; \varrho) = (x)^2 + \frac{1}{\varrho}(\min\{0, x - 1\})^2$$

$$x^*(\varrho) = \frac{1}{\varrho+1} \rightarrow 1 \quad \text{für } \varrho \rightarrow 0. \quad x^*(\varrho) < 1 \text{ ist } \mathbf{unzulässig}$$

für das Ausgangsproblem

$$\nabla^2 \Phi(x; \varrho) = \begin{cases} 2 & x > 1 \\ 2 + 2/\varrho & \text{für } x < 1 \end{cases}$$

Die Funktion ist also nicht überall zweimal stetig differenzierbar, jedoch einmal mit Lipschitzstetigem Gradienten.

$$\text{Ferner gilt } -\frac{2}{\varrho} \min\{0, x^*(\varrho) - 1\} = \frac{2}{\varrho+1} \rightarrow \lambda^*.$$

2. $n = 2$, $p = 1$, $m = 0$, $f(x) = (x_1)^2 + 4x_1x_2 + 5(x_2)^2 - 10x_1 - 20x_2$

$$h_1(x) = 2 - (x_1 + x_2).$$

$$x_1^* = \frac{1}{2}, \quad x_2^* = \frac{3}{2}, \quad \mu_1^* = 3$$

$$\Phi(x; \varrho) = f(x) + \frac{1}{\varrho}(h_1(x))^2 \quad \text{quadratische streng konvexe Funktion}$$

$$\nabla^2 \Phi(x; \varrho) = \begin{pmatrix} 2 & 4 \\ 4 & 10 \end{pmatrix} + \frac{1}{\varrho} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

mit den Eigenwerten: $10 + \frac{2}{\varrho} + \mathcal{O}(\varrho)$ und $2 + \mathcal{O}(\varrho)$

$$x^*(\varrho) = \frac{1}{2+\varrho} \begin{pmatrix} 1 + 5\varrho \\ 3 \end{pmatrix} \rightarrow x^* \quad \text{für } \varrho \rightarrow 0$$

$$-\frac{2}{\varrho} h_1(x^*(\varrho)) = \frac{6}{2+\varrho} \rightarrow \mu_1^* \quad \text{für } \varrho \rightarrow 0.$$

Konditionszahl von $\nabla^2 \Phi(x; \varrho) = \frac{1}{\varrho} + \mathcal{O}(1)$ für $\varrho \rightarrow 0$.

3. $f(x) = (x)^2$, $g(x) = x - 1$, $n = m = 1$, $p = 0$.

Lösung: $x^* = 1$, $\lambda^* = 2$.

$$B(x; \varrho) = (x)^2 - \varrho \ln(x - 1) \quad \text{für } x > 1.$$

$$x^*(\varrho) = \frac{1}{4}(2 + \sqrt{4 + 8\varrho}) = 1 + \frac{\varrho}{2} + \mathcal{O}((\varrho)^2) \rightarrow x^* \quad \text{für } \varrho \rightarrow 0.$$

Konvergenzaussage für das Penaltyverfahren:

Satz B.14. Für jedes reelle α sei die Niveaumenge

$$\mathcal{L}_f(\alpha) = \{x \in \mathbb{R}^n, f(x) \leq \alpha\}$$

beschränkt und abgeschlossen. Ferner sei $\mathfrak{S} \neq \emptyset$ und $\varrho_k \searrow 0$.
Dann gilt für die aus

$$x^*(\varrho) \in \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \Phi(x; \varrho)$$

bestimmten Werte $x^*(\varrho_k)$:

1. $\Phi(x^*(\varrho_{k+1}); \varrho_{k+1}) \geq \Phi(x^*(\varrho_k); \varrho_k)$
2. $f(x^*(\varrho_{k+1})) \geq f(x^*(\varrho_k))$
3. $\psi(x^*(\varrho_{k+1})) \leq \psi(x^*(\varrho_k))$
4. Jeder Häufungspunkt von $\{x^*(\varrho_k)\}$ ist Lösung von NLO.
5. Sind $f, g, h \in C^1(\mathbb{R}^n)$ und ist $\operatorname{argmin} \{f(x) : x \in \mathfrak{S}\}$ einpunktig (d.h. Lösung von NLO eindeutig), dann gilt $x^*(\varrho_k) \rightarrow x^*$.
6. Falls x^* regulärer Punkt ist, (d.h. $(\nabla h(x^*), \nabla g_{\mathcal{A}}(x^*))$ ist spaltenregulär), und die strikte Komplementarität gilt, d.h. $\lambda^* + g(x^*) > 0$, dann

$$\begin{aligned} \lambda_i^*(\varrho_k) &:= -\frac{2}{\varrho_k} \min\{0, g_i(x^*(\varrho_k))\} \rightarrow \lambda_i^* \quad i = 1, \dots, m \\ \mu_j^*(\varrho_k) &:= -\frac{2}{\varrho_k} h_j(x^*(\varrho_k)) \rightarrow \mu_j^* \quad j = 1, \dots, p \end{aligned}$$

wobei λ^*, μ^* die eindeutig bestimmten Multiplikatoren sind.

7. Gilt zusätzlich die hinreichende Bedingung zweiter Ordnung in x^* , ($f, g, h \in C^2(\mathcal{U}(x^*))$), dann gilt weiter

$$\begin{aligned} \|x^*(\varrho) - x^*\| &\leq C\varrho \\ \|\lambda^*(\varrho) - \lambda^*\| &\leq C\varrho \\ \|\mu^*(\varrho) - \mu^*\| &\leq C\varrho \end{aligned}$$

mit einer geeigneten Konstanten C .

8. Unter den Voraussetzungen aus 6. und 7. divergieren genau $p + |\mathcal{A}|$ Eigenwerte von $\nabla_{xx}^2 \Phi(x; \varrho)$ wie c/ϱ gegen unendlich, während die übrigen beschränkt bleiben. □

Ein analoger Satz gilt für $B(x; \varrho)$.

NUMAWWW

Die Aussage 8 des voranstehenden Satzes besagt, daß die Minimierung von Φ mit kleiner werdendem ϱ immer schwieriger wird. Nur das Newton-Verfahren selbst (und spezielle, aber aufwendige Ansätze für $\{A_k\}$ in Satz A.31) sind dagegen weitgehend unempfindlich. Also darf ϱ nicht wirklich "klein" werden!

Die direkte Anwendung von Minimierungsmethoden bei willkürlich gewählter Nullfolge $\{\varrho_k\}$ ist daher in der Regel nicht erfolgreich. Stellt man aber die notwendigen Extremalbedingungen (mit ϱ als Parameter) auf, dann erhält man ein parameterabhängiges (nichtlineares) Gleichungssystem, für das man Pfadverfolgungsmethoden erfolgreich anwenden kann, indem man ϱ langsam variiert, jedenfalls solange die Lagrangefunktion des Problems gleichmäßig konvex in x ist. Dieses nichtlineare System lautet

$$\begin{aligned} \nabla_x f(x) - \nabla h(x)\mu - \nabla g(x)\lambda &= 0, \\ h(x) &= 0, \\ g_i(x)\lambda_i &= \rho > 0, \quad i = 1, \dots, m \end{aligned}$$

Wichtige spezielle Anwendungen: lineare, quadratische Optimierung mit einem Problem in Standardform:

$$\begin{aligned} f(x) &= -a^T x + \frac{1}{2}x^T A x, \quad A \text{ pos. semidefinit } (A = 0 \text{ möglich}) \\ h(x) &= B^T x - b = 0, \quad B \text{ von vollem Spaltenrang} \\ g(x) &= x \geq 0 \\ B(x; \varrho) &= f(x) - \varrho \sum_{i=1}^n \ln(x_i) \end{aligned}$$

$$x^*(\varrho) = \operatorname{argmin} \{B(x; \varrho) : x > 0, \underline{h(x) = 0}\} \quad (\text{B.3})$$

(die linearen Gleichungen werden also stets exakt erfüllt) Dies ist die Grundidee der "innere Punkte-Verfahren" für LP nach Karmarkar etc. für LP bzw. Monteiro und Adler, Goldfarb & Liu etc für QP. Hierbei bezieht sich "Inneres" nur auf die Schrankenrestriktionen. Die Vorgehensweise erzeugt eine Näherungsfolge $x^k \rightarrow x^*$ für die Lösung des Ausgangsproblems. (In dieser Verfahrensklasse gibt es Verfahren von in n polynomialem Gesamtaufwand für LP, QP in Standardform)

Man kann auf die Einhaltung der Gleichungsrestriktionen während der Iteration sogar verzichten und erhält dann die inneren-Punkte-Methoden mit unzulässigen (bzgl. der Gleichungen) Punkten. Ausgangspunkt der Überlegungen sind wieder die im konvexen Fall mit Slaterpunkt notwendigen und hinreichenden Kuhn-Tucker-Bedingungen zusammen mit den

Zulässigkeitsbedingungen

$$\begin{aligned} Ax - a - B\mu - \lambda &= 0, \\ \lambda &\geq 0, \\ x &\geq 0, \\ \lambda_i x_i &= 0, \quad i = 1, \dots, n \\ B^T x - b &= 0. \end{aligned}$$

Man kann dies System als nichtlineares Gleichungssystem in x, μ und λ interpretieren. Dieses System wird nun abgeändert zu

$$\begin{aligned} Ax - a - B\mu - \lambda &= 0, \\ \lambda_i x_i &= \epsilon, \quad i = 1, \dots, n, \\ B^T x - b &= 0. \end{aligned}$$

Auf der Menge

$$x > 0, \quad \lambda > 0$$

(dies ist die Menge der “inneren Punkte” bezüglich der Ungleichungsrestriktionen) ist die Jacobi-Matrix dieses Systems immer regulär wegen der Rangbedingung für B . Die Lösung beschreibt eine differenzierbare, in ϵ parametrisierte Kurve, die für $\epsilon \rightarrow 0$ gegen einen Lösungspunkt des Ausgangsproblems konvergiert.

Bemerkung B.15. *Man beachte, daß im LP- und auch im streng konvexen QP-Fall die Lösung nicht notwendig eindeutig ist, da die Multiplikatoren nicht notwendig eindeutig sind, denn $(B, -I_A)$ hat nicht notwendig vollen Rang. Die Lösungskurve konvergiert also gegen eine spezielle Lösung des Problems. Nur wenn auch die Regularitätsbedingung und die Bedingung der strikten Komplementarität erfüllt ist, ist auch im LP-Fall die Lösung eindeutig.*

□

Diese Lösungskurve wird nun approximativ verfolgt, wobei das gedämpfte Newtonverfahren als Lösungsverfahren für festes ϵ zum Einsatz kommt. Bei der Dämpfung ist zusätzlich die strenge Einhaltung der Nebenbedingungen $x > 0$, $\lambda > 0$ zu beachten. Dies führt bei naiver Vorgehensweise in der Homotopie zu bedeutenden numerischen Schwierigkeiten und die eigentliche Kunst beim Entwurf der Algorithmen besteht darin, mit möglichst geringem Aufwand in “genügender” Nähe der Trajektorie zu bleiben und ϵ einerseits nicht zu schnell und andererseits auch nicht zu langsam zu variieren. Es stellt sich heraus, daß die Folge ϵ_k an die sogenannte Dualitätslücke

$$(x^k)^T \lambda^k / n$$

gekoppelt werden muß. Ein typischer Algorithmus ist der folgende von Kojima, Mizuno und Yoshise:

$$\begin{aligned}
\mathcal{J}_F(z^k)\Delta z_N^k &= -F(z^k) \text{ zu lösen, Newtonschritt} \\
\mathcal{J}_F(z^k)\Delta z_C^k &= \varrho_k \hat{e} \text{ zu lösen, Zentrierungsschritt} \\
\varrho_k &= \sigma_k((x^k)^T \lambda^k)/n, \\
\Delta z^k &= \Delta z_N^k + \Delta z_C^k \stackrel{\text{def}}{=} (\Delta x^k, \Delta \mu^k, \Delta \lambda^k)^T \\
\hat{\alpha}_k &= \min\{\min\{x_i^k/(-\Delta x_i^k) : \Delta x_i^k < 0\}, \min\{\lambda_i^k/(-\Delta \lambda_i^k) : \Delta \lambda_i^k < 0\}\}, \\
\alpha_k &= \min\{1, \tau_k \hat{\alpha}_k\}, \\
z^{k+1} &= z^k + \alpha_k \Delta z^k.
\end{aligned}$$

Dabei ist $z = (x, \mu, \lambda)$, $\tau_k \in]0, 1[$, $\sigma_k \in [0, 1[$ und $\hat{e} = (0, \dots, 0, e)^T$ mit $e = (1, \dots, 1)^T \in \mathbb{R}^n$. Die Funktion F ist definiert durch

$$F(x, \mu, \lambda) = \begin{pmatrix} Ax - a - B\mu - \lambda \\ -B^T x + b \\ X\lambda \end{pmatrix}$$

mit $X = \text{diag}(x_1, \dots, x_n)$. Im LP- und konvexen QP-Fall kann man polynomiale Komplexität mit einer Schrittzahl $\mathcal{O}(\sqrt{n}L)$ erreichen, wobei L die sogenannte Informationslänge der Eingabedaten bedeutet, also etwa für ganzzahlige Koeffizienten im LP-Fall

$$L = n^2 \max_{i,j} \{\log_2(|b_{ij}| + 1), \log_2(|b_j| + 1), \log_2(|a_i| + 1)\}$$

bzw. Reduktion der Dualitätslücke $x^T \lambda$ auf einen Wert δ in $\mathcal{O}(\sqrt{n} |\ln \delta|)$ Schritten. Dies sind theoretische Aussagen. In der Praxis gilt für diese Varianten die Faustregel "30 Schritte genügen für volle Genauigkeit" (in der üblichen Rechengenauigkeit). Ist $A = 0$, dann ist die Lösung des Gleichungssystems mit der Matrix

$$\mathcal{J}_F(z^k) = \begin{pmatrix} A & -B & -I \\ -B^T & 0 & 0 \\ \Lambda_k & 0 & X_k \end{pmatrix}, \quad \Lambda = \text{diag}(\lambda_i),$$

noch vergleichsweise einfach möglich, da sie auf die Lösung eines Systems mit der nach Voraussetzung positiv definiten und in der Dimension viel kleineren Matrix $B^T \Lambda_k^{-1} X_k B$ zurückgeführt werden kann. Für allgemeines semidefinites A ist dies nicht mehr möglich und man muß auf Methoden für große indefinite symmetrische Systeme ausweichen. Wenn man ein nichtkonvexes Problem vorliegen hat (d.h. an die Stelle der Matrix A tritt nun eine indefinite Hessematrix einer Lagrangefunktion) oder die Gradienten der Gleichungsrestriktionen linear abhängig werden, bricht dieser Lösungsansatz zusammen, da nun die Lösungskurve des Systems gar nicht mehr notwendig existiert. Modifikationen der Vorgehensweise, die auch in diesem (in der Praxis häufigen Fall) noch durchführbar sind, sind Gegenstand gegenwärtiger Forschung.

B.3.2 Die Multiplikator-Methoden von Powell, Hestenes und Rockafellar

Unser Hauptziel ist nun die Vermeidung des Grenzübergangs $\varrho \rightarrow 0$ beim Penalty-Parameter ϱ , d.h. die Divergenz $\frac{1}{\varrho} \rightarrow \infty$ und damit die (einiger) Eigenwerte von $\nabla^2 \Phi(\cdot, \varrho) \rightarrow \infty$ soll vermieden werden.

Als Lösungsansatz schwebt uns nun die Sattelpunktaussage für die Lagrangefunktion vor, d.h. wir wollen $\nabla_{xx}^2 L(x, \lambda, \mu)$ positiv definit machen durch Hinzufügen einer Ableitung des Strafterms, d.h. Konvexifizierung des Problems, sodaß der Sattelpunktsatz wenigstens lokal gilt. Ausgangspunkt ist also hier nicht die Zielfunktion, sondern die Lagrangefunktion des Problems. Im Übrigen wollen wir aber die Vorgehensweise nachahmen, die wir bei der Besprechung der Sattelpunkteigenschaft schon angedeutet haben: Bei gewähltem Lagrange- und Penaltyparameter wird die primale Variable (durch unrestringierte Minimierung) zu einer Funktion des Lagrangeparameters und die Maximierung bezüglich dieser dualen Variablen wird die eigentliche treibende Kraft zur Problemlösung. Dabei muß der Penaltyparameter dann schon "richtig" gewählt sein.

B.3.2.1 Gleichungsrestringierte Probleme

Die erweiterte Lagrange-Funktion ist also :

$$L_A(x, \mu; \varrho) = f(x) - \mu^T h(x) + \frac{\varrho}{2} (h(x)^T h(x))$$

Bemerkung B.16. Wir benutzen hier nur einen Parameter ϱ . Man kann alles auch mit individuellen Strafparametern für die einzelnen Funktionen durchführen. Dies ist in der Praxis angemessener, da es durchaus schwierig sein kann, eine vernünftige gemeinsame Skalierung für die einzelnen Restriktionen zu finden

Beispiele

1.

$$\begin{aligned} f(x) &= (x_1)^2 - x_2, & \text{konvex} \\ h(x) &= (3x_1 - 2x_2 + 1), & \text{affinlinear} \end{aligned}$$

d.h. $n = 2$, $p = 1$

$$\nabla_{xx}^2 L_A(x, \mu; \varrho) = \begin{pmatrix} 2 + 9\varrho & -6\varrho \\ -6\varrho & 4\varrho \end{pmatrix}$$

ist positiv definit für alle $\varrho > 0$, (aber f ist nicht nach unten beschränkt auf \mathbb{R}^2), d.h. L_A ist bzgl. x gleichmäßig konvex. Zu beliebigem $\mu \in \mathbb{R}$ existiert genau ein $x(\mu)$ mit $\nabla_x L(x(\mu), \mu; \varrho) = 0$:

$$\begin{aligned} \begin{pmatrix} 2x_1 \\ -1 \end{pmatrix} - \begin{pmatrix} 3 \\ -2 \end{pmatrix} \mu + \varrho \begin{pmatrix} 3 \\ -2 \end{pmatrix} (3x_1 - 2x_2 + 1) &= 0 \\ \Rightarrow h(x) = (\mu - 1/2)/\varrho, \quad x_1 \equiv \frac{3}{4}, \quad x_2 = \frac{13}{8} + (\frac{1}{2} - \mu)/(2\varrho). \end{aligned}$$

Für $\mu^* = \frac{1}{2}$ also $h(x(\mu^*)) = 0$, $x(\mu^*) = x^*$

$$L_A(x(\mu), \mu; \varrho) = \frac{9}{16} - \frac{13}{8} - \frac{\frac{1}{2} - \mu}{2\varrho} - \mu(\mu - \frac{1}{2})/\varrho + (\mu - \frac{1}{2})^2/(2\varrho)$$

hat in $\mu^* = \frac{1}{2}$ ein Maximum!

2. Nichtkonvexe Problemstellung

$$f(x) = (x)^3, \quad h(x) = x + 1, \quad x^* = -1, \quad \mu^* = 3 \quad (n = 1, p = 1)$$

$$L_A(x, \mu; \varrho) = (x)^3 - \mu(x + 1) + \frac{\varrho}{2}(x + 1)^2$$

$x(\mu)$ wird definiert durch $\frac{\partial}{\partial x} L_A(x, \mu; \varrho) = 0$; also

$$x(\mu) = \frac{1}{6} \left(-\varrho \pm \sqrt{\varrho^2 - 12(\varrho - \mu)} \right)$$

und $\frac{\partial^2}{(\partial x)^2} L_A(x, \mu; \varrho) > 0$ für $x > -\varrho/6$.

Nun muß also $\varrho > 6$ und $\varrho^2 - 12(\varrho - \mu) > 0$ gelten. Wir haben also Bedingungen an ϱ **und** μ . Dann existiert eine **lokale** Minimalstelle von $L_A(\frac{1}{6}(-\varrho + \sqrt{\dots}), \dots, (\frac{1}{6}(-\varrho - \sqrt{\dots})))$ ergibt lokales Maximum von L_A bzgl. x für $\mu \rightarrow 3$ geht $x(\mu) \rightarrow -1 = x^*$

$L_A(x(\mu), \mu; \varrho)$ hat in $\mu^* = 3$ ein **lokales** Maximum.

3. Nichtkonvexe Problemstellung

$$n = 2, p = 1, f(x) = -x_1 x_2, \quad h_1(x) = (x_1 - 3)^2 + (x_2)^2 - 10.$$

$$\nabla_x L_A(x, \mu; \varrho) = \begin{pmatrix} -x_2 \\ -x_1 \end{pmatrix} - \mu \begin{pmatrix} 2(x_1 - 3) \\ 2x_2 \end{pmatrix} + \varrho \begin{pmatrix} 2(x_1 - 3) \\ 2x_2 \end{pmatrix} ((x_1 - 3)^2 + (x_2)^2 - 10) = 0$$

$$\begin{aligned} \nabla_{xx}^2 L_A(x, \mu; \varrho) &= \begin{pmatrix} -2\mu + 4\varrho(x_1 - 3)^2 & , & -1 + 4\varrho x_2(x_1 - 3) \\ -1 + 4\varrho x_2(x_1 - 3) & , & -2\mu + 4\varrho(x_2)^2 \end{pmatrix} \\ &+ 2\varrho \left((x_1 - 3)^2 + (x_2)^2 - 10 \right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

NLO hat die (globale) Lösung $x^* = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$, $\mu^* = -1$. Nur in einer kleinen Umgebung von $\mu^* = -1$ hat $\nabla_x L_A(x, \mu; \varrho) = 0$ eine Lösung, die in einer Umgebung von $\begin{pmatrix} 4 \\ 3 \end{pmatrix}$ liegt und die eine lokale Minimalstelle von L_A ist, für jedes $\varrho \geq 0$!

Aus den Beispielen ergibt sich, daß man im konvexen Fall, also f konvex und h affinlinear, keine Probleme haben wird, wenn die Niveaumengen $\mathcal{L}_f(\alpha) \cap \mathfrak{S}$ beschränkt sind, **sonst** aber **nur lokale Konvergenz** erwarten kann. Da man in der Praxis sowohl den Lagrangeparameter (hier μ) und den Penaltyparameter ϱ schätzen muss, kann es vorkommen, daß

man bei zu kleinem ϱ Divergenz eines Minimierungsverfahrens bezüglich x beobachtet. Es ist aber durchaus problematisch, diese Divergenz richtig zu diagnostizieren und dann auch richtig darauf zu reagieren. Soll z.B. der Penaltyparameter schnell oder langsam vergrößert werden? Soll man mit dem "alten" x -Wert neu starten oder einfach den letzten Wert benutzen? etc.

Einige theoretische Resultate zeigen, daß unter günstigen Bedingungen diese Vorgehensweise tatsächlich erfolgreich sein kann:

Satz B.17. *Es sei x^* eine lokale Minimalstelle des Problems $x^* \in \operatorname{argmin} \{f(x) : h(x) = 0\}$. f, h_1, \dots, h_p seien zweimal stetig differenzierbar auf einer Umgebung von x^* . Es gelte $\operatorname{Rang}(\nabla h(x^*)) = p$. μ^* sei der zugehörige eindeutig bestimmte Multiplikator. Ferner gelte die hinreichende Bedingung zweiter Ordnung, d.h.*

$$z^T (\nabla^2 f(x^*) - \sum_{i=1}^p \mu_i^* \nabla^2 h_i(x^*)) z \geq \alpha z^T z \quad \text{mit } \alpha > 0$$

für alle z mit $\nabla h(x^*)^T z = 0$. Dann gibt es eine Umgebung \mathcal{U}_1 von x^* und eine Umgebung \mathcal{U}_2 von μ^* , sodaß für jedes $\mu \in \mathcal{U}_2$ die Gleichung

$$\nabla_x L_A(x, \mu, \varrho) = 0$$

für hinreichend großes festes ϱ genau eine Lösung $x(\mu)$ in \mathcal{U}_1 hat, für die $\nabla_{xx}^2 L_A(x(\mu), \mu; \varrho)$ positiv definit ist. ($x(\mu)$ somit strenge lokale Minimalstelle von L_A bzgl. x bei festem μ). Die implizit definierte Funktion

$$\varphi(\mu) := L_A(x(\mu), \mu; \varrho)$$

besitzt an der Stelle μ^* eine strenge lokale (unrestringierte) Maximalstelle. Es gilt

$$\begin{aligned} \nabla_{\mu} \varphi(\mu) &= -h(x(\mu)) \\ \nabla_{\mu\mu}^2 \varphi(\mu) &= -\frac{1}{\varrho} I + \mathcal{O}\left(\left(\frac{1}{\varrho}\right)^2\right) \quad \text{für } \varrho \rightarrow \infty. \end{aligned}$$

□

Verfahren von Hestenes und Powell:

Vorgehensweise: Parameter $\beta > 1$, $\gamma > 1$, $\varrho = \varrho_0 > 0$, $0 < \varepsilon \ll 1$. x^0 (Startwert für x^*) gegeben. Löse $\|\nabla f(x^0) - \nabla h(x^0)\mu\| = \min_{\mu}$ (Ausgleichsaufgabe). Resultat: Startwert μ^0 .

Die äussere Iteration leistet die Maximierung bezüglich μ .

Für $k = 0, 1, \dots$,

1. Berechne $x(\mu^k)$ durch unrestringierte Minimierung von $L_A(x, \mu^k; \varrho)$ bzgl. x . Während dieser Minimierung Test auf angemessene Wahl von ϱ (Konvexitätstest) : Sei $x^{k,j}$ die

j -te Näherung für $x(\mu^k)$, $j \geq 1$, mit $x^{k,0} := x^k$. Wenn

$$(\nabla_x L_A(x^{k,j}, \mu^k; \varrho) - \nabla_x L_A(x^{k,j-1}, \mu^k; \varrho))^T (x^{k,j} - x^{k,j-1}) \leq \varepsilon \|x^{k,j} - x^{k,j-1}\|^2$$

(d.h. L_A ist nicht genügend konvex), dann $\varrho := \beta\varrho$, setze Minimierung fort. (Der Penaltyterm wird verstärkt.) Wenn

$$\|h(x^{k,j})\| \geq \gamma \|h(x^{k,0})\|, \quad \text{dann} \quad \varrho := \beta\varrho, \quad j := 0, \quad x^{k,0} := x^k$$

(Wenn die Unzulässigkeit der x -Werte in der inneren Iteration zu stark zunimmt, dann starte Minimierung neu mit grösserem Penalty-Parameter).

2.

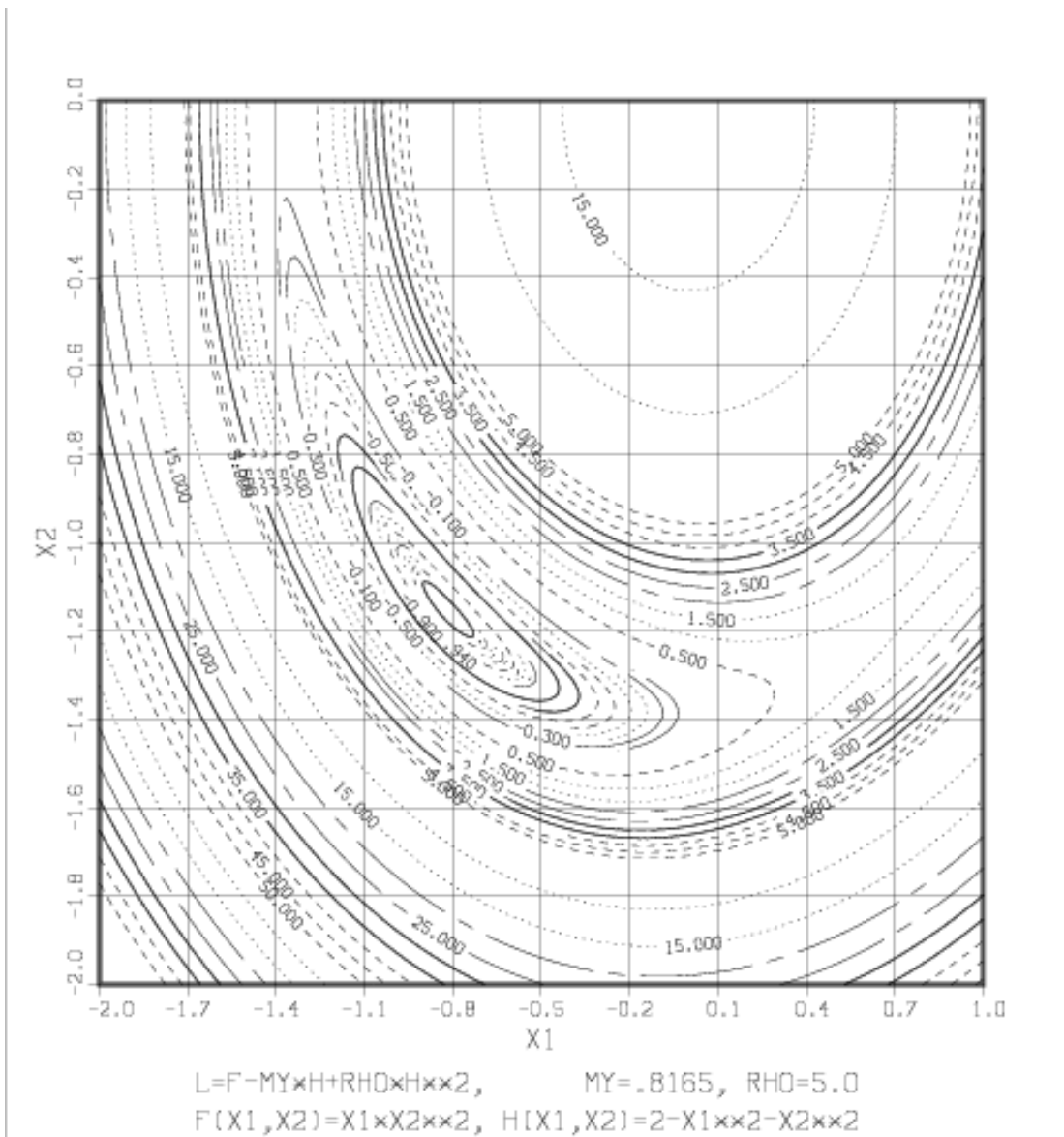
$$\begin{aligned} \mu^{k+1} &:= \mu^k - \varrho h(x(\mu^k)) \\ x^{k+1} &:= x(\mu^k) \end{aligned}$$

Wegen Satz B.17 stellt die Korrektur für μ^k einen approximativen Newtonschritt dar, d.h. man kann in dieser äusseren Iteration schnelle Konvergenz erwarten.

3. Wenn $\|h(x^{k+1})\| \geq \|h(x^k)\|$, dann $\varrho := \beta\varrho$. (Eigentlich sollte die Unzulässigkeit abnehmen).
4. Abbruch des Verfahrens, wenn $\|h(x^{k+1})\|$ und $\|\nabla_x L(x^{k+1}, \mu^{k+1})\|$ genügend klein.

NUMAWWW

Die folgende Abbildung zeigt die erweiterte Lagrangefunktion bei ‘‘richtig‘‘ gewähltem Lagrange- und Penaltyparameter. Sie wirkt hier wie eine ‘‘exakte‘‘ differenzierbare Penaltyfunktion, d.h. eine einzige unrestringierte Minimierung einer differenzierbaren Funktion löst das restringierte Problem.



Die Multiplikator-Methode von Hestenes und Powell ist auch die Basis eines interessanten und praktisch bewährten Ansatzes zur Lösung hochdimensionaler nichtlinearer Optimierungsprobleme. Ein allgemeines NLO-Problem wird zunächst in eines mit Gleichungen und ausschließlich Schrankenrestriktionen umgewandelt:

$$\begin{aligned}
 f(x) &= \min, & g(x) &\geq 0, & h(x) &= 0 \Leftrightarrow \\
 \hat{f}(x,y) &= \min, & \hat{h}(x,y) &= 0, & y &\geq 0 \text{ mit} \\
 \hat{f}(x,y) &= f(x), & \hat{h}(x,y) &= \begin{pmatrix} g(x) - y \\ h(x) \end{pmatrix}.
 \end{aligned}
 \tag{B.4}$$

Für das Problem (B.4) setzt man nun L_A mit \hat{h} an, behält aber die Schrankenrestriktionen explizit bei:

$$\hat{x} := \begin{pmatrix} x \\ y \end{pmatrix}$$

$\hat{x}(\hat{\mu})$ definiert durch

$$\hat{x}(\hat{\mu}) \in \operatorname{argmin} \left\{ \hat{f}(\hat{x}) - \hat{\mu}^T \hat{h}(\hat{x}) + \frac{\varrho}{2} (\hat{h}^T \hat{h})(\hat{x}) : y \geq 0 \right\}$$

($\hat{\mu}$ sind die Multiplikatoren zu den Restriktionen $\hat{h} = 0$).

Danach dann

$$\hat{f}(\hat{x}(\hat{\mu})) - \hat{\mu}^T \hat{h}(\hat{x}(\hat{\mu})) + \frac{\varrho}{2} (\hat{h}^T \hat{h})(\hat{x}(\hat{\mu})) \stackrel{!}{=} \max_{\hat{\mu}}.$$

Schrankenrestringierte Probleme lassen sich genauso einfach lösen wie unrestringierte Probleme. Die Bestimmung von $\hat{\mu}^*$ (optimal) wird in einer Doppeliteration wie bei Hestenes /Powell umgesetzt. Dies ist die Grundidee von LANCELOT (large and nonlinear constrained extended Lagrangian optimization technique) von Conn, Gould & Toint. Es gilt folgender Konvergenzsatz:

Satz B.18. *Seien die Voraussetzungen von Satz B.17 erfüllt und $\|\mu^0 - \mu^*\|$ hinreichend klein. Dann bleibt die Folge der ϱ -Werte beschränkt, und es gibt Konstanten $0 < L < 1$ und C , sodaß*

$$\left. \begin{aligned} \|\mu^{k+1} - \mu^*\| &\leq L \|\mu^k - \mu^*\| \leq L^{k+1} \|\mu^0 - \mu^*\| \quad (\rightarrow 0) \\ \|\hat{x}^{k+1} - \hat{x}^*\| &\leq C \|\mu^k - \mu^*\| \leq CL^k \|\mu^0 - \mu^*\| \quad (\rightarrow 0) \end{aligned} \right\} \quad (\text{B.5})$$

□

Satz B.19. *Sei $f \in C^2(\mathbb{R}^n)$ gleichmäßig konvex, h affin linear und ∇h von Rang p . Dann konvergiert das obige Verfahren für jedes $\varrho = \varrho_0 > 0$ und jedes x^0 im Sinne der Aussage (B.5) von Satz B.18.*

□

B.3.2.2 Ungleichungsrestringierte Probleme (Methode von Rockafellar)

Die Methode von Rockafellar entsteht aus derjenigen von Hestenes und Powell, wenn man Ungleichungsrestriktionen mittels vorzeichenbehafteter Schlupfvariablen in Gleichungen überführt und die Schlupfvariablen dann mittels der Kuhn-Tucker-Gleichungen für das Problem

$$L_A(x, z, \mu, \lambda; \varrho) \stackrel{!}{=} \min_{x, z} \quad z \geq 0$$

mit

$$L_A(x, z, \mu, \lambda; \varrho) := f(x) - \lambda^T (g(x) - z) - \mu^T h(x) + \frac{\varrho}{2} (\|h(x)\|^2 + \|g(x) - z\|^2)$$

wieder eliminiert. Es entsteht die erweiterte Lagrange-Funktion von Rockafellar

$$L_R(x, \mu, \lambda; \varrho) := f(x) - \mu^T h(x) + \frac{\varrho}{2} \|h(x)\|^2 - \frac{1}{2\varrho} \|\lambda\|^2 + \frac{\varrho}{2} \sum_{i=1}^m \left(\min\{0, g_i(x) - \lambda_i/\varrho\} \right)^2.$$

Man beachte, daß diese Funktion nur einmal stetig differenzierbar ist, wenn x und λ frei variieren, aber so oft wie die Funktionen f, g, h , wenn man x und λ einschränkt auf Umgebungen, wo keiner der Terme $g_i(x) - \lambda_i/\varrho$ einen Nulldurchgang hat. Im folgenden Satz wird genau diese Situation durch die Forderung der strikten Komplementarität erzeugt. Für diese Funktion gelten analoge Aussagen zu den Sätzen B.17-B.19 und auch der zugehörige Algorithmus läuft völlig analog. Die einzige Besonderheit bezieht sich auf die Berechnung der ersten partiellen Ableitung der äusseren Funktion φ bezüglich λ . Der Algorithmus läuft völlig analog ab, $\left\| \begin{pmatrix} h(x) \\ (g(x) - \lambda/\varrho)^- \end{pmatrix} \right\|$ ersetzt dabei $\|h(x)\|$, (d.h. von $g(x) - \lambda/\varrho$ zählen nur die negativen Komponenten. ($z_i^- = 0$ falls $z_i \geq 0$, sonst $z_i^- = z_i$.)

$$\begin{pmatrix} \lambda^{k+1} \\ \mu^{k+1} \end{pmatrix} = \begin{pmatrix} \lambda^k \\ \mu^k \end{pmatrix} + \varrho \nabla_{(\lambda, \mu)} \varphi(\lambda^k, \mu^k)$$

ersetzt $\mu^{k+1} := \mu^k - \varrho h(x(\mu^k))$. Gleichungen für $\nabla \varphi$ siehe Satz B.20. Die Funktion von Rockafellar ist nur unter der Bedingung des strikten komplementären Schlupfes und nur lokal zweimal stetig differenzierbar. Kiwiel (J.O.T.A. 88, 1996, 233-236) hat eine analoge Funktion angegeben, die global zweimal stetig differenzierbar (bezüglich x) ist und für die unsere Algorithmen also uneingeschränkt anwendbar sind:

$$L_K(x, \mu, \lambda; \varrho) = f(x) - \mu^T h(x) + \varrho \|h(x)\|^2 + \frac{1}{3\varrho} \sum_{i=1}^m t_i$$

mit

$$t_i = \max\{0, (\text{sign}(\lambda_i) \sqrt{|\lambda_i|}) - \varrho g_i(x)\}^3 - |\lambda_i|^{3/2}.$$

Die Aufdatierungsformel für die Multiplikatoren ist dann

$$\begin{aligned} \mu^{k+1} &= \mu^k - 2\varrho h(x^{k+1}), \\ \lambda_i^{k+1} &= \max\{0, |\lambda_i^k| - \varrho g_i(x^{k+1})\}^2 \end{aligned}$$

NUMAWWW

Satz B.20. x^* sei lokale Minimalstelle von f auf $\mathfrak{S} = \{x \in \mathcal{D} : h(x) = 0, g(x) \geq 0\}$, \mathcal{D} sei offen, \mathfrak{S} sei beschränkt und abgeschlossen, f, g, h seien zweimal stetig differenzierbar auf \mathcal{D} . x^* sei weiterhin regulärer Punkt, d.h. mit l aktiven Ungleichungsrestriktionen und

$$\mathcal{A}(x^*) = \{i : g_i(x) \equiv 0\} =: \{i_1, \dots, i_l\}$$

ist

$$N^* := (\nabla h_1(x^*), \dots, \nabla h_p(x^*), \nabla g_{i_1}(x^*), \dots, \nabla g_{i_l}(x^*))$$

vom Rang $p + l$. Ferner gelte in x^* die strikte Komplementarität, d.h. $\lambda^* + g(x^*) > 0$, sowie die hinreichende Bedingung zweiter Ordnung, d.h. $z^T \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) z \geq \alpha z^T z$ mit $\alpha > 0$ für alle z mit $(N^*)^T z = 0$. Dann gibt es ein $\varrho_0 > 0$, sodaß für alle $\varrho \geq \varrho_0$ L_R für (λ, μ) aus einer geeigneten Umgebung \mathcal{U}_2 von (λ^*, μ^*) eine eindeutige Gradientennullstelle $x(\lambda, \mu)$ in einer Umgebung \mathcal{U}_1 von x^* (bezüglich x) besitzt.

Ferner gilt

$$\nabla_{xx}^2 L_R(x, \lambda, \mu; \varrho) \text{ positiv definit für } x \in \mathcal{U}_1, (\lambda, \mu) \in \mathcal{U}_2,$$

d.h. $x(\lambda, \mu)$ ist strenges lokales (unrestringiertes) Minimum von L_R bzgl. x . Die implizit definierte Funktion

$$\varphi(\lambda, \mu) := L_R(x(\lambda, \mu), \lambda, \mu; \varrho)$$

besitzt auf \mathcal{U}_2 die eindeutige Gradientennullstelle (λ^*, μ^*) ,

$$-\nabla_{(\lambda, \mu)}^2 \varphi(\lambda, \mu)$$

ist positiv definit (d.h. (λ^*, μ^*) ist **unrestringiertes Maximum**) und es gilt

$$\nabla_{(\lambda, \mu)}^2 \varphi(\lambda, \mu) = -\frac{1}{\varrho} I + \mathcal{O}\left(\left(\frac{1}{\varrho}\right)^2\right) \text{ für } \varrho \rightarrow \infty.$$

Ferner

$$\begin{aligned} \frac{\partial}{\partial \lambda_i} \varphi(\lambda, \mu) &= \begin{cases} -\frac{1}{\varrho} \lambda_i & \text{falls } g_i(x(\lambda, \mu)) - \frac{\lambda_i}{\varrho} \geq 0 \\ -g_i(x(\lambda, \mu)) & \text{falls } g_i(x(\lambda, \mu)) - \frac{\lambda_i}{\varrho} < 0 \end{cases} \\ \frac{\partial}{\partial \mu_j} \varphi(\lambda, \mu) &= -h_j(x(\lambda, \mu)), \quad j = 1, \dots, p. \end{aligned}$$

Falls $\mathcal{D} = \mathbb{R}^n$, f streng konvex, h affin linear und g_i konkav, $i = 1, \dots, m$, dann kann man $\mathcal{U}_1 = \mathbb{R}^n$, $\mathcal{U}_2 = \mathbb{R}^m \times \mathbb{R}^p$ nehmen, d.h. alle Aussagen gelten global. \square

Bemerkung B.21. Wie im nur gleichungsrestringierten Fall gilt obiger Satz für eine konvexe Problem global und für jedes feste $\varrho > 0$. In diesem Fall kann man also einen Penaltyparameter in vernünftiger Größenordnung (z.B. 100 für Probleme, bei denen die Skalierung alle Gradienten in die Größenordnung 1 bringt) fest wählen und dann die Iteration ablaufen lassen, ohne heuristische Zusatzmaßnahmen in Betracht ziehen zu müssen. Alles, was man

benötigt, ist dann ein leistungsfähiger unrestringierter Optimierungscodes.

B.3.3 Exakte differenzierbare Penalty-Funktionen (ERG)

Hier wird NLO in eine einzige unrestringierte Minimierungsaufgabe mit einer differenzierbaren Zielfunktion umgewandelt. Diese sind jedoch sehr kompliziert aufgebaut und enthalten schwer zu bestimmende Parameter. Ausser in einigen Spezialfällen haben sie daher kein praktisches Interesse gefunden.

B.3.3.1 Primale exakte differenzierbare Penalty-Funktionen

Diese Funktionen enthalten nur die primale Variable x . Sie entstehen, wenn man die Multiplikatoren mit Hilfe der Kuhn-Tucker-Bedingungen als Funktionen von x ausdrückt und diese Ausdrücke in die erweiterten Lagrange-Funktionen L_A bzw. L_R einsetzt.

Der einfachere Fall liegt vor, wenn nur Gleichungen zu berücksichtigen sind.

Definiere μ als Funktion von x durch die Forderung

$$\|\nabla f(x) - \nabla h(x)\mu\| \stackrel{!}{=} \min_{\mu} \iff \mu(x) = (\nabla h(x)^T \nabla h(x))^{-1} \nabla h(x)^T \nabla f(x)$$

Einsetzen in L_A : exakte Penalty-Funktion von **Fletcher** :

$$\Phi(x; \varrho) = f(x) - h(x)^T (\nabla h(x)^T \nabla h(x))^{-1} \nabla h(x)^T \nabla f(x) + \frac{\varrho}{2} h(x)^T h(x).$$

Für den Fall gemischter Gleichungen und Ungleichungen geht man aus von den modifizierten Kuhn-Tucker-Bedingungen:

$$\begin{aligned} \nabla f(x) - \nabla h(x)\mu - \nabla g(x)\lambda &= 0, \\ \gamma g_i(x)\lambda_i &= 0, \quad i = 1, \dots, m \quad \text{mit } \gamma > 0, \\ \gamma h_j(x)\mu_j &= 0, \quad j = 1, \dots, p. \end{aligned}$$

Dieses System fasst man bei gegebenem x auf als lineares System von $n + m + p$ linearen Gleichungen in den $m+p$ Unbekannten μ , λ und löst es im Sinne der Methode der kleinsten Quadrate nach μ und λ und setzt das Resultat in die Funktion von Rockafellar ein.

Man erhält die exakte Penalty-Funktion von **di Pillo und Grippo**:

$$\begin{aligned} \begin{pmatrix} \mu(x) \\ \lambda(x) \end{pmatrix} &:= \left(\underbrace{(\nabla h(x), \nabla g(x))^T (\nabla h(x), \nabla g(x))}_{(m+p) \times (m+p)\text{-Matrix}} \right. \\ &\quad \left. + \gamma^2 \begin{pmatrix} H^2(x) & 0 \\ 0 & G^2(x) \end{pmatrix} \right)^{-1} \begin{pmatrix} \nabla h(x)^T \nabla f(x) \\ \nabla g(x)^T \nabla f(x) \end{pmatrix} \\ H(x) &= \text{diag}(h_i(x)) \\ G(x) &= \text{diag}(g_i(x)), \quad \gamma \neq 0 \text{ fest gewählt.} \\ \Phi(x; \varrho) &= f(x) - \mu(x)^T h(x) - \frac{1}{2\varrho} \|\lambda(x)\|^2 \\ &\quad + \frac{\varrho}{2} \left(\|h(x)\|^2 + \sum_{i=1}^m \left(\min(0, g_i(x) - \lambda_i(x)/\varrho) \right)^2 \right) \end{aligned}$$

Unter einschränkenden Regularitätsvoraussetzungen entsprechen die strengen lokalen Minimalstellen von NLO strengen lokalen unrestringierten Minimalstellen von Φ , wenn ϱ hinreichend groß (aber endlich, fest) ist.

Der offensichtliche Vorteil dieser Vorgehensweise besteht darin, daß nur eine einzige unrestringierte Minimierung geleistet werden muß. Der Nachteil liegt in der schwierigen Steuerung des Parameters ϱ . Die Funktionen sind auch nur sehr aufwendig auszuwerten. Es gibt zusätzliche stationäre Punkte, die nicht lokalen Lösungen von NLO entsprechen.

B.3.3.2 Primal-duale exakte Penalty-Funktionen

Diese entstehen aus den erweiterten Lagrange-Funktionen durch Hinzufügen eines Strafterms für die Verletzung der Bedingung $\nabla_x L(x, \lambda, \mu) = 0$, z.B.

$$\begin{aligned} \Phi(x, \lambda, \mu; \varrho) &= f(x) - \frac{1}{2\varrho} \lambda^T \lambda - \mu^T h(x) \\ &\quad + \frac{\varrho}{2} \left(\|h(x)\|^2 + \sum_{i=1}^m \left(\min(0, g_i(x) - \lambda_i/\varrho) \right)^2 \right) \\ &\quad + \eta \left\| \begin{pmatrix} \nabla h(x)^T \\ \nabla g(x)^T \end{pmatrix} \nabla f(x) - \begin{pmatrix} \nabla h(x), \nabla g(x) \end{pmatrix}^T (\nabla h(x), \nabla g(x)) \right. \\ &\quad \left. + \gamma^2 \begin{pmatrix} H^2(x) & 0 \\ 0 & G^2(x) \end{pmatrix} \begin{pmatrix} \mu \\ \lambda \end{pmatrix} \right\|^2 \end{aligned}$$

$\gamma > 0$ fest, η, ϱ fest hinreichend groß. H, G s.o.

Hier sind nun x, λ, μ **Minimierungsvariablen**. Es ist nur eine unrestringierte Minimierung (Dimension: $n + m + p$) zu leisten. Die Auswertung von Φ , insbesondere von $\nabla \Phi$ ist aber

sehr aufwendig. Die Wahl der Parameter ϱ, η erweist sich als schwierig. Die praktischen Resultate sind sehr enttäuschend. Deshalb spielen Funktionen dieser Art gegenwärtig keine praktische Rolle.

B.3.4 Primale Verfahren für linear restringierte Probleme

Bei linear restringierten Problemen hat man die Möglichkeit, einen zulässigen Punkt zu bestimmen bzw. zu entscheiden, ob ein solcher existiert, während man bei nichtlinearen Restriktionen selbst diese Aufgabe nur unter sehr einschränkenden Bedingungen (z.B. der globalen Gültigkeit einer verallgemeinerten Mangasarian-Fromowitz-Bedingung, bei der an die Stelle der Menge der aktiven Restriktionen die der aktiven oder verletzten Restriktionen tritt) immer gelöst werden kann. Von dieser Besonderheit macht man häufig Gebrauch und konstruiert Folgen von Näherungen an die (bzw. eine) Lösung, die stets zulässig sind und in der Regel alle auf dem Rand der zulässigen Menge liegen. In diesem Abschnitt besprechen wir kurz das Simplexverfahren der linearen Optimierung, einen QP-Löser und schliesslich ein Abstiegsverfahren für linear restringierte Probleme mit allgemeiner nichtlinearer Zielfunktion.

Im Folgenden sei:

$$\begin{aligned} g(x) &= G^T x + g^0 \\ h(x) &= H^T x + h^0 \\ \mathcal{A}(x) &= \{i : g_i(x) = 0\} \\ N_{\mathfrak{B}} &= (H, G_{\mathfrak{B}}) \quad \text{für } \mathfrak{B} \subset \mathcal{A}(x) \end{aligned}$$

wobei

$$\begin{aligned} G_{\mathfrak{B}} &= (g^{i_1}, \dots, g^{i_l}) \quad \text{für } \mathfrak{B} = \{i_1, \dots, i_l\} \\ g^j &= \nabla g_j(x) \quad (\text{unabhängig von } x). \end{aligned}$$

B.3.4.1 Das LP-Problem: Simplexverfahren von Dantzig

In diesem Unterabschnitt behandeln wir den Fall

$$f(x) = c^T x$$

unter (affin) linearen Restriktionen. Dieser Fall ist theoretisch besonders leicht behandelbar, wenn man ihn in eine gewisse Standardform bringt, nämlich: Gesucht

$$\begin{aligned} x^* : \quad f(x^*) &= c^T x^* = \min \{c^T x : x \in \mathfrak{S}\} \\ \mathfrak{S} &= \{x \in \mathbb{R}^n : Ax = b, \quad x \geq 0\}. \end{aligned} \tag{B.6}$$

Dabei soll gelten $A \in \mathbb{R}^{p \times n}$, $b \geq 0$, $\text{Rang}(A) = p$. Ist $\text{Rang}(A) < p$, dann kann man im Prinzip redundante Gleichungen durch Elimination entfernen, da die Verträglichkeit der Gleichungen ja gegeben ist, bis die Matrix vollen Zeilenrang hat. Andere Formen von LO–Aufgaben, die auf mannigfache Art auftreten können, werden wir stets in diese Normalform überführen, obwohl dies unter dem Gesichtspunkt von Speicher– und Rechenaufwand nicht immer besonders sinnvoll ist. In der in der Praxis eingesetzten Software wird dies nicht durchgeführt. Die Standardform erlaubt aber eine erhebliche Vereinfachung in der Beschreibung der Methode. Die Technik der Transformation wird an Beispielen erläutert.

- a) Ungleichungsnebenbedingungen, die keine reinen Vorzeichenbedingungen sind, werden durch Einführung von zusätzlichen vorzeichenbeschränkten Variablen (“Schlupfvariablen”) in Gleichungsform gebracht:

$$\begin{aligned} Bx \leq b &\iff Bx + y = b, \quad y \geq 0 &\iff (B, I_m) \begin{pmatrix} x \\ y \end{pmatrix} = b, \quad y \geq 0 \\ Bx \geq b &\iff Bx - y = b, \quad y \geq 0 &\iff (B, -I_m) \begin{pmatrix} x \\ y \end{pmatrix} = b, \quad y \geq 0. \end{aligned}$$

Statt $f(x) = c^T x$ wird dann $f(\tilde{x}) = (c^T, 0)\tilde{x}$ mit $\tilde{x} = \begin{pmatrix} x \\ y \end{pmatrix}$ minimiert.

- b) Nach Anwendung der Technik aus a) erhält man eine LO–Aufgabe, bei der eventuell ein Teil der Variablen keiner Vorzeichenbeschränkung unterliegt. Wir zerlegen den Variablenvektor x entsprechend in Teilvektoren y und z , von denen y schon vorzeichenbeschränkt sei. Wir betrachten also Nebenbedingungen der Form

$$p \left\{ \underbrace{B_1}_{n_1}, \underbrace{B_2}_{n_2} \right\} \begin{pmatrix} y \\ z \end{pmatrix} = b, \quad y \geq 0.$$

Den nicht vorzeichenbeschränkten Vektor $z \in \mathbb{R}^{n_2}$ schreiben wir als Differenz zweier vorzeichenbeschränkter Vektoren:

$$z = z^+ - z^-, \quad z^+ \geq 0, \quad z^- \geq 0, \quad z^+, z^- \in \mathbb{R}^{n_2}.$$

Wegen

$$B_2 z = B_2(z^+ - z^-) = (B_2, -B_2) \begin{pmatrix} z^+ \\ z^- \end{pmatrix}$$

erhalten wir so die transformierte Aufgabe

$$\begin{aligned} \tilde{f}(\tilde{x}) &= \tilde{c}^T \tilde{x} \stackrel{!}{=} \min \\ \tilde{B} \tilde{x} &= b, \quad \tilde{x} \geq 0 \\ \tilde{B} &= (B_1, B_2, -B_2), \quad \tilde{x}^T = (y^T, (z^+)^T, (z^-)^T), \quad \tilde{c}^T = (c^{1T}, c^{2T}, -c^{2T}). \end{aligned}$$

Bemerkung B.22. Man kann nichtvorzeichenbeschränkte Variablen aus den Gleichungsnebenbedingungen auch durch Elimination entfernen, wodurch man eine Aufgabe geringerer Dimension erhält. Wegen des einfacheren Zugangs haben wir hier darauf verzichtet. Rechentechnisch verfährt man in der Praxis jedoch anders. \square

Beispiel B.23.

$$\begin{aligned}
 f(x) = 4x_1 + 5x_2 + 2x_3 &\stackrel{!}{=} \min \\
 3x_1 + 5x_2 &= 3 \\
 2x_1 - 4x_2 + x_3 &\leq 4 \\
 7x_1 + 6x_2 - 3x_3 &\geq 3 \\
 x_1 &\geq 0.
 \end{aligned}$$

Einführung von Schlupfvariablen:

$$\begin{aligned}
 3x_1 + 5x_2 &= 3 \\
 2x_1 - 4x_2 + x_3 + y_1 &= 4 \\
 7x_1 + 6x_2 - 3x_3 - y_2 &= 3 \\
 x_1 \geq 0, \quad y_1 \geq 0, \quad y_2 \geq 0.
 \end{aligned}$$

Zusätzliche Einführung vorzeichenbeschränkter Variablen für x_2, x_3 :

$$\begin{aligned}
 3x_1 + 5x_2^+ - 5x_2^- &= 3 \\
 2x_1 - 4x_2^+ + 4x_2^- + x_3^+ - x_3^- + y_1 &= 4 \\
 7x_1 + 6x_2^+ - 6x_2^- - 3x_3^+ + 3x_3^- - y_2 &= 3 \\
 x_1, x_2^+, x_2^-, x_3^+, x_3^-, y_1, y_2 &\geq 0
 \end{aligned}$$

$$\tilde{f}(\tilde{x}) = 4x_1 + 5x_2^+ - 5x_2^- + 2x_3^+ - 2x_3^- \stackrel{!}{=} \max; \quad \tilde{x} = (x_1, x_2^+, x_2^-, x_3^+, x_3^-, y_1, y_2)^T.$$

□

c) Maximierungsaufgaben kann man lösen, indem man $-f$ minimiert und umgekehrt.

Wir diskutieren nun die Struktur der zulässigen Menge \mathfrak{S} und der Lösungsmenge \mathcal{M} der linearen Optimierungsaufgabe.

Es ist unmittelbar klar, daß \mathfrak{S} und die Lösungsmenge \mathcal{M} konvex sind, sofern sie nicht leer sind.

Beispiel B.24.

$$\text{a) } \mathfrak{S} = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : -x_1 - x_2 = 1, \quad x_1 \geq 0, \quad x_2 \geq 0 \right\} = \emptyset$$

$$\text{b) } \mathfrak{S} = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_1 + x_2 = 1, \quad x_1 \geq 0, \quad x_2 \geq 0 \right\} \neq \emptyset$$

$$\text{c) } \mathfrak{S} = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_1 - x_2 = 2, \quad x_1 \geq 0, \quad x_2 \geq 0 \right\} \neq \emptyset, \quad f(x) = x_1 : \mathcal{M} = \emptyset$$

$$\text{d) c) mit } f(x) = -x_2 : \mathcal{M} = \left\{ \begin{pmatrix} 2 \\ 0 \end{pmatrix} \right\}$$

$$e) \quad \mathfrak{S} = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_1 + x_2 = 1, \quad x_1 \geq 0, \quad x_2 \geq 0 \right\}, \quad f(x) = x_1 + x_2 : \mathcal{M} = \mathfrak{S}.$$

□

Definition B.25. Eine Menge \mathfrak{S} gemäß (B.6) heißt ein **konvexes Polyeder**.
Ein konvexes und kompaktes Polyeder heißt **konvexes Polytop**.

□

Definition B.26. Sind x_1, \dots, x_N Punkte im \mathbb{R}^n und $\alpha_1, \dots, \alpha_N \geq 0$ mit

$$\sum_{i=1}^N \alpha_i = 1,$$

dann heißt

$$\sum_{i=1}^N \alpha_i x_i$$

eine **Konvexkombination** dieser Punkte.

Definition B.27. Sei $x \in \mathcal{D}$ und \mathcal{D} konvex. Dann heißt x **Extrempunkt** von \mathcal{D} , wenn x nicht als Konvexkombination anderer Punkte von \mathcal{D} dargestellt werden kann.

Wir wollen nun nicht voraussetzen, daß \mathfrak{S} beschränkt ist. Es gibt also unter Umständen Richtungen in \mathfrak{S} , die ins Unendliche durchlaufen werden können. Dies ist Gegenstand der folgenden

Definition B.28. $d \neq 0$ heißt **Richtung** in \mathfrak{S} , wenn für alle $x \in \mathfrak{S}$ und alle $\tau \geq 0$ $x + \tau d \in \mathfrak{S}$ gilt. d heißt **extremale Richtung** in \mathfrak{S} , wenn d Richtung in \mathfrak{S} ist und nicht aus anderen Richtungen in \mathfrak{S} linear kombiniert werden kann.

□

Wir geben nun den Darstellungssatz für konvexe Polyeder der hier vorliegenden Form an, der besagt, daß $x \in \mathfrak{S}$ dargestellt werden kann als Summe einer Konvexkombination der Extrempunkte von \mathfrak{S} und einer Summe von positiven Vielfachen extremer Richtungen von \mathfrak{S} , falls es solche gibt.

Satz B.29. Die Menge \mathcal{M} der Extrempunkte von \mathfrak{S} ist endlich und nicht leer, etwa $\{x^1, \dots, x^s\}$. Die Menge der extremalen Richtungen von \mathfrak{S} ist leer oder endlich, etwa $\{d^1, \dots, d^r\}$. Ist $x \in \mathfrak{S}$ beliebig, dann gilt eine Darstellung

$$x = \sum_{i=1}^s \alpha_i x^i + \sum_{j=1}^r \tau_j d^j \quad \text{mit} \quad \alpha_i \in [0, 1], \quad \sum_{i=1}^s \alpha_i = 1, \quad \tau_j \geq 0.$$

Beweis: siehe z.B. bei Stoer&Witzgall.

□

Die wesentliche Aussage dieses Satzes ist einmal die Existenz mindestens eines Extrempunktes und die Darstellungsformel, mit deren Hilfe nun die Funktionswerte der linearen

Funktion $f(x) = c^T x$ durch die Werte von f auf der Menge der Extrempunkte und der Menge der extremalen Richtungen beschrieben werden können:

$$c^T x = \sum_{i=1}^s \alpha_i (c^T x^i) + \sum_{j=1}^r \tau_j (c^T d^j) \quad \text{für } x \in \mathfrak{S},$$

mit

$$\alpha_i \in [0, 1], \quad \sum_{i=1}^s \alpha_i = 1, \quad \tau_j \geq 0 \quad \text{für } j = 1, \dots, r.$$

Hieraus folgt unmittelbar

Satz B.30. *Das LO-Problem habe die Standardform (B.6). \mathfrak{S} sei nicht leer. Dann gilt: Entweder ist $f(x) = c^T x$ auf \mathfrak{S} nicht nach unten beschränkt oder es existiert ein Extrempunkt von \mathfrak{S} , an dem f sein Infimum annimmt.*

Beweis: Ist die Menge der extremalen Richtungen $\{d^j\}$ von \mathfrak{S} nicht leer und gibt es ein $d^{j'}$ mit $c^T d^{j'} < 0$, so ist f nicht nach unten beschränkt auf \mathfrak{S} (man betrachte $\tau_{j'} \rightarrow \infty$). Ist $c^T d^j \geq 0$ für alle extremalen Richtungen von \mathfrak{S} , dann setze man $\tau_1 = \dots = \tau_r = 0$. Dadurch wird f bezüglich der τ_j minimiert. Dann aber folgt aus der Darstellungsformel

$$f(x) \geq \min_{i=1, \dots, s} c^T x^i = f(x^{i_0}) \quad \text{für ein } i_0 \in \{1, \dots, s\}$$

wo $\{x^1, \dots, x^s\} \neq \emptyset$ die Extrempunktmenge von \mathfrak{S} ist. □

Die Lösung des LP-Problems liegt also immer auch in einem Extrempunkt, aber unter Umständen auch auf einem vollständigen Randstück von \mathfrak{S} . Der Simplexalgorithmus arbeitet nun so, daß er, ausgehend von einem Extrempunkt zu einem "benachbarten" Extrempunkt mit verkleinertem (bzw. im Entartungsfall nicht vergrößerten, s.h.) Funktionswert fortschreitet. Da es nur endlich viele Extrempunkte gibt, ist das Verfahren damit notwendig endlich. Es kann aber exponentiell viele (in n) Schritte benötigen. Wir benötigen als nächstes eine Charakterisierung der Extrempunkte von \mathfrak{S} . Dazu gilt

Satz B.31. *$x \in \mathfrak{S}$ ist Extrempunkt von \mathfrak{S} genau dann, wenn die Matrix $A_{\mathcal{B}_+}$ spaltenregulär ist, mit $\mathcal{B}_+ = \{i \in \{1, \dots, m\} : x_i > 0\}$. (Es ist dann also auch $|\mathcal{B}_+| \leq p$).*

Man bezeichnet in diesem Zusammenhang eine Extrempunkt auch als "Ecke". In einer solchen Ecke kann man also durch Hinzunahme von Spalten aus A (wegen der Annahme $\text{Rang}(A) = p$) eine invertierbare $p \times p$ -Matrix aufbauen. Diese kann man benutzen, um p der Variablen x_i aus den Gleichungen $Ax = b$ mittels der übrigen $n - p$ auszudrücken und das Problem lokal zu einem nur vorzeichenrestringierten Problem in $n - p$ Variablen zu machen. Das ist der Ansatz für den Simplexalgorithmus. Im Sinne der zuvor entwickelten

Theorie haben wir aufgrund der speziellen Struktur der Aufgabe

$$\begin{aligned}\nabla h(x) &\equiv A^T, \\ \nabla g(x) &\equiv I \in \mathbb{R}^{n \times n}, \\ \mathcal{A}(x) &= \{i : x_i = 0\}.\end{aligned}$$

Ist $|\mathcal{A}(x)| > n - p$, dann sind mehr als n Restriktionen bindend und es liegt ein nicht-regulärer Punkt vor. Im Folgenden wollen wir dies ausschliessen. Dann ist also wegen Satz B.31 in einem Extrempunkt (einer Ecke) immer $|\mathcal{A}(x)| = n - p$, $\mathcal{B}_+ = p$ und $A_{\mathcal{B}_+}$ invertierbar. In diesem Fall nennt man \mathcal{B}_+ eine "Basis". Allgemeiner:

Definition B.32. Sei $\mathcal{B} \subset \{1, \dots, n\}$, $|\mathcal{B}| = p$ und $\tilde{A} := (a^i)_{i \in \mathcal{B}}$ regulär. Dann heißt \mathcal{B} eine Basis, die Variablen x_i , $i \in \mathcal{B}$ heißen Basisvariablen und die übrigen x_i Nicht-basisvariablen. □

Beispiel B.33. Es sei

$$\mathfrak{S} = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x \in \mathbb{R}^3 : x \geq 0, \quad x_1 + 2x_2 + 3x_3 = 6 \right\}.$$

Hier ist also $p = 1$, $n = 3$, $A = (1, 2, 3)$, $\text{Rang}(A) = 1 = p$. Mögliche Basismengen $\mathcal{B}_1 = \{1\}$, $\mathcal{B}_2 = \{2\}$, $\mathcal{B}_3 = \{3\}$. Es können nicht zwei Komponenten einer Ecke > 0 sein. Die drei Ecken sind $(6, 0, 0)^T$, $(0, 3, 0)^T$, $(0, 0, 2)^T$. □

Wie bereits erwähnt, wollen wir also von Ecke zu Ecke fortschreiten mit Verkleinerung des Zielfunktionswertes. Hier die Definition "benachbart":

Definition B.34. Seien $x^1, x^2 \in \mathfrak{S}$ zwei Ecken. x^1, x^2 heißen benachbart, falls

$$|\mathcal{B}(x^1) \cap \mathcal{B}(x^2)| = p - 1.$$

□

Wir gehen davon aus, daß eine zulässige Ecke von \mathfrak{S} bekannt ist. Es sei $A = (a^1, \dots, a^n)$ und \mathcal{B}_0 die zu x^0 gehörende Basis. Dann ist $A_{\mathcal{B}_0} = (a^i)_{i \in \mathcal{B}_0}$ regulär. Mit dieser Information lösen wir nun den Gleichungsteil der Kuhn-Tucker-Bedingungen:

$$\begin{aligned}c - A^T \mu - \lambda &= 0 \\ Ax - b &= 0 \\ x_i \lambda_i &= 0, \quad i = 1, \dots, n\end{aligned}$$

Aufgeteilt nach Indizes in \mathcal{B}_0 und $\bar{\mathcal{B}}_0$ lautet dies

$$\begin{aligned}A_{\mathcal{B}_0} x_{\mathcal{B}_0}^0 &= b \\ \lambda_{\mathcal{B}_0}^0 &= 0 \\ c_{\mathcal{B}_0} - (A_{\mathcal{B}_0})^T \mu^0 &= 0 \\ \lambda_{\bar{\mathcal{B}}_0} &= c_{\bar{\mathcal{B}}_0} - (A_{\bar{\mathcal{B}}_0})^T \mu^0.\end{aligned}$$

Für beliebiges $x \in \mathfrak{S}$ ist

$$Ax = b \iff A_{\mathcal{B}_0}x_{\mathcal{B}_0} = b - A_{\bar{\mathcal{B}}_0}x_{\bar{\mathcal{B}}_0} \iff x_{\mathcal{B}_0} = A_{\mathcal{B}_0}^{-1}b - A_{\mathcal{B}_0}^{-1}A_{\bar{\mathcal{B}}_0}x_{\bar{\mathcal{B}}_0}$$

und daher

$$\begin{aligned} f(x)|_{Ax=b} &= c_{\mathcal{B}_0}^T \left(A_{\mathcal{B}_0}^{-1}b - A_{\mathcal{B}_0}^{-1}A_{\bar{\mathcal{B}}_0}x_{\bar{\mathcal{B}}_0} \right) + c_{\bar{\mathcal{B}}_0}^T x_{\bar{\mathcal{B}}_0} \\ &= c_{\mathcal{B}_0}^T A_{\mathcal{B}_0}^{-1}b + \lambda_{\bar{\mathcal{B}}_0}^T x_{\bar{\mathcal{B}}_0}, \end{aligned}$$

so daß wir f nunmehr als Funktion der $n - p$ Nicht-Basisvariablen dargestellt haben, die keinen Gleichungs-, sondern nur noch den Vorzeichenrestriktionen unterliegen. Für x^0 ist $x_{\bar{\mathcal{B}}_0}^0 = 0$ und daher

$$\begin{aligned} f(x^0) &= c_{\mathcal{B}_0}^T A_{\mathcal{B}_0}^{-1}b \\ f(x) &= f(x^0) + (\lambda^0)_{\bar{\mathcal{B}}_0}^T \underbrace{x_{\bar{\mathcal{B}}_0}}_{\geq 0}. \end{aligned}$$

Also folgt

Satz B.35. Sei x^0 Ecke von \mathfrak{S} und $\lambda_{\mathcal{B}_0}^0 := 0$, $\lambda_{\bar{\mathcal{B}}_0}^0 := c_{\bar{\mathcal{B}}_0} - A_{\bar{\mathcal{B}}_0}^T (A_{\mathcal{B}_0}^{-1})^T c_{\mathcal{B}_0}$. Genau dann ist x^0 Lösung der Aufgabe $f(x) = c^T x = \min!$ $x \in \mathfrak{S}$, wenn $\lambda^0 \geq 0$. \square

Sei nun x^0 nicht optimal, d.h.

$$\exists l \in \bar{\mathcal{B}}_0 \text{ mit } \lambda_l^0 < 0.$$

Dann besagt obige Darstellung, daß f verkleinert werden kann, wenn man x_l von null weg vergrößert und gleichzeitig die Zulässigkeit von x bezüglich der Gleichungen erhält. Dies bedeutet, daß man x wie folgt abändern muss:

$$\begin{aligned} x_j &= 0 \quad \text{für } j \in \bar{\mathcal{B}}_0 \setminus \{l\} \\ x_l &= \tau \geq 0 \\ x_{\mathcal{B}_0} &= x_{\mathcal{B}_0}^0 - \tau d_{\mathcal{B}_0}^0 \quad \text{mit } A_{\mathcal{B}_0} d_{\mathcal{B}_0}^0 = a^l \end{aligned}$$

x_l heisst die **eintretende** Variable (engl. entering, incoming variable) und die Auswahlregel für den Index l "pricing". In der Praxis sind die Dimensionen dieser Probleme oft riesig und daher der Aufwand für die Bestimmung von l , das ja nicht eindeutig bestimmt ist, nicht vernachlässigbar. Nun sind folgende Fälle möglich:

1.

$$d_{\mathcal{B}_0}^0 \leq 0 \text{ komponentenweise}$$

Dann kann man f entlang der durch τ parametrisierten Geraden

$$x^0 - \tau d^0,$$

wo wir $d_l^0 = -1$ und $d_i^0 = 0$ für $i \in \bar{\mathcal{B}}_0 \setminus \{l\}$ gesetzt haben, (d^0 ist eine "Richtung" in \mathfrak{S}) unbegrenzt verkleinern, ohne \mathfrak{S} zu verlassen: f ist nicht nach unten beschränkt auf \mathfrak{S} . Dies bedingt natürlich einen Abbruch der Rechnung.

2. Es gibt mindestens eine Komponente $d_k^0 > 0$. Dann nimmt die entsprechende Komponente von $x_{\mathcal{B}_0}$ streng monoton ab mit wachsendem τ und es gibt wegen der Vorzeichenbedingung eine maximal erlaubte Schrittweite τ_0 . Da f linear fällt, wird diese Schrittweite benutzt. Es ist

$$\tau_0 = \min\left\{\frac{x_{j_i}}{d_{j_i}^0} : j_i \in \mathcal{B}_0 \text{ und } d_{j_i}^0 > 0\right\}.$$

Wird das Minimum für $i = k$ angenommen, dann ist nun $x_{j_k} = 0$, d.h. j_k verlässt die Basis, x_{j_k} ist eine "outgoing" Variable. x^1 ist nun eine neue Ecke und ein neuer Schritt des Verfahrens beginnt. Wird das Minimum für mehrere Indizes gleichzeitig angenommen, dann ist die neue Ecke "entartet" (d.h. man hat mehr als n aktive Restriktionen).

$$\mathcal{B}_1 = (\mathcal{B}_0 \setminus \{j_k\}) \cup \{l\}.$$

Algorithmus Simplexverfahren von Dantzig

Gegeben eine Ecke x^0 von \mathfrak{S} . $A_{\mathcal{B}_0}$ sei invertierbar (Wenn x^0 entartet ist, muss man geeignete Spalten von A ergänzen, um dies zu erreichen, was aber wegen der Rangannahme für A möglich ist.)

Für $k = 0, 1, \dots$

1.

$$\bar{\mathcal{B}}_k = \{1, \dots, n\} \setminus \mathcal{B}_k.$$

2. Löse

$$A_{\bar{\mathcal{B}}_k}^T \mu^k = c_{\mathcal{B}_k}.$$

3. Setze

$$\lambda_{\mathcal{B}_k}^k = 0 \quad \lambda_{\bar{\mathcal{B}}_k}^k = c_{\bar{\mathcal{B}}_k} - (A_{\bar{\mathcal{B}}_k})^T \mu^k.$$

4. Falls $\lambda^k \geq 0$, dann ist x^k optimal: STOP.

5. Bestimme $l \in \bar{\mathcal{B}}_k$ mit $\lambda_l^k < 0$. Regel von Bland: Bestimme das kleinste solche l .

6. Löse

$$A_{\mathcal{B}_k} d_{\mathcal{B}_k}^k = a^l.$$

und setze

$$d_l^k = -1, \quad d_i^k = 0 \text{ für } i \in \bar{\mathcal{B}}_k \setminus \{l\},$$

7. Falls $d_{\mathcal{B}_k}^k \leq 0$, dann ist f nicht nach unten beschränkt auf \mathfrak{S} : STOP.

8. Setze

$$\tau = \min\left\{\frac{x_{j_i}^k}{d_{j_i}^k} : j_i \in \mathcal{B}_k \text{ und } d_{j_i}^k > 0\right\}.$$

9. Regel von Bland: Wähle s als kleinsten Index, für den das Minimum in der Bestimmung von τ angenommen wird. (Im Nichtentartungsfall ist s eindeutig bestimmt)

10.

$$\mathcal{B}_{k+1} = (\mathcal{B}_k \setminus \{s\}) \cup \{l\} .$$

11.

$$x^{k+1} = x^k - \tau d^k .$$

Wegen der Sätze B.31 und B.30 folgt sofort

Satz B.36. *Unter den Voraussetzungen dieses Abschnittes liefert das beschriebene Verfahren nach endlich vielen Schritten eine Optimallösung, falls eine solche existiert. Ist f nicht nach unten beschränkt, so wird dies ebenfalls nach endlich vielen Schritten festgestellt. Pro Schritt nimmt f im strengen Sinne ab. Im Entartungsfall liefert die Regel von Bland ebenfalls einen finiten Algorithmus, nun nimmt aber f nicht in jedem Schritt streng ab, der Algorithmus kann einige Schritte in einer Ecke stehen bleiben bis ein Indexaustausch zu einer Nachbarecke führt. \square*

Bemerkung B.37. *Die Regel von BLAND, im Entartungsfall jeweils die outgoing- und incoming-Variable mit dem kleinsten Index zu wählen, ist die einfachste "Anticycling"-Regel. Diese Regel führt aber u.U. zu unnötigem Aufwand und auch numerischen Schwierigkeiten. Bezüglich besserer Vorgehensweisen konsultiere man die Spezialliteratur, Stichwort "lexikographische Austauschregel". \square*

Der Rechengang sei am folgenden Beispiel demonstriert

Beispiel B.38. $-30x_1 - 20x_2 - 0x_3 - 0x_4 - 0x_5 \stackrel{!}{=} \min$
 $x_i \geq 0, \quad i = 1, \dots, 5;$

$$\begin{aligned} 5x_1 + x_2 + x_3 &= 60 \\ 3x_1 + 4x_2 + x_4 &= 60 \\ 4x_1 + 3x_2 + x_5 &= 60 \end{aligned}$$

also

$$A = \begin{pmatrix} 5 & 1 & 1 & 0 & 0 \\ 3 & 4 & 0 & 1 & 0 \\ 4 & 3 & 0 & 0 & 1 \end{pmatrix} \quad \text{Rang}(A) = 3 = p.$$

Mit $\mathcal{B}_0 = \{3, 4, 5\}$, $\bar{\mathcal{B}}_0 = \{1, 2\}$ wird

$$\begin{aligned} x^0 &= (0, 0, 60, 60, 60)^T, \\ \mu^0 &= 0 \text{ weil } c_{\mathcal{B}_0} = 0, \\ \lambda_{\bar{\mathcal{B}}_0}^0 &= (-30, -20)^T \end{aligned}$$

Wir wählen $l = 1$. Dies ergibt

$$\begin{aligned} d_{\mathcal{B}_0}^0 &= (5, 3, 4)^T, \\ \tau_0 &= \min\left\{\frac{60}{5}, \frac{60}{3}, \frac{60}{4}\right\} = 12 \\ x^1 &= (12, 0, 0, 24, 12)^T, \\ \mathcal{B}_1 &= \{1, 4, 5\}, \\ \bar{\mathcal{B}}_1 &= \{2, 3\}. \end{aligned}$$

Nun ist

$$\begin{aligned} A &= \begin{pmatrix} 5 & 0 & 0 \\ 3 & 1 & 0 \\ 4 & 0 & 1 \end{pmatrix}, \\ A\mu^1 &= (-30, 0, 0)^T \Rightarrow \mu^1 = (-6, 0, 0), \\ \lambda_{\bar{\mathcal{B}}_1}^1 &= \begin{pmatrix} -20 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 & 4 & 3 \\ 1 & 0 & 0 \end{pmatrix} \mu^1 = \begin{pmatrix} -14 \\ 6 \end{pmatrix} \end{aligned}$$

Es ist nun $l = 2$ und daher

$$\begin{aligned} \begin{pmatrix} 5 & 0 & 0 \\ 3 & 1 & 0 \\ 4 & 0 & 1 \end{pmatrix} d_{\mathcal{B}_1}^1 &= \begin{pmatrix} 1 \\ 4 \\ 3 \end{pmatrix} \\ d_{\mathcal{B}_1}^1 &= \frac{1}{5}(1, 17, 11)^T \\ \tau_1 &= \min\left\{\frac{12}{5}, \frac{24}{17}, \frac{12}{11}\right\} = \frac{60}{11} \\ \mathcal{B}_2 &= \{1, 2, 4\} \\ x^2 &= \frac{1}{11}(120, 60, 0, 60, 0)^T. \end{aligned}$$

Nun ist

$$\begin{aligned} A_{\mathcal{B}_2}^T \mu^2 &= \begin{pmatrix} 5 & 3 & 4 \\ 1 & 4 & 3 \\ 0 & 1 & 0 \end{pmatrix} \mu^2 = \begin{pmatrix} -30 \\ -20 \\ 0 \end{pmatrix}, \\ \mu^2 &= \begin{pmatrix} -\frac{10}{11} \\ 0 \\ -\frac{70}{11} \end{pmatrix} \\ A_{\bar{\mathcal{B}}_2}^T &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ c_{\bar{\mathcal{B}}_2} &= (0, 0)^T \\ \lambda_{\bar{\mathcal{B}}_2}^2 &= \left(\frac{10}{11}, \frac{70}{11}\right)^T \end{aligned}$$

und daher ist x^2 optimal. □

Für die praktische Durchführung des Verfahrens ist es wichtig, daß man die linearen Gleichungen mit einer numerisch stabilen Matrixfaktorisierung (LR-Zerlegung mit Pivotisierung, unter Berücksichtigung einer eventuell vorhandenen Dünnbesetztheit der Matrix) lösen kann. Da von Schritt zu Schritt nur jeweils eine Spalte in der Basismatrix ausgetauscht wird, kann man diese Zerlegungen mit geringem Aufwand ineinander umrechnen. Um eine zu grosse Rundungsfehlerakkumulation zu vermeiden, wird nach einer festen Anzahl solcher Aufdatierungsschritte eine völlig neue Zerlegung berechnet.

Der Algorithmus setzt voraus, daß eine zulässige Ecke bereits bekannt ist. In gewissen Spezialfällen ist es leicht, einen zulässigen Eckpunkt für eine LO-Aufgabe in Standardform zu finden, wenn nämlich die Gleichungsnebenbedingungen alle durch die Einführung von Schlupfvariablen entstanden sind, d.h. $Ax = b$ hat die Form $Ix^1 + A_2x^2 = b$. Hier kann nämlich stets $x^0 = \begin{pmatrix} b \\ 0 \end{pmatrix}$ gewählt werden mit $A = (I, A_2)$, $\mathcal{B}_0 = \{1, \dots, p\}$, wenn man den Schlupfvariablen diese Indizes zuordnet. In anderen Situationen kann die Bestimmung einer Ausgangsecke jedoch sehr schwierig sein, so daß man zu einem geeigneten numerischen Verfahren greifen muß. Als solches erweist sich die Simplexmethode selbst. (Dies ist die sogenannte "Phase I" des Verfahrens. Wir betrachten dazu die Aufgabe

$$\left. \begin{aligned} -\sum_{i=1}^m y_i &= -e^T y \stackrel{!}{=} \max, & e &= (1, \dots, 1)^T \\ Ax + y &= b, & x \geq 0, y \geq 0 & \text{(o.B.d.A. } b \geq 0). \end{aligned} \right\} \quad (\text{B.7})$$

Dies ist eine LO-Aufgabe in der Standardform, für die man sofort eine zulässige Ecke $\begin{pmatrix} x^0 \\ y^0 \end{pmatrix}$, nämlich $\begin{pmatrix} 0 \\ b \end{pmatrix}$, angeben kann. Man löst nun (B.7) mit dem Simplexverfahren. Hierbei sind viele Varianten möglich, z.B. kann man versuchen, auch hier schon die Zielfunktion mit ins Spiel zu bringen, um eine möglichst günstige zulässige Ausgangsecke zu finden. Details siehe in der Spezialliteratur.

B.3.4.2 Ein Algorithmus für das definite quadratische Optimierungsproblem QP

$$f(x) = \frac{1}{2}x^T Ax - b^T x, \quad A = A^T \quad \text{positiv definit}$$

a) Nur Gleichungen:

$$h(x) = H^T x + h^0 = 0$$

Die Multiplikator-Regel ist hinreichend und notwendig für Optimalität. Sie lautet hier

$$\begin{aligned} Ax^* - b - H\mu^* &= 0 \\ H^T x^* + h^0 &= 0. \end{aligned}$$

Sei $x^0 \in \mathbb{R}^n$ beliebig. Dann

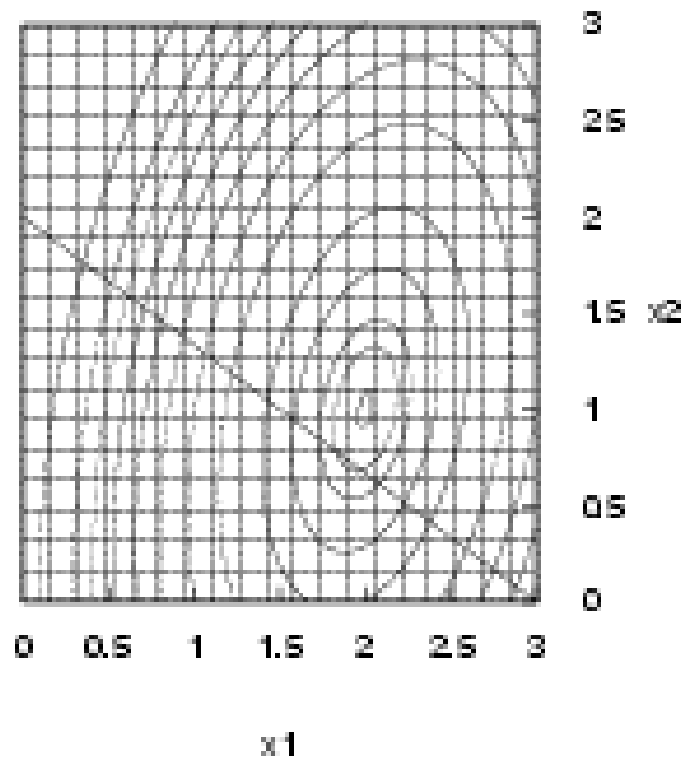
$$\begin{aligned} A \underbrace{(x^* - x^0)}_{-d} - H\mu^* &= b - Ax^0 = -\nabla f(x^0) \\ H^T(x^* - x^0) &= -H^T x^0 - h^0 = -h(x^0) \end{aligned}$$

d.h. bei beliebigem x^0 liefert die Lösung (d, μ^*) **eines** linearen Gleichungssystems

$$\begin{pmatrix} A & H \\ H^T & 0 \end{pmatrix} \begin{pmatrix} d \\ \mu^* \end{pmatrix} = \begin{pmatrix} \nabla f(x^0) \\ h(x^0) \end{pmatrix}$$

die exakte Lösung $x^* = x^0 - d, \mu^*$.

gleichungsrestriktiertes OP



Ist x^0 selbst zulässig, dann kann d gedeutet werden als schiefe Projektion von $\nabla f(x^0)$ auf die Hyperebene parallel zur linearen Mannigfaltigkeit $h(x) = 0$. Ist $A = I$, dann handelt es sich um eine orthogonale Projektion.

b) Gleichungen und Ungleichungen gemischt:

Der Algorithmus beruht auf der iterativen Anwendung von Fall a). Zunächst werden die "aktiven" Ungleichungen ($g_i(x)$ mit $i \in \mathcal{A}(x^k)$) als Gleichungen behandelt und Fall a) angewendet. Ist $d^k \neq 0$, wird $x^{k+1} = x^k - \sigma_k d^k$ gesetzt mit optimalem oder maximal zulässigem σ_k . Ist $d^k = 0$ und $\lambda_{\mathcal{A}}^k \geq 0$, ist x^k optimal, da die Multiplikator-Regel notwendig und hinreichend für Optimalität ist. Ist $\lambda_t^k < 0$ für ein $t \in \mathcal{A}(x^k)$,

setzt man $\mathfrak{B} := \mathcal{A}(x^k) \setminus \{t\}$ "Inaktivierungsschritt" und wendet wieder Fall a) an. Jetzt wird $d^k \neq 0$ und man setzt wieder $x^{k+1} = x^k - \sigma_k d^k$ mit optimalem oder maximal zulässigen σ_k .

Algorithmus:

$x^0 \in \mathfrak{S}$ gegeben.

$k = 0, 1, 2, \dots$:

1. $\mathfrak{B} := \mathcal{A}(x^k)$

2. Löse das lineare Gleichungssystem

$$\begin{pmatrix} A & N_{\mathfrak{B}} \\ N_{\mathfrak{B}}^T & 0 \end{pmatrix} \begin{pmatrix} d^k \\ + \begin{pmatrix} \mu^k \\ \lambda_{\mathfrak{B}}^k \end{pmatrix} \end{pmatrix} = + \begin{pmatrix} \nabla f(x^k) \\ 0 \end{pmatrix}$$

3. Falls $d^k \neq 0$ gehe zu 6.

4. Falls $d^k = 0$ und $\lambda_{\mathfrak{B}}^k \geq 0$: STOP, $x^k = x^*$

5. Falls $d^k = 0$ und $\lambda_t^k < 0$ für irgend ein $t \in \mathfrak{B}$, setze

$$\mathfrak{B} := \mathfrak{B} \setminus \{t\};$$

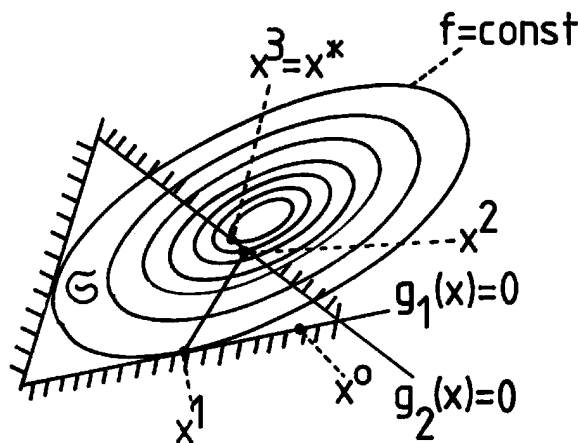
gehe zu 2.

6. $\sigma_k^* = \min\{g_i(x^k)/(d^k)^T \nabla g_i : i \notin \mathcal{A}(x^k) \text{ und } (d^k)^T \nabla g_i > 0\}$ ³

$$\sigma_k = \min\{1, \sigma_k^*\}$$

$$x^{k+1} = x^k - \sigma_k d^k$$

7. Gehe zu 1.



³Konvention: $\min\{\emptyset\} = +\infty$

Satz B.39. Sei $f(x) = \frac{1}{2}x^T Ax - b^T x$ mit $A = A^T$ positiv definit.
 $g(x) = G^T x + g^0$, $h(x) = H^T x + h^0$, $\mathfrak{S} = \{x \in \mathbb{R}^n : g(x) \geq 0, h(x) = 0\} \neq \emptyset$.
 Ferner gelte

$$\left. \begin{array}{l} \text{Für alle } x \in \mathfrak{S} \text{ sei } N_{\mathcal{A}(x)} = (H, G_{\mathcal{A}(x)}) = (H, g^{i_1}, \dots, g^{i_l}) \\ \text{mit } \mathcal{A}(x) = \{i_1, \dots, i_l\} \text{ spaltenregulär, d.h. vom Rang } p+l. \end{array} \right\} \quad (\text{B.8})$$

Dann bestimmt obiger Algorithmus $x^* = \operatorname{argmin} \{f(x) : x \in \mathfrak{S}\}$ in endlich vielen Schritten. \square

NUMAWWW

Bemerkung B.40. Man kann auf die einschränkende Voraussetzung (B.8) verzichten. \mathfrak{B} wird dann aber anders gesteuert (dieser Fall ist wesentlich komplizierter). Man kann auch nur semidefinites A zulassen, solange nur A auf dem Nullraum aller Matrizen $N_{\mathfrak{B}}^T$ positiv definit ist

Eine rechnerische Vereinfachung bietet die Transformation auf den Spezialfall $A = I$:

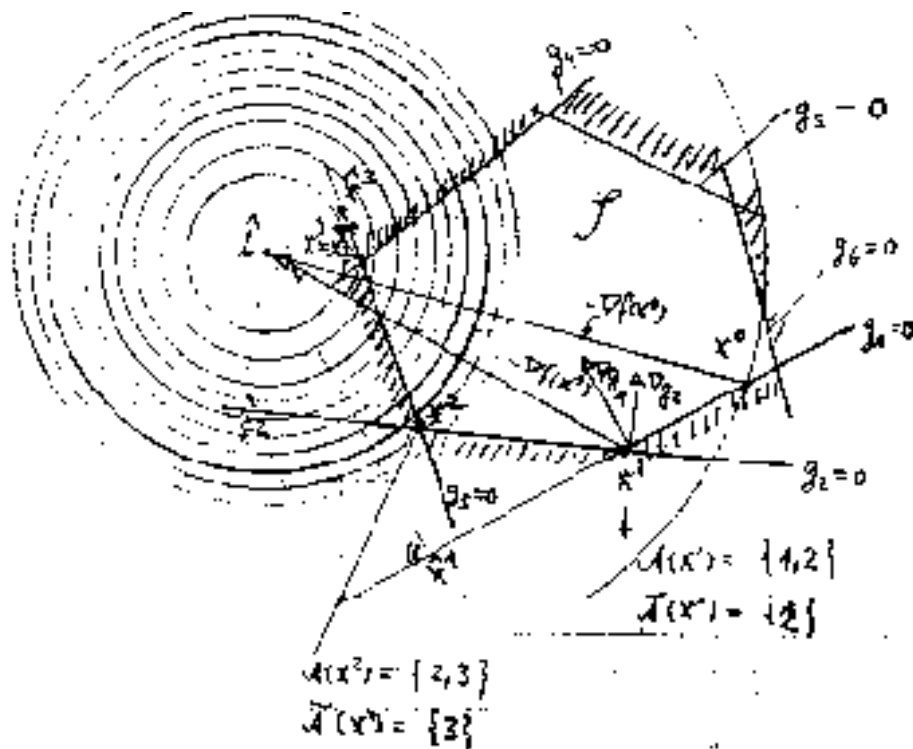
$$\begin{aligned} A &= LL^T && \text{Cholesky-Zerlegung} \\ \hat{x} &:= L^T x &\Leftrightarrow & x = (L^T)^{-1} \hat{x} \\ \hat{b} &:= L^{-1} b \\ \hat{H} &:= L^{-1} H \\ \hat{G} &:= L^{-1} G \\ \hat{f}(\hat{x}) &:= \frac{1}{2} \hat{x}^T \hat{x} - \hat{b}^T \hat{x}, \\ \hat{\mathfrak{S}} &:= \{\hat{x} \in \mathbb{R}^n : \hat{H}^T \hat{x} + h^0 = 0, \hat{G}^T \hat{x} + g^0 \geq 0\} \end{aligned}$$

Gesucht ist also der Punkt von $\hat{\mathfrak{S}}$, der am nächsten bei \hat{b} liegt (im Sinne des euklidischen Abstands). Dieser Fall wird rechnerisch besonders einfach, wenn man dann mit der QR-Zerlegung von $\hat{N}_{\mathfrak{B}} = (\hat{H}, \hat{G}_{\mathfrak{B}})$ arbeitet, die man von Schritt zu Schritt aufdatieren kann:

$$\begin{aligned} \begin{pmatrix} I & \hat{N}_{\mathfrak{B}} \\ \hat{N}_{\mathfrak{B}}^T & 0 \end{pmatrix} \begin{pmatrix} d \\ +(\lambda_{\mathfrak{B}}^{\mu}) \end{pmatrix} &= + \begin{pmatrix} \nabla f \\ 0 \end{pmatrix} &\Leftrightarrow & s = \begin{pmatrix} s^1 \\ s^2 \end{pmatrix} = Q^T d \\ \begin{pmatrix} I & 0 & R \\ 0 & I & 0 \\ R^T & 0 & 0 \end{pmatrix} \begin{pmatrix} s^1 \\ s^2 \\ +(\lambda_{\mathfrak{B}}^{\mu}) \end{pmatrix} &= + \begin{pmatrix} b^1 \\ b^2 \\ 0 \end{pmatrix} && Q \nabla f = b = \begin{pmatrix} b^1 \\ b^2 \end{pmatrix} && \text{also} \\ &&& Q \hat{N}_{\mathfrak{B}} = \begin{pmatrix} R \\ 0 \end{pmatrix} \end{aligned}$$

$$\boxed{s^1 = 0, \quad s^2 = b^2, \quad R(\lambda_{\mathfrak{B}}^{\mu}) = b^1, \quad d = Qs.}$$

Die Partitionierung der Vektoren entspricht dabei der Partitionierung von $Q\hat{N}_{\mathfrak{B}}$.



Man kann auch bei positiv definitem A x mit Hilfe der Lagrangebedingung als Funktion der Multiplikatoren bestimmen und nun die Lagrangefunktion bezüglich der dualen Variablen maximieren. Dies ist der Ansatzpunkt des Verfahrens von Goldfarb und Idnani. Siehe dazu die Spezialliteratur. Hochdimensionale konvexe QP-Probleme kann man mit diesen Matrixfaktorisierungsmethoden nicht behandeln. Hier benutzt man iterative Löser vom cg-Typ, verbunden mit Projektionsmethoden. Siehe auch dazu die Spezialliteratur.

B.3.4.3 Minimierung einer allgemeinen Funktion unter linearen Gleichungs- und Ungleichungsrestriktionen

Der Algorithmus entsteht aus einer Modifikation des Algorithmus aus C3.4.2. Es handelt sich also um eine sogenannte "active set"-Methode. Er ist jetzt nicht mehr notwendig finit. Im Folgenden werden Gleichungssysteme der gleichen Struktur wie in C3.4.2 betrachtet.

Es genügt, daß

$$z^T A z > 0 \quad \text{für alle } z \text{ mit } N_{\mathfrak{B}}^T z = 0,$$

damit $-d$ aus dem folgenden System

$$\begin{pmatrix} A & N_{\mathfrak{B}} \\ N_{\mathfrak{B}}^T & 0 \end{pmatrix} \begin{pmatrix} d \\ +(\mu_{\mathfrak{B}}) \end{pmatrix} = + \begin{pmatrix} \nabla f(x) \\ 0 \end{pmatrix}$$

eine zulässige Abstiegsrichtung für f in x wird, d.h. mit geeigneten $\sigma^*, \sigma^{**} > 0$ ist

$$f(x - \sigma d) < f(x) \quad \text{für } 0 < \sigma \leq \sigma^{**}$$

und

$$x - \sigma d \in \mathfrak{S} \quad \text{für } 0 < \sigma \leq \sigma^*.$$

Für σ^* gilt die gleiche Formel wie in Abschnitt C.3.4.1!

Für den Abstieg von f benutzt man wieder Anfangsschrittweitenschätzung und Abstiegstest aus dem Goldstein-Armijo-Test wie im unrestringierten Fall. Bei iterativer Anwendung kann man z.B. wählen

- $A_k = I$ \mapsto Gradientenprojektionsverfahren von Rosen
- A_k aus BFGS-Formel \mapsto projiziertes BFGS-Verfahren
- $A_k = \nabla^2 f(x^k)$
wenn f gleichmäßig konvex \mapsto projiziertes Newtonverfahren.

Es kommt noch eine Komplikation hinzu: Es wird nun nicht $d^k = 0$ nach stets endlicher Schrittzahl, sondern $d^k \rightarrow 0$, $\lambda_t^k \leq c < 0$ für ein $t \in \mathfrak{B}_k$ und eine unendliche Folge $\{k\}$ ist möglich. Man muß also inaktivieren, auch wenn d^k noch nicht null, aber klein geworden ist im Sinne von

$$0 \leq \nabla f(x^k)^T d^k < (\min_{i \in \mathfrak{B}_k} \lambda_i^k)^2 \cdot c,$$

mit einer geeignet gewählten Konstanten c . Wenn man aber inaktiviert hat, darf man nicht sofort ein zweites Mal inaktivieren, es sei denn es ist $d^k = 0$ in diesem Schritt. So vermeidet man sogenanntes ‘‘Zick-zack-Laufen‘‘ (engl.: ‘‘anti-zig-zag strategy’’). Sonst könnte die Folge der maximal zulässigen Schrittweiten $\{\sigma_k^*\}$ gegen null konvergieren ohne daß $\{x^k\}$ gegen einen Kuhn-Tucker-Punkt konvergiert. Das Resultat dieser Überlegungen ist der nachstehende Algorithmus:

Gegeben $x^0 \in \mathfrak{S}$.

Verfahrensparameter: $0 < \beta < 1$, $0 < \delta < \frac{1}{2}$, $c_1 \ll 1 \ll c_2$, $0 < c_3$.

$\{A_k\}$ eine beschränkte Folge symm. Matrizen mit $z^T A_k z \geq \varrho_1 z^T z$ für alle z mit $N_{\mathfrak{B}_k}^T z = 0$ mit einem geeigneten ϱ_1 (die auch u.U. erst im Laufe des Verfahrens mitkonstruiert wird. ϱ_1 muß nicht explizit bekannt sein).

Setze $\omega_0 = 0$.

$k = 0, 1, 2, \dots$

1. $\mathfrak{B}_k := \mathcal{A}(x^k)$, $\vartheta_k := 0$, $\alpha_k := 0$
2. Löse das lineare System

$$\begin{pmatrix} A_k & N_{\mathfrak{B}_k} \\ N_{\mathfrak{B}_k}^T & 0 \end{pmatrix} \begin{pmatrix} d^k \\ \begin{pmatrix} \mu^k \\ \lambda_{\mathfrak{B}_k}^k \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \nabla f(x^k) \\ 0 \end{pmatrix}$$

$\lambda_i^k := 0$ für $i \notin \mathfrak{B}_k$.

Falls $\alpha_k = 1$ gehe zu Schritt 7.

3. Falls $d^k = 0$ und $\lambda^k \geq 0$, dann $x^k = x^*$: Stop.

4. Bestimme

$$\gamma_k = \min_{1 \leq i \leq m} \lambda_i^k, \quad i_0 \text{ so, daß } \lambda_{i_0}^k = \gamma_k \quad (\text{Index, für den das Minimum angenommen wird.})$$

5. Test, ob eine Inaktivierung lohnend wäre: Falls $\gamma_k \leq -c_3 \sqrt{\nabla f(x^k)^T d^k}$ setze $\vartheta_k := 1$

6. Falls ($\omega_k = 0$ und $\vartheta_k = 1$) oder ($d^k = 0$) und ($\alpha_k = 0$)
(Eine Inaktivierung ist erlaubt und lohnend oder zwingend notwendig und wurde noch nicht ausgeführt:) Setze

$$\mathfrak{B}_k := \mathfrak{B}_k \setminus \{i_0\}, \quad \alpha_k := 1$$

und gehe zu Schritt 2.

7. Setze

$$\hat{\sigma}_k = \begin{cases} 1 & \text{falls } f(x^k) - f(x^k - d^k) \geq \nabla f(x^k)^T d^k \\ \max\left\{c_1, \min\left\{c_2, \frac{\nabla f(x^k)^T d^k}{2(f(x^k - d^k) - f(x^k) + \nabla f(x^k)^T d^k)}\right\}\right\} & \text{sonst.} \end{cases}$$

unrestringierte Anfangsschrittweite

$$\sigma_k^* = \min\{g_i(x^k)/\nabla g_i^T d^k : i \notin \mathcal{A}(x^k), \nabla g_i^T d^k > 0\}$$

maximal zulässige Schrittweite

$$\sigma_{k,0} = \min\{\hat{\sigma}_k, \sigma_k^*\}$$

tatsächliche Anfangsschrittweite

$$\sigma_k = \max\{\beta^j \sigma_{k,0} : j \in \mathbb{N}_0 \text{ und } f(x^k) - f(x^k - \beta^j \sigma_{k,0} d^k) \geq \delta \beta^j \sigma_{k,0} \nabla f(x^k)^T d^k\}$$

Abstiegstest

$$x^{k+1} = x^k - \sigma_k d^k$$

$$8. \quad \omega_{k+1} = \begin{cases} 1 & \text{falls } \mathcal{A}(x^{k+1}) \neq \mathfrak{B}_k \\ 0 & \text{sonst.} \end{cases}$$

Bem.: Zur Bedeutung der drei Steuerparameter: In Schritt k ist $\omega_k = 0$, wenn in diesem Schritt uneingeschränkt inaktiviert werden durfte, sonst $\omega_k = 1$. $\vartheta_k = 1$, wenn in diesem Schritt Inaktivierung lohnend gewesen wäre, sonst $\vartheta_k = 0$. $\alpha_k = 1$, wenn in diesem Schritt tatsächlich inaktiviert wurde, sonst $\alpha_k = 0$. \square

Es gilt der folgende Konvergenzsatz:

Satz B.41. Sei $f \in C^2(\mathcal{D})$, $\mathcal{D} \supset \mathfrak{S}$, $\mathcal{L}_f(f(x^0)) \cap \mathfrak{S}$ kompakt, $N_{\mathcal{A}(x)}$ spaltenregulär für jedes $x \in \mathcal{L}_f(f(x^0)) \cap \mathfrak{S}$. Dann erfüllt jeder Häufungspunkt der vom Algorithmus erzeugten Folge die Bedingungen der Multiplikatorregel. Gibt es nur endlich viele solcher Punkte, dann konvergiert die Gesamtfolge. Erfüllt in diesem Fall der Grenzwert die Bedingung der strikten Komplementarität und die hinreichende Bedingung zweiter Ordnung, und gilt für die gewählten bzw. konstruierten Matrizen A_k die Bedingung

$$\frac{(I - N_{\mathfrak{B}_k}(N_{\mathfrak{B}_k}^T N_{\mathfrak{B}_k})^{-1} N_{\mathfrak{B}_k}^T)(\nabla^2 f(x^*) - A_k)d^k}{\|d^k\|} \rightarrow 0 \quad (\text{B.9})$$

dann gilt $\|x^{k+1} - x^*\| \leq \varepsilon_k \|x^k - x^*\|$ für eine geeignete Nullfolge $\{\varepsilon_k\}$, d.h. es liegt sogar superlineare Konvergenz vor. \square

NUMAWWW

Bemerkung: (B.9) ist eine Zusatzbedingung an A_k und ist z.B. für gleichmäßig konvexes f und $A_k = \nabla^2 f(x^k)$ erfüllt. Sie ist auch erfüllt, wenn f gleichmäßig konvex ist und A_k nach der BFGS-Formel bestimmt wird. Wegen

$$d^k = (I - N_{\mathfrak{B}_k}(N_{\mathfrak{B}_k}^T N_{\mathfrak{B}_k})^{-1} N_{\mathfrak{B}_k}^T)d^k$$

handelt es sich um eine Zusatzbedingung an die auf die Tangentialmannigfaltigkeit an \mathfrak{S} in x^* projizierte quadratische Form $z^T A_k z$. Die Bedingung ist leer, wenn $|\mathcal{A}(x^*)| + p = n$. \square

Bemerkung: Der Algorithmus ist auf nichtlineare Restriktionen übertragbar unter Hinzunahme der schon früher erwähnten Restoration der aktiven Restriktionen. (Dies führt zu den sogenannten grg-Verfahren (generalized reduced gradient)).

NUMAWWW

Im Fall einer linearen Zielfunktion fällt diese auf jeder linearen Randmannigfaltigkeit unbeschränkt, d.h. man kann immer den maximal zulässigen Schritt σ_k^* nehmen. Sobald die aktive Menge die Mächtigkeit n hat, spielt die Wahl von A nur noch für die Skalierung von d^k eine Rolle, ist im Zusammenspiel mit der Schrittweitenwahl also irrelevant bis auf die Definitheit. Man kann sich dann vorstellen, $A_k = \tau I$ genommen zu haben mit einem sehr kleinen positiven τ . Das bedeutet, daß man ein LP-Problem auch als QP-Problem mit einem sehr kleinen quadratischen Term lösen kann. So entsteht dann eine spezielle Variante des dualen Simplexalgorithmus.

Ein wesentlicher Nachteil dieser Methode ist es, daß pro Schritt höchstens eine Restriktion inaktiviert werden kann. Hat man ein Problem mit nur Schrankenrestriktionen, dann kann

man wesentlich effizienter arbeiten, wenn man bei einem Wechsel der aktiven Menge einen sogenannten Gradientenprojektionsschritt einschiebt. Dabei wird f auf der Projektion des Strahls $x - \sigma \nabla f(x)$ auf die zulässige Menge approximativ minimiert. Diese Projektion ist ein stückweise linearer Pfad. Die Projektion berechnet sich aus

$$\mathcal{P}_{\mathfrak{C}}(x) = \hat{x}$$

mit

$$\hat{x}_i = \begin{cases} (x_u)_i & \text{falls } x_i < (x_u)_i \\ x_i & \text{falls } (x_u)_i \leq x_i \leq (x_o)_i \\ (x_o)_i & \text{falls } x_i > (x_o)_i \end{cases} .$$

Man kann linear restringierte auch mit den Vertrauensbereichsmethoden behandeln. Benutzt man dann als Norm die Maximumnorm, dann erhält man spezielle Varianten der SQP-Verfahren.

B.3.5 SLP- und SQP-Verfahren

Grundlage: NLO wird lokal durch ein lineares oder quadratisches Optimierungsproblem approximiert. Aus dieser Approximation erhält man eine Korrekturrichtung d^k . Der Übergang $x^k \mapsto x^{k+1} = x^k + \sigma d^k$ wird wieder durch eine Kontrollfunktion, und zwar eine exakte Penaltyfunktion, gesteuert, indem man verlangt, daß diese Funktion monoton abnimmt. Wegen der leichten Bestimmbarkeit der Parameter ist folgende Funktion dabei besonders bewährt :

$$\Phi(x; \vec{\beta}, \vec{\gamma}) := f(x) - \sum_{i=1}^m \beta_i \min\{0, g_i(x)\} + \sum_{j=1}^p \gamma_j |h_j(x)|.$$

(“gewichtete l_1 -Penalty-Funktion”, Zangwill, Pietrzykowski(1968)) Für diese Funktion ist folgendes beweisbar:

Satz B.42. Seien f, g, h auf der offenen Menge $\mathcal{D} \subset \mathbb{R}^n$ definiert und zweimal stetig differenzierbar. Ferner sei für ein $\tau_0 > 0$

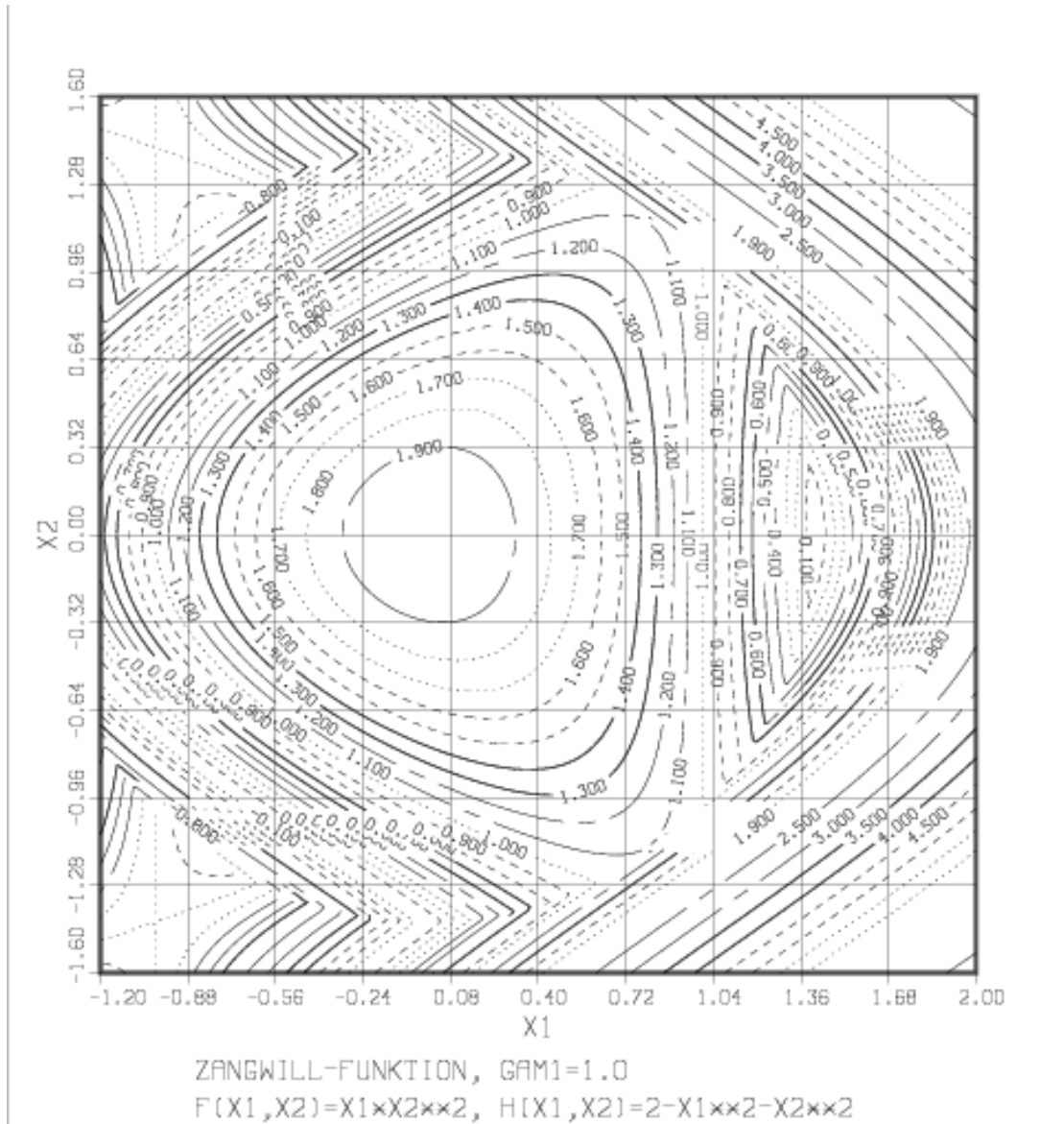
$$\mathfrak{S}(\tau_0) = \{x \in \mathcal{D} : |h_j(x)| \leq \tau_0, j = 1, \dots, p, g_i(x) \geq -\tau_0, i = 1, \dots, m\}$$

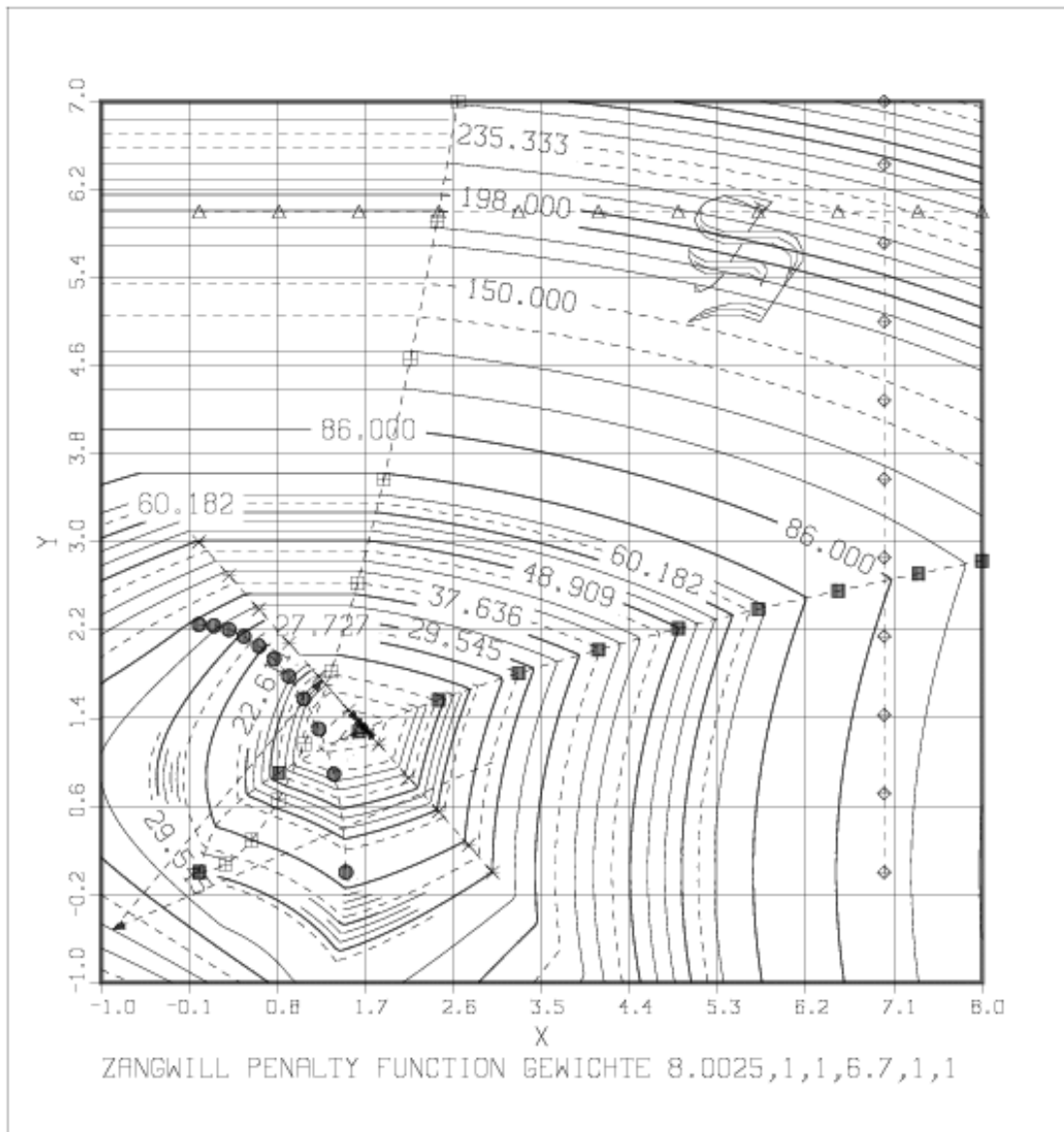
kompakt (insbesondere ist also die zulässige Menge von NLO, $\mathfrak{S}(0)$ kompakt). Für jedes $x \in \mathfrak{S}(\tau_0)$ gelte: $\nabla h(x)$ ist spaltenregulär und es gibt ein $z \in \mathbb{R}^n$ mit

$$\begin{aligned} \nabla h(x)^T z &= 0, & \nabla g_{\mathcal{A} \cup \mathcal{V}}(x)^T z &> 0 \\ \mathcal{A} \cup \mathcal{V} &:= \{i : g_i(x) \leq 0\} \end{aligned}$$

(Dies ist also eine Erweiterung der Mangasarian-Fromowitz-Bedingung auf unzulässige Punkte.) Dann gibt es Gewichte $\beta_i > 0, i = 1, \dots, m, \gamma_j > 0, j = 1, \dots, p$, sodaß jedes strenge restringierte Minimum von f auf \mathfrak{S} auch strenges unrestringiertes Minimum von Φ auf $\mathfrak{S}(\tau_0)$ ist und umgekehrt jedes strenge unrestringierte Minimum von Φ auf $\mathfrak{S}(\tau_0)$ zulässig ($\in \mathfrak{S}$) und strenge lokale Minimalstelle von f auf \mathfrak{S} ist. Gilt in einer solchen Minimalstelle die Regularitätsbedingung und die hinreichende Bedingung zweiter Ordnung, dann genügt es (lokal) $\beta_i > \lambda_i^*$ und $\gamma_j > |\mu_j^*|$ zu wählen. \square

Andere Varianten benutzen $f(x) + \gamma \|h(x), g(x)^-\|_\infty$ oder $f(x) + \gamma \|h(x), g(x)^-\|_2$. Für diese gelten analoge Sätze. Algorithmisch werden die Penaltyparameter lokal adaptiv aus den Werten der Lagrangemultiplikatoren für die Subprobleme bestimmt. Die folgenden Abbildungen zeigen diese exakte nichtdifferenzierbare Penaltyfunktion, einmal für einen gleichungsrestringierten und das andere Mal für einen ungleichungsrestringierten Fall.





Die Abbildung oben zeigt die Penalty-Funktion für das Problem mit

$$\begin{aligned}
 f(x) &= (x_1)^2 + (2.25x_2)^2 \\
 g_1 &= x_1 + x_2 - 3 \\
 g_2 &= 2.25(x_1)^2 + (x_2)^2 - 2.25^2 \\
 g_3 &= (x_1)^2 - x_2 \\
 g_4 &= (x_2)^2 - x_1 \\
 g_5 &= 7 - x_1 \\
 g_6 &= 6 - x_2
 \end{aligned}$$

Die unrestringierte Minimierung von Φ bei "geeignet gewählten", festen Gewichten $\vec{\beta}, \vec{\gamma}$ ist also in gewissem Sinne äquivalent zur Lösung von NLO. Die direkte Minimierung dieser Funktion, etwa unter Einsatz spezieller Methoden der nichtglatten Optimierung, führt nicht zu effizienten Lösungsverfahren. Es zeigt sich aber weiter, daß man durch Ersetzung von NLO durch ein lineares oder quadratisches Programm (lokal) Abstiegsrichtungen für Φ erzeugen kann, die Konvergenz gegen einen Kuhn-Tucker-Punkt von NLO erzwingen. Insbesondere mit quadratischen Approximationen kann man dann schnelle Konvergenz erhalten und in der Tat gehören diese sogenannten SQP (sequential quadratic programming) Methoden zu den effizientesten bekannten Optimierungsmethoden. Bei der Methode der sequentiellen quadratischen Optimierung geht man vor wie folgt:

1. Linearisierung der Restriktionen an der Stelle x^k :

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

definiert die zulässige Menge $\mathfrak{F}(x^k)$ des QP-Problems.

2. Quadratische Approximation von f definiert die Zielfunktion

$$q_k(d) := f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T A_k d.$$

Dabei ist A_k positiv definit.

(Im Idealfall einer konvexen Optimierungsaufgabe ist wählbar

$$A_k = \nabla_{xx}^2 L(x^k, \lambda^{k-1}, \mu^{k-1})$$

Unter Zusatzvoraussetzungen entspricht dann das Verfahren dem Newtonverfahren für die Kuhn-Tucker-Gleichungen) Man kann A_k z.B. auch durch eine BFGS-Aufdatierung der Hessematrix einer erweiterten Lagrangefunktion definieren.

3. d^k ist definiert als Lösung des QP-Problems

$$d^k = \operatorname{argmin} \{q_k(d) : d \in \mathfrak{F}(x^k)\}.$$

λ^k, μ^k seien die Multiplikatoren zu den Ungleichungs- und Gleichungsrestriktionen dieses quadratischen Optimierungsproblems.

4. Anpassung der Gewichte $\vec{\beta}, \vec{\gamma}$: Wähle

$$\vec{\beta} \geq \lambda^k + \varepsilon \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \vec{\gamma} \geq |\mu^k| + \varepsilon \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

und zwar so, daß $\vec{\beta}, \vec{\gamma}$ sich nur endlich oft ändern. (z.B. $\beta_i^k = \lambda_i^k + 2\varepsilon$ falls $\beta_i^k < \lambda_i^k + \varepsilon$. Man muss auch vorsehen, daß die Penaltyparameter wieder verkleinert werden können)

5. Bestimmung der Schrittweite:

Wähle σ_k , sodaß $x^k + \sigma_k d^k \in \mathfrak{S}(\tau_0)$ (s.o.) und

$$\Phi(x^k; \vec{\beta}, \vec{\gamma}) - \Phi(x^k + \sigma_k d^k; \vec{\beta}, \vec{\gamma}) \geq \sigma_k \delta \left((d^k)^T A_k d^k + \varepsilon \psi(x^k) \right)$$

mit $\psi(x) := \sum_{j=1}^p |h_j(x)| - \sum_{i=1}^m \min\{0, g_i(x)\}$.

$0 < \delta < \frac{1}{2}$ fest, $\varepsilon > 0$ fest. (z.B. $\sigma_k \in \{1, \frac{1}{2}, \frac{1}{4}, \dots\}$ maximal)

6. $x^{k+1} = x^k + \sigma_k d^k$.

Hauptprobleme bei dieser Vorgehensweise sind:

1. Im nichtkonvexen Fall ist die zulässige Menge des Subproblems häufig leer, also $\mathfrak{F}(x^k) = \emptyset$, das Verfahren dann nicht definiert, sodaß man eine Modifikation einführen muss. Dies ist insbesondere dann häufig der Fall, wenn die Unzulässigkeit noch groß ist, auch bei sonst gutartigen Problemen.
2. Die Konstruktion von A_k ist kompliziert, wenn das Ausgangsproblem nicht konvex ist und schnelle Konvergenz erreicht werden soll. Zur Erzwingung (langsamer) globaler Konvergenz genügt theoretisch $A_k = I$, aber das zugehörige Verfahren ist inakzeptabel langsam. Wenn man exakte zweite Ableitungen zur Verfügung hat, bietet sich

$$A_k = \nabla_{xx}^2 L(x^k, \lambda^{k-1}, \mu^{k-1})$$

an mit einer eventuell zusätzlichen Regularisierung oder einem Lösungsverfahren für nichtkonvexe QP's.

3. Die Konstruktion $x^{k+1} = x^k + \sigma_k d^k$ mit dem oben angegebenen d^k genügt nicht, um $\sigma_k = 1$ lokal zu gewährleisten, auch wenn man ein konvexes Problem hat und BFGS- oder Hesse-Matrizen benutzt. (letzteres ist notwendig, um schnelle (superlineare) Konvergenz zu erreichen). Man muß dann entweder eine differenzierbare exakte Penalty-Funktion einsetzen oder die Konstruktion abändern zu

$$\begin{aligned} x^{k+1} &= x^k + \sigma_k d^k - (\sigma_k)^2 z^k \\ z^k &:= N_k (N_k^T N_k)^{-1} \begin{pmatrix} h(x^k + d^k) \\ g_{\mathcal{A}}(x^k + d^k) \end{pmatrix} \quad N_k = (\nabla h(x^k), \nabla g_{\mathcal{A}}(x^k)) \end{aligned}$$

mit $\mathcal{A} = \mathcal{A}(x^k)$. Dies ist bekannt als Vermeidung des "Maratos-Effekt"-s durch "second order correction".

Schwierigstes Problem ist 1. Verschiedene Abhilfen sind bekannt, z.B. Einführung sogenannter Schlupfvariablen $u, v, w \geq 0$ (die man möglichst klein machen will und die die Verträglichkeit der Restriktionen stets sicherstellen)

Schritt 1.

$$\begin{aligned} g(x^k) + \nabla g(x^k)^T d + u &\geq 0 \\ h(x^k) + \nabla h(x^k)^T d + v &\geq 0 \\ -h(x^k) - \nabla h(x^k)^T d + w &\geq 0 \\ u \geq 0, \quad v \geq 0, \quad w &\geq 0 \end{aligned}$$

definiert die zulässige Menge $\mathfrak{F}(x^k)$. (Sie ist nie leer).

Schritt 2:

$$\begin{aligned} q_k(d, u, v, w) &= f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T A_k d + \alpha(u^T e + v^T e + w^T e) \\ &\quad + \beta(u^T u + v^T v + w^T w) \end{aligned}$$

mit β in der Größenordnung von $\|A_k\|$ und $\alpha \gg \beta$ definiert die Zielfunktion. $e = (1, \dots, 1)^T$ in passender Dimension.

q_k wird bzgl. d, u, v, w minimiert. Bei hinreichend großem α werden u, v, w zu null, wenn die ursprüngliche zulässige Menge des QP nicht leer war. In dieser Form ist das Verfahren unter den Voraussetzungen von Satz B.42 global konvergent auf $\mathfrak{S}(\tau_0)$.

NUMAWWW

Für hochdimensionale Probleme löst man das QP-Subproblem zweckmässig mit einer inneren-Punkte-Methode.

Bei der **sequentiellen linearen Optimierung (SLP)** wird f nur linear approximiert; also

$$l(d, u, v, w) = f(x^k) + \nabla f(x^k)^T d + \alpha(u^T e + v^T e + w^T e) \stackrel{!}{=} \min_{d, u, v, w \in \mathfrak{F}_k}$$

$\alpha \gg 1$

wird auf $\mathfrak{F}(x^k)$ minimiert, wobei noch für die Größe von d

Schranken eingeführt werden, um zu erreichen, daß l in jedem Fall nach unten beschränkt ist. Es werden wieder Schlupfvariablen eingeführt, um die linearisierten Restriktionen in jedem Fall verträglich zu machen.

$$\left. \begin{aligned} -\varrho \leq d_i \leq \varrho \quad & i = 1, \dots, n \\ g(x^k) + \nabla g(x^k)^T d + u &\geq 0 \\ h(x^k) + \nabla h(x^k)^T d + v &\geq 0 \\ -h(x^k) - \nabla h(x^k)^T d + w &\geq 0 \\ u \geq 0, \quad v \geq 0, \quad w &\geq 0 \end{aligned} \right\} \mathfrak{F}(x^k)$$

Dies ergibt ein lineares Optimierungsproblem zur Bestimmung von d^k . Für hinreichend großes α, ρ ist d wieder eine Abstiegsrichtung für Φ . Da es sehr leistungsfähige Verfahrensvarianten für lineare Optimierungsprobleme hoher Dimension gibt, bietet sich ein Zugang zur Lösung nichtlinearer Optimierungsprobleme hoher Dimension auf diesem Weg an. Allerdings ist die Konvergenzgeschwindigkeit der Methode in diesem Fall immer nur R-linear, d.h. es gibt ein $c_1 > 0$ und $0 < c_2 < 1$, sodaß $\|x^k - x^*\| \leq c_1(c_2)^k$ (unter den Voraussetzungen von Satz B.42).

Man kennt auch Varianten der Vorgehensweise, bei denen man auf eine Abstiegskontrollfunktion verzichtet und verlangt, daß das Paar $(f(x^{k+1}), \psi(x^{k+1}))$ mit

$$\psi(x) = \|(h(x), g(x)^-)\|$$

im Sinne der Halbordnung des \mathbb{R}^2 genügend unterhalb eines achsenparallelen Streckenzuges liegt, der alle Werte $(f(x^j), \psi(x^j))$, $j \leq k$ die im Sinne dieser Halbordnung nicht vergleichbar sind, einhüllt (die sogenannten "Filtermethoden".) Dies läuft auf die Verwendung einer Abstiegskontrollfunktion hinaus, deren Gewichtung sich in jedem Schritt ändern darf. Anstelle von Schrittweitenverfahren benutzt man häufig auch Vertrauensbereichvarianten zur Festlegung des Schrittes. Viele Details findet man nur in der neueren Zeitschriftenliteratur.

B.4 Semidefinite Optimierung

Dies ist ein noch sehr junges Gebiet, hervorgegangen aus der Aufgabenstellung der Eigenwertmaximierung einer symmetrischen Matrix. Es hat bedeutende Anwendungen in den Ingenieurwissenschaften, aber auch in der kombinatorischen Optimierung.

Literatur:

LIEVEN VANDENBERGHE, STEPHEN BOYD: Semidefinite Programming. SIAM Review 38, (1996), 49-95.

WOLKOWICZ, HENRY (ED.); SAIGAL, ROMESH (ED.); VANDENBERGHE, LIEVEN (ED.): Handbook of semidefinite programming. Theory, algorithms, and applications. Dordrecht: Kluwer Academic Publishers. (1999)

Definition B.43. A, B seien reell symmetrische Matrizen. Dann ist definiert

$$A \succcurlyeq B \iff A - B \text{ positiv semidefinit.}$$

□

Bemerkung B.44. Mit der Halbordnung \succcurlyeq wird die Menge der positiv semidefiniten Matrizen zu einem konvexen Kegel, d.h. mit jedem Element ist auch jedes positive Vielfache in dieser Menge enthalten und die Menge ist konvex. Dieser Kegel ist jedoch nicht polyedrisch.

Ein semidefinites Programm hat folgende Form:

$$\left. \begin{array}{l} \text{Minimiere } c^T x = f(x) \\ \text{unter der N.B. } F(x) \succcurlyeq 0 \text{ mit } F(x) = F_0 + \sum_{i=1}^n x_i F_i \end{array} \right\} \quad (\text{SDP})$$

Dabei sind die F_i gegebene symmetrische Matrizen der Dimension $m \times m$. Dies ist offensichtlich eine konvexe Optimierungsaufgabe und die allgemeine Theorie konvexer Aufgaben ist anwendbar.

(SDP) umfaßt viele andere Optimierungsaufgaben, die sich so umformulieren lassen:

1. LP mit der Nebenbedingung

$$\begin{aligned} Ax \geq b &\iff \sum_{i=1}^n x_i a^i - b \geq 0 \\ &\iff -\text{diag}(b_j) + \sum_{i=1}^n x_i \text{diag}(a_1^i, \dots, a_n^i) \succcurlyeq 0 \end{aligned}$$

wo $A = (a^1, \dots, a^n)$. Hier ist also $m = n$.

2. quadratisch restringierte QP's:

$$f_0(x) = (A_0x - b_0)^T(A_0x + b_0) = \min$$

$$f_i(x) = (A_ix + b_i)^T(A_ix + b_i) - d_i^T x - \gamma_i \leq 0 \quad i = 1, \dots, m$$

\iff

min t

$$\begin{pmatrix} I & A_0x + b_0 \\ (A_0x + b_0)^T & t \end{pmatrix} \succcurlyeq 0$$

$$\begin{pmatrix} I & A_ix + b_i \\ (A_ix + b_i)^T & d_i^T x + \gamma_i \end{pmatrix} \succcurlyeq 0 \quad i = 1, \dots, m, \tag{\Delta}$$

was offensichtlich wieder die Form (SDP) hat, wenn man (Δ) als Diagonalblöcke einer einzigen $(mn) \times (mn)$ -Matrix interpretiert.

3. Minimierung des maximalen Eigenwerts einer symmetrischen Matrix:

$$\min t, \quad tI - A(x) \succcurlyeq 0$$

wenn $A(x)$ linear von x abhängt. Probleme dieses Typs treten in der Kontrolltheorie, in der Strukturmechanik und auch in der kombinatorischen Optimierung auf.

4. Norm-Minimierung

$$\|A(x)\|_2 = \min_x \iff$$

$$t \stackrel{!}{=} \min$$

$$\begin{pmatrix} tI & A(x) \\ A(x)^T & tI \end{pmatrix} \succcurlyeq 0$$

($A(x)$ muss linear von x abhängen.)

5. Strukturoptimierung eines Tragwerkes

- f Knotenkräfte
- d Knotenverschiebungen
- l_i Balkenlänge, x_i Querschnitte (gesucht)

Optimaler Entwurf

$$f^T d = \min$$

$$f = A(x)d$$

$$\sum_{i=1}^n l_i x_i \leq v$$

$$(x_u)_i \leq x_i \leq (x_o)_i$$

Steifigkeitsmatrix $A(x)$ hängt linear von den Querschnitten ab.

\iff

min t mit

$$\begin{aligned} \begin{pmatrix} t & f^T \\ f & A(x) \end{pmatrix} \succcurlyeq 0 \\ \sum_{i=1}^n l_i x_i \leq v \\ (x_u)_i \leq x_i \leq (x_o)_i . \end{aligned}$$

6. Mustererkennung

Gegeben: Zwei Punktmenge $\{x^1, \dots, x^k\}$, $\{y^1, \dots, y^l\} \subset \mathbb{R}^d$.

Diese Punktmenge sollen durch eine quadratische Fläche getrennt werden. D.h. gesucht ist

$$f(x) = x^T A x + b^T x + \gamma$$

mit

$$\left. \begin{aligned} f(x^i) &\leq 0 & i = 1, \dots, k \\ f(y^j) &\geq 0 & j = 1, \dots, l \end{aligned} \right\} \oplus$$

Mit der Forderung, f solle ein möglichst sphärisches Ellipsoid sein, gelangen wir zum Problem

$$\begin{aligned} \lambda &= \min \\ \lambda I &\succcurlyeq A \succcurlyeq I \end{aligned}$$

und \oplus . (Die Unbekannten sind A, b, γ, λ , also liegt wieder solch ein konvexes (SDP) vor).

Man könnte dieses Problem mit klassischen inneren-Punkte-Methoden behandeln. Z.B. ist

$$\phi(x) = -\ln \det(F(x))$$

eine geeignete Barrierefunktion für die Semidefinitheitsrestriktion $F(x) \succcurlyeq 0$. Denn

$$\det(F(x)) = \prod_{i=1}^n \lambda_i \quad \text{mit den Eigenwerten } \lambda_i \text{ von } F(x)$$

Die partiellen Ableitungen sind (man beachte $F(x) = F_0 + \sum_{i=1}^n x_i F_i$)

$$\begin{aligned} \frac{\partial}{\partial x_i} \phi(x) &= -\text{spur} (F_i \cdot (F(x))^{-1}) \\ \frac{\partial^2}{\partial x_i \partial x_j} \phi(x) &= \text{spur} (F_i \cdot (F(x))^{-1} \cdot F_j \cdot (F(x))^{-1}). \end{aligned}$$

Eine Beweisskizze für die erste Formel ist die folgende: Wir benutzen die Definition der partiellen Ableitung

$$\frac{\partial}{\partial x_i} \phi(x) = \lim_{\tau \rightarrow 0} \frac{\phi(x + \tau e^i) - \phi(x)}{\tau}.$$

Hierin ist e^i der i -te Koordinateneinheitsvektor. Wir berechnen nun zunächst den Zähler des Bruches rechts:

$$\begin{aligned} & -\ln(\det(F(x + \tau e^i))) + \ln(\det(F(x))) \\ = & -\ln\left(\frac{\det(F(x + \tau e^i))}{\det(F(x))}\right) \\ = & -\ln\left(\frac{\det((I + \tau F_i(F(x))^{-1})F(x))}{\det(F(x))}\right) \\ = & -\ln(\det(I + \tau F_i(F(x))^{-1})) \quad \text{nach dem Produktsatz für Determinanten} \\ = & -\sum_{i=1}^m \ln(\lambda_i(I + \tau F_i(F(x))^{-1})) \quad \text{Determinante = Produkt der Eigenwerte} \\ = & -\sum_{i=1}^m \ln(1 + \lambda_i(\tau F_i(F(x))^{-1})) \\ = & -\sum_{i=1}^m \tau \lambda_i(F_i(F(x))^{-1}) + \mathcal{O}(\tau^2) \end{aligned}$$

weil

$$\ln(1 + x) = x - x^2/2 + x^3/3 \dots$$

und Einsetzen in die Grenzwertgleichung liefert die Behauptung. Die zweite Formel beweist man mit dem gleichen Trick, nun angewandt auf die Formel für eine partielle Ableitung.

Die direkte Anwendung dieser Formeln, etwa im Zusammenhang mit den primalen inneren-Punkte-Verfahren erfordert die Lösung einer großen Anzahl $(n(n+1)/2 + n)n$ von Gleichungssystemen mit der Matrix $F(x)$, was sehr aufwendig ist. Dies läßt sich jedoch vermeiden, wenn man die Sattelpunkteigenschaften der konvexen Optimierung zum Einsatz bringt. Der Restriktion $F(x) \succcurlyeq 0$ entspricht im Kegel der positiv semidefiniten Matrizen ein LAGRANGE-Multiplikator $\Lambda \succcurlyeq 0$ und das Skalarprodukt ist

$$A \bullet B = \sum_{i,j=1}^m a_{ij} b_{ij} = \text{spur}(A^T B).$$

Das duale Problem erhält man aus

$$L(x, \Lambda) = c^T x - \sum_{i,j} \lambda_{ij} (F_{0,ij} + \sum_{k=1}^n x_k F_{k,ij}) \stackrel{!}{=} \max_{\Lambda \succcurlyeq 0}$$

mit der Nebenbedingung

$$\frac{\partial}{\partial x_k} L(x, \Lambda) = 0, \quad \text{also} \quad c_k - \Lambda \bullet F_k = 0 \quad k = 1, \dots, n$$

d.h. es lautet

$$\begin{aligned} -\Lambda \bullet F_0 &= \max \\ \Lambda \bullet F_k &= c_k \quad k = 1, \dots, n \\ \Lambda &\succcurlyeq 0. \end{aligned}$$

Benutzt man die Gleichungsrestriktionen zu einer Teilelimination der Unbekannten, dann erhält dieses Problem die gleiche Struktur wie das primale. Dies soll hier aber nicht weiter verfolgt werden. Wir wollen vielmehr den Sattelpunktsatz der konvexen Optimierung anwenden, um zu primal-dualen Methoden zu gelangen.

Sind x und Λ primal bzw. dual zulässig, dann gilt auch hier

$$c^T x + \Lambda \bullet F_0 = \Lambda \bullet \left(\sum_{k=1}^n x_k F_k + F_0 \right) \geq 0. \quad (\text{B.10})$$

(Weil ein Produkt positiv semidefiniter Matrizen wieder semidefinit ist, obwohl es nicht symmetrisch ist im allgemeinen!

Beweis:

$$\begin{aligned} (B + \varepsilon I)^{1/2} (AB) (B + \varepsilon I)^{-1/2} &= (B + \varepsilon I)^{1/2} A (B + \varepsilon I - \varepsilon I) (B + \varepsilon I)^{-1/2} \\ &= (B + \varepsilon I)^{1/2} A (B + \varepsilon I)^{1/2} - \varepsilon (B + \varepsilon I)^{1/2} A (B + \varepsilon I)^{-1/2} \\ &\sim AB, \quad \varepsilon \rightarrow 0. \end{aligned}$$

Die Dualitätslücke ist eine **lineare** Funktion in x und Λ , siehe linke Seite von (B.10).

In analoger Weise liefern primal bzw. dual zulässige Werte \bar{x} bzw. $\tilde{\Lambda}$ obere bzw. untere Schranken für einen Optimalwert. Um einen Dualitätssatz zu beweisen, benötigt man hier die SLATER-Bedingung sowohl für das primale wie für das duale Problem:

Die Optimalmengen können hier leer sein, obwohl Infimum/Supremum existieren (der Kegel ist ja nichtpolyedrisch).

Beispiel:

$$\begin{aligned} t &= \min \\ &\begin{pmatrix} x & 1 \\ 1 & t \end{pmatrix} \succcurlyeq 0 \\ \iff \\ t &= \min, \quad t \geq 0, \quad xt \geq 1 \\ \inf_t &= 0 \end{aligned}$$

□

Satz B.45. *Es gilt*

$$\inf\{c^T x : F(x) \succcurlyeq 0\} = \sup\{-\Lambda \bullet F_0 : \Lambda \succcurlyeq 0, \Lambda \bullet F_k = c_k, k = 1, \dots, n\}$$

falls

(a) $\exists x : F(x) \succ 0$ *oder*

(b) $\exists \Lambda : \Lambda \succ 0, \Lambda \bullet F_k = c_k, k = 1, \dots, n.$

Gilt sowohl (a) als auch (b), dann werden inf und sup beide angenommen.

Zum Beweis siehe NESTEROV und NEMIROVSKII § 4.2.

□

Beispiele:

(a) Das folgende Beispiel zeigt, daß man auf die SLATER-Bedingung nicht verzichten kann:
 $n = 2, m = 3$

min x_1 mit

$$\begin{pmatrix} 0 & x_1 & 0 \\ x_1 & x_2 & 0 \\ 0 & 0 & x_1 + 1 \end{pmatrix} \succcurlyeq 0.$$

Dieses Problem hat die zulässige Menge $x_1 = 0, x_2 \geq 0$, also 0 als Optimalwert. Das duale Problem lautet

$$\begin{aligned} - \begin{pmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{12} & \lambda_{22} & \lambda_{23} \\ \lambda_{13} & \lambda_{23} & \lambda_{33} \end{pmatrix} \bullet \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} &\stackrel{!}{=} \max \\ \Lambda \bullet \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} &= 1 \\ \Lambda \bullet \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} &= 0 \\ \Lambda &\succcurlyeq 0. \end{aligned}$$

Dies wird zu

$$\begin{aligned}
 -\lambda_{33} &= \max \\
 \lambda_{22} &= 0 \\
 2\lambda_{12} + \lambda_{33} &= 1 \\
 \lambda_{11} &\geq 0 \\
 -\lambda_{12}^2 &\geq 0 \quad \Rightarrow \quad \lambda_{12} = 0, \quad \Rightarrow \quad \lambda_{33} = -1 \\
 -\lambda_{23}^2 &\geq 0 \quad \Rightarrow \quad \lambda_{23} = 0 \\
 \lambda_{11} - \lambda_{13}^2 &\geq 0
 \end{aligned}$$

mit dem Optimalwert -1.

(b) Das Normminimierungsproblem:

$$\|A(x)\|_2 = \min$$

mit

$$A(x) = A_0 + \sum_{i=1}^{\tilde{n}} x_i A_i \quad A_i \in \mathbb{R}^{p \times q}$$

lautet als semidefinites Programm

min t mit

$$\begin{pmatrix} tI & A(x) \\ A(x)^T & tI \end{pmatrix} \succcurlyeq 0.$$

Es ist jetzt $m = p + q$ und $n = \tilde{n} + 1$. Es ist strikt zulässig mit $x = 0$ und $t > \|A_0\|$.

Das duale Programm ist

$$\begin{aligned}
 - \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}^T & \Lambda_{22} \end{pmatrix} \bullet \begin{pmatrix} 0 & A_0 \\ A_0^T & 0 \end{pmatrix} &= \max & \iff & -2 \operatorname{spur}(A_0^T \Lambda_{12}) = \max \\
 \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}^T & \Lambda_{22} \end{pmatrix} \bullet \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} &= 1 & \iff & \operatorname{spur}(\Lambda_{11}) + \operatorname{spur}(\Lambda_{22}) = 1 \\
 \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}^T & \Lambda_{22} \end{pmatrix} \bullet \begin{pmatrix} 0 & A_i \\ A_i^T & 0 \end{pmatrix} &= 0 & \iff & \operatorname{spur}(A_i^T \Lambda_{12}) = 0 \\
 & & & i = 1, \dots, n \\
 \Lambda \succcurlyeq 0, \quad \Lambda &= \Lambda^T.
 \end{aligned}$$

Dies läßt sich noch vereinfachen zu einem Problem in Λ_{12} allein:

$$\begin{aligned}
 -2 \operatorname{spur}(A_0^T \Lambda_{12}) &= \max \\
 \operatorname{spur}(A_i^T \Lambda_{12}) &= 0 \\
 \|\Lambda_{12}\|_F &\leq \frac{1}{2}
 \end{aligned}$$

welches natürlich strikt zulässig ist mit $\Lambda_{12} = 0$

□

Für das semidefinite Problem sind mehrere primal-duale Verfahren vorgeschlagen worden. Das zur Zeit wohl beste ist das von ALIZADEH, HAEBERLY und OVERTON:

{ ALIZADEH,F.; HAEBERLY, J.P.A.; OVERTON,M.L.: Primal-dual interior-point methods for semidefinite Programming: convergence rates, stability and numerical results. SIOPT 8, (1998), 746-768,siehe auch MING GU: primal-dual interior-point methods for semidefinite programming in finite precision. SIOPT 10, (2000), 462-502 }

Das Verfahren geht aus von den Optimalitätsbedingungen mit

$$Z \hat{=} F(x)$$

$$Z = F_0 + \sum_{i=1}^n x_i F_i \tag{B.11}$$

$$\Lambda \bullet Z = 0 \quad (\text{Dualitätslücke} = 0) \tag{B.12}$$

$$\Lambda \bullet F_i = c_i \quad i = 1, \dots, n \tag{B.13}$$

$$\Lambda = \Lambda^T \succcurlyeq 0, \quad Z = Z^T \succcurlyeq 0. \tag{B.14}$$

Wegen (B.14) und (B.12) gilt auch

$$\Lambda Z = 0 \quad \text{Komplementarität} \tag{B.15}$$

Dazu beachte man, daß wegen $\Lambda = \Lambda^T$ und $Z = Z^T$ beide positiv semidefinit auch ΛZ positiv semidefinit ist und wegen $\text{spur}(\Lambda^T Z) = 0$ alle Eigenwerte =0 hat. (B.11, B.13, B.15) stellen ein System von $n(n+1) + n$ Gleichungen für die $n(n+1) + n$ Unbekannten Λ, Z, x dar (man beachte die Symmetrie von Λ und Z). Der zentrale Pfad (die Homotopiekurve) zu diesem System ist definiert durch

$$Z = F_0 + \sum_{i=1}^n x_i F_i$$

$$\Lambda \bullet F_i = c_i, \quad i = 1, \dots, n$$

$$\Lambda Z = \varepsilon I, \quad \varepsilon > 0$$

$$\Lambda = \Lambda^T \succcurlyeq 0, \quad Z = Z^T \succcurlyeq 0.$$

Im Prinzip möchte man die Gleichungen für den zentralen Pfad wieder mit $\varepsilon \rightarrow 0$ lösen, und zwar mit dem NEWTON-Verfahren. Dabei tritt das Problem auf, daß die NEWTON-Korrektur nicht notwendig symmetrisch ist, man aber mit symmetrischen Matrizen arbeiten muß. Von ZHANG stammt die Idee, einen Symmetrisierungsoperator einzuführen und statt

$$\Lambda Z = \varepsilon I$$

$$H_P(\Lambda Z) = \varepsilon I$$

zu betrachten mit einer regulären Matrix P und

$$H_P(\Lambda Z) \stackrel{\text{def}}{=} \frac{1}{2}(P\Lambda ZP^{-1} + (P\Lambda ZP^{-1})^T).$$

$P = I$ ergibt die ALIZADEH-HABERLY-OVERTON-Richtung AHO. Um das NEWTON-System für die obigen Gleichungen aufzustellen, benötigt man die einzelnen partiellen Ableitungen. Um eine kurze übersichtliche Schreibweise zu erhalten, bedient man sich dabei zweckmäßig einiger Matrixhilfsoperationen:

$$H \in \mathbb{R}_{\text{symm}}^{n \times n} \rightarrow \text{svec}(H) = \begin{pmatrix} h_{11} \\ \sqrt{2}h_{12} \\ \vdots \\ \sqrt{2}h_{1n} \\ h_{22} \\ \vdots \\ \sqrt{2}h_{2n} \\ \vdots \\ h_{nn} \end{pmatrix} \in \mathbb{R}^{n(n+1)/2}$$

$$v \in \mathbb{R}^{n(n+1)/2} \rightarrow \text{smat}(v) = \begin{pmatrix} v_1 & \frac{1}{\sqrt{2}}v_2 & \cdots & \frac{1}{\sqrt{2}}v_n \\ \frac{1}{\sqrt{2}}v_2 & v_{n+1} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{2}}v_n & \cdots & \cdots & v_{n(n+1)/2} \end{pmatrix} \quad \text{symmetrisch}$$

(Es gilt $\text{smat}(\text{svec}(H)) = H$.) Man führt ein symmetrisiertes Kroneckerprodukt ein:

$$G, K \in \mathbb{R}^{n \times n} \rightarrow G \otimes_s K \in \mathbb{R}_{\text{symm}}^{(n(n+1))/2}$$

durch

$$(G \otimes_s K) \text{svec}(H) = \frac{1}{2} \text{svec}(KHG^T + GHK^T) \quad \text{für } H \in \mathbb{R}_{\text{symm}}^{n \times n}$$

(d.h. die Matrix wird definiert durch ihre Wirkung auf einen beliebigen Vektor $\text{svec}(H)$). Damit lautet das NEWTON-System (Aufstellung als Übung)

$$\begin{pmatrix} Z_k & L_k & 0 \\ 0 & I_{m(m+1)/2} & \mathcal{A}^T \\ \mathcal{A} & 0 & 0 \end{pmatrix} \begin{pmatrix} \text{svec}(\Delta \Lambda_k) \\ \text{svec}(\Delta Z_k) \\ \Delta x^k \end{pmatrix} = \begin{pmatrix} \text{svec}(\sigma_k \varepsilon_k I_n - \frac{1}{2}(Z_k \Lambda_k + \Lambda_k Z_k)) \\ \text{svec}(Z_k - F_0 - \sum_{i=1}^n x_i^k F_i) \\ c - \mathcal{A} \text{svec}(\Lambda_k) \end{pmatrix}$$

$$\stackrel{\text{def}}{=} \begin{pmatrix} r_{\text{compl}}^k \\ r_{\text{dual}}^k \\ r_{\text{primal}}^k \end{pmatrix}$$

weil

$$\Lambda \bullet F_i = \text{svec}(\Lambda)^T \text{svec}(F_i).$$

Dabei gilt

$$\mathcal{A} = \begin{pmatrix} \text{svec} (F_1)^T \\ \vdots \\ \text{svec} (F_n)^T \end{pmatrix}$$

$$\mathcal{Z}_k = Z_k \otimes_s I_m$$

$$\mathbf{L}_k = \Lambda_k \otimes_s I_m$$

$$\varepsilon_k = (Z_k \bullet \Lambda_k)/n, \quad 0 < \sigma_k < 1.$$

Dieses System läßt sich wie im LP-Fall reduzieren:

$$\begin{pmatrix} \mathcal{Z}_k & \mathbf{L}_k & O \\ O & \mathcal{I} & \mathcal{A}^T \\ \mathcal{A} & O & O \end{pmatrix} = \begin{pmatrix} \mathcal{I} & O & O \\ O & \mathcal{I} & O \\ \mathcal{A}\mathcal{Z}_k^{-1} & -\mathcal{A}\mathcal{Z}_k^{-1}\mathbf{L}_k & \mathcal{I} \end{pmatrix} \begin{pmatrix} \mathcal{Z}_k & \mathbf{L}_k & O \\ O & \mathcal{I} & \mathcal{A}^T \\ O & O & C_{33} \end{pmatrix}$$

mit

$$\mathcal{I} = I_{m(m+1)/2}, \quad C_{33} = \mathcal{A}\mathcal{Z}_k^{-1}\mathbf{L}_k\mathcal{A}^T \quad n \times n.$$

Danach hat man dann

$$\begin{pmatrix} \mathcal{Z}_k & \mathbf{L}_k & O \\ O & I & \mathcal{A}^T \\ O & O & C_{33} \end{pmatrix} \begin{pmatrix} \text{svec} (\Delta\Lambda_k) \\ \text{svec} (\Delta Z_k) \\ \Delta x^k \end{pmatrix} = \begin{pmatrix} r_{\text{compl}}^k \\ r_{\text{dual}}^k \\ r_{\text{primal}}^k + \mathcal{A}\mathcal{Z}_k^{-1}(\mathbf{L}_k r_{\text{dual}}^k - r_{\text{compl}}^k) \end{pmatrix}.$$

Um C_{33} und die rechte Seite zu berechnen, genügt es, Systeme der Form

$$\mathcal{Z}_k v = u$$

zu lösen. Wegen

$$\mathcal{Z}_k = Z_k \otimes_s I_m$$

also

$$(Z_k \otimes_s I_n) \text{svec} (V_i) = \frac{1}{2}(V_i Z_k + Z_k V_i)$$

reduziert sich dies auf

$$V_i Z_k + Z_k V_i = 2F_i \quad i = 1, \dots, n, \quad \rightarrow V_i \in \mathbb{R}_{\text{symm}}^{m \times m}.$$

Eine Gleichung dieses Typs heißt eine LYAPUNOV-(Matrix-)Gleichung. Mit der vollständigen Spektralzerlegung von Z_k

$$Z_k = U_k D_k U_k^T$$

wird dies zu

$$U_k^T V_i U_k D_k + D_k U_k^T V_i U_k = 2U_k^T F_i U_k \stackrel{\text{def}}{=} 2\bar{F}_i$$

oder

$$(\bar{V}_i)_{jl} = 2 \left(\frac{(\bar{F}_i)_{jl}}{(D_k)_{jj} + (D_k)_{ll}} \right)$$

wobei

$$\bar{V}_i = U_k^T V_i U_k, \quad \bar{F}_i = U_k^T F_i U_k.$$

An dieser Darstellung erkennt man auch die Regularität von Z_k für positiv definites Z_k . Analog verfährt man bei der Berechnung der rechten Seite. Das System für Δx^k löst man direkt:

$$C_{33}\Delta x^k = r_{\text{primal}}^k + \mathcal{A}Z_k^{-1}(\mathbb{L}_k r_{\text{dual}}^k - r_{\text{compl}}^k)$$

und dann

$$\begin{aligned} \Delta Z_k &= \text{smat} (r_{\text{dual}}^k - \mathcal{A}^T \Delta x^k) \\ \Delta \Lambda_k &= \text{smat} (Z_k^{-1}(r_{\text{compl}}^k - \mathbb{L}_k \text{svec}(\Delta Z_k))) \end{aligned}$$

d.h. die Korrekturen sind nun auch symmetrisch. Wie bei primal-dualen Methoden üblich hat man nun die Schrittweiten zu bestimmen. Primal und dual kann man verschiedene Schrittweiten bestimmen.

Dies geschieht wörtlich wie im LP/QP-Fall:

$$\begin{aligned} \Lambda_{k+1} &= \Lambda_k + \alpha_k \Delta \Lambda_k \\ Z_{k+1} &= Z_k + \beta_k \Delta Z_k \\ x^{k+1} &= x^k + \beta_k \Delta x^k \end{aligned}$$

wobei mit $0 < \tau < 1$

$$\begin{aligned} \alpha_k &= \begin{cases} 1 & \text{falls } \lambda_{\min}(\Lambda_k^{-1/2}(\Delta \Lambda_k)\Lambda_k^{-1/2}) \geq 0 \\ \min\left\{1, \frac{\tau}{\lambda_{\min}(\Lambda_k^{-1/2}(\Delta \Lambda_k)\Lambda_k^{-1/2})}\right\} & \text{sonst} \end{cases} \\ \beta_k &= \begin{cases} 1 & \text{falls } \lambda_{\min}(Z_k^{-1/2}(\Delta Z_k)Z_k^{-1/2}) \geq 0 \\ \min\left\{1, -\frac{\tau}{\lambda_{\min}(Z_k^{-1/2}(\Delta Z_k)Z_k^{-1/2})}\right\} & \text{sonst.} \end{cases} \end{aligned}$$

Die erforderlichen minimalen Eigenwerte erhält man zweckmäßig aus den allgemeinen definierten Eigenwertproblemen

$$\Lambda_k u = \lambda \Delta \Lambda_k u \quad \lambda \text{ Eigenwert, } u \neq 0 \text{ Eigenvektor}$$

bzw.

$$Z_k u = \lambda \Delta Z_k u .$$

Da nur der kleinste Eigenwert gesucht ist, ist dies vergleichsweise kostengünstig, sodaß der Hauptaufwand in der Aufstellung und Faktorisierung von C_{33} und der Spektralzerlegung von Z_k zu sehen ist.

Mit $0 < \sigma_u \leq \sigma_k \leq \sigma_o < 1$ ist die allgemeine Konvergenztheorie anwendbar und führt auf eine polynomiale Komplexität.

Bem.: Es gibt mehrere gute public-domain Programme für SDP-Probleme, siehe den Online-Software-Guide.

Zum Komplexitätsbeweis siehe:

R.D.C. MONTEIRO: Polynomial convergence of primal-dual algorithms for semidefinite programming based on the MONTEIRO and ZHANG family of directions. SIOPT8, (1998), 797-812.

Kapitel C

Anhang: Die Bunch-Parlett-Zerlegung

Die LDL^T -Zerlegung kann formal auch für $d_i < 0$ durchgeführt werden. Sie ist dann jedoch numerisch instabil, weshalb man dies niemals tun sollte. Die Bunch-Parlett-Zerlegung ist auch für indefinite Matrizen stabil. Diese erfordert allerdings die Einführung von Vertauschungen und lautet

$$P^T A P = LDL^T .$$

Dabei ist P eine Permutationsmatrix, L eine untere Dreiecksmatrix mit Diagonale $(1, \dots, 1)$ und D eine Block-Diagonalmatrix mit 1×1 und 2×2 Blöcken. Diese setzt selbstverständlich die Symmetrie von A voraus. (Für uns steht A ja für die Hessematrix einer Funktion, d.h. hier ist die Symmetrie gegeben). Wird im laufenden Zerlegungsschritt ein 1×1 Block D_j benutzt, dann entspricht dies einem normalen Gauß-Eliminationsschritt mit dem Pivot D_j . Eine 2×2 Untermatrix in D entspricht einem 2×2 -Pivot. Die Elimination und damit Erzeugung der Elemente in L (2 Spalten) und die Umrechnung der "Restmatrix" erfolgt dann gemäß

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} I & O \\ A_{21}A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} & O \\ O & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{pmatrix} \begin{pmatrix} I & A_{11}^{-1}A_{12} \\ O & I \end{pmatrix}$$

Hier entspricht A der laufenden "Restmatrix", A_{11} ist der 2×2 -Pivot, $A_{21}A_{11}^{-1}$ entspricht zwei in L erzeugten Spalten, wobei in L darüber der Diagonalblock

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

eingetragen wird und $A_{22} - A_{21}A_{11}^{-1}A_{12}$ ist die neue, in der Dimension um 2 kleinere neue "Restmatrix". Die Auswahl der Pivots und die Festlegung der Vertauschungen geschieht nach folgendem Schema: Teilschritt i : Elemente der "Restmatrix" haben die Indizes i, \dots, n .

1. $\alpha_i = \max\{|a_{jj}| : j = i, \dots, n\}$
2. $\beta_i = \max\{|a_{jk}| : j, k = i, \dots, n\}$
3. wenn $\alpha_i \geq \beta_i$, dann wähle i_0 mit $|a_{i_0, i_0}| = \alpha_i$, tausche Spalte i_0 mit Spalte i und Zeile i_0 mit Zeile i .
4. Eliminiere mit dem 1×1 -Pivot a_{i_0, i_0} :
 $l_{j,i} = a_{j,i}/a_{i_0, i_0}, j = i+1, \dots, n, \quad i = i+1,$
5. Sei $|a_{rs}| = \beta_i > \alpha_i$. (Jetzt ist natürlich $r \neq s$). Tausche Zeile i mit Zeile r und Spalte i mit Spalte r .
Tausche Zeile $i+1$ mit Zeile s und Spalte $i+1$ mit Spalte s .
(a_{rs} steht jetzt auf Platz $i, i+1$.)
6. Wähle den 2×2 -Pivot aus den Positionen (i, i) , $(i, i+1)$, $(i+1, i)$, $(i+1, i+1)$, eliminiere gemäss obiger Block-Formel und erhöhe i um 2.

Wegen der Nichtsingularität von L besitzt A nach dem Trägheitssatz von Sylvester genauso viele Eigenwerte > 0 , $= 0$, < 0 wie D , und die letzteren sind leicht zu bestimmen. Hat D negative Eigenwerte, so sind dies ja Eigenwerte der einzelnen Blöcke, und auch deren Eigenvektoren sind unmittelbar anzugeben. Ist

$$D_i v_i = \lambda_i v_i, \quad v_i^T v_i = 1,$$

dann setzt man durch Auffüllen mit Nullen an den übrigen Positionen

$$v = (0, \dots, 0, v_i^T, 0, \dots, 0)^T$$

und hat dann in der Lösung von

$$L^T P^T z = v$$

eine Richtung z mit

$$z^T A z = \lambda_i$$

Für $\lambda_i < 0$ nennt man dies eine Richtung negativer Krümmung. Diese sind nützlich bei der Konstruktion von Verfahren, die auch die notwendige Bedingung zweiter Ordnung bei der Minimierung garantiert erfüllen sollen.

Kapitel D

Literatur

- 1 Avriel, M. und Golany, B. (Eds.) : Mathematical Programming for Industrial Engineers. Marcel Dekker 1996. – Gute Übersichtsartikel über den gesamten Bereich, auch diskrete und stochastische Probleme.
- 2 Baldrick, R.: Applied Optimization. Formulation and Algorithms for Engineering Systems. Cambridge University Press 2006.
- 3 Geiger, C. und Kanzow, Chr.: Numerische Verfahren zur Lösung unrestringierter Minimierungsaufgaben. Springer 1999. Sehr theorieorientiert.
- 4 Geiger, C. und Kanzow, Chr.: Theorie und Numerik restringierter Optimierungsaufgaben. Springer 2002. Sehr theorieorientiert.
- 5 Nazareth, L.: Computer solution of linear programs. Oxford Univ. Press 1987. – gut geschriebene Einführung in die praktische Lösung von LP's.
- 6 Nesterov, J.E., Nemirovsky, A.S.: Interior Point Polynomila Methods in Convex Programming: Theory and Applications. SIAM Philadelphia, 1994. Sehr theorieorientiert.
- 7 Nocedal, Jorge und Wright, Stephen: Numerical Optimization. Berlin: Springer 1999. – Modernes einführendes Lehrbuch, sehr gut geschrieben.
- 8 Rao, S.S.: Engineering Optimization, Theory and Practice. J. Wiley, 1996. – Breit gehaltene Einführung in das Gesamtgebiet, viele Beispiele
- 9 Ruszczyński, A.: Nonlinear Optimization. Princeton University Press 2006. Eher theorieorientiert, aber sehr gut lesbar geschrieben.
- 10 Wright, Stephen: Primal-Dual-Interior Point Methods. Philadelphia: SIAM 1997 – Sehr speziell, beschreibt detailliert die zur Zeit vielversprechendste Methode für grosse LP's.

Zeitschriftenliteratur wird an der jeweils benutzten Stelle direkt zitiert.

Kapitel E

Wichtige Quellen in Internet

<http://numawww.mathematik.tu-darmstadt.de:8081>

(einige numerische Optimierungscodes interaktiv anwendbar)

<http://plato.la.asu.edu/guide.html>

Sammlung von Information insbesondere über frei zugängliche Software und Hilfsmittel. Hinweise auf weitere Quellen.

<http://www.mcs.anl.gov/otc/Guide>

Der SIAM Optimization Software Guide

<http://www.mcs.anl.gov/NEOS>

Network optimization submission tool. Man muss hier jedoch die die Modellierungssprachen AMPL bzw. GAMS beherrschen. Die meisten dieser Codes haben ein Interface dazu. Dann muss man nicht das spezielle Eingabeformat des jeweiligen Programms kennen. NEOS erlaubt Online-Erprobung einer Fülle moderner Optimierungssoftware.

In der Regel findet man im Netz bereits vorgefertigte Softwarelösungen, die meisten davon für akademischen Gebrauch kostenfrei: Die bei weitem grösste und wichtigste Quelle ist die

NETLIB: <http://www.netlib.org/>

eine Sammlung von Programmbibliotheken in f77, f90 , c, c++ für alle numerischen Anwendungen: Man kann nach Stichworten suchen (“search“) und bekommt auch Informationen aus dem NaNet (Numerical Analysis Net) Die Bibliotheken findet man unter “browse repository“ Die wichtigsten Bibliotheken sind:

1. specfunc, cephes, amos : spezielle Funktionen
2. fftpack : Fast Fourier Transform

3. fitpack , diercks : Spline Approximation und Interpolation
4. lapack clapack, lapack90 : die gesamte Numerische Lineare Algebra (voll besetzte und Band-Matrizen) inklusive Eigenwerte und lineare Ausgleichsrechnung, sehr gute Qualität
5. lanz, lanczos : Eigenwerte/Eigenvektoren grosser dünn besetzter Matrizen
6. toms: Transactions on Mathematical Software. Sammlung von Algorithmen für verschiedene Aufgaben, sehr gute Qualität u.a. auch automatische Differentiation, Arithmetik beliebiger Genauigkeit, mehrere Optimierungscodes, Nullstellenbestimmung, cubpack (Kubatur)
7. linalg: Iterative Verfahren für lineare Systeme, sonstige lineare Algebra, auch direkte Löser für grosse dünn besetzte lineare Systeme.
8. templates: Iterative Verfahren für lineare Systeme
9. quadpack : Quadratur (bestimmte Integrale, 1-dimensional)
10. odepack, ode : numerische Integration von gewöhnlichen Dglen, auch Randwertaufgaben
11. fishpack : Helmholtzgleichung mit Differenzenverfahren
12. opt,minpack1 : Optimierungssoftware

Suchen nach software, wenn der Name des Programmmoduls bekannt ist, erfolgt mit

xarchie

sonst mit Hilfe des Service

<http://math.nist.gov/HotGAMS/>

Dort findet man ein Suchmenu, wo man nach Problemklassen geordnet durch einen Entscheidungsbaum geführt wird bis zu einer Liste verfügbarer software (auch in den kommerziellen Bibliotheken IMSL und NAG) Falls der code als public domain vorliegt, wird er bei "Anklicken" sofort geliefert.

Andere wichtige Quelle

<http://elib.zib.de/>

Dort gibt es auch Bibliotheken, teilweise mit guten Eigenentwicklungen der Gruppe um P. Deuffhard (die Codelib), sowie sonstige weitere Verweise. Wichtig für Ingenieursanwendungen: Die Finite-Element-Resources

http://www.engr.usask.ca/~macphedran/finite/fe_resources/fe_resources.html

Dort findet man Links zu freien FEM-Codes und viele weitere Quellen, das hier auch installierte Felt - System .

Hat man Fragen, z.B. nach Software, Literatur oder auch zu spezifischen mathematischen Fragestellungen, kann man in einer der News-groups eine Anfrage platzieren. Häufig bekommt man sehr schnell qualifizierte Hinweise. Zugang zu Newsgroups z.B. über

xrn

mit "subscribe".

sci.math.num-analysis
sci.op-research

sind die wichtigsten für diesen Bereich. (Es gibt natürlich auch im Bereich Informatik bzw. Software und Ingenieurwissenschaften eine Fülle solcher News-groups)

Im xrn-Menue kann man mit "post" eine Anfrage abschicken und dabei die Zielgruppe frei wählen.

Kapitel F

Notation, Formeln

$$\|x\| = (x^T x)^{1/2} \quad \text{euklidische Vektornorm, Länge von } x \quad (l_2\text{-Norm})$$

{ Andere gebräuchliche Längenmaße:

$$\|x\|_\infty = \max_i |x_i| \quad \text{Maximumnorm} \quad (l_\infty\text{-Norm})$$

$$\|x\|_1 = \sum_{i=1}^n |x_i| \quad \text{Betragssummennorm} \quad (l_1\text{-Norm}) \}$$

x, y, z, \dots (Spalten)vektoren $\hat{=}(n \times 1)$ -Matrix

$$e^i = (0, \dots, 0, \underbrace{1}_i, 0, \dots, 0)^T \in \mathbb{R}^n$$

x^T, y^T, \dots Zeilenvektoren $\hat{=}(1 \times n)$ -Matrix

x^0, x^k, \dots Elemente von Vektorfolgen

$\{\alpha_k\}, \{\beta_k\}$ skalare Folgen

$x_i, y_j,$ Vektorkomponenten

$(\sigma)^2, \|x\|^2, \dots$ Exponentiationen nur für arithmetische Ausdrücke

$\xi, \alpha, \beta, \dots$ Skalare

A, B, C, Δ, Σ Matrizen, Diagonalmatrizen

I, I_n Einheitsmatrix (immer quadratisch)

0 Nullmatrix

Dimension: aus dem Zusammenhang, eventuell als Index, z.B. $0_{n \times m}$.

1. Ist f eine vektorwertige Funktion von n Veränderlichen x ,
 $f : \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, so bezeichnet

$$\mathcal{J}_f(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} f_1 & , \dots , & \frac{\partial}{\partial x_n} f_1 \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_1} f_m & , \dots , & \frac{\partial}{\partial x_n} f_m \end{pmatrix}$$

die **Jacobimatrix** von f .

Jacobimatrix: Zeilennummer $\hat{=}$ Funktionsnummer
 Spaltennummer $\hat{=}$ Variablennummer

2. Der **Gradient** ist stets die transponierte Jacobimatrix:

$$\nabla f(x) = (\mathcal{J}_f(x))^T$$

Gradient: Zeilennummer $\hat{=}$ Variablennummer
 Spaltennummer $\hat{=}$ Funktionsnummer

3. Für eine skalare Funktion $f : \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}$ bezeichnet

$$\nabla^2 f(x) = \left(\frac{\partial^2}{\partial x_i \partial x_j} f(x) \right) = \left((\nabla \nabla^T) f \right)(x)$$

die **Hessematrix** von f .

Für vektorwertige Funktionen kommt diese Konstruktion nur im Zusammenhang mit der Taylorentwicklung vor, d.h. als Vektor

$$\begin{bmatrix} d^T \nabla^2 f_1(x) d \\ \vdots \\ d^T \nabla^2 f_m(x) d \end{bmatrix} =: \underbrace{d^T (\nabla^2 f(x)) d}_{\text{symbolisch, nur für } m=1 \text{ echte Matrix-Vektor-Notation}}$$

mit einem Inkrementvektor $d \in \mathbb{R}^n$

Intervall im \mathbb{R}^n :

$$[x^0, x^0 + d] = \{x^0 + td, \quad 0 \leq t \leq 1\}$$

Hilfsmittel

Mittelwertsätze

$f \in C^1(\mathcal{D})$

$$\begin{aligned} f(x^0 + d) &= f(x^0) + \nabla f(x^0)^T d + o(\|d\|) \quad *) \\ f(x^0 + d) &= f(x^0) + \nabla f(x^0 + \vartheta d)^T d \quad \text{falls } f \text{ skalar, } 0 < \vartheta < 1 \end{aligned}$$

$$f(x^0 + d) = f(x^0) + \left(\underbrace{\int_0^1 \nabla f(x^0 + td)^T dt}_{\text{Integral ist komponentenweise zu nehmen}} \right) d$$

Taylorentwicklung

$f \in C^2(\mathcal{D}), x^0 \in \mathcal{D}$

$$\begin{aligned} f(x^0 + d) &= f(x^0) + \nabla f(x^0)^T d + \frac{1}{2} d^T \nabla^2 f(x^0) d + o(\|d\|^2) \quad *) \\ f(x^0 + d) &= f(x^0) + \nabla f(x^0)^T d + \frac{1}{2} d^T \nabla^2 f(x^0 + \vartheta d) d \quad \text{falls } f \text{ skalar} \\ f(x^0 + d) &= f(x^0) + \nabla f(x^0)^T d + d^T \left(\int_0^1 (1-t) \nabla^2 f(x^0 + td) dt \right) d \end{aligned}$$

*) $o(\cdot)$ Landau-Symbol (klein-o)

$o(1)$ bezeichnet eine Größe, die bei einem (in der Regel implizit definierten) Grenzübergang gegen null geht. $o(\|d\|^k)$ bezeichnet eine Größe, die schneller gegen null geht als $\|d\|^k$, d.h.

$$o(\|d\|^k) / \|d\|^k \rightarrow 0 \quad \text{für } d \rightarrow 0.$$

Formeln, Rechnen mit $\frac{d}{d\sigma}, \nabla$ bei Vektorfunktionen:

$$\frac{d}{d\sigma} f(x - \sigma d)|_{\sigma=0} = -(\nabla f(x))^T d$$

$$\frac{d^2}{(d\sigma)^2} f(x - \sigma d)|_{\sigma=0} = d^T \nabla^2 f(x) d$$

$$\nabla(f(x)g(x)) = g(x)\nabla f(x) + f(x)\nabla g(x)$$

$$\nabla(f(g(x))) = \nabla g(x)\nabla f(y)|_{y=g(x)}$$

Insbesondere für: $f(y) = y^T y$ ergeben sich

$$\nabla(\|g(x)\|^2) = 2(\nabla g(x))g(x)$$

$$\nabla^2(\|g(x)\|^2) = 2(\nabla g(x))(\nabla g(x))^T + 2 \sum_{i=1}^m g_i(x) \nabla^2 g_i(x) \quad \text{für } g: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

Differentiation einer inversen Matrix nach einem Parameter:

$$\frac{d}{d\sigma} (A(\sigma))^{-1} = -(A(\sigma))^{-1} \left(\frac{d}{d\sigma} A(\sigma) \right) (A(\sigma))^{-1}$$

Index

- A-konjugiert, 46
- A-orthogonal, 46
- Abstiegsrichtung, 31
- Armijo, 34
- Ausgleichsrechnung, 54
- Austauschregel, lexikographische, 101

- backtracking, 34
- Barriere-Verfahren, 75
- BFGS, 39
- BFGS, partitioniertes, 44
- Bland, 101
- box constraints, 63
- Broyden-Fletcher-Goldfarb-Shanno, 39

- cg-Verfahren, 47
- Cholesky-Zerlegung, 13

- Dantzig, 93
- Davidon-Fletcher-Powell, 39
- DFP, 39
- Differentiation, automatische, 30
- Differentiation, numerische, 30
- DiPillo, Grippo, 92
- duale Variablen, 74
- Dualitätslücke, 74, 81

- Ecke, 70
- Extremalbedingung, hinreichende, 11
- Extremalbedingung, notwendige, 11
- Extrempunkt, 96

- Fletcher, 91
- Fletcher-Reeves, 48
- Fritz-John-Kriterium, 65
- Funktion, gleichmässig konvexe, 17
- Funktion, konvexe, 13
- Funktion, streng konvexe, 13

- Gauvin, 71
- Gauß-Newton, 57
- goldener Schnitt, 20
- Goldfarb-Idnani, 108
- Goldstein-Armijo, 34
- gradientenbezogen, gleichmässig, 31
- Gradientenprojektionsverfahren, 69, 112
- Gradientenverfahren, 41
- grg, 111

- Halbordung, 120
- Hestenes-Powell, 85
- Householdermatrix, 56

- innere Punkte-Verfahren, 80

- Kegel, nichtpolyedrisch, 120
- Kiwiel, 89
- Kojima-Mizuno-Yoshise, 82
- Komplemetarität, strikte, 70
- konvexe Optimierung, 64
- Konvexitätskriterien, 17
- Konvexkombination, 96
- Kroneckerprodukt, symmetrisiertes, 128
- Kuhn-Tucker-Bedingung, 66

- L1-penalty, 112
- Lagrangefunktion, 69
- Lagrangefunktion, erweiterte, 85, 88
- LANCELOT, 88
- LDLT-Zerlegung, 12
- Levenberg – Marquardt, 59
- limited memory Verfahren, 45

- Line search, 30
- Liniensuche, 30
- LP, 93
- Lyapunov-Matrix-Gleichung, 129

- Mangasarian-Fromowitz-Bedingung, 65
- Menge, konvexe, 13
- Minimierung, ableitungsfreie, 51
- Multiplikatormethode, 85
- Multiplikatorregel, 66
- Multiplikatorregel von Lagrange, 66

- Nelder-Mead, 51
- Newton-ähnliche Verfahren, 38
- Newton-Verfahren, 38
- Normalgleichungen, 55

- Parameteranpassung, 54
- Penalty-Verfahren, 75
- Penaltyfunktion, exakte, primal-duale, 92
- Penaltyfunktionen, exakte, 91
- Pietrzykowski, 112
- positiv definit, 12
- Powell-Wolfe, 35
- Präkonditionierung, 48
- Problem, semidefinites, 120

- QP, 103
- QR-Zerlegung, 56
- quadratische Interpolation, fortgesetzte, 29
- quasi-Newton-Verfahren, 38
- quasikonvex, 18
- quasikonvex, streng, 18

- Regression, orthogonale, 60
- Regularitätsbedingung, 65
- Restriktionen, affin lineare, 64
- Restriktionsqualifikationen, 65
- Richtung, extremale, 96
- Rockafellar, 88
- Rosen, 69

- Sattelpunkt, 73
- Schlupfvariablen, 64

- Schrankenrestriktionen, 63
- Simplexverfahren, 93
- Slater-Bedingung, 66
- SR1, 39
- Subgradient, 63
- Symmetrisierungsoperator, 128

- Tangentialmannigfaltigkeit, 70

- Verfahren, primal-duales, 127
- Vertrauensbereich, 49

- Zangwill, 112