

Nichtlineare Optimierung: Linear restringierte, insbesondere Quadratische Optimierung

- Entwurf -

Ernst-Peter Beisel
Bergische Universität Wuppertal
Fachbereich C
Oktober 2009

(**Achtung:** fast alle Skizzen fehlen in diesem Skript)

Kapitel 1

Einleitung

Die Modellierung praktisch gegebener Optimierungsprobleme führt in der Regel zu Optimierungsaufgaben, bei denen eine Zielfunktion $f : U \rightarrow \mathbf{R}$ über einem Teilbereich $X \subseteq U$ des \mathbf{R}^n zu minimieren ist. Dabei ist f zumindest als stetig vorausgesetzt, oft aber als (zweimal) stetig differenzierbar. Der zulässige Bereich wird dann durch Ungleichungen und Gleichungen beschrieben

$$X = \{x \in U : g(x) \leq 0, h(x) = 0\},$$

die über (stetig differenzierbare) Funktionen $g : U \rightarrow \mathbf{R}^m$, $h : U \rightarrow \mathbf{R}^p$ ausgedrückt sind.

Dabei sind all diese Funktionen in einer zumeist vereinfachenden Modellierung aus Gründen der (schnellen) Lösbarkeit der entstehenden Optimierungsaufgaben häufig affin-linear, es liegen Aufgaben der **Linearen Optimierung** vor. Um diese Linearität zu erreichen, werden allerdings teilweise erhebliche Ungenauigkeiten in der Beschreibung der Wirklichkeit hingenommen. Dabei ist die Ungenauigkeit bei der Modellierung der Zielfunktion oft größer als die der Restriktionen.

Deshalb ist ein interessantes Modell die **linear restringierte Optimierungsaufgabe**. Diese behält die Linearität der Restriktionen bei, lässt aber bei der Zielfunktion beliebige Nichtlinearität zu. Mit diesem Aufgabentyp soll sich diese Vorlesung hauptsächlich beschäftigen.

Zum Beispiel können die Koeffizienten einer zunächst linear formulierten Ziel-

funktion (von den Abnahmemengen x_i abhängige Kosten)

$$f(x_1, \dots, x_n) = \sum_{i=1}^n c_i x_i$$

in Wirklichkeit nicht konstant sein, sondern von der Größe der Werte x_i abhängen: beschreiben z. B. die c_i *Kosten pro Mengeneinheit*, so ist allgemein davon auszugehen, dass bei steigender Abnahmemenge diese Mengenkosten sinken. Je größer $|x_i|$ oder $\|x\|$ ist, desto niedriger wird in der Praxis c_i sein, was durch ein lineares Modell nicht wiedergegeben wird. Also sollte man realistisch $c_i = c_i(x_1, \dots, x_n)$ ansetzen. Dies macht die Zielfunktion nichtlinear.

Kennt man die Abhängigkeit der Kostenkoeffizienten c_i von den Variablen x_1, \dots, x_n nicht, so wird man sie ggf. in bewährter Manier als affin-linear ansetzen,

$$c_i = \bar{c}_i + \sum_{j=1}^n r_{ij} x_j$$

mit geschätzten Größen \bar{c}_i und r_{ij} . In diesem Fall wird die Zielfunktion zu einem quadratischen Polynom, das Modell zu einer **quadratischen Optimierungsaufgabe**. Quadratische Optimierungsaufgaben sind der nächst einfache Modellierungsfall nach den linearen Optimierungsaufgaben und es ist von großem praktischen Interesse zu wissen, wie man diesen Aufgabentyp am besten löst. Dies ist auch eine primäre Zielsetzung dieser Vorlesung.

Nun ist die theoretische und algorithmische Betrachtung quadratischer Optimierungsaufgaben nicht wirklich von allgemeineren nichtlinearen Modellen zu trennen, deshalb wird im Folgenden auch von allgemeineren nichtlinearen Optimierungsaufgaben die Rede sein. Wir wollen aber versuchen, diese größer Allgemeinheit, wann immer sinnvoll, auf linear restringierte Optimierungsaufgaben zu beschränken, d. h. auf Aufgaben mit nichtlinearer Zielfunktion, aber linearen Restriktionen. Bei diesen Aufgaben sind manche Dinge einfacher zu beschreiben als im allgemeineren Fall. Gleichzeitig können die grundlegenden Techniken doch auch hier schon dargestellt werden. Einige der angeführten Fakten gelten allerdings im allgemeinen nichtlinearen Fall nur unter zusätzlichen Einschränkungen.

Im Folgenden betrachten wir also eine *linear restringierte Optimierungsaufgabe* mit *stetig differenzierbarer (nicht-)linearer* Zielfunktion über einem Po-

lyeder $M \subset \mathbf{R}^n$

$$(P) \quad \begin{cases} \text{minimiere} & f(x) \\ \text{s.d.} & x \in M \end{cases}$$

Gehen wir davon aus, dass M wie folgt beschrieben ist:

$$M = \{x \in \mathbf{R}^n : Ax \leq a, Hx = h\}$$

mit $A \in \mathbf{R}^{m \times n}$ und $H \in \mathbf{R}^{p \times n}$. Diese Darstellung beinhaltet auch die *Standarddarstellungen* des zulässigen Bereichs, $\{x \in \mathbf{R}^n : Hx = h, x \geq 0\}$ und $\{x \in \mathbf{R}^n : Ax \leq a, x \geq 0\}$, bei denen die Matrix H maximalen Zeilenrang hat.

Besonders interessieren wir uns für die *quadratische Optimierungsaufgabe*

$$(QP) \quad \begin{cases} \text{minimiere} & f(x) = \frac{1}{2}x^T Qx + p^T x \\ \text{s.d.} & Ax \leq a, Hx = h \end{cases}$$

mit *symmetrischer* Matrix $Q \in \mathbf{R}^{n \times n}$.

Bemerkung: Die Matrix Q der Koeffizienten quadratischer Terme in der Darstellung eines quadratischen Polynoms f ist nicht eindeutig. Man kann sie oBdA als *symmetrisch* annehmen. Beispiel: $f(x) = x_1x_2 + 3x_2x_1$ lässt sich auch als $f(x) = 2x_1x_2 + 2x_2x_1$ schreiben.

Wir werden zunächst eine *Einführung in der Theorie der linear restringierten Optimierungsaufgaben* und klassische, mit dem Simplexverfahren verwandte *Verfahren der quadratischen Optimierung* kennenlernen. Wir werden über *grundlegende Techniken* zur Konstruktion von Verfahren reden und im Weiteren diese Techniken zur konkreten Konstruktion von *Verfahren zur linear restringierten Optimierung* verwendet. Wann immer möglich, sollen die dabei besprochenen Verfahren der quadratischen Optimierung von *MATLAB Programmen* begleitet werden, um so einen Eindruck von der Leistungsfähigkeit der Verfahren zu erhalten. Insgesamt ist die Vorlesung eher verfahrensorientiert. Dabei geht es uns vorrangig um die grundsätzliche Konvergenz der Verfahren. Gelegentlich geben wir ohne Beweis auch eine lokale Konvergenzgeschwindigkeit zu den Verfahren an.

Überhaupt wird aus Zeitgründen an vielen Stellen der Vorlesung, an denen eine Aussage aus sich heraus verständlich oder einsichtig ist, auf einen formalen Beweis verzichtet und lediglich eine Literaturstelle angegeben, an der der Beweis nachgelesen werden kann. Würden alle Beweise ausgeführt, wäre

der Zeitbedarf für die angesprochenen Inhalte sicherlich doppelt so groß, die angestrebte Zusammenschau der Ergebnisse, Techniken und Konstruktionen aber erheblich behindert.

Grundlage für die Vorlesung und als Begleitliteratur empfohlen sind die folgenden sehr lesenswerten Bücher, auf die wir auch immer wieder in Form von Hinweisen zurückgreifen müssen, wenn der zeitliche Rahmen der Vorlesung die ausführliche Darstellung eines Beweises nicht erlaubt:

- [BSS] Bazarraa/Sherali/Shetty: *Nonlinear Programming - Theory and Algorithms*, Third Edition, Wiley Interscience 2006
- [GK2] Geiger/Kanzow: *Theorie und Numerik restringierter Optimierungsaufgaben*, Springer 2002
- [GK1] Geiger/Kanzow: *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*, Springer 1999
- [A] Alt: *Nichtlineare Optimierung*, Vieweg Verlag 2002

Ferner wurden auch Informationen zu Verfahren aus den folgenden älteren Büchern entnommen:

- [KKR] Künzi/Krelle/von Randow: *Nichtlineare Optimierung*, Springer 1979
- [H] Horst: *Nichtlineare Optimierung*, Carl Hanser Verlag 1979

In all diesen Büchern sind auch viele konkrete Beispiele und Modelle enthalten, die anzusehen wir dem Zuhörer empfehlen.

Zusätzlich werden verwendet:

- [B1] Beisel: *Affin transformierende global konvergente Innere-Punkte-Verfahren der Linearen Optimierung mit langen Schrittweiten*, Shaker Verlag 1998
- [M1] Mendel: *Primal-duale pfadorientierte Innere- und Äußere-Punkte-Verfahren zur Lösung linearer Optimierungsaufgaben*, Shaker Verlag 1998

Notwendige Voraussetzung zum Verständnis dieser speziellen Ausprägung einer Vorlesung zur nichtlinearen Optimierung ist das Vertrautsein mit den Techniken der Linearen Optimierung, ausgewiesen durch einen erfolgreichen Besuch der entsprechenden Bachelorvorlesung. Hilfreich zu einem späteren Zeitpunkt der Vorlesung, aber nicht notwendig, sind Kenntnisse aus dem Bereich der "Innere-Punkte Verfahren", weshalb ggf. auch der gleichzeitige Besuch dieser Vorlesung empfohlen wird.

Kapitel 2

Theorie linear restringierter Optimierung

Im Folgenden werden einige grundlegende Begriffe und Aussagen zu linear restringierten Optimierungsaufgaben zusammengestellt. Da vieles davon bekannt oder unmittelbar einsichtig sein dürfte, wird weitgehend auf Beweise verzichtet. Wir unterstützen stattdessen die angeführten Aussagen und Begriffe weitmöglichst durch Skizzen, um deren Akzeptanz zu erhöhen. Wichtig ist uns vorrangig die "Zusammenschau der Dinge".

2.1 Konvexität

Wir betrachten in diesem Abschnitt eine Funktion $f : X \subseteq \mathbf{R}^n \rightarrow \mathbf{R}$ mit konvexem Definitionsbereich X . Dabei ist X **konvex**, wenn die Verbindungsstrecke zwischen je zwei Punkten aus X ganz in X enthalten ist.

Man beachte: Es gibt offene und abgeschlossene konvexe Mengen sowie solche, die weder offen noch abgeschlossen sind. Der Durchschnitt beliebig vieler konvexer Mengen ist konvex. Lineare Halbräume der Form $\{x \in \mathbf{R}^n : a^T x \leq \alpha\}$ zu vorgegebenen $a \in \mathbf{R}^n$ und $\alpha \in \mathbf{R}$ sind konvex und damit auch die Hyperebenen $\{x \in \mathbf{R}^n : a^T x = \alpha\}$. Daher ist der zulässige Bereich der linear restringierten Aufgabe (P) konvex.

Die Funktion f heißt **konvex**, wenn für je zwei Punkte $x, y \in X$ gilt:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad \text{für beliebiges } \lambda \in (0, 1)$$

Sie heißt **streng konvex**, wenn dabei bei $x \neq y$ immer das $<$ -Zeichen gilt.

Sie heißt **gleichmäßig konvex**, wenn es ein $\mu > 0$ gibt, so dass sogar für je zwei Punkte $x, y \in X$ und beliebiges $\lambda \in (0, 1)$ gilt

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

Die Funktion f heißt **(streng, gleichmäßig) konkav**, wenn $(-f)$ (streng, gleichmäßig) konvex ist.

Beachte: Gleichmäßig konvexe Funktionen sind streng konvex, streng konvexe Funktionen sind konvex, wie sich unmittelbar aus den Definitionen ergibt.

2.1.1 Charakterisierung konvexer Funktionen

Die Konvexität von f kann wie folgt charakterisiert werden:

Satz 2.1 *Es sei X offen und f stetig differenzierbar.*

1. *Genau dann ist f konvex, wenn für für je zwei Punkte $x, y \in X$ gilt*

$$f(x) - f(y) \geq \nabla f(y)^T (x - y)$$

Genau dann ist dabei die Konvexität streng, wenn bei $x \neq y$ immer das $>$ -Zeichen gilt.

Genau dann ist die Konvexität gleichmäßig, wenn es sogar ein $\mu > 0$ gibt mit

$$f(x) - f(y) \geq \nabla f(y)^T (x - y) + \mu\|x - y\|^2 \quad \text{für alle } x, y \in X$$

2. *Genau dann ist f konvex, wenn für für je zwei Punkte $x, y \in X$ gilt*

$$(x - y)^T (\nabla f(x) - \nabla f(y)) \geq 0$$

Genau dann ist dabei die Konvexität streng, wenn bei $x \neq y$ immer das $>$ -Zeichen gilt.

Genau dann ist die Konvexität gleichmäßig, wenn es sogar ein $\mu > 0$ gibt mit

$$(x - y)^T (\nabla f(x) - \nabla f(y)) \geq \mu\|x - y\|^2 \quad \text{für alle } x, y \in X$$

Beweis: [GK1], Satz 3.5 und Satz 3.7

Unter schärferer Voraussetzung gilt der

Satz 2.2 *Es sei X offen und f zweimal stetig differenzierbar.*

1. *f ist genau dann konvex, wenn $\nabla^2 f(x)$ für alle $x \in X$ positiv semidefinit ist.*
2. *Ist $\nabla^2 f(x)$ für alle $x \in X$ positiv definit, so ist f streng konvex.*
3. *Genau dann ist f gleichmäßig konvex, wenn es sogar ein $\mu > 0$ gibt mit*

$$d^T \nabla^2 f(x) d \geq \mu \|d\|^2 \quad \text{für alle } x \in X \text{ und alle } d \in \mathbf{R}^n$$

Beweis: [GK1], Satz 3.8

Man beachte, dass die strenge Konvexität hier nicht äquivalent charakterisiert wird. Das liegt z.B. an der Funktion $f(x) = x^4$, die streng konvex ist, aber im Nullpunkt eine verschwindende zweite Ableitung hat.

Nun betrachten wir eine Optimierungsaufgabe, bei der f über einer Teilmenge X des \mathbf{R}^n minimiert werden soll.

$$(KO) \quad \min \{f(x) : x \in X\}$$

Der folgende Satz stellt grundlegende Ergebnisse für die Optimierung zusammen.

Satz 2.3 *Es sei f stetig differenzierbar. Dann gelten die folgenden Aussagen:*

1. *Ist f konvex, X offen und $x^* \in X$ ein stationärer Punkt von f , so ist x^* Optimallösung von (KO).*
2. *Ist f konvex, so ist die Menge der Optimallösungen von (KO) konvex (evtl leer).*
3. *Ist f streng konvex, so besitzt (KO) höchstens eine Optimallösung.*
4. *Ist f gleichmäßig konvex und X nichtleer und abgeschlossen, so besitzt (KO) genau eine Optimallösung.*

Beweis: [GK1], Satz 3.10 und Satz 3.12

2.1.2 Spezialfall: Quadratische Optimierung

Betrachten wir nun die Aufgabe (QP) mit quadratischer Zielfunktion und linearen Restriktionen. Die Zielfunktion

$$f(x) = \frac{1}{2}x^T Qx + p^T x$$

ist zweimal stetig differenzierbar mit $\nabla f(x) = Qx + p$ und $\nabla^2 f(x) = Q$. Deswegen ist f genau dann konvex, wenn Q positiv semidefinit ist und genau dann streng konvex, wenn Q positiv definit ist. Letzteres kann wie folgt eingesehen werden. Wir haben nur noch die Implikation zu zeigen: Ist f streng konvex, so ist Q positiv definit.

Sei $d \in \mathbf{R}^n$ mit $d \neq 0$ vorgegeben. Ferner sei $x \in \mathbf{R}^n$ beliebig. Dann setze $y := x - d$. Weil f streng konvex ist, gilt

$$\begin{aligned} (x - y)^T (\nabla f(x) - \nabla f(y)) &> 0 \iff \\ (x - y)^T (Qx + p - [Qy + p]) &> 0 \iff d^T Qd > 0 \end{aligned}$$

also ist Q positiv definit.

Da $\nabla^2 f(x)$ konstant, also für alle x gleich ist, ist f konvex, genau dann, wenn f auf einer offenen konvexen Teilmenge $X \subset \mathbf{R}^n$ konvex. Letzteres ist nicht der Fall, wenn Q indefinit ist. Dies ist nicht selbstverständlich, wie die Funktion $g(x) = x^3$ zeigt. Diese ist auf \mathbf{R}_{++}^n streng konvex, aber auf \mathbf{R}_{--}^n streng konkav.

Ferner ist die quadratische Funktion f genau dann streng konvex, wenn sie gleichmäßig konvex ist, wie man leicht einsieht. Denn ist λ_{\min} der kleinste Eigenwert von Q , so ist dieser bei positiv definitem Q positiv und es gilt

$$d^T \nabla^2 f(x) d = d^T Qd \geq \lambda_{\min} \|d\|^2 \quad \text{für alle } d \in \mathbf{R}^n$$

Ist also f streng konvex, so besitzt die Aufgabe (QP) genau eine Optimallösung.

2.1.3 Übungsaufgaben

Aufgabe 1

Zeigen Sie, dass sich jedes quadratische Polynom in n Variablen über eine orthogonale Transformation und eine Verschiebung des Koordinatensystems auf die folgende Form bringen lässt:

$$f(x) = \frac{1}{2}x^T Qx + p^T x$$

mit einer Diagonalmatrix $Q = (q_{ij})$ und einem Vektor p , der höchstens in den Komponenten j ungleich null ist, für die $q_{jj} = 0$ gilt. Argumentieren Sie, warum Eigenschaften wie Konvexität oder strenge Konvexität bei der Transformation nicht verloren gehen.

Aufgabe 2

Überlegen Sie sich ein Konzept, auf der Grundlage von Aufgabe 1 ein Beispielerzeugungprogramm in MATLAB zu schreiben, das quadratische Optimierungsaufgaben in allgemeiner Lage und allgemeiner Struktur erstellt. Stellen Sie einen Katalog von Forderungen auf, die eine zu erzeugende Aufgabe erfüllen soll (Konvexität, Lösbarkeit, Lage der Optimallösung, Typ und Beschränktheit des zulässigen Bereichs,...) und legen Sie Parameter fest, die diese Eigenschaften bei der Erzeugung steuern. Erzeugen Sie schließlich eine Serie unterschiedlicher Aufgaben, die es mit später zu besprechenden Verfahren zu lösen gilt.

Aufgabe 3

Erzeugen sie ein paar niederdimensionale linear restringierte Aufgaben, die aus der Literatur bekannte, schwer zu lösende nichtlineare Funktionen als Zielfunktion haben. (siehe etwa [GK1])

2.2 Optimalität

Da die Zielfunktion der Aufgabe (P) stetig und nichtlinear ist, existiert zwar bei kompaktem zulässigen Bereich eine Optimallösung der Aufgabe (*globaler Minimierer*), wir müssen aber davon ausgehen, dass daneben noch weitere *lokale Minimierer* existieren, die jeweils den kleinsten Zielfunktionswert in ihrer unmittelbaren Umgebung besitzen, nicht aber den absolut kleinsten.

Zu beachten ist allerdings, dass bei konvexer Zielfunktion f jeder lokale Minimierer von (P) automatisch globaler Minimierer von (P) ist, wie auch anschaulich leicht einzusehen ist.

Die folgenden Ausführungen in diesem Abschnitt beziehen sich alle auf die linear restringierte Aufgabe (P) mit dem konvexen Polyeder M als zulässigem Bereich. Viele der im Folgenden aufgeführten Fakten sind anschaulich klar, deshalb werden auch in diesem Abschnitt nur wenige Beweise geführt.

2.2.1 Subminimale Punkte

Wir wollen versuchen, lokale Minimierer zu charakterisieren. Dazu betrachten wir *zulässige Richtungen*: Vorgegeben sei ein zulässiger Punkt $x \in M$. Ein Vektor $d \in \mathbf{R}^n$ heißt **zulässige Richtung im Punkte x bzgl M** , wenn es ein $\kappa > 0$ gibt mit $x + \lambda d \in M$ für alle $0 \leq \lambda \leq \kappa$. Wegen der vorausgesetzten Konvexität des zulässigen Bereichs gilt auch die schwächere Charakterisierung, dass $x + \kappa d \in M$ gilt für ein $\kappa > 0$. Im Falle allgemeiner nichtlinearer Restriktionen gilt dies i.a. nicht.

Man beachte: Ist d eine zulässige Richtung in x , so ist auch αd zulässige Richtung in x für alle $\alpha > 0$: zulässige Richtungen in x bilden einen Kegel. Wir bezeichnen mit $\mathcal{K}(M, x)$ den **Kegel der zulässigen Richtungen in x bzgl M** .

Mithilfe zulässiger Richtungen können lokale Minima *notwendig* charakterisiert werden.

Lemma 2.4 *Sei x ein lokaler Minimierer der Aufgabe (P) . Dann gilt für alle im Punkte x zulässigen Richtungen d :*

$$\nabla f(x)^T d \geq 0$$

Beweis: vgl. [GK2] Lemma 2.30

Zu beachten ist, dass der Term $\nabla f(x)^T d$ die *Richtungsableitung* von f längs d angibt.

Wir nennen zulässige Punkte, die die im Lemma genannten Eigenschaft besitzen, **subminimale Punkte**. Lokale Minimierer sind also subminimal, aber nicht alle subminimalen Punkte sind lokale Minimierer, sondern können z.B. auch lokale Maximierer sein. Ist die Aufgabe lösbar, so besitzt sie durch den Minimierer einen subminimalen Punkt.

Der im Lemma angesprochene Sachverhalt führt noch zu einer weiteren Begriffsbildung, die an dieser Stelle genannt werden soll. Ist d eine zulässige Richtung, so führt diese nach Definition zumindest ein kleines Stück durch den zulässigen Bereich, d.h. es gibt ein $y \in M$ so, dass d positives Vielfache des Vektors $(y - x)$ ist. Ist M konvex, so erhält man zulässige Richtungen in x gerade durch Differenzbildung $(y - x)$ für beliebiges $y \in M$.

Insofern kann die im Lemma angesprochene notwendige Charakterisierung eines lokalen Minimums für (P) an der Stelle x auch wie folgt formuliert werden: Es gilt

$$\nabla f(x)^T (y - x) \geq 0 \quad \text{für alle } y \in M$$

Dadurch motiviert **definieren wir:**

Sei $X \subseteq \mathbf{R}^n$ nichtleer und abgeschlossen und $F : X \rightarrow \mathbf{R}^n$ gegeben. Dann verstehen wir unter dem zu dieser Situation gehörigen **Variationsungleichungs-Problem** die Aufgabenstellung $VIP(X, F)$, einen Vektor x^* zu finden, für den gilt

$$F(x^*)^T (y - x) \geq 0 \quad \text{für alle } y \in X$$

Subminimale Punkte zu (P) sind also genau die Lösungen von $VIP(M, \nabla f)$. Für Variationsungleichungs-Probleme gibt es eigenständige Verfahren, die zur Ermittlung subminimaler Punkte herangezogen werden können. Wir werden später darauf zurückkommen.

Ist x kein subminimaler Punkt, so gibt es mindestens eine zulässige Richtung in x mit $\nabla f(x)^T d < 0$, entlang dieser Richtung fällt also die Zielfunktion tendenziell ab. Wegen der Stetigkeit der Ableitung der Zielfunktion und wegen der Definition der zulässigen Richtung fällt die Zielfunktion entlang d für ein

kleines Stück sogar echt ab, d.h es gibt ein $\lambda > 0$ so, dass $f(x + \tau d) < f(x)$ und $x + \tau d \in M$ für alle $0 \leq \tau \leq \lambda$ gilt. Wir nennen dann d eine **zulässige Abstiegsrichtung in x** bzgl (P) .

Zulässige Abstiegsrichtungen bekommt man offensichtlich auf folgende Weise: Man löst die Aufgabe

$$(ZA1) \quad \begin{cases} \text{minimiere} & \nabla f(x)^T d \\ \text{s.d.} & A_1 d \leq 0, Hd = 0 \\ & \|d\| \leq 1 \end{cases} \quad (2.1)$$

Dabei ist $\| \cdot \|$ eine beliebig gewählte Norm und A_1 fasst diejenigen Zeilen A_i von A zusammen, für die gilt: $A_i x = a_i$. Diese Zeilen entsprechen den sogenannten **aktiven (Ungleichungsrestriktionen-)Restriktionen**, die durch x mit Gleichheitszeichen erfüllt werden. **Zu beachten** ist dabei, dass die anderen Ungleichungsrestriktionen bei kleiner Abänderung von x nicht Gefahr laufen, verletzt zu werden.

Skizze:

Trifft man im Zuge der Bearbeitung der Aufgabe (ZA1) auf eine zulässige Lösung d mit *negativem* Zielfunktionswert, so hat man eine zulässige Abstiegsrichtung für (P) gefunden, existiert keine solche, ist x subminimal.

Die Bedeutung der Restriktion $\|d\| \leq 1$ in der Aufgabe (ZA1) ist die folgende: Hat man eine Lösung d der Bedingungen $A_1 d \leq 0, Hd = 0$ gefunden, so ist auch λd Lösung dieser Ungleichungen für jedes $\lambda > 0$. Dies ist insbesondere richtig für d mit $\nabla f(x)^T d < 0$, so dass im diesem Falle die Aufgabe (ZA1) ohne Beschränkung der Länge von d eine unbeschränkte Zielfunktion hätte.

Natürlich könnte man auf die Restriktion $\|d\| \leq 1$ verzichten, wenn man im Falle des Abbruchs des Simplexverfahrens wegen unbeschränkter Zielfunktion die entsprechende Kegelerzeugende d als Abstiegsrichtung nimmt (beachte:

der Nullpunkt ist einzige Ecke von (ZA1)). Nur ist deren Qualität in der Regel nicht so gut.

Alternativ kann man die Länge der Lösung d auch wie folgt begrenzen

$$(ZA2) \quad \begin{cases} \text{minimiere} & \nabla f(x)^T d \\ \text{s.d.} & A_1 d \leq 0, Hd = 0 \\ & \nabla f(x)^T d \geq -1 \end{cases}$$

Beide Aufgaben zur Berechnung einer zulässigen Abstiegsrichtung gehen auf ZOUTENDIJK zurück. Wir werden später auf sie genauer zu sprechen kommen.

2.2.2 KKT-Punkte

Eine weitere *notwendige* Eigenschaft lokaler Minimierer liefern die KARUSH-KUHN-TUCKER Bedingungen.

Satz 2.5 *Sei x^* ein lokaler Minimierer der Aufgabe (P). Dann existieren Lagrange-Multiplikatoren $\lambda^* \in \mathbf{R}^m$ und $\mu^* \in \mathbf{R}^p$ so, dass (x^*, λ^*, μ^*) den **KKT-Bedingungen** von (P) genügt:*

$$(KKT) \quad \begin{cases} \nabla f(x) + A^T \lambda + H^T \mu = 0, \lambda \geq 0 \\ Hx = h, Ax + y = a, y \geq 0 \\ 0 = \lambda^T y \end{cases}$$

Beweis: [GK2] Satz 2.42

Wir nennen einen Punkt $x \in M$ **KKT-Punkt**, wenn er zusammen mit geeigneten $\lambda \in \mathbf{R}^m$, $\mu \in \mathbf{R}^p$ die Bedingungen (KKT) erfüllt.

Bemerkung: Wird die Zielfunktion f der Aufgabe (P) in einem lokalen Minimierer x^* linearisiert, d.h. durch ihre lineare Approximation

$$T_f(x) = f(x^*) + \nabla f(x^*)^T (x - x^*)$$

ersetzt, so sind zulässigen Richtungen in diesem Punkte bzgl der Originalaufgabe und der linearisierten Aufgabe genau die gleichen. Wegen $\nabla T_f(x) = \nabla f(x)$ stimmen zulässigen Abstiegsrichtungen und sogar die KKT Bedingungen beider Aufgaben in x^* überein.

Zu beachten ist, dass die KKT-Bedingungen sich auf ein reines Gleichungssystem reduzieren, wenn keine Ungleichungsbedingungen gegeben sind. In diesem Fall können also KKT-Punkte durch Lösen eines (nichtlinearen) Gleichungssystems ermittelt werden (siehe unten).

Auch die KKT-Bedingungen charakterisieren lokale Optimalität für Punkte aus (P) nur notwendig, sie sind nicht hinreichend, wie wiederum lokale Maximierer zeigen. Immerhin gilt: KKT-Punkte sind subminimal.

Denn sei etwa $x \in M$ betrachtet und d eine in x zulässige Richtung. Dann gilt sicher $Hd = 0$ und $A_1 d \leq 0$, wobei $A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$ eine Aufteilung der Matrix A in aktive und inaktive Zeilen ist. Damit folgt aus den KKT-Bedingungen

$$\begin{aligned} \nabla f(x) &= -A_1^T \lambda_1 - A_2^T \lambda_2 - H^T \mu, \quad \lambda_1 \geq 0, \lambda_2 = 0 \Rightarrow \\ \nabla f(x)^T d &= -\lambda_1^T A_1 d - \lambda_2^T A_2 d - \mu^T Hd = -\lambda_1^T A_1 d \geq 0 \end{aligned}$$

Interessanterweise gilt sogar eine Äquivalenz der Begriffe "subminimaler Punkt" und "KKT-Punkt". (Dies gilt so i.a. nicht bei nichtlinearen Restriktionen)

Satz 2.6 $x \in M$ ist genau dann ein subminimaler Punkt von (P) , wenn x ein KKT-Punkt von (P) ist.

Beweis: [GK2], Satz 2.36, Satz 2.42 und die bereits oben gemachte Feststellung, dass KKT-Punkte subminimal sind.

Unter der zusätzlichen Voraussetzung der Konvexität der Zielfunktion sind die KKT-Bedingungen und damit die Subminimalität auch hinreichend für lokale Minimierer.

Satz 2.7 Die Zielfunktion der Aufgabe (P) sei konvex. Dann ist jeder KKT-Punkt der Aufgabe auch ein lokaler Minimierer.

Beweis: [GK2], Satz 2.46.

In Anbetracht der Tatsache, dass jeder lokale Minimierer der Aufgabe (P) mit konvexer Zielfunktion bereits ein globaler Minimierer ist und dass die KKT-Bedingungen bereits als notwendige Optimalitätsbedingungen erkannt wurden, gilt das

Korollar 2.8 Die Zielfunktion der Aufgabe (P) sei konvex. Dann ist ein Punkt $x \in \mathbf{R}^n$ genau dann ein (zulässiger globaler) Minimierer von (P) , wenn x ein KKT-Punkt bzw subminimal ist.

Bedingungen 2. Ordnung

In gewisser Weise sind die KKT Bedingungen noch zu schwach, um lokale Minima scharf zu charakterisieren. Diesem Ziel kommt man näher, wenn man zu den KKT Bedingungen noch weitere Bedingungen hinzunimmt. Während die KKT Bedingungen sich nur der nullten und ersten Ableitung von f bedienen, kommen durch Verwendung der zweiten Ableitung, also der Hessematrix von f , stärkere Bedingungen zustande:

Satz 2.9 *Es sei $x^* \in M$ ein lokaler Minimierer der Aufgabe (P) . Ist f zweimal stetig differenzierbar, so gelten neben den KKT Bedingungen in x^* auch die folgenden notwendigen Bedingungen zweiter Ordnung:*

$$d^T \nabla^2 f(x^*) d \geq 0 \quad \text{für alle } d \in \mathcal{K}(M, x^*) \text{ mit } \nabla f(x^*)^T d = 0$$

Beweis: [A], Satz 5.3.1

Dieser Satz lässt sich so interpretieren, dass in einem lokalen Minimierer die Funktion entlang einer zulässigen, zum steilsten Anstieg senkrecht stehenden Richtung konvex verlaufen muss.

Skizze:

Ähnlich kann lokale Minimalität auch hinreichend charakterisiert werden.

Satz 2.10 *Es sei $x^* \in M$ zur Aufgabe (P) beliebig vorgegeben. Ist f zweimal stetig differenzierbar und gelten neben den KKT Bedingungen in x^* auch die folgenden hinreichenden Bedingungen zweiter Ordnung:*

$$d^T \nabla^2 f(x^*) d \geq \alpha \|d\|^2 \quad \text{für alle } d \in \mathcal{K}(M, x^*) \text{ mit } \nabla f(x^*)^T d = 0$$

für ein $\alpha > 0$, so liegt in x^ ein strikter lokaler Minimierer zu (P) vor.*

Beweis: [A], Satz 5.3.3

Dieser Satz kann analog dem vorherigen interpretiert werden und auch die obige Skizze passt.

2.2.3 Fritz John Bedingungen

Ziel der Konstruktion von Verfahren zur Lösung einer Optimierungsaufgabe ist, Punkte zu ermitteln, von denen gezeigt werden kann, dass sie Minimierer der Aufgabe sind. Über die Optimalitätsbedingungen 2. Ordnung ist es uns fast gelungen, lokale Minimierer eindeutig zu charakterisieren. Leider erweisen sich diese Bedingungen in der Regel als viel zu scharf, als dass man sie für einen Lösungspunkt x^* des Verfahrens nachweisen könnte. So muss man sich meist damit begnügen, theoretisch nachzuweisen, dass x^* ein KKT Punkt ist, um dann im konkreten Einzelfall nachzuprüfen, ob tatsächlich ein (lokaler) Minimierer vorliegt.

Zuweilen erweisen sich sogar die KKT Bedingungen als zu scharf. Eine Abschwächung der KKT-Bedingungen sind die *Fritz John-Bedingungen*:

Satz 2.11 Sei x^* ein lokaler Minimierer der Aufgabe (P) . Dann existieren Lagrange-Multiplikatoren $\lambda^* \in \mathbf{R}^m$ und $\mu^* \in \mathbf{R}^p$ und ein $\tau \geq 0$ so, dass $(x^*, \lambda^*, \mu^*, \tau)$ den **Fritz John-Bedingungen** von (P) genügt:

$$(FJ) \quad \begin{cases} \tau \nabla f(x) + A^T \lambda + H^T \mu = 0, \lambda \geq 0 \\ Ax + y = a, y \geq 0 \\ 0 = \lambda^T y \\ (\tau, \lambda, \mu) \neq (0, 0, 0) \end{cases}$$

Natürlich ist jeder KKT-Punkt ein Fritz John-Punkt, das Umgekehrte gilt in der Regel nicht. Dies hängt u.U. sehr stark vom Format der Aufgabenstellung ab.

Beispiele:

1. Wir betrachten die Aufgabe

$$\min \{x_2 : x_1 + x_2 \leq 1, -x_1 - x_2 \leq -1, -x_2 \leq 0\}$$

ohne Gleichungsrestriktionen.

Skizze:

Zu dieser Aufgabe ist jeder zulässige Punkt ein Fritz John Punkt: Setzen wir $f(x_1, x_2) := x_2$, $g_1(x_1, x_2) := x_1 + x_2 - 1$, $g_2(x_1, x_2) := -x_1 - x_2 + 1$, $g_3(x_1, x_2) := -x_2$ so lauten die zugehörigen Gradienten

$$\begin{aligned}\nabla f(x_1, x_2) &= (0, 1), \quad \nabla g_1(x_1, x_2) = (1, 1), \\ \nabla g_2(x_1, x_2) &= (-1, -1), \quad \nabla g_3(x_1, x_2) = (0, -1)\end{aligned}$$

Für ein zulässiges $x = (x_1, x_2)$ gilt dann $g_1(x_1, x_2) = 0$ und $g_2(x_1, x_2) = 0$, die Fritz John Bedingungen werden daher erfüllt durch

$$\tau = 0, \lambda_1 = 1, \lambda_2 = 1, \lambda_3 = 0, y_1 = 0, y_2 = 0, y_3 = x_2 \geq 0.$$

Dagegen prüft man leicht, dass nur der Punkt $(x_1, x_2) = (1, 0)$ ein KKT-Punkt ist, nämlich z. B. mit den Parametern

$$\tau = 1, \lambda_1 = 1, \lambda_2 = 1, \lambda_3 = 1, y_1 = 0, y_2 = 0, y_3 = 0.$$

Dieser Punkt gibt offensichtlich ein globales Minimum an der Aufgabe an.

2. Formulieren wir nun die obige Aufgabe unter Zuhilfenahme einer Gleichungsrestriktion:

$$\min \{x_2 : x_1 + x_2 = 1, -x_2 \leq 0\}$$

Sie besitzt offenbar den gleichen zulässigen Bereich und dieselbe Zielfunktion wie die vorherige Aufgabe. Hier beschreibt sich mit den obigen Abkürzungen der zulässige Bereich durch $g_1(x_1, x_2) = x_1 + x_2 - 1 = 0$, $g_3(x_1, x_2) = -x_2 \leq 0$. Ist (x_1, x_2) ein zulässiger Punkt dieser Aufgabe, so muss in

$$\tau \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \lambda_3 \begin{pmatrix} 0 \\ -1 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \tau, \lambda_3 \geq 0, \quad (\tau, \lambda_3, \mu) \neq (0, 0, 0)$$

zwangsläufig $\mu = 0$ und $\tau = \lambda_3 > 0$ gelten. Für einen Fritz John Punkt muss dann ferner die Bedingung $\lambda_3 \cdot y_3 = 0$ erfüllt sein, also $y_3 = 0 = x_2$ gelten. Damit ist nur der Punkt $(x_1, x_2) = (1, 0)$ ein Fritz John-Punkt. Dieser ist gleichzeitig der einzige KKT-Punkt.

2.2.4 Spezialfall: Quadratische Optimierung

Betrachten wir wieder die Aufgabe (QP) und bestimmen wir zu dieser die KKT-Bedingungen.

Die Ableitung der Zielfunktion lautet

$$\nabla f(x) = Qx + p$$

daher ergeben sich die KKT-Bedingungen zu

$$(QKKT) \quad \begin{cases} Qx + p + A^T \lambda + H^T \mu = 0, \lambda \geq 0 \\ Hx = h, Ax + y = a, y \geq 0 \\ 0 = \lambda^T y \end{cases}$$

Zu beachten ist, dass dieses Gleichungs-/Ungleichungssystem in eine quadratische Optimierungsaufgabe in den Variablen $(x, y, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}^p$ umgeschrieben werden kann

$$(QPD) \quad \begin{cases} \text{minimiere} & \lambda^T y \\ \text{s.d.} & Qx + p + A^T \lambda + H^T \mu = 0, \lambda \geq 0 \\ & Hx = h, Ax + y = a, y \geq 0 \end{cases},$$

bei der die Variablen x und μ unbeschränkt bleiben.

Die Aufgabe (QPD) ist genau dann mit Zielwert null lösbar, wenn es einen KKT-Punkt zur Aufgabe (QP) gibt. Löst man also die Aufgabe (QPD) mit einem Verfahren und erhält einen Minimierer mit Zielwert null, so hat man damit einen KKT Punkt der Aufgabe (QP) gefunden.

Wenn die gegebene Aufgabe (QP) lösbar ist und damit einen KKT-Punkt besitzt, ist der optimale Zielfunktionswert der Aufgabe (QPD) gleich null.

Zu beachten ist, dass es Aufgaben (QP) gibt, die einen KKT-Punkt besitzen und die trotzdem nicht lösbar sind (Beispiel: minimiere $f(x) = -x^2$ über \mathbf{R}).

Auch wenn die Aufgabe (QP) konvex ist, muss sie nicht notwendig einen KKT-Punkt besitzen (z. B. bei linearer Zielfunktion). Ist allerdings der zulässige Bereich kompakt oder ist die Zielfunktion streng konvex, so besitzt die Aufgabe einen Minimierer und damit einen KKT-Punkt. In diesem Fall ist die Aufgabe (QPD) mit optimalem Zielwert 0 lösbar.

2.3 Dualität

Die KKT-Bedingungen der Aufgabe (P)

$$(KKT) \quad \begin{cases} \nabla f(x) + A^T \lambda + H^T \mu = 0, \lambda \geq 0 \\ Hx = h, Ax + y = a, y \geq 0 \\ 0 = \lambda^T y \end{cases}$$

zerfallen in drei Teile, die in jeweils einer Zeile gelistet sind. Die zweite Zeile entspricht den Bedingungen der Zulässigkeit der Aufgabe (P) , sie beschreibt die **Primale Zulässigkeit**. Die dritte Zeile benennt die **Komplementaritätsbedingungen (complementary slackness)**. Die erste Zeile hat zunächst keine Bedeutung. Allerdings besteht eine gewisse formale Ähnlichkeit zur zweiten Zeile. *Wir stellen die Frage*, ob die erste Zeile auch die Zulässigkeit einer Optimierungsaufgabe beschreibt, die in Verwandtschaft zur gegebenen Aufgabe (P) steht.

2.3.1 Lagrange Dual

Im Folgenden betrachten wir eine etwas allgemeinere Aufgabe als (P)

$$(NP) \quad \begin{cases} \text{minimiere} & f(x) \\ \text{s.d.} & g(x) \leq 0 \\ & h(x) = 0 \\ & x \in X \end{cases}$$

wobei $f : \mathbf{R}^n \rightarrow \mathbf{R}$, $g : \mathbf{R}^n \rightarrow \mathbf{R}^m$, $h : \mathbf{R}^n \rightarrow \mathbf{R}^p$ gegebene Funktionen sind und X eine gegebene nichtleere Teilmenge des \mathbf{R}^n . Die Menge X kann selbst ein Polyeder sein oder z.B. \mathbf{R}^n oder \mathbf{R}_+^n oder \mathbf{R}_{++}^n oder aber \mathbf{Z}^n .

Man beachte, dass die Wahl der Nebenbedingungen zur Beschreibung eines zulässigen Bereichs in keinster Weise eindeutig ist, der gleiche zulässige Bereich kann z.B. mit verschiedenen X beschrieben werden.

Wir bezeichnen (NP) als **primale Aufgabe**. Offensichtlich ist die Aufgabe (P) von diesem Typ. Wir können (NP) wie folgt eine **duale Aufgabe** zuordnen:

$$(ND) \quad \begin{cases} \text{maximiere} & \phi(u, v) \\ \text{s.d.} & u \geq 0 \end{cases}$$

wobei $\phi(u, v) := \inf \{f(x) + u^T g(x) + v^T h(x) : x \in X\}$ als **duale Lagrange Funktion** festgesetzt wurde. Man bezeichnet die Aufgabe

$$\inf \{f(x) + u^T g(x) + v^T h(x) : x \in X\}$$

auch als **Lagrange Relaxation** von (NP) , bei der die Restriktionen $g(x) \leq 0$ und $h(x) = 0$ **relaxiert**, d.h. als Nebenbedingungen entfernt und dafür mit Multiplikatoren in die Zielfunktion aufgenommen wurden. $\phi(u, v)$ gibt also den optimalen Zielwert dieser Aufgabe im Abhängigkeit der **Lagrange-Multiplikatoren** u und v an.

Man beachte aber, dass die Funktion ϕ nicht notwendig reellwertig ist, sie kann durchaus den Wert $(-\infty)$ annehmen.

Wir legen fest: Die duale Aufgabe besitzt keine zulässige Lösung, wenn $\phi(u, v) = -\infty$ gilt für *alle* (u, v) mit $u \geq 0$ und schreiben dies als $\sup \{\phi(u, v) : u \geq 0\} = -\infty$. Die duale Aufgabe ist unbeschränkt, wenn $\sup \{\phi(u, v) : u \geq 0\} = \infty$ gilt.

Beispiel: (vgl [BSS], Beispiel 6.1.1)

Wir betrachten die quadratische Aufgabe

$$\begin{array}{ll} \text{minimiere} & z = x_1^2 + x_2^2 \\ \text{s.d.} & x_1 + x_2 \geq 4 \\ & x_1, x_2 \geq 0 \end{array}$$

Skizze:

Diese Aufgabe hat offensichtlich die Optimallösung $(\bar{x}_1, \bar{x}_2) = (2, 2)$ mit Zielwert $\bar{z} = 8$.

Zur Dualisierung setzen wir $g(x_1, x_2) = -x_1 - x_2 + 4$ und $X = \{(x_1, x_2) : x_1, x_2 \geq 0\}$. Dann gilt

$$\begin{aligned} \phi(u) &= \inf \{x_1^2 + x_2^2 + u(-x_1 - x_2 + 4) : x_1, x_2 \geq 0\} \\ &= \inf \{x_1^2 - ux_1 : x_1 \geq 0\} + \inf \{x_2^2 - ux_2 : x_2 \geq 0\} + 4u \end{aligned}$$

Zur Lösung der Infimum-Aufgaben machen wir eine Fallunterscheidung:

Ist $u \geq 0$, so werden die Infima angenommen bei $x_1 = x_2 = \frac{u}{2}$, ist dagegen $u \leq 0$, so werden die Infima angenommen bei $x_1 = x_2 = 0$. Also gilt

$$\phi(u) = \begin{cases} -\frac{1}{2}u^2 + 4u, & \text{für } u \geq 0 \\ 4u, & \text{für } u \leq 0 \end{cases}$$

und die zur gegebenen Aufgabe duale Aufgabe lautet dementsprechend

$$\max \left\{ -\frac{1}{2}u^2 + 4u : u \geq 0 \right\},$$

deren Maximierer offenbar $\bar{u} = 4$ mit Zielwert $\phi(\bar{u}) = 8$ lautet.

Skizze:

■

2.3.2 Dualitätssätze

Zu bemerken ist, dass die optimalen Zielwerte der primalen und der dualen Aufgabe im letzten Beispiel übereinstimmen. Dies ist kein Zufall, wie die folgenden Ergebnisse zeigen.

Satz 2.12 (Schwache Dualität)

Sei \bar{x} eine zulässige Lösung der Aufgabe (NP) und (\bar{u}, \bar{v}) eine zulässige Lösung der Aufgabe (ND). Dann gilt

$$f(\bar{x}) \geq \phi(\bar{u}, \bar{v})$$

Beweis: Es gilt

$$\begin{aligned} \phi(\bar{u}, \bar{v}) &= \inf \{ f(x) + \bar{u}^T g(x) + \bar{v}^T h(x) : x \in X \} \\ &\leq f(\bar{x}) + \bar{u}^T g(\bar{x}) + \bar{v}^T h(\bar{x}) \leq f(\bar{x}) \end{aligned}$$

□

Man nennt die Differenz

$$D := \inf \{ f(x) : g(x) \leq 0, h(x) = 0, x \in X \} - \sup \{ \phi(u, v) : u \geq 0 \}$$

die **Dualitätslücke** zwischen (NP) und (ND). Diese ist nach dem Satz immer nichtnegativ, kann aber durchaus positiv sein.

Wir vereinbaren $D = 0$ auch für die pathologischen Fälle, dass die primale Aufgabe unbeschränkt und die duale leer ist

$$\inf \{ f(x) : g(x) \leq 0, h(x) = 0, x \in X \} = -\infty, \sup \{ \phi(u, v) : u \geq 0 \} = -\infty$$

oder dass die duale Aufgabe unbeschränkt und die primale leer ist

$$\sup \{ \phi(u, v) : u \geq 0 \} = \infty, \inf \{ f(x) : g(x) \leq 0, h(x) = 0, x \in X \} = \infty.$$

sowie $D = \infty$, wenn die primale und die duale Aufgabe leer sind

$$\inf \{ f(x) : g(x) \leq 0, h(x) = 0, x \in X \} = \infty, \sup \{ \phi(u, v) : u \geq 0 \} = -\infty$$

Korollar 2.13

1. *Besitzen beide Aufgaben, die primale Aufgabe (NP) und die duale Aufgabe (ND), zulässige Lösungen, so sind beide Zielfunktionen beschränkt. Haben beide Aufgaben zulässige Lösungen mit gleichem Zielfunktionswert, so sind diese beide auch Optimallösungen.*

2. Ist die primale Aufgabe (NP) zulässig, aber unbeschränkt, so besitzt die duale Aufgabe (ND) keine zulässige Lösung.
3. Ist die duale Aufgabe (ND) zulässig, aber unbeschränkt, so besitzt die primale Aufgabe (NP) keine zulässige Lösung.

Unter schärferen Voraussetzungen kann sogar eine stärkere Dualitätsaussage gemacht werden.

Satz 2.14 (Starke Dualität)

Vorgegeben sei die Aufgabe (NP) mit konvexer Zielfunktion f . Es seien X nichtleer und konvex und die Restriktionsfunktionen g_i konvex, die Restriktionsfunktionen h_j affin linear. Es gebe einen Vektor $\hat{x} \in X$, der zum relativen Inneren des zulässigen Bereichs gehört, d. h. die Eigenschaften

$$g(\hat{x}) < 0 \quad \text{und} \quad h(\hat{x}) = 0 \quad (\text{Slaterbedingung})$$

besitzt. Dann gilt für die Dualitätslücke

$$D = \inf \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\} - \sup \{\phi(u, v) : u \geq 0\} = 0$$

Ferner gilt: Ist das Infimum endlich, so ist die duale Aufgabe lösbar, d.h. es gibt eine Optimallösung (\bar{u}, \bar{v}) von (ND) mit

$$\phi(\bar{u}, \bar{v}) = \max \{\phi(u, v) : u \geq 0\}$$

Wird das Infimum sogar angenommen, etwa bei \bar{x} , so gilt $\bar{u}^T g(\bar{x}) = 0$ und \bar{x} ist primale Optimallösung.

Beweis: [BSS], Theorem 6.2.4 und anschließende Bemerkung; [GK2]; Satz 6.13.

Beachte: Gemäß der Bemerkung in [GK2] am Ende von Abschnitt 6.2.2 kann auf die Slaterbedingung verzichtet werden, wenn (wie bei Aufgabe (P)) die Restriktionen g_i ebenfalls linear sind und X durch endlich viele lineare Ungleichungen beschrieben wird.

Das folgende Beispiel zeigt, dass eine verschwindende Dualitätslücke auch bei nichtkonvexer Zielfunktion und nichttrivialem Dual möglich ist und wie das Dual von der Wahl von X abhängig ist.

Beispiel: Betrachten wir als Beispiel die Aufgabe

$$\begin{aligned} \min \quad & f(x) = -\frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 \\ \text{s.d.} \quad & -1 \leq x_1, x_2 \leq 2 \end{aligned}$$

mit den Unbekannten $x_1, x_2 \in \mathbf{R}$. Die Zielfunktion ist offensichtlich nicht konvex über dem zulässigen Bereich. Der eindeutige Minimierer der Aufgabe liegt bei $x_1 = 2, x_2 = 0$ mit $f(2) = -2$, wie man leicht einsieht, da beide Variablen unabhängig variiert werden können (die Aufgabe ist *separabel*).

1. Zur Konstruktion einer dualen Aufgabe wählen zunächst wir $X = \mathbf{R}^2$. Damit beschreibt sich das Lagrange Dual über

$$\begin{aligned} f(x) &= \frac{1}{2}x^T Q x + p^T x \quad \text{mit } Q = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, p = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \text{und } Ax - a &\leq 0 \quad \text{mit } A = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}, a = \begin{pmatrix} 2 \\ 1 \\ 2 \\ 1 \end{pmatrix}, \end{aligned}$$

zu $\max \{ \phi(\lambda) : \lambda \geq 0 \}$ mit

$$\begin{aligned} \phi(\lambda) &= \inf \left\{ \frac{1}{2}x^T Q x + p^T x + \lambda^T (Ax - a) : x \in X \right\} \\ &= \inf \left\{ \frac{1}{2}x^T \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} x + \lambda^T \left(\begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix} x - \begin{pmatrix} 2 \\ 1 \\ 2 \\ 1 \end{pmatrix} \right) : x \in \mathbf{R}^2 \right\} \\ &= \inf \left\{ \begin{aligned} &-\frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + \lambda_1(x_1 - 2) + \lambda_2(-x_1 - 1) \\ &+ \lambda_3(x_2 - 2) + \lambda_4(-x_2 - 1) : x_1, x_2 \in \mathbf{R} \end{aligned} \right\} \end{aligned}$$

Die Funktion $\varphi(x_1) := -\frac{1}{2}x_1^2 + \lambda_1(x_1 - 2) + \lambda_2(-x_1 - 1)$ stellt bei festen λ_1, λ_2 entlang der x_1 -Achse eine nach unten geöffnete Parabel dar, daher ist $\inf \varphi(x_1) = -\infty$, unabhängig von der konkreten Wahl der λ ! Die duale Aufgabe ist also leer und die Dualitätslücke ∞ .

2. Nun wählen wir $X = [-1, 2] \times \mathbf{R}$ zur Konstruktion des Duals. Dann ist

$\phi(\lambda)$ gegeben durch

$$\begin{aligned}\phi(\lambda) &= \inf \left\{ -\frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + \lambda_3(x_2 - 2) + \lambda_4(-x_2 - 1) : x_1 \in [-1, 2], x_2 \in \mathbf{R} \right\} \\ &= \inf \left\{ -\frac{1}{2}x_1^2 : x_1 \in [-1, 2] \right\} + \inf \left\{ \frac{1}{2}x_2^2 + \lambda_3x_2 - \lambda_4x_2 : x_2 \in \mathbf{R} \right\} - 2\lambda_3 - \lambda_4\end{aligned}$$

Das erste Infimum wird bei $x_1 = 2$ angenommen, ist also gleich -2 . Die zweite Funktion beschreibt eine nach oben geöffnete Parabel, deren Minimum am Scheitelpunkt angenommen wird. Wir setzen die Ableitung gleich null

$$x_2 + \lambda_3 - \lambda_4 = 0 \implies x_2 = \lambda_4 - \lambda_3$$

Einsetzen in die Funktion ergibt den minimalen Funktionswert in Abhängigkeit von λ_3, λ_4

$$\begin{aligned}&\frac{1}{2}(\lambda_4 - \lambda_3)^2 + \lambda_3(\lambda_4 - \lambda_3) - \lambda_4(\lambda_4 - \lambda_3) \\ &= -\frac{1}{2}\lambda_3^2 - \frac{1}{2}\lambda_4^2 + \lambda_3\lambda_4\end{aligned}$$

Damit ergibt sich insgesamt als Dualaufgabe

$$\max \left\{ -\frac{1}{2}\lambda_3^2 - \frac{1}{2}\lambda_4^2 + \lambda_3\lambda_4 - 2\lambda_3 - \lambda_4 - 2 : \lambda_3, \lambda_4 \geq 0 \right\}$$

Zur Bestimmung des Maximums leiten wir die Funktion ab und setzen null

$$\begin{aligned}\text{nach } \lambda_3 &: & -\lambda_3 + \lambda_4 - 2 &= 0 \\ \text{nach } \lambda_4 &: & -\lambda_4 + \lambda_3 - 1 &= 0\end{aligned}$$

Dieses System ist offensichtlich nicht zu erfüllen. Also kann das Maximum nur auf dem Rand angenommen werden. Wir setzen zunächst $\lambda_3 = 0$ in die Funktion ein und erhalten die Funktion $-\frac{1}{2}\lambda_4^2 - \lambda_4 - 2$. Deren Ableitung ist $-\lambda_4 - 1$. Dies ist null für $\lambda_4 = -1$, welches außerhalb des zulässigen Bereichs liegt. Auf dem zulässigen Rand fällt die Funktion ab bei steigenden λ_4 .

Nun betrachte noch $\lambda_4 = 0$. Dies ergibt als Funktion auf dem Rand $-\frac{1}{2}\lambda_3^2 - 2\lambda_3 - 2$. Diese Funktion leitet sich ab zu $-\lambda_3 - 2$. Nullsetzen

ergibt den Wert $\lambda_3 = -2$, was wiederum ausserhalb des zulässigen Bereichs liegt. Auf dem zulässigen Rand ist die Funktion abfallend.

Also wird das Maximum bei $\lambda_3 = \lambda_4 = 0$ angenommen mit Funktionswert -2 . Damit haben wir als Dualitätslücke $D = 0$ gefunden.

2.3.3 Spezialfall: Quadratische Optimierung

Betrachten wir wieder die quadratische Optimierungsaufgabe (QP) . Dann kann man der gegebenen Aufgabe auch eine Umformulierung der dualen Aufgabe zuordnen, die *ebenfalls quadratisch* ist. Um dies einzusehen unterstellen wir, dass bei der Festlegung von (QP) im Format (NP) : $X = \mathbf{R}^n$ gewählt wird. Der zulässige Bereich von (QP) sei *nichtleer*.

Die duale Aufgabe

Zunächst lautet die duale Aufgabe gemäß der Lagrange Dualität:

$$(ND) \quad \begin{cases} \text{maximiere} & \phi(\lambda, \mu) \\ \text{s.d.} & \lambda \geq 0 \end{cases}$$

mit $\phi(\lambda, \mu) = \inf \{ p^T x + \frac{1}{2} x^T Q x + \lambda^T (Ax - a) + \mu^T (Hx - h) : x \in \mathbf{R}^n \}$.

Es gilt nun das folgende

Lemma 2.15 *Es sei $f(x) = \frac{1}{2} x^T Q x + p^T x + q$ eine gegebene quadratische Funktion. Dann gilt:*

1. *Genau dann besitzt f einen Minimierer, wenn f konvex und nach unten beschränkt ist.*
2. *Ist die Funktion konvex und nach unten beschränkt, so haben alle stationären Punkte von f den gleichen (minimalen) Zielfunktionswert.*
3. *Ist die Funktion f nicht konvex, so ist nach unten unbeschränkt.*

Beweis: Wir unterwerfen f einer Hauptachsentransformation mit einer orthogonalen Matrix B , gebildet aus Eigenvektoren von Q . Setze also $x = B\bar{x}$. Dann ist

$$\begin{aligned} f(B\bar{x}) &= \frac{1}{2} (B\bar{x})^T Q B\bar{x} + p^T B\bar{x} + q \\ &= \frac{1}{2} \bar{x}^T \bar{Q} \bar{x} + (B^T p)^T \bar{x} + q \end{aligned}$$

mit Diagonalmatrix $\bar{Q} := B^T Q B$. Nachträglich können wir noch über quadratische Ergänzung erreichen, dass die Koeffizienten von $B^T p$ zumindest dann null sind, wenn die entsprechenden Eigenwerte von \bar{Q} ungleich null sind. Damit können wir oBdA davon ausgehen, dass f in einem geeigneten gedrehten und verschobenen Koordinatensystem von der eingangs gegebenen Form ist, nur dass Q eine Diagonalmatrix ist und die Koeffizienten von p höchstens dann ungleich null sind, wenn die entsprechenden Diagonalelemente von Q null sind.

Berechnet man nun den Gradienten und setzt ihn gleich null, so lauten die n Gleichungen entweder

$$p_i = 0 \quad \text{oder} \quad x_i = 0$$

je nachdem, ob der Diagonalwert von Q gleich null ist oder nicht. Eine Lösung des Systems kann es nur geben, wenn alle betroffenen p_i tatsächlich null sind. In diesem Fall haben alle stationären Punkte den gleichen Funktionswert und die stationären Punkte sind genau dann Minima, wenn f konvex ist.

□

Betrachten wir für vorgegebenes (λ, μ) die in x quadratische Funktion

$$\varphi(x) := p^T x + \frac{1}{2} x^T Q x + \lambda^T (Ax - a) + \mu^T (Hx - h)$$

Aus dem Lemma können wir folgern: Die Funktion φ nimmt nur dann an einer Stelle \bar{x} ein Minimum an, wenn sie konvex und nach unten beschränkt ist. Besitzt sie einen Minimierer, so nimmt sie für alle Minimierer denselben Wert an. **Zu beachten** ist, dass φ genau dann konvex ist, wenn die Ausgangsfunktion f konvex ist.

Es gilt nun bei vorgegebenen (λ, μ) : entweder das Infimum wird angenommen bei einem \bar{x} oder es gilt $\phi(\lambda, \mu) = -\infty$. \bar{x} erfüllt dann als Minimierer der Funktion $\varphi(x) = p^T x + \frac{1}{2} x^T Q x + \lambda^T (Ax - a) + \mu^T (Hx - h)$ die KKT-Bedingung für φ : $Q\bar{x} + p + A^T \lambda + H^T \mu = 0$.

Will man die duale Aufgabe über KKT-Bedingungen an die Funktion φ beschreiben, so sollte man Q als positiv semidefinit und die Zielfunktion der Aufgabe (QP) als beschränkt voraussetzen. Dann nämlich besitzt die duale Aufgabe nach dem starken Dualitätssatz eine Optimallösung und es gilt $\phi(\lambda, \mu) > -\infty$ für mindestens ein zulässiges (λ, μ) . Andererseits: Ist f konvex und über dem zulässigen Bereich von (QP) unbeschränkt, so ist die duale Aufgabe nach dem schwachen Dualitätssatz leer, d.h. die Funktion φ besitzt für kein zulässiges (λ, μ) ein Minimum, also auch keinen KKT-Punkt.

Ist also die Funktion f als konvex **vorausgesetzt**, so beschreibt sich die duale Aufgabe in jedem Fall, jetzt einschließlich der Variablen x , gleichwertig zu

$$(QD) \quad \begin{cases} \text{maximiere} & w(x, \lambda, \mu) := p^T x + \frac{1}{2} x^T Q x + \lambda^T (Ax - a) + \mu^T (Hx - h) \\ \text{s.d.} & p + Qx + A^T \lambda + H^T \mu = 0, \lambda \geq 0 \end{cases}$$

Hierin sind die Variablen x und μ nichtvorzeichenbeschränkt! Wir nennen (QD) die **Standardform der dualen Aufgabe zu (QP)** .

Zu beachten ist: Besitzt die duale Aufgabe eine Optimallösung (x, λ, μ) , so ist *nicht zwangsläufig* davon auszugehen, dass x dabei primal zulässig ist. Es stellt sich spontan die **Frage**, ob eine primale Aufgabe (QP) existiert, deren Zielfunktion konvex und nach unten beschränkt ist und die trotzdem keinen KKT-Punkt besitzt. Wir werden später sehen, dass der Korrektheitsbeweis zum Verfahren von Wolfe hierauf die Antwort "nein" gibt. Ist also (QD) lösbar, so besitzt (QP) zwar eine Optimallösung, diese muss aber nicht mit der x -Komponente der Optimallösung von (QD) übereinstimmen.

Die primal-duale Aufgabe

Mit dieser Form der dualen Aufgabe ist, zumindest im Fall quadratischer Optimierungsaufgaben mit positiv semidefiniter Matrix Q , unsere eingangs gestellte Frage positiv beantwortet: die erste Zeile der KKT-Bedingungen für (QP) beschreibt die Zulässigkeit der dualen Aufgabe im neuen Format.

Es sei Q als positiv semidefinit *vorausgesetzt*. Betrachten wir nun noch einmal die Aufgabe

$$(QPD) \quad \begin{cases} \text{minimiere} & \lambda^T y \\ \text{s.d.} & Qx + p + A^T \lambda + H^T \mu = 0, \quad \lambda \geq 0 \\ & Hx = h, \quad Ax + y = a, \quad y \geq 0 \end{cases}$$

die wir nun **Standardform der primal-dualen Aufgabe zu (QP)** nennen.

Im Gegensatz zu dem, was wir eben über die duale Aufgabe festgestellt haben gilt: Ist (x, λ, μ) ein zulässiger Punkt zu dieser Aufgabe, so ist x zulässig zur primalen Aufgabe und (x, λ, μ) zulässig zur dualen Aufgabe. Die Zielfunktion $\lambda^T y$ erhält folgende Bedeutung:

Sei (x, λ, μ) ein zulässiger Punkt zu (QPD) . Dann gilt

$$\begin{aligned} & f(x) - w(x, \lambda, \mu) \\ = & p^T x + \frac{1}{2} x^T Q x - \left(p^T x + \frac{1}{2} x^T Q x + \lambda^T (Ax - a) + \mu^T (Hx - h) \right) \\ = & -\lambda^T (Ax - a) = \lambda^T y \end{aligned}$$

wegen $Hx = h$ und $y = a - Ax$. Die Zielfunktion gibt also die **Dualitätslücke** zwischen der primalen Lösung x und der dualen Lösung (x, λ, μ) an. Diese ist wegen der geltenden starken Dualität immer null. Damit sind die Optimallösungen der Aufgabe (QPD) genau die KKT-Punkte der primalen Aufgabe (QP) und entsprechen somit genau deren Optimallösungen.

Interpretation: (QPD) entsteht durch "Zusammenführung" der primalen und der dualen Aufgabe: Zunächst wird die Zielfunktion der dualen Aufgabe mit (-1) multipliziert, damit diese auch als Minimierungsaufgabe formuliert ist. Löst man nun beide Aufgaben simultan (bei gleichem x) mit der Summe der Einzelzielfunktionen als neuer Zielfunktion, so entsteht (QPD) .

Dualitätsaussagen

Ist die Zielfunktion f *konvex*, ist also Q positiv semidefinit, so sind die Voraussetzungen des starken Dualitätssatzes (mit $X = \mathbf{R}^n$) erfüllt und es gilt, dass die Dualitätslücke gleich null ist. In diesem Fall gilt *genau eine* der folgenden Aussagen:

1. die primale Aufgabe besitzt keine zulässige Lösung.
2. die primale Aufgabe besitzt eine zulässige Lösung und ihre Zielfunktion ist unbeschränkt. Dann ist die duale Aufgabe ist leer.
3. die primale Aufgabe besitzt einen KKT-Punkt. Dann besitzen beide Aufgaben, die primale, (QP) und die duale, (QD) , eine Optimallösung mit gleichem Zielfunktionswert sowie gleichem x und die primal-duale Aufgabe (QPD) besitzt eine Optimallösung mit Zielwert null. Jede Optimallösung der primal-dualen Aufgabe gibt gleichzeitig eine primale und eine duale Optimallösung an.

Alternative Formulierungen

Es sei f wieder als *konvex* vorausgesetzt.

1. Setzt man in (QD) die Nebenbedingung in die Zielfunktion ein, so ergibt sich die duale Aufgabe zu (QP) im Format

$$(QD_1) \quad \begin{cases} \text{maximiere} & w = -\frac{1}{2}x^T Qx - a^T \lambda - h^T \mu \\ \text{s.d.} & Qx + p + A^T \lambda + H^T \mu = 0, \quad \lambda \geq 0 \end{cases}$$

Man beachte, dass für $Q = 0$ die Dualität linearer Optimierungsaufgaben beschrieben wird. Hervorzuheben ist: Die Zielfunktion dieser Aufgabe ist *konkav*, da die Originalzielfunktion f konvex ist. Multipliziert man sie mit (-1) , so entsteht wieder eine *konvexe quadratische Minimierungsaufgabe*.

2. Wenn \bar{x} Optimallösung des primalen Problems ist, so ist \bar{x} auch der x -Teil einer Optimallösung des dualen Problems. Die Umkehrung dieser Aussage gilt i.A. nicht, d.h. der x -Teil einer dualen Optimallösung muss nicht primal zulässig sein.

Wenn Q *positiv definit* (invertierbar) ist, so besitzt die primale Aufgabe eine eindeutige Optimallösung und die duale Aufgabe mindestens eine Optimallösung, die in ihrem x Teil wegen $x = -Q^{-1}(p + A^T \lambda + H^T \mu)$ bei festen λ und μ eindeutig ist.

Die letzte Gleichung kann man dann verwenden, um der dualen Aufgabe noch ein anderes Format zu geben. In diesem Fall lautet die Zielfunktion der dualen Ausgabe

$$\begin{aligned} w &= -\frac{1}{2}x^T Qx - a^T \lambda - h^T \mu \\ &= -\frac{1}{2}(p + A^T \lambda + H^T \mu)^T Q^{-1} (p + A^T \lambda + H^T \mu) - a^T \lambda - h^T \mu \end{aligned}$$

so dass die duale Aufgabe nun die Form hat

$$(QD_2) \quad \max \left\{ -\frac{1}{2}(p + A^T \lambda + H^T \mu)^T Q^{-1} (p + A^T \lambda + H^T \mu) - a^T \lambda - h^T \mu : \lambda \geq 0 \right\}$$

Es ist klar, dass die Zielfunktion nach Ausmultiplizieren wieder quadratisch in $\begin{pmatrix} \lambda \\ \mu \end{pmatrix}$ ist. Zu beachten ist, dass diese Formulierung der Dualen Aufgabe sehr einfache Nebenbedingungen besitzt.

3. Zur Vereinfachung der weiteren Analyse von (QD_2) betrachten wir den Spezialfall, dass in (QP) keine Gleichungsrestriktionen $Hx = h$ auftreten. Man kann jede primale Aufgabe (QP) auf diese Form bringen.

Dann gilt für die duale Aufgabe

$$(QD_2) \quad \max \left\{ -\frac{1}{2}(p + A^T \lambda)^T Q^{-1} (p + A^T \lambda) - a^T \lambda : \lambda \geq 0 \right\}$$

Der quadratische Term innerhalb der Klammern schreibt sich um zu

$$\begin{aligned} &(p + A^T \lambda)^T Q^{-1} (p + A^T \lambda) \\ &= p^T Q^{-1} p + 2p^T Q^{-1} A^T \lambda + \lambda^T A Q^{-1} A^T \lambda \end{aligned}$$

Setzt man zur Abkürzung

$$G := A Q^{-1} A^T \quad \text{und} \quad g := A Q^{-1} p + a$$

so lautet, bei Vernachlässigung des konstanten Terms die duale Aufgabe nun

$$(QD_2) \quad \min \left\{ \frac{1}{2} \lambda^T G \lambda + g^T \lambda : \lambda \geq 0 \right\}$$

Aus der Form von G ist klar, dass die Aufgabe (QD_2) eine konvexe Zielfunktion besitzt, da G positiv semidefinit ist. Besitzt A maximalen Zeilenrang, so ist die Aufgabe sogar streng konvex.

Ferner lauten die ursprünglichen KKT Bedingungen von (QP)

$$(QKKT) \quad \begin{cases} Qx + p + A^T \lambda = 0, \lambda \geq 0 \\ Ax + y = a, y \geq 0 \\ 0 = \lambda^T y \end{cases}$$

Setzt man

$$x = -Q^{-1} (p + A^T \lambda) \quad (2.2)$$

auch in die KKT Bedingungen ein, so ergibt sich

$$(QKKT) \quad \begin{cases} \lambda \geq 0 \\ -G\lambda - g + y = 0, y \geq 0 \\ 0 = \lambda^T y \end{cases}$$

Auf der anderen Seite lauten die KKT Bedingungen von (QD_2) in der letzten Formulierung

$$(QD_2KKT) \quad \begin{cases} G\lambda + g - y = 0, y \geq 0 \\ \lambda \geq 0 \\ 0 = y^T \lambda \end{cases}$$

woraus man $(QKKT) = (QD_2KKT)$ erkennt. Da die KKT Bedingungen bei konvexen Aufgaben notwendige und hinreichende Optimalitätsbedingungen sind, ist klar, dass $\lambda = \lambda^*$ genau dann Optimallösung von (QD_2) ist, wenn $x = x^* = -Q^{-1} (p + A^T \lambda^*)$ Optimallösung von (QP) ist. Da Q positiv definit ist, liefert die Formel (2.2) für alle Optimallösungen λ^* der dualen Aufgabe denselben x -Wert x^* .

Folgerung: Es kann also unter der Voraussetzung Q positiv definit alternativ zur primalen Aufgabe auch die duale Aufgabe gelöst werden und das Ergebnis über (2.2) in eine Optimallösung der primalen Aufgabe zurückgerechnet werden. Wegen der besonders einfachen Nebenbedingungen der dualen Aufgabe ist dies sehr attraktiv.

4. Betrachten wir nun eine andere spezielle Form der Ausgangsaufgabe (QP) . Es werde angenommen, dass keine expliziten Gleichungsrestriktionen auftreten und dass alle Variablen vorzeichenbeschränkt sind, was

explizit ausgewiesen sei

$$(QPS) \quad \begin{cases} \text{minimiere} & f(x) = \frac{1}{2}x^T Qx + p^T x \\ \text{s.d.} & Ax \leq a, x \geq 0 \end{cases}$$

Auch dies ist natürlich keine Beschränkung der Allgemeinheit, da die Aufgabe (QP) gleichwertig in (QPS) umgeschrieben werden kann (siehe auch nächste Kapitel)

Die KKT-Bedingungen zu (QPS) lauten

$$(QPSKKT) \quad \begin{cases} Qx + p + A^T \lambda - \mu = 0, \lambda, \mu \geq 0 \\ Ax + y = a, x, y \geq 0 \\ 0 = \lambda^T y + \mu^T x \end{cases}$$

Zu beachten ist, dass hier *sämtliche Variablen vorzeichenbeschränkt* sind.

In Matrixform schreibt sich $(QPSKKT)$ zu

$$\begin{pmatrix} \mu \\ y \end{pmatrix} - \begin{pmatrix} Q & A^T \\ -A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} p \\ a \end{pmatrix}$$

$$\begin{pmatrix} x \\ \lambda \end{pmatrix}^T \begin{pmatrix} \mu \\ y \end{pmatrix} = 0, \begin{pmatrix} x \\ \lambda \end{pmatrix} \geq 0, \begin{pmatrix} \mu \\ y \end{pmatrix} \geq 0$$

Diese Formulierung ist vom Format eines **Linear Complementary Problem (LCP)**

$$\begin{aligned} w - Bz &= q \\ w &\geq 0, z \geq 0 \\ w^T z &= 0 \end{aligned}$$

(mit quadratischer Matrix B). Zur Lösung eines LCP gibt es eigene Verfahren, die sich also auf (QP) anwenden lassen. Wir werden später darauf zurückkommen.

Teil I

Simplexverfahren

Kapitel 3

Dreiphasenalgorithmen

Der zulässige Bereich der uns interessierenden Optimierungsaufgaben ist ein konvexes Polyeder. Dessen Eckpunkte sind prinzipiell mit Simplexschritten erreichbar. Für Aufgaben mit linearer Zielfunktion ist dies vorteilhaft, denn bekanntermaßen liegt dann in der Menge der Optimallösungen immer eine Ecke.

Bei nichtlinearer Zielfunktion ist dem i. a. nicht so. Dennoch sind auch in diesem Fall die Optimallösungen mit dem Simplexverfahren ansprechbar, z. B. dann, wenn die Optimallösungen als Schnittpunkte von linearen Restriktionen mit nichtvorzeichenbeschränkten Schlupfvariablen beschrieben werden. Wie dies funktioniert, sollen die folgenden Überlegungen zeigen.

Die Techniken des Simplexverfahrens sind, wie sich zeigen wird, nicht auf die lineare Optimierung beschränkt. Wir werden im Folgenden simplexartige Verfahren sowohl für die quadratische Optimierung als auch für die linear restrinigierte Optimierung erarbeiten.

3.1 Lineare Techniken

Viele Verfahren der nichtlinearen Optimierung nutzen das Simplexverfahren als Werkzeug. Dies ist erst recht der Fall, wenn die Restriktionen der gegebenen nichtlinearen Aufgabe linear sind.

3.1.1 Transformation auf Standardformat

In der linearen Optimierung sind verschiedene Standardformate für den zulässigen Bereich üblich. Diese vereinfachen die Beschreibung einschlägiger Verfahren, z.B. des Simplexverfahrens. Das trifft auch zu für die linear restringierte, nichtlineare Optimierung. Deshalb sollen solche Standardformate angesprochen werden.

Wichtig sind vorallem zwei Formate

$$\begin{aligned} (SPA) \quad & \begin{cases} \text{minimiere} & f(x) \\ \text{s.d.} & Ax \leq a, x \geq 0 \end{cases} \quad \text{und} \\ (SPB) \quad & \begin{cases} \text{minimiere} & f(x) \\ \text{s.d.} & Hx = h, x \geq 0 \end{cases} \end{aligned}$$

wobei die Matrix H maximalen Zeilenrang habe. Bei der Transformation auf eines dieser beiden Formate ist Hauptaugenmerk auf die Transformation der (nicht-) linearen Zielfunktion zu legen.

Vorgegeben sei wieder die Aufgabe

$$(P) \quad \begin{cases} \text{minimiere} & f(x) \\ \text{s.d.} & Ax \leq a, Hx = h \end{cases}$$

Eine erste Möglichkeit, diese Aufgabe auf das erste Standardformat zu bringen ist, zunächst die Gleichungen $Hx = h$ in zwei Ungleichungen unzuschreiben: $Hx \leq h$, $-Hx \leq -h$ und dann die noch vorzeichen-unbeschränkten Variablen über die Variablentransformation $x =: x^1 - x^2$ mit $x^1, x^2 \geq 0$ durch vorzeichenbeschränkte Variablen zu ersetzen. Dies vergrößert die Anzahl der Variablen und der Restriktionen erheblich. Die Aufgabe lautet dann in der ersten Standardform

$$(SP1) \quad \begin{cases} \text{minimiere} & g(x^1, x^2) = f(x^1 - x^2) \\ \text{s.d.} & \begin{pmatrix} A & -A \\ H & -H \\ -H & H \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \leq \begin{pmatrix} a \\ h \\ -h \end{pmatrix}, \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \geq 0 \end{cases}$$

deren Systemmatrix die Dimension $(m + 2p) \times (2n)$ hat.

Eine nachträgliche Einführung von Schlupfvariablen liefert auch die zweite Standardform

$$(SP2) \quad \left\{ \begin{array}{l} \text{minimiere} \quad g(x^1, x^2, y^1, y^2, y^3) = f(x^1 - x^2) \\ \text{s.d.} \quad \begin{pmatrix} A & -A \\ H & -H \\ -H & H \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} + \begin{pmatrix} y^1 \\ y^2 \\ y^3 \end{pmatrix} = \begin{pmatrix} a \\ h \\ -h \end{pmatrix} \\ \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \geq 0, \quad \begin{pmatrix} y^1 \\ y^2 \\ y^3 \end{pmatrix} \geq 0 \end{array} \right.$$

in der Dimension $(m + 2p) \times (2n + m + 2p)$.

Besitzt H von vorneherein maximalen Zeilenrang, so besteht eine zweite Möglichkeit der Transformation darin, *zunächst* Schlupfvariablen einzuführen, um nur noch Gleichungsrestriktionen zur Verfügung zu haben. Danach werden alle noch nicht vorzeichenbeschränkten Variablen durch zwei vorzeichenbeschränkte Variablen ersetzt. Dies liefert das zweite Standardformat

$$(SP3) \quad \left\{ \begin{array}{l} \text{minimiere} \quad g(x^1, x^2, y) = f(x^1 - x^2) \\ \text{s.d.} \quad \begin{pmatrix} A & -A & I \\ H & -H & 0 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ y \end{pmatrix} = \begin{pmatrix} a \\ h \end{pmatrix}, \quad \begin{pmatrix} x^1 \\ x^2 \\ y \end{pmatrix} \geq 0 \end{array} \right.$$

mit Systemmatrix die Dimension $(m + p) \times (2n + m)$, welches durch Umschreiben der Gleichungen in Ungleichungen ins erste Standardformat transferiert werden kann

$$(SP4) \quad \left\{ \begin{array}{l} \text{minimiere} \quad g(x^1, x^2, y) = f(x^1 - x^2) \\ \text{s.d.} \quad \begin{pmatrix} A & -A & I \\ H & -H & 0 \\ -A & A & -I \\ -H & H & 0 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ y \end{pmatrix} \leq \begin{pmatrix} a \\ h \\ -a \\ -h \end{pmatrix}, \quad \begin{pmatrix} x^1 \\ x^2 \\ y \end{pmatrix} \geq 0 \end{array} \right.$$

Dimension $(2m + 2p) \times (2n + m)$.

Eine dritte Möglichkeit, die (fast) ohne Ausweitung der Anzahl der Variablen und Restriktionen auskommt, beruht auf der Anwendung der ersten Phasen des **Dreiphasenalgorithmus der Linearen Programmierung**. Dieser werde im Folgenden kurz dargestellt.

3.1.2 Linearer Dreiphasenalgorithmus

Gegeben sei eine allgemeine lineare Aufgabenstellung der folgenden Form mit expliziten Vorzeichenbeschränkungen

$$(LP) \quad \begin{cases} \min & f(x) = c^T x + q \\ s.d. & Ax \leq a \\ & Hx = h \\ & x_j \geq 0, \quad j \in J \end{cases}$$

wobei $A \in \mathbf{R}^{m_1 \times n}$, $H \in \mathbf{R}^{m_2 \times n}$ und J eine vorgegebene Teilmenge von $\{1, \dots, n\}$ ist. Es sind also ein Teil der Variablen x_j explizit vorzeichenbeschränkt, nämlich die mit $j \in J$. Die restlichen Variablen x_j , $j \in J$, sind nicht explizit vorzeichenbeschränkt. Wir nennen sie **freie Variablen**.

Wir bringen diese Aufgabenstellung nach Einführung von Schlupfvariablen in ein (zunächst i. a. nicht zulässiges) **verkürztes Starttableau** ein

$$\begin{array}{c|c|c} & x^T & \min \\ \hline f & -c^T & q \\ y & A & a \\ 0 & H & h \end{array}$$

das wie gewohnt zu lesen ist (der Doppelstrich symbolisiert das Gleichheitszeichen in den Zeilen, die Kopf-Nullen in der Basis dürfen als *künstliche Variablen mit Wert 0* interpretiert werden). Es gilt als vereinbart, dass die Vorzeichenbeschränkungen der x_j , $j \in J$, nicht als explizite Restriktionen in das Tableau eingebracht werden, sondern wie gewohnt implizit gelten. (Zur Kennzeichnung der freien Variablen könnten diese mit einem speziellen Zeichen markiert sein, z.B. einer Tilde, einem Hut etc.)

Bemerkung: Anstelle der verkürzten können natürlich auch **Normaltableaus** verwendet werden. Zu beachten ist, dass verkürztes Tableau und Normaltableau *jederzeit äquivalent ineinander umgerechnet* werden können.

In der folgenden Beschreibung des Dreiphasenalgorithmus wird davon ausgegangen, dass Spalten mit freien Variablen x_i in der Nichtbasis jederzeit mit (-1) multipliziert werden können. In diesem Fall wird die freie Variable x_i durch eine künstliche Variable ersetzt, die das Negative der Variablen x_i angibt. Wir nennen sie $(-x_i)$.

Dreiphasenalgorithmus (DPA)

Phase 0 (Bereinigen) Zunächst sollen *durch gezielte Austauschschritte* "Kopf-Nullen" aus der Basis in die Nichtbasis gebracht werden. (Hierbei kann bei der Pivotwahl nachrangig der Effekt auf die Zielfunktion beachtet werden.) Dies reduziert die Dimension des Tableaus u. U. erheblich. Gleichzeitig können freie Variablen von der Nichtbasis in die Basis gebracht werden.

Dabei ist folgendes zu beachten:

- Spalten mit Kopf-Nullen in der Nichtbasis können sofort gestrichen werden, da diese als künstliche Variablen interpretiert werden, die den Wert null haben und diesen auch behalten sollen. Sofern in Phase 0 nicht alle künstlichen Variablen "0" über ATS aus der Basis in die Nichtbasis gebracht werden können, kann dies nur daran liegen, dass kein entsprechendes Pivotelement vorhanden ist, d. h. dass die entsprechende Zeile nur Nullen enthält. Ist dann die rechte Seite der Zeile ungleich null, so liegt eine widersprüchliche Gleichung vor und das System besitzt keine zulässige Lösung. Ist die rechte Seite auch 0, so entspricht die Zeile einer Gleichung " $0 = 0$ ", die problemlos gestrichen werden kann. Man kann also davon ausgehen, dass nach Abschluss der Phase 0 sich keine "Kopf-0" mehr in Basis oder Nichtbasis befindet.
- Nichtvorzeichenbeschränkte Variable in der Basis beschreiben Restriktionen, die eigentlich gar keine sind, da Punkte auf **beiden** Seiten der entsprechenden Hyperebene zulässig sind (sie vereinigen zwei Halbräume). Also könnten auch diese Zeilen aus dem Tableau gestrichen werden. Der Wert der entsprechenden Basisvariablen muss dann bei Bedarf nachträglich berechnet werden.

(Es ist bei Handrechnung u. U. sinnvoll, die entsprechenden Zeilen im Tableau zu belassen, damit der Wert der zugehörigen Variablen in der letzten Basislösung leicht abgelesen werden kann. Die Zeilen gelten dann als für jede Quotientenminimierung gesperrt - sie sind quasi nicht existent -, werden im ATS aber mit umgerechnet.)

Phase 1 (Zulässig machen) Nach Abschluß der Phase 0 liegt ein in der Regel noch unzulässiges Tableau (ohne Kopf-Nullen in der Basis, s. u.) vor,

sagen wir

	x_N^T	min
f	\bar{a}_0^T	\bar{b}_0
x_B	\bar{A}	\bar{b}

Dieses kann wie folgt sukzessive zulässig gemacht werden. Zunächst wird eine *Hilfszielfunktionszeile* benutzt und die eigentliche Zielfunktionszeile wird außer acht gelassen. Dazu wird einfach eine Zeile l mit $\bar{b}_l < 0$ ausgewählt und als neue Zielfunktionszeile angesehen. Die zugehörige Variable $x_{\beta(l)}$ wird nun solange *maximiert*, bis ihr Wert in der Basislösung nichtnegativ geworden ist. Die alte Zielfunktionszeile wird (wegen der gewünschten Äquivalenz aller Tableaus) mit umgerechnet.

Dazu verfähre, solange $\bar{b}_l < 0$ gilt, wie folgt:

- Wähle einen negativen Eintrag in der Zeile l , etwa \bar{a}_{lk} . Dieses k gibt dann den *Pivotspaltenindex* an.

Ist diese Wahl (auch nach Übergang von einer freien Nichtbasis-Variablen x_i zu $(-x_i)$) nicht möglich, **stop3**, es existiert keine zulässige Lösung von (LP) (denn die l -te Zeile fordert dann die Gleichheit von $\bar{b}_l < 0$ mit einem nichtnegativem Wert "links vom Gleichheitszeichen").

- Berechne

$$\alpha_{\max} = \frac{\bar{b}_l}{\bar{a}_{lk}} = \min \left\{ \frac{\bar{b}_j}{\bar{a}_{jk}} \mid \bar{b}_j \geq 0, \bar{a}_{jk} > 0, j = 1, \dots, m \right\}$$

(ohne gesperrte Zeilen, s.o.) und führe mit dem Pivotelement \bar{a}_{lk} den ATS durch.

Kann α_{\max} nicht berechnet werden, so führe stattdessen den ATS mit dem Pivot \bar{a}_{lk} durch. (Diese Pivotwahl macht \bar{b}_l mit *einem einzigen* ATS nichtnegativ.)

Zu beachten ist, dass *generell* α_{\max} so berechnet wird, dass beim ATS keine der Zeilen, die bereits zulässig sind, unzulässig wird, hingegen potentiell auch unzulässige Zeilen zulässig werden können.

Wende diese Vorgehensweise solange an, bis keine negative rechte Seite mehr existiert oder das Verfahren über **stop3** terminiert. Phase 1 endet also mit einem zulässigen Tableau oder der Information, daß ein solches nicht existiert.

Phase 2 (Optimieren) In dieser Phase liegt nun ein zulässiges Tableau vor, dieses soll optimiert werden. Dazu wird der normale Simplex-Algorithmus (unter Einbeziehung möglicher freier Variablen in der Nichtbasis, die mit (-1) multipliziert werden können) verwendet. Die Phase endet daher mit einem Optimaltableau (**stop1**) oder einem Abbruchtableau, aus dem eine Kante abgelesen werden kann, auf der die Zielfunktion unbeschränkt abfällt (**stop2**). Da jede zulässige Lösung zu der Aufgabe, die dem Abbruchtableau unterliegt, einer zulässigen Lösung der ursprünglichen Aufgabe entspricht, ist damit eine Optimallösung der Aufgabe (LP) gefunden oder nachgewiesen, dass keine Optimallösung existiert.

Das folgende Beispiel erläutert die beschriebene Vorgehensweise:

Beispiel 3.1 zum DPA

Wir betrachten die folgende Aufgabe

$$\begin{array}{rcllcl}
 \min & 42x_1 & +38x_2 & +43x_3 & +53x_4 & + & 400 \\
 & 3x_1 & +2x_2 & +4x_3 & +6x_4 & \leq & 500 \\
 & 4x_1 & +4x_2 & +5x_3 & +3x_4 & \leq & 500 \\
 & 6x_1 & +7x_2 & +2x_3 & +4x_4 & \leq & 500 \\
 & x_1 & +x_2 & & & = & 50 \\
 & & & x_3 & +x_4 & = & 70 \\
 & x_i & \geq 0, & i = & 2, 3, 4 & &
 \end{array}$$

Mit Hilfe der Schlupfvariablen x_5, x_6, x_7 erhalten wir als Starttableau für den Dreiphasenalgorithmus (das "Dach" auf der Variablen x_1 soll andeuten, daß diese Variable frei ist)

O	\hat{x}_1	x_2	x_3	x_4	min	
f	-42	-38	-43	-53	400	
x_5	3	2	4	6	500	
x_6	4	4	5	3	500	\rightsquigarrow
x_7	6	7	2	4	500	
0	(1)	1	0	0	50	
0	0	0	1	1	70	

Die Pivotwahl erfolgt gemäß Phase 0. Die freie Nichtbasis-Variable x_1 wird gegen die erste künstliche Basis-Variable 0 getauscht (Pivotelement eingeklammert). Nach dem ATS werden die 4. Zeile und die 1. Spalte aus dem Tableau gestrichen. Wir merken uns: $x_1 + x_2 = 50$.

In einem weiteren Schritt der Phase 0 wird auch noch die zweite Basis-Null in die Nichtbasis gebracht und anschließend die zugehörige Spalte gestrichen.

I	x_2	x_3	x_4	
f	4	-43	-53	2500
x_5	-1	4	6	350
x_6	0	5	3	300
x_7	1	2	4	200
0	0	(1)	1	70

 \rightsquigarrow

II	x_2	x_4	
f	4	-10	5510
x_5	-1	2	70
x_6	0	-2	-50
x_7	1	(2)	60
x_3	0	1	70

Mit Tableau II ist Phase 0 beendet. Die Pivotwahl in Tableau II wird gemäß Phase 1 vorgenommen. Da nun x_6 einen negativen Wert zugewiesen bekommt, wird die zu x_6 gehörige Zeile als Hilfszielfunktion ausgewählt. Diese enthält als einzigen negativen Eintrag: -2 . Daher wird $k = 2$ der Pivotspaltenindex. Quotientenminimierung liefert anschließend $i = 3$ als Pivotzeilenindex. Nach dem ATS ist das Tableau zulässig, Phase 2 wird eingeleitet und ein weiterer ATS genügt, das Tableau auch optimal zu machen:

III	x_2	x_7	
f	9	5	5810
x_5	-2	-1	10
x_6	(1)	1	10
x_4	0.5	0.5	30
x_3	-0.5	-0.5	40

 \rightsquigarrow

IV	x_6	x_7	
f	-9	-4	5720
x_5	2	1	30
x_2	1	1	10
x_4	-0.5	0	25
x_3	0.5	0	45

Die Optimallösung lautet:

$$x_1^* = 40, \quad x_2^* = 10, \quad x_3^* = 45, \quad x_4^* = 25, \quad f^* = 5720.$$

□

Warmstart

Ein Vorteil des Dreiphasenalgorithmus ist, dass er von "beliebig vorgegebener Lösung gestartet" werden kann. Dies ist wie folgt zu verstehen: Vorgegeben sei ein beliebiges Tableau der Aufgabe (LP), z.B. das Starttableau oder das

Endtableau der Bereinigungsphase des DPA

T_1	x_N^T	\min
f	\bar{a}_0^T	\bar{b}_0
x_B	\bar{A}	\bar{b}

Hierin umfassen die Variablenvektoren x_N und x_B je eine Teilmenge der Originalvariablen einschließlich anfänglicher Schlupfvariablen. Seien dabei die Variablen oBdA wie folgt abgezählt:

$$x_N := (x_1, \dots, x_p) \quad \text{und} \quad x_B := (x_{p+1}, \dots, x_{p+s})$$

Seien nun etwa $\bar{x}_1, \dots, \bar{x}_p$ beliebig vorgegebene Werte für die Variablen x_N . Dann können der Aufgabe die Restriktionen $z_i - x_i = -\bar{x}_i$ mit *freien künstlichen* Variablen z_i hinzugefügt werden, $i = 1, \dots, p$. Diese verändern die Aufgabenstellung zunächst einmal nicht, weil die künstlichen Variablen frei sind. Nähmen die z_i alle den Wert 0 an, so wäre $x_i = \bar{x}_i$ für alle $i = 1, \dots, p$.

Bringt man diese zusätzlichen Restriktionen in das Tableau T_1 ein, so entsteht

T_1	x_N^T	\min
f	\bar{a}_0^T	\bar{b}_0
x_B	\bar{A}	\bar{b}
z	$-I$	$-\bar{x}$

wobei I die $p \times p$ -Einheitsmatrix beschreibe. In diesem Tableau können zunächst einmal p ATS gemacht werden mit den letzten p Zeilen.

\bar{T}_1	z^T	\min
f	\bar{a}_0^T	$\bar{b}_0 - \bar{a}_0^T \bar{x}$
x_B	\bar{A}	$\bar{b} - \bar{A} \bar{x}$
x_N	$-I$	\bar{x}

Damit hat man ein zur ursprünglichen Aufgabenstellung äquivalentes Tableau erhalten, dessen Basislösung $x_N = \bar{x}$ und $x_B = \bar{b} - \bar{A}\bar{x}$ lautet.

Kennt man also eine Lösung der gegebenen Aufgabe, die geometrisch in der Nähe einer Optimallösung liegt, so kann der DPA dort gestartet werden. Diese Idee kann insbesondere zu einem **Restart** des DPA verwendet werden, um numerische Ungenauigkeiten auszubügeln, die sich bei langem Lauf des Verfahrens eingeschlichen haben. Dazu nimmt man die zuletzt verfügbaren Werte $\bar{x}_1, \dots, \bar{x}_p$ für x_1, \dots, x_p und startet den DPA vom Tableau \bar{T}_1 neu.

Beachte: Man kann diese Methode auch zur Nachiteration verwenden, wenn man einen Restart mit dem Abschlusstableau des DPA vornimmt. z. B. nach einer nachträglichen Veränderung der Ausgangsdaten.

Nichtlineare Zielfunktion

Ist statt der linearen Optimierungsaufgabe (LP) die linear restringierte Aufgabe (P) vorgegeben, so können die ATS der Phase 0 und der Phase 1 bei Vernachlässigung der Zielfunktion genau so durchgeführt werden, da diese die Originalzielfunktion nicht benötigen. Zum Abschluss liegt ein zulässiges Tableau der gegebenen Aufgabe vor. Allerdings muss danach die Zielfunktion noch an die vorliegende Darstellung angepasst werden. Dies kann geschehen, indem die affin-linearen Verknüpfungen der Variablen aus dem zuletzt vorliegenden Tableau abgelesen und in die Zielfunktion eingesetzt werden. Da beim Dreiphasenalgorithmus die Menge der Variablen im Prinzip unverändert bleibt, ist dies problemlos möglich. Die Zielfunktion wird dann nur in Abhängigkeit von den aktuellen Nichtbasisvariablen angegeben. Dies hat zur Folge, dass die Zielfunktion ihre "geometrischen Eigenschaften" wie z.B. eine Konvexität behält.

Beispiel 3.2 *Wir betrachten noch einmal das obige Beispiel, diesmal allerdings mit der nichtlinearen Zielfunktion $f(x) = 42x_1^2 + 38x_2^2 + 43x_3 + 53x_4 + 400$. Dies schreibt sich in Normalform zu*

$$f(x) = \frac{1}{2} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}^T \begin{pmatrix} 84 & 0 & 0 & 0 \\ 0 & 76 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 43 \\ 53 \end{pmatrix}^T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + 400$$

Wir verwenden die Phase 0 des DPA, um die gegebene Aufgabenstellung auf Normalformat (SPB) zu bringen. Die transformierte Aufgabe wird dann durch obiges Tableau II beschrieben. In Tableau II lesen wir auch die Transformationen ab:

$$\begin{aligned}x_1 + x_2 &= 50 \implies x_1 = 50 - x_2 \\x_3 + x_4 &= 70 \implies x_3 = 70 - x_4\end{aligned}$$

Dies eingesetzt in die Zielfunktion ergibt

$$\begin{aligned}f(x) &= 42 \cdot (50 - x_2)^2 + 38x_2^2 + 43 \cdot (70 - x_4) + 53x_4 + 400 \\&= 80x_2^2 - 4200x_2 + 10x_4 + 108410\end{aligned}$$

oder in Normalform

$$f(x) = \frac{1}{2} \begin{pmatrix} x_2 \\ x_4 \end{pmatrix}^T \begin{pmatrix} 160 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_2 \\ x_4 \end{pmatrix} + \begin{pmatrix} -4200 \\ 10 \end{pmatrix}^T \begin{pmatrix} x_2 \\ x_4 \end{pmatrix} + 108410$$

Damit lautet die transformierte Aufgabenstellung (Standardform (SPB)):

$$\begin{aligned}\min & 80x_2^2 - 4200x_2 + 10x_4 + 108410 \\s.d. & x_5 - x_2 + 2x_4 = 70 \\& x_6 - 2x_4 = -50 \\& x_7 + x_2 + 2x_4 = 60 \\& x_3 + x_4 = 70 \\& x_1 + x_2 = 50 \\& x_2, \dots, x_7 \geq 0\end{aligned}$$

oder nur in den Variablen x_2 und x_4 ausgedrückt (Standardform (SPA)):

$$\begin{aligned}\min & 80x_2^2 - 4200x_2 + 10x_4 + 108410 \\s.d. & -x_2 + 2x_4 \leq 70 \\& -2x_4 \leq -50 \\& x_2 + 2x_4 \leq 60 \\& 0 \leq x_4 \leq 70, 0 \leq x_2\end{aligned}$$

Die gegebene Aufgabe ist damit unter Benutzung zweier Simplexschritte ohne Dimensionsausweitung auf Standardformat gebracht.

Bemerkung: Wären beim Bringen in ein Standardformat zum Schluss von den Nichtbasisvariablen einige frei, so müsste man diese doch noch durch Variablensplitting vorzeichenbeschränkt machen, um eine richtige Standardform zu haben. Dies gilt aber nur, wenn diese Variablen partout nicht in die Basis getauscht werden können. Zu beachten ist, dass solche Variablen, die sich nicht in die Basis bringen lassen, bei *nicht*linearer Zielfunktion nicht auf Unbeschränktheit schließen lassen.

3.1.3 Spezialfall: quadratische Optimierung

Im Falle quadratischer Optimierungsaufgaben ist eine geeignete Formulierung der Zielfunktion in den Standardformen wichtig. Deshalb wollen wir für die oben anfänglich angesprochenen vier Transformationen in ein Standardformat die Zielfunktionen explizit angeben.

Die Zielfunktion bei Transformation auf Standardformat

Vorgegeben sei die Zielfunktion der allgemeinen Aufgabestellung: $f(x) = \frac{1}{2}x^T Q x + p^T x$. Die Zielfunktion ist in allen Fällen der Transformation auf Standardformat im Wesentlichen die gleiche:

$$\tilde{f}(x^1, x^2) = f(x^1 - x^2) = \frac{1}{2}(x^1 - x^2)^T Q (x^1 - x^2) + p^T (x^1 - x^2)$$

Zuweilen kommen noch Schlupf-Variablen vor, die aber keine direkte Auswirkung auf die Zielfunktion haben (s.o.), sagen wir Variablen, die in einem Vektor y zusammengefasst sind. Dann schreibt sich die Zielfunktion

$$\tilde{f}(x^1, x^2, y) = \frac{1}{2} \begin{pmatrix} x^1 \\ x^2 \\ y \end{pmatrix}^T \begin{pmatrix} Q & -Q & 0 \\ -Q & Q & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ y \end{pmatrix} + \begin{pmatrix} p \\ -p \\ 0 \end{pmatrix}^T \begin{pmatrix} x^1 \\ x^2 \\ y \end{pmatrix},$$

damit hat sie wieder das gewünschte Standardformat.

Eine gegebene Konvexität der Zielfunktion f überträgt sich auf \tilde{f} : Sei etwa $\begin{pmatrix} x^1 \\ x^2 \\ y \end{pmatrix} \in \mathbf{R}^{n+n+m}$ beliebig vorgegeben. Dann gilt bei vorausgesetzter

positiver Semidefinitheit von Q :

$$\begin{pmatrix} x^1 \\ x^2 \\ y \end{pmatrix}^T \begin{pmatrix} Q & -Q & 0 \\ -Q & Q & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ y \end{pmatrix} = (x^1 - x^2)^T Q (x^1 - x^2) \geq 0$$

Die Zielfunktion bei Anwendung des Dreiphasenalgorithmus

Will man den Vorteil der Verwendung des Dreiphasenalgorithmus ausnutzen, nämlich den zulässigen Bereich auf Standardformat zu bringen, ohne dabei die Dimension des Problems auszuweiten, so hat man die Zielfunktion auf die entstehende Nichtbasis umzurechnen. Dies wollen wir im folgenden näher untersuchen.

Gegeben sei ein Tableau ohne Zielfunktion, das anfänglich den zulässigen Bereich beschreibt

$$\frac{T \mid x^T \mid 1}{y \mid A \mid b}$$

und eine anfängliche Darstellung der Zielfunktion in Abhängigkeit der Variablen x

$$f(x) = \frac{1}{2}x^T Qx + p^T x + q$$

Ferner sei gegeben ein entsprechendes Tableau \tilde{T} , das aus T durch einen oder mehrere ATS hervorgegangen ist.

$$\frac{\tilde{T} \mid \tilde{x}^T \mid 1}{\tilde{y} \mid \tilde{A} \mid \tilde{b}}$$

wobei $\begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix}$ den ursprünglichen Variablenvektor $\begin{pmatrix} x \\ y \end{pmatrix}$ in einer anderen Reihenfolge beschreibt.

Im Tableau \tilde{T} kann eine Umrechnungsfunktion $x = \varphi(\tilde{x})$ abgelesen werden. Diese Abbildung ist affin-linear und sei etwa durch $\varphi(\tilde{x}) = G \cdot \tilde{x} + g$ angesetzt mit einer $(n \times n)$ -Matrix G und $g \in \mathbf{R}^n$.

Damit lässt sich die Zielfunktion auf die neuen Nichtbasisvariablen umrechnen:

$$\begin{aligned}
 & f(G \cdot \tilde{x} + g) \\
 = & \frac{1}{2} (G \cdot \tilde{x} + g)^T Q (G \cdot \tilde{x} + g) + p^T (G \cdot \tilde{x} + g) + q \\
 = & \frac{1}{2} [\tilde{x}^T G^T Q G \tilde{x} + 2g^T Q G \tilde{x} + g^T Q g] + p^T G \tilde{x} + p^T g + q \\
 = & \frac{1}{2} \tilde{x}^T [G^T Q G] \tilde{x} + [G^T Q g + G^T p]^T \tilde{x} + \left[\frac{1}{2} g^T Q g + p^T g + q \right]
 \end{aligned}$$

Offensichtlich hat die Zielfunktion nun wieder das Standardformat.

Damit eröffnen sich grundsätzlich zwei Möglichkeiten zur Berechnung der Zielfunktion:

- (i) Man rechnet bei jedem Austauschschritt die Zielfunktion um aus den Daten des letzten Tableaus.
- (ii) Man berechnet erst am Ende mehrerer ATS die Zielfunktion um.

Tableaueform Die zweite Möglichkeit haben wir oben bereits im allgemeineren Fall betrachtet. Wir wollen noch auf die erste der genannten Möglichkeiten näher eingehen. Unterstellen wir also, dass \tilde{T} durch nur einen *einzigsten* ATS aus T hervorgeht.

Notieren wir die wesentlichen Daten der Zielfunktion *willkürlich* in einer Matrix

$$(Q, -p),$$

so machen wir folgende **Beobachtung**

$$\begin{aligned}
 G^T (Q, -p) \begin{pmatrix} G & -g \\ 0 & 1 \end{pmatrix} &= (G^T Q, -G^T p) \begin{pmatrix} G & -g \\ 0 & 1 \end{pmatrix} \\
 &= (G^T Q G, -[G^T Q g + G^T p])
 \end{aligned}$$

Dies sind gerade die entsprechenden Daten der umgerechneten Zielfunktion in der gleichen Notierung!

Weiterhin gilt, wenn a_{ik} als Pivotelement aus der Matrix T für den Umrechnungsschritt $T \rightsquigarrow \tilde{T}$ angenommen wird und x_l im Kopf der Pivotspalte k steht:

$$\begin{pmatrix} G & -g \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & & & 0 & 0 \\ 0 & 1 & & & & 0 \\ \dots & & \dots & & & \dots \\ 0 & & & 1 & 0 & 0 \\ -\tilde{a}_{i1} & & & -\tilde{a}_{il} & & -\tilde{a}_{in} & -\tilde{b}_i \\ 0 & & & 0 & 1 & & 0 \\ \dots & & & & & \dots & \dots \\ 0 & & & & & 1 & 0 \\ 0 & & & & & 0 & 1 \end{pmatrix}$$

Dies ist die $(n \times n)$ -Einheitsmatrix, in der die l -te Zeile durch die i -te Zeile der Matrix von \tilde{T} ersetzt wurde.

Man kann nun leicht nachrechnen, dass die Multiplikation

$$(Q, -p) \cdot \begin{pmatrix} G & -g \\ 0 & 1 \end{pmatrix}$$

nichts anderes bedeutet, als die Matrix $(Q, -p)$ mit der Pivotzeile von T (!) als zusätzlich eingeführte (*externe*) Pivotzeile (bei gleicher Pivotwahl) nach den Simplexregeln umzurechnen.

Entsprechend bedeutet eine Multiplikation von links mit G^T , die Pivotzeile von T (!) ohne die letzte Komponente als zusätzliche Pivotspalte *negativ* einzuführen und unter Beibehaltung des Pivotelements nach den Regeln des Simplexverfahrens umzurechnen.

Damit haben wir erkannt:

Die Umrechnung der quadratischen Zielfunktion im ATS erfolgt durch zweimalige Simplexumrechnung!

Der Einfachheit halber rechnet man bei Handrechnung daher folgendes "erweiterte Simplextableau" fortlaufend um

$$\begin{array}{c|c||c} T & x & \min \\ * & Q & -p \\ y & A & b \end{array}$$

Mit diesem kann die zweimalige Simplex-Umrechnung leicht bewerkstelligt werden. Das Feld * ist dafür vorgesehen, bei der zweiten Umrechnung des oberen Teils des Tableaus die um den Rechte-Seite-Wert amputierte, negative alte Pivotzeile als Pivotspalte aufzunehmen.

Beispiel 3.3 *Wir betrachten noch einmal das obige Zahlenbeispiel:*

O	\hat{x}_1	x_2	x_3	x_4	min	
	84	0	0	0	0	
	0	76	0	0	0	
	0	0	0	0	-43	
	0	0	0	0	-53	
x_5	3	2	4	6	500	↔
x_6	4	4	5	3	500	
x_7	6	7	2	4	500	
0	(1)	1	0	0	50	
0	0	0	1	1	70	
I	x_2	x_3	x_4			
	160	0	0	4200		
	0	0	0	-43		
	0	0	0	-53		
x_5	-1	4	6	350		↔
x_6	0	5	3	300		
x_7	1	2	4	200		
0	0	(1)	1	70		
\hat{x}_1	1	0	0	50		

I	0	x_2	x_3	x_4	
(-1)	-84	84	0	0	-4200
-1	0	76	0	0	0
0	0	0	0	0	-43
0	0	0	0	0	-53
x_5	-3	-1	4	6	350
x_6	-4	0	5	3	300
x_7	-6	1	2	4	200
0	0	0	1	1	70
\hat{x}_1	1	1	0	0	50
II	x_2	0	x_4		
	0	160	0	4200	
	(-1)	0	0	-43	
	-1	0	0	-53	
x_5	-1	-4	2	70	
x_6	0	-5	-2	-50	
x_7	1	-2	2	60	
x_3	0	1	1	70	
\hat{x}_1	1	0	0	50	

II	x_2	x_4	
	160	0	4200
	0	0	-10
x_5	-1	2	70
x_6	0	-2	-50
x_7	1	2	60
x_3	0	1	70
\hat{x}_1	1	0	50

Der absolute Wert q der Zielfunktion muss aus der Originalzielfunktion berechnet werden

$$q = 42 \cdot 50^2 + 38 \cdot 0^2 + 43 \cdot 70 + 53 \cdot 0 \cdot +400 = 108010$$

Das Ergebnis stimmt genau mit der obigen Vorabrechnung überein.

Man beachte: *Hätte man bei diesem Beispiel noch die Phase 1 des DPA angeschlossen, so hätte man eine Standarddarstellung der Aufgabenstellung mit nichtnegativer rechter Seite erhalten.*

Zu beachten ist, dass am Ende dieser Tableaurechnung, sollte sie im Rahmen der Phase 0 des DPA eingesetzt werden und dabei eine Variable "0" in die Nichtbasis gebracht worden sein, die entsprechende Spalte *und gleichzeitig* auch die zugehörige Zeile gestrichen werden. Ferner ist beim Multiplizieren einer *freien Variablen* mit (-1) sowohl die zugehörige Spalte *als auch die entsprechende Zeile* mit (-1) zu multiplizieren. Beides ist für die Symmetrie der Matrix Q unumgänglich!

Außerdem sei betont, dass eine Konstante der Zielfunktion vernachlässigt und nicht mit umgerechnet wird. Dieser Wert ist für den Optimierungsprozess irrelevant und kann am Ende des Verfahrens aus den Originaldaten ausgerechnet werden.

Es sei jedoch darauf hingewiesen, dass diese Tableaurechnungen wie im Simplexverfahren nur von Ecke zu Ecke des Polyeders springen können. Dies wird der Lösung der Quadratischen Optimierungsaufgabe i. a. nicht gerecht.

3.1.4 Übungsaufgaben

Aufgabe 1

Man programmiere den Dreiphasenalgorithmus der *Linearen Optimierung* im MATLAB so, dass er von vorgegebener (zulässiger oder unzulässiger) Lösung starten kann und überprüfe den Warmstart (Restart) an mehreren selbsterzeugten Beispielen

Aufgabe 2

Erstellen Sie ein MATLAB Programm, das eine gegebene allgemeine quadratische Optimierungsaufgabe parametergesteuert in eine der beiden Standardformen überführt, wobei beim Format (*SPA*) die rechte Seite a nichtnegativ sein soll. Testen Sie Ihr Programm an mehreren selbsterzeugten Testaufgaben unterschiedlichen Typs.

3.2 Quadratisches Simplexverfahren

Gegeben sei die folgende allgemeine quadratische Optimierungsaufgabe

$$(Q) \quad \begin{cases} \text{minimiere} & z(x) = \frac{1}{2}x^T Qx + p^T x + q \\ \text{so dass} & Ax \leq a, \quad Hx = h \\ & x_j \geq 0, \quad j \in J \end{cases}$$

mit symmetrischer Matrix $Q \in \mathbf{R}^{n \times n}$ und vorgegebener Menge J von Indizes. Ausgangspunkt der Überlegungen ist das **notwendige Optimalitätskriterium**, mit dem subminimale Punkte beschrieben werden:

Ist x ein zulässiger Punkt und ein lokaler Minimierer zur Aufgabe (Q), so gilt

$$D_x z(x) d \geq 0$$

für alle zulässigen Richtungen d an x .

Hierbei beschreiben wir mit $D_x z(x)$ den Gradienten von z an der Stelle x , als Zeile geschrieben:

$$D_x z(x) = \left(\frac{\partial z}{\partial x_1}(x), \dots, \frac{\partial z}{\partial x_n}(x) \right) = \nabla z(x)^T$$

wobei $x^T = (x_1, \dots, x_n)$ unterstellt wurde. Zu beachten ist, dass alle Komponenten dieses Vektors nach Voraussetzung potentiell von x_1, \dots, x_n abhängen.

Zu beachten ist ferner, dass Punkte, die die genannte Bedingung erfüllen, nicht notwendig *lokale Minimierer* darstellen, sondern ggf. auch Sattelpunkte oder sogar Maximierer sein können. Ist die Zielfunktion konvex, so charakterisiert die Bedingung *globale Minimierer*.

Ziel des im Folgenden zu formulierenden *Verfahrens von Beale* ist es, einen subminimalen Punkt zur Aufgabe (Q) zu bestimmen oder nachzuweisen, dass kein solcher existiert.

3.2.1 Beschreibung des Verfahrens

Im folgenden beschreiben wir ein simplexartiges Verfahren zur Lösung der Aufgabe (Q) . Der "Trick" bei diesem Verfahren ist, das auch andere Punkte als Original-Basislösungen angelaufen werden können, weil im Laufe des Verfahrens künstliche, nicht vorzeichenbeschränkte Variablen eingeführt werden, die es erlauben, diese Punkte als Basislösungen eines Tableaus darzustellen, ohne dass dabei der zulässige Bereich beschnitten wird.

Es sei noch folgende **Vorbetrachtung** angestellt: Ist in einem zulässigen Punkt \bar{x} eine zulässige Abstiegsrichtung gegeben, gilt also

$$D_x z(\bar{x}) \cdot d < 0$$

so fällt die Zielfunktion, einmal unabhängig von einem einschränkenden Polyeder betrachtet, entlang des Vektors $x(\mu) = \bar{x} + \mu d$ mit $\mu \geq 0$ bei ansteigendem μ immer weiter ab, solange bis $D_x z(x(\mu)) \cdot d = 0$ gilt. Dies liegt daran, dass die Zielfunktion eingeschränkt auf die Gerade $g : x(\mu), \mu \in \mathbf{R}$ eine Parabel oder eine Gerade darstellt. Wird dabei $D_x z(x(\mu)) \cdot d = 0$ gar nicht erreicht, so fällt die Zielfunktion unbegrenzt.

Skizze:

Zu bemerken ist an dieser Stelle ebenfalls, dass die Menge der Punkte des \mathbf{R}^n , deren Richtungsableitung in Richtung d verschwindet, eine *Hyperebene* L in \mathbf{R}^n bildet:

$$d^T \nabla z(x) = 0 \iff d^T (Qx + p) = 0 \iff (Qd)^T x + p^T d = 0$$

Dies ist eine Besonderheit quadratischer Zielfunktionen!

Dabei kann der Fall, dass diese Gleichung für alle $x \in \mathbf{R}^n$ erfüllt ist, nicht eintreten, da es ja mit \bar{x} ein x gibt mit $D_x z(x) \cdot d < 0$. Über die Lage von d zu L kann folgendes ausgesagt werden:

- entweder die Gerade g durch \bar{x} mit Richtung d schneidet L in einem eindeutigen Punkt. Das ist genau dann der Fall, wenn die Zielfunktion, eingeschränkt auf die Gerade eine echte Parabel darstellt. Diese hat einen eindeutigen Scheitelpunkt. In diesem Fall ist d linear unabhängig zu allen Vektoren parallel zu L .
- oder die genannte Gerade schneidet L in unendlich vielen Punkten. Dann verlief sie innerhalb von L , insbesondere der Punkt \bar{x} müsste in L liegen, was nicht geht wegen $D_x z(\bar{x}) \cdot d < 0$.
- oder die Gerade g schneidet L nicht, dann liegt sie parallel zu L . In diesem Fall gibt es keinen Punkt auf der Geraden g , in dem die Richtungsableitung verschwindet. Das ist der Fall der Geraden in der obigen Skizze. Die Zielfunktion fällt hier unbeschränkt auf der Geraden.

Wir wollen Simplextableaus (*zunächst ohne Zielfunktion*) nutzen, um den zulässigen Bereich der Aufgabe (Q) in unterschiedlicher Basisdarstellung zu beschreiben. Dazu betrachten wir das (nicht notwendig zulässige) Ausgangstableau (in verkürzter Darstellung)

$$\begin{array}{c|c|c} T_0 & x^T & \\ \hline y & A & a \\ 0 & H & h \end{array}$$

das im Laufe des Verfahrens über Austauschschritte äquivalent zum Tableau

$$\begin{array}{c|c|c} \tilde{T} & \tilde{x} & 1 \\ \hline \tilde{y} & \tilde{A} & \tilde{b} \end{array}$$

umgeformt wurde. Wir wollen dieses Tableau durch einen weiteren Austauschschritt in das Tableau

$$\begin{array}{c|c|c} \bar{T} & \bar{x} & 1 \\ \hline \bar{y} & \bar{A} & \bar{b} \end{array}$$

überführen. In all diesen Tableaus beschreibe NBV , $NB\tilde{V}$, $NB\bar{V}$ jeweils die im Sinne der Anordnung des Tableaus geordnete Menge der Nichtbasisvariablen und BV , $B\tilde{V}$, $B\bar{V}$ entsprechend die Menge der Basisvariablen. $k \in NB\tilde{V}$ beschreibe sin Kurzschreibweise die Variable \tilde{x}_k , die im Tableau \tilde{T} an k -ter Stelle in den Nichtbasisvariablen steht (Kopfvariable in der k -ten

Spalte), $i \in B\tilde{V}$ die Variable \tilde{y}_i , die im Tableau \tilde{T} an i -ter Stelle in den Basisvariablen steht (Kopfvariable in der i -ten Zeile).

Es sei $x = \varphi(\tilde{x})$ die beim Übergang von T_0 zu \tilde{T} entstandene Darstellung von x in Abhängigkeit von \tilde{x} . Dann ist $\tilde{z} = z \circ \varphi$ die Darstellung der Zielfunktion in Abhängigkeit der Variablen \tilde{x} . Wir wissen: die Funktion φ kann im Tableau \tilde{T} abgelesen werden.

Wir können oBdA **davon ausgehen**, dass das vorliegende Tableau \tilde{T} *zulässig* ist. Etwa im Nachhinein ins Tableau eingebrachte **künstliche (freie)** Variablen können sich in \tilde{T} allenfalls in der Nichtbasis aufhalten (siehe Vorschrift des untenstehenden Verfahrens).

Im Folgenden wollen wir die Originalvariablen aus der Aufgabenstellung von den künstlichen Variablen unterscheiden, wir nennen sie **eigentliche** Variablen. Eigentliche Variablen können frei oder vorzeichenbeschränkt sein.

Quadratisches Simplexverfahren von Beale

- (0) **Start:** Ausgangspunkt sei das zulässige und in der Basis von freien Variablen freie Tableau \tilde{T} (*ohne Zielfunktion*).
- (1) **Pivotspalte:** Wähle ein *künstliches (freies)* $k \in NB\tilde{V}$ mit $D_{\tilde{x}}\tilde{z}(0)_k \neq 0$ als Pivotspaltenindex. Gilt dabei $D_{\tilde{x}}\tilde{z}(0)_k > 0$, so multipliziere die k -te Spalte des Tableaus mit (-1) . (Vorsicht: da hierbei die Nichtbasisvariable in der k -ten Spalte in ihr Negatives übergeht, ändert sich die Funktion φ .)

Ist diese Wahl nicht möglich, wähle eine *eigentliches* $k \in NB\tilde{V}$ mit

$$\begin{aligned} D_{\tilde{x}}\tilde{z}(0)_k &\neq 0, & \text{falls } k \text{ frei oder} \\ D_{\tilde{x}}\tilde{z}(0)_k &< 0, & \text{falls } k \text{ vorzeichenbeschränkt} \end{aligned}$$

als Pivotspaltenindex. Gilt im ersten Fall $D_{\tilde{x}}\tilde{z}(0)_k > 0$, so multipliziere die k -te Spalte des Tableaus mit (-1) . (Vorsicht: da hierbei die Nichtbasisvariable in der k -ten Spalte in ihr Negatives übergeht, ändert sich die Funktion φ .)

Ist diese Wahl nicht möglich, **stopp1:** Die aktuelle Lösung ist *suboptimal*.

(2) **Neue Restriktion:** Füge dem Tableau die (lineare) Restriktion

$$u + D_{\tilde{x}} \tilde{z} (\tilde{x})_k = 0$$

mit (*neuer*) *künstlicher freier* Variablen u hinzu. (Das Tableau bleibt dabei zulässig).

(3) **Pivotzeile:** Wähle eine Pivotzeile wie im Simplexverfahren, wobei die neu eingefügte Zeile mit einbezogen wird. Wird dabei das Minimum bei mehreren Zeilen gleichzeitig angenommen, von denen eine die mit der künstlichen Kopfvariablen ist, so wähle eine Zeile ohne künstliche Kopfvariable.

Ist eine solche Wahl nicht möglich, **stopp2:** Es existiert *keine Optimal-lösung* zum Problem.

(4) **Austauschschritt:** Führe den normalen Simplex-Austauschschritt $\tilde{T} \rightsquigarrow \bar{T}$ durch. Streiche alle Zeilen aus \bar{T} , in deren Kopf eine künstliche freie Variable steht. Gehe zu (1).

Die Korrektheit des Verfahrens zur Bestimmung eines subminimalen Punktes der Aufgabe (Q) zeigt *im Falle der Nichtentartung* der angelaufenen Basislösungen der folgende

Satz 3.4 *Das Verfahren von Beale ermittelt im nichtentarteten Fall nach endlich vielen Schritten eine suboptimale Lösung von (Q) oder bricht ab wegen Nichtexistenz eines globalen Minimums.*

Beweis: Es sei ein Tableau \tilde{T} vorausgesetzt, das die Bedingung der Ausgangssituation des Verfahrens erfüllt.

(i) Es sei zunächst $D_{\tilde{x}} \tilde{z} (0)_k = 0$ angenommen für alle freien $k \in NB\tilde{V}$ und $D_{\tilde{x}} \tilde{z} (0)_k \geq 0$ für alle vorzeichenbeschränkten $k \in NB\tilde{V}$. Sei d_N eine beliebige zulässige Richtung an die Basislösung von \tilde{T} . Dann lässt sich d_N bzgl der aktuellen Basiseinheitsvektoren \tilde{e}_k darstellen als

$$d_N = \sum_{k \in NB\tilde{V}} \lambda_k \tilde{e}_k$$

mit

$$\begin{aligned}\lambda_k &\in \mathbf{R}, \text{ falls } k \text{ frei} \\ \lambda_k &\geq 0, \text{ falls } k \text{ vorzeichenbeschränkt}\end{aligned}$$

Deshalb gilt in diesem Falle: $D_{\tilde{x}}\tilde{z}(0) \cdot d_N \geq 0$ und die aktuelle Basislösung ist suboptimal.

- (ii) Es sei eine Pivotspalte k gemäß den Vorschriften des Verfahrens gewählt und es kann nun eine Pivotzeile gewählt werden. Nach Konstruktion können wir davon ausgehen, dass $D_{\tilde{x}}\tilde{z}(0)_k < 0$ gilt. Daraus können wir zunächst schließen, dass die rechte Seite u_0 der unter (2) neu hinzugekommenen Zeile im Tableau positiv ist: wir erhalten diese rechte Seite u_0 , indem wir in die zusätzliche Restriktion $\tilde{x} = 0$ einsetzen.

$$u + D_{\tilde{x}}\tilde{z}(\tilde{x})_k = 0, \quad \implies u_0 = -D_{\tilde{x}}\tilde{z}(0)_k > 0$$

Weil u freie Variable ist, bleibt das Tableau bei Hinzunahme der neuen Restriktion zulässig. Die rechte Seite des Tableaus bleibt also durchgehend positiv (Nichtentartung!).

Zur Pivotzeilenwahl bilden wir

$$\lambda = \min \left\{ \frac{b_l}{\tilde{a}_{lk}} : l \in B\tilde{V}, \tilde{a}_{lk} > 0 \right\}$$

Betrachten wir nun bei beliebigem $\mu \geq 0$ den Vektor $\tilde{x}(\mu)$ mit

$$\begin{aligned}\tilde{x}_k(\mu) &= \mu, \quad \tilde{x}_j(\mu) = 0 \quad \text{für alle } j \in NB\tilde{V}, j \neq k \\ \tilde{x}_l(\mu) &= \tilde{b}_l - \mu\tilde{a}_{lk} \quad \text{für alle } l \in B\tilde{V}.\end{aligned}$$

Für $\mu = 0$ ist dies gerade die zu \tilde{T} zugehörige Basislösung.

Die oben berechnete Zahl $\lambda > 0$ gibt nun an, wie weit μ anwachsen darf, ohne dass der zulässige Bereich verlassen wird und ohne dass dabei die Grenze $D_{\tilde{x}}\tilde{z}(\tilde{x}(\mu))_k = 0$ überschritten wird. Dies bedeutet, dass das Verfahren beim folgenden ATS zur nächsten Ecke übergeht (wenn die u -Zeile *nicht* Pivotzeile ist) oder zu einer künstlichen neuen Basislösung (wenn die u -Zeile Pivotzeile ist), wobei der Wert der Zielfunktion (bei Nichtentartung) nach Konstruktion echt absinkt.

(iii) Wenden wir uns nun der Situation zu, wo eine Pivotspalte k gewählt ist, eine Pivotzeile aber *nicht* gewählt werden kann. In dieser Situation erfüllt der in (ii) definierte Vektor $\tilde{x}(\mu)$ für alle $\mu > 0$ alle Restriktionen einschließlich der neu hinzugekommenen Restriktion. Da dabei die Bedingung $D_{\tilde{x}}\tilde{z}(\tilde{x}(\mu))_k = 0$ nicht erreicht wird, fällt die Zielfunktion gemäß obiger Vorbetrachtung unbegrenzt. Also gibt es kein globales Minimum der gestellten Aufgabe.

(iv) Wir machen als nächstes die folgende **Beobachtung**:

Wurden *nacheinander* die zusätzlichen Restriktionen mit den künstlichen freien Variablen u_1, \dots, u_s jeweils als Pivotzeile gewählt, so sind u_1, \dots, u_s danach weiterhin in der Nichtbasis und es gilt im dann entstandenen Tableau \bar{T}

$$D_{\tilde{x}}\bar{z}(0)_k = 0$$

für alle $k \in NB\bar{V}$, die zu u_1, \dots, u_s gehören.

Zum Beweis dieser Behauptung stellen wir folgende Überlegung an, die auf unserer obigen Vorbetrachtung basiert: die künstliche Restriktion wird gebildet, indem im zulässigen Punkt $\tilde{x} = 0$ eine zulässige Abstiegsrichtung d (in der aktuellen Darstellung eine Koordinatenrichtung) ausgesucht und dazu mit der Richtungsableitung

$$u_1 + D_{\tilde{x}}\tilde{z}(\tilde{x})d = 0$$

gebildet wird. Nach dem ATS gilt (beschrieben im Tilde-System)

$$0 = D_{\tilde{x}}\tilde{z}(\tilde{x})d = d^T (\tilde{Q}\tilde{x} + \tilde{p})$$

und die Hyperebene L , die neu eingeführt ist, beschreibt genau alle Punkte x , deren Richtungsableitung in Richtung d verschwindet. d ist linear unabhängig zu allen Vektoren, die parallel sind zu der Hyperebene.

Danach wird übergegangen zu einem neuen zulässigen Punkt \tilde{x}' , der in L liegt. (Die Richtung d liegt nun als Vielfaches eines Einheitsvektors in linearen Raum der späteren u_1, \dots, u_s .)

Es wird eine neue zulässige Abstiegsrichtung d' ausgesucht, die in (parallel zu) L liegt, da nun $u_1 = 0$ gilt. Nun wird die bisherige Konstruktion wiederholt.

Es entsteht dadurch, solange *nacheinander* jeweils die künstliche Restriktion als Pivotspalte gewählt wird, eine Folge linear unabhängiger zulässiger Abstiegsrichtungen, entlang derer die Richtungsableitung im aktuellen zulässigen Punkt alle verschwinden.

Diese Richtungen liegen nach Konstruktion alle im von u_1, \dots, u_s aufgespannten linearen Unterraum (bzgl der aktuellen Koordinatendarstellung) und bilden dort eine Basis. Daher verschwindet die Ableitung von z auf dem gesamten Unterraum, insbesondere in Richtung der u_1, \dots, u_s .

- (v) Bleibt zu zeigen, dass das Verfahren nach endlich vielen Schritten abbricht.

Es gibt nur *endlich viele Kombinationen von eigentlichen Variablen*, die als *die* eigentlichen Variablen der Nichtbasis eines zum Problem gehörigen Tableaus auftreten können. Nichtentartung vorausgesetzt, fällt die Zielfunktion bei jedem Umrechnungsschritt strikt ab.

Ferner kann das Verfahren zu jeder der genannten Kombinationen höchstens einmal ein Tableau anlaufen, aus dem beim nächsten Schritt eine Pivotspalte gewählt werden *muss*, die zu einer *eigentlichen* Variablen gehört. Dies ist wie folgt einzusehen:

Wir wollen ein solches Tableau \tilde{T} ein **kritisches Tableau** nennen. Das Tableau ist dadurch charakterisiert, dass für alle künstlichen Nichtbasisvariablen k die Richtungsableitung $D_{\tilde{x}}\tilde{z}(0)_k$ verschwindet. Die Basislösung eines kritischen Tableaus ist wesentlich durch seine eigentlichen Nichtbasisvariablen bestimmt. Werden diese auf null gesetzt, so beschreiben sie eine Hyperebene H , Stützhyperebene für ein Randpolyeder des zulässigen Bereichs, die Basislösung ist nach Konstruktion ein innerer Punkt des Randpolyeders, bei dem die Ableitung der Zielfunktion, eingeschränkt auf H , verschwindet. Wir wissen, dass alle stationären Punkte einer quadratischen Zielfunktion denselben Zielfunktionswert haben. Deshalb ist der Zielfunktionswert eines kritischen Tableaus nur abhängig von der Kombination der eigentlichen Variablen in der Nichtbasis.

Es genügt daher zu zeigen, dass ausgehend von einem beliebigen Tableau nach endlich vielen Schritten immer ein kritisches Tableau angefahren wird.

Sei also ein beliebiges Tableau \tilde{T} vorgegeben, das im Laufe des Verfahrens erreicht wird. Die Regeln des Verfahrens sehen vor, dass vorrangig Pivotspalten mit künstlichen Kopf-Variablen gewählt werden. Solange kein kritisches Tableau erreicht ist, nimmt die Zahl der künstlichen Variablen in der Nichtbasis also ab, es sei denn, eine künstliche Nichtbasis-Variablen wird mit einer künstlichen Basisvariablen getauscht. Bei jedem solchen ATS nimmt aber nach Beobachtung (iv) die Zahl der *wählbaren* künstlichen Nichtbasisvariablen ab, so dass nach endlich vielen solcher Schritte entweder ein kritisches Tableau erreicht ist oder wieder die Anzahl der künstlichen Variablen in der Nichtbasis wieder um eins reduziert wurde.

□

Das Verfahren von Beale ist sehr allgemein anwendbar und verallgemeinert bekannte Verfahren. Zum Abschluss des theoretischen Teils seien noch einige **Bemerkungen** gemacht.

- Das Verfahren kann praktisch bei jedem Punkt $x^0 \in \mathbf{R}^n$ (etwa einer bekannten Näherung der Optimallösung, ob zulässig oder nicht) gestartet werden, wenn durch Einführung von zusätzlichen freien Variablen der Punkt als Basislösung dargestellt wird.
- Ist die Matrix Q symmetrisch und positiv definit und ist die Menge der Restriktionen der Aufgabe leer, so ermittelt das Verfahren von Beale, ausgehend von einem beliebigen Punkt, in endlich vielen Schritten ein globales Minimum der Zielfunktion!
- Ist die Matrix Q die Nullmatrix, so handelt es sich bei der Aufgabe um eine lineare Optimierungsaufgabe. In diesem Fall läuft das Verfahren wie das lineare Simplexverfahren ab.
- Ist die Matrix Q indefinit, so ermittelt das Verfahren ein Subminimum, das z.B. ein lokales Minimum sein kann.
- Trifft das Verfahren auf eine Basislösung, deren rechte Seite eine Null enthält, so sind Techniken der linearen Optimierung anzuwenden, um

einen Basiswechsel vorzunehmen, der ein Entkommen aus der Entartung erlaubt.

- Im Beweis zum Verfahren von Beale haben wir den affinen Unterraum L betrachtet, der durch die Gleichung $d^T(Qx + p) = 0$ festgelegt ist, d eine zulässige Richtung. Betrachtet man dann speziell einen Punkt $\bar{x} \in L$ und eine Richtung g in L , so liegt auch $\bar{x} + g$ in L und es gilt $d^T(Q(\bar{x} + g) + p) = 0$. Multipliziert man aus und beachtet $d^T(Q\bar{x} + p) = 0$, so folgt $d^TQg = 0$. Dies ist ein besonderer Zusammenhang zwischen den beiden Richtungen. Wir nennen linear unabhängige Richtungen d und g mit der Eigenschaft $d^TQg = 0$ zueinander **schwach Q -konjugiert**. Wir werden diese Eigenschaft später präzisieren und genauer studieren.

3.2.2 Beispiele

Beginnen wir unsere Überlegungen zur praktischen Durchführung des Verfahrens von Beale mit Beispielen.

Beispiel 3.5 (*Beale*) Gegeben sei die folgende Aufgabenstellung:

$$\begin{array}{ll} \text{minimiere} & z(x_1, x_2) = -6x_1 + 2x_1^2 - 2x_1x_2 + 2x_2^2 \\ \text{s.d.} & x_1 + x_2 \leq 2 \\ & x_1, x_2 \geq 0 \end{array}$$

Ausgangstableau zu unserem Verfahren ist also

$$\begin{array}{c|cc|c} T_0 & x_1 & x_2 & \\ \hline y & 1 & 1 & 2 \end{array}$$

Dieses Tableau ist zulässig, so dass keine Phase 1 angewendet werden muss.

1. Schritt: Wir berechnen die Ableitungen

$$D_x z(x) = (-6 + 4x_1 - 2x_2, -2x_1 + 4x_2)$$

also $D_x z(0) = (-6, 0)$. Es kann daher nur die eigentliche Variable x_1 als Pivotspaltenvariable gewählt werden.

Die neu einzuführende Zeile im Tableau lautet mit freier Variablen u_1

$$u_1 + (-6 + 4x_1 - 2x_2) = 0$$

also $u_1 + 4x_1 - 2x_2 = 6$. Dies liefert als erweitertes Tableau

$$\begin{array}{c|cc|c} T_0 & x_1 & x_2 & \\ \hline y & 1 & 1 & 2 \\ u_1 & (4) & -2 & 6 \end{array}$$

Pivotzeile wird die zweite Zeile, Umrechnung mit dem gekennzeichneten Pivot liefert

$$\begin{array}{c|cc|c} T_1 & u_1 & x_2 & \\ \hline y & -\frac{1}{4} & \frac{3}{2} & \frac{1}{2} \\ x_1 & \frac{1}{4} & -\frac{1}{2} & \frac{3}{2} \end{array}$$

2. Schritt: Wir haben die Zielfunktion auf die neue Nichtbasis umzurechnen: Aus dem Tableau lesen wir ab

$$x_1 = -\frac{1}{4}u_1 + \frac{1}{2}x_2 + \frac{3}{2}$$

$$\begin{aligned} \tilde{z}(\tilde{x}) &= \tilde{z}(u_1, x_2) \\ &= -6 \left(-\frac{1}{4}u_1 + \frac{1}{2}x_2 + \frac{3}{2} \right) + 2 \left(-\frac{1}{4}u_1 + \frac{1}{2}x_2 + \frac{3}{2} \right)^2 \\ &\quad - 2 \left(-\frac{1}{4}u_1 + \frac{1}{2}x_2 + \frac{3}{2} \right) x_2 + 2x_2^2 \end{aligned}$$

so dass nach der Kettenregel

$$D_{\tilde{x}}\tilde{z}(\tilde{x})_1 = \frac{3}{2} + 2 \cdot 2 \left(-\frac{1}{4}u_1 + \frac{1}{2}x_2 + \frac{3}{2} \right) \cdot \left(-\frac{1}{4} \right) - 2x_2 \cdot \left(-\frac{1}{4} \right)$$

also $D_{\tilde{x}}\tilde{z}(0) = \frac{3}{2} - \frac{3}{2} = 0$, wie zu erwarten war. Ferner ergibt sich

$$\begin{aligned} D_{\tilde{x}}\tilde{z}(\tilde{x})_2 &= -3 + 2 \cdot 2 \left(-\frac{1}{4}u_1 + \frac{1}{2}x_2 + \frac{3}{2} \right) \cdot \frac{1}{2} \\ &\quad + 2 \cdot 2x_2 - 2 \left(-\frac{1}{4}u_1 \right) - 2 \cdot \frac{1}{2} \cdot 2x_2 - 3 \end{aligned}$$

somit

$$D_{\tilde{x}}\tilde{z}(0)_2 = -3 + 3 - 3 = -3 < 0$$

die neue Pivotspaltenvariable ist also x_2 und das Tableau T_1 erweitert sich zu

$$\begin{array}{c|cc|c} \tilde{T}_1 & u_1 & x_2 & \\ \hline y & -\frac{1}{4} & \left(\frac{3}{2}\right) & \frac{1}{2} \\ x_1 & \frac{1}{4} & -\frac{1}{2} & \frac{3}{2} \\ u_2 & 0 & 3 & 3 \end{array}$$

wie man leicht nachrechnet. Als Pivotzeile ergibt sich dann die erste Zeile, wodurch die neu eingeführte Zeile gleich wieder gestrichen werden kann. Der Austauschschritt liefert

$$\begin{array}{c|cc|c} T_2 & u_1 & y & \\ \hline x_2 & -\frac{1}{6} & \frac{2}{3} & \frac{1}{3} \\ x_1 & \frac{1}{6} & \frac{1}{3} & \frac{5}{3} \end{array}$$

3. Schritt: Aus dem Tableau lesen wir ab

$$x_2 = \frac{1}{6}u_1 - \frac{2}{3}y + \frac{1}{3}, \quad x_1 = -\frac{1}{6}u_1 - \frac{1}{3}y + \frac{5}{3}$$

dies eingesetzt in die Originalzielfunktion ergibt

$$\begin{aligned} \tilde{z}(\tilde{x}) &= -6 \left(-\frac{1}{6}u_1 - \frac{1}{3}y + \frac{5}{3} \right) - 2 \left(\frac{1}{6}u_1 - \frac{2}{3}y + \frac{1}{3} \right) \left(-\frac{1}{6}u_1 - \frac{1}{3}y + \frac{5}{3} \right) \\ &\quad + 2 \left(-\frac{1}{6}u_1 - \frac{1}{3}y + \frac{5}{3} \right)^2 + 2 \left(\frac{1}{6}u_1 - \frac{2}{3}y + \frac{1}{3} \right)^2 \end{aligned}$$

Damit ergibt sich

$$\begin{aligned} D_{\tilde{x}}\tilde{z}(\tilde{x})_1 &= 1 - 2 \cdot \frac{1}{6} \left(-\frac{1}{6}u_1 - \frac{1}{3}y + \frac{5}{3} \right) - 2 \left(\frac{1}{6}u_1 - \frac{2}{3}y + \frac{1}{3} \right) \left(-\frac{1}{6} \right) \\ &\quad + 2 \cdot 2 \left(-\frac{1}{6}u_1 - \frac{1}{3}y + \frac{5}{3} \right) \left(-\frac{1}{6} \right) + 2 \cdot 2 \left(\frac{1}{6}u_1 - \frac{2}{3}y + \frac{1}{3} \right) \frac{1}{6} \end{aligned}$$

und

$$D_{\tilde{x}}\tilde{z}(0)_1 = 1 - \frac{10}{18} + \frac{2}{18} - \frac{20}{18} + \frac{4}{18} = -\frac{6}{18} = -\frac{1}{3} < 0$$

Daraus ergibt sich eine neue Restriktion und folgende Umrechnung des erweiterten Tableaus, bei der die letzte Zeile aus dem Tableau wieder gestrichen wird:

$$\begin{array}{c|cc|c} \tilde{T}_2 & u_1 & y & \\ \hline x_2 & -\frac{1}{6} & \frac{2}{3} & \frac{1}{3} \\ x_1 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \\ u_2 & \left(\frac{1}{3}\right) & -\frac{1}{3} & \frac{1}{3} \end{array} \rightsquigarrow \begin{array}{c|cc|c} T_3 & u_2 & y & \\ \hline x_2 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ x_1 & -\frac{1}{2} & \frac{1}{2} & \frac{3}{2} \end{array}$$

4. Schritt: Wir lesen ab

$$x_2 = -\frac{1}{2}u_2 - \frac{1}{2}y + \frac{1}{2}, \quad x_1 = \frac{1}{2}u_2 - \frac{1}{2}y + \frac{3}{2}$$

$$\begin{aligned} \tilde{z}(\tilde{x}) &= -6 \left(\frac{1}{2}u_2 - \frac{1}{2}y + \frac{3}{2} \right) - 2 \left(-\frac{1}{2}u_2 - \frac{1}{2}y + \frac{1}{2} \right) \left(\frac{1}{2}u_2 - \frac{1}{2}y + \frac{3}{2} \right) \\ &\quad + 2 \left(\frac{1}{2}u_2 - \frac{1}{2}y + \frac{3}{2} \right)^2 + 2 \left(-\frac{1}{2}u_2 - \frac{1}{2}y + \frac{1}{2} \right)^2 \end{aligned}$$

$$D_x \tilde{z}(0)_1 = 0, \quad D_x \tilde{z}(0)_2 = 1$$

Es ist kein weiterer Schritt mehr durchführbar, das Tableau T_3 ist optimal.

Endergebnis: $(x_1, x_2) = \left(\frac{3}{2}, \frac{1}{2}\right)$ ist Minimierer mit Zielfunktionswert $z = -\frac{11}{2}$.

Skizze:

Berechnen wir nun den minimalen Wert einer leichten Abänderung der Zielfunktion aus dem vorigen Beispiel, diesmal ohne Restriktionen. Wird das Verfahren zur unrestringierten Optimierung angewendet, ist das Tableau anfänglich leer. Es füllt sich aber im Laufe des Verfahrens. Die Originalvariablen

x_1, \dots, x_n werden als *freie* Variablen aufgefasst, damit die Aufgabenstellung wirklich unrestringiert ist. Ließe man diese Variablen vorzeichenbeschränkt, so würde eine Optimierung nur in \mathbf{R}_+^n durchgeführt.

Beispiel 3.6 *Gesucht ist ein Minimierer der Funktion $z(x_1, x_2) = 6x_1 + 2x_1^2 + 2x_1x_2 + 2x_2^2$. Zu beachten ist, dass hier die beiden Variablen frei sind.*

Wir beginnen mit dem trivialen Anfangstableau

$$\frac{T_0 \mid x_1 \mid x_2 \mid}{\hline}$$

dessen zugehörige Anfangslösung der Nullpunkt ist.

1. Schritt: *Wir berechnen die Ableitungen*

$$D_x z(x) = (6 + 4x_1 + 2x_2, 2x_1 + 4x_2)$$

also $D_x z(0) = (6, 0)$. Es kann daher nur die freie Variable x_1 als Pivotspaltenvariable gewählt werden. Diese muss allerdings durch $\bar{x}_1 := (-x_1)$ ersetzt werden. Dadurch ändert sich die Zielfunktion zu $\bar{z}(\bar{x}_1, x_2) = -6\bar{x}_1 + 2\bar{x}_1^2 - 2\bar{x}_1x_2 + 2x_2^2$ und die Ableitung zu $D_{\bar{x}} \bar{z}(\bar{x}) = (-6 + 4\bar{x}_1 - 2x_2, -2\bar{x}_1 + 4x_2)$

Die neu einzuführende Zeile im Tableau lautet mit freier Variablen u_1

$$u_1 + (-6 + 4\bar{x}_1 - 2x_2) = 0$$

also $u_1 + 4\bar{x}_1 - 2x_2 = 6$. Dies liefert als erweitertes Tableau

$$\frac{T_0 \mid \bar{x}_1 \mid x_2 \mid}{u_1 \mid (4) \mid -2 \mid 6}$$

Pivotzeile wird die einzige Zeile, Umrechnung mit dem gekennzeichneten Pivot liefert

$$\frac{T_1 \mid u_1 \mid x_2 \mid}{\bar{x}_1 \mid \frac{1}{4} \mid -\frac{1}{2} \mid \frac{3}{2}}$$

(wir lassen die Zeile mit der freien Variablen \bar{x}_1 im Tableau, weil es sich um eine "eigentliche" Variable handelt)

2. Schritt: *Wir haben die Zielfunktion auf die neue Nichtbasis umzurechnen: Aus dem Tableau lesen wir ab*

$$\bar{x}_1 = -\frac{1}{4}u_1 + \frac{1}{2}x_2 + \frac{3}{2}$$

$$\begin{aligned}\tilde{z}(\tilde{x}) &= \tilde{z}(u_1, x_2) \\ &= -6 \left(-\frac{1}{4}u_1 + \frac{1}{2}x_2 + \frac{3}{2} \right) + 2 \left(-\frac{1}{4}u_1 + \frac{1}{2}x_2 + \frac{3}{2} \right)^2 \\ &\quad - 2 \left(-\frac{1}{4}u_1 + \frac{1}{2}x_2 + \frac{3}{2} \right) x_2 + 2x_2^2\end{aligned}$$

so dass nach der Kettenregel

$$D_{\tilde{x}}\tilde{z}(\tilde{x})_1 = \frac{3}{2} + 2 \cdot 2 \left(-\frac{1}{4}u_1 + \frac{1}{2}x_2 + \frac{3}{2} \right) \cdot \left(-\frac{1}{4} \right) - 2x_2 \cdot \left(-\frac{1}{4} \right)$$

also $D_{\tilde{x}}\tilde{z}(0)_1 = \frac{3}{2} - \frac{3}{2} = 0$, wie zu erwarten war. Ferner ergibt sich

$$\begin{aligned}D_{\tilde{x}}\tilde{z}(\tilde{x})_2 &= -3 + 2 \cdot 2 \left(-\frac{1}{4}u_1 + \frac{1}{2}x_2 + \frac{3}{2} \right) \cdot \frac{1}{2} \\ &\quad + 2 \cdot 2x_2 - 2 \left(-\frac{1}{4}u_1 \right) - 2 \cdot \frac{1}{2} \cdot 2x_2 - 3\end{aligned}$$

somit

$$D_{\tilde{x}}\tilde{z}(0)_2 = -3 + 3 - 3 = -3 < 0$$

die neue Pivotspaltenvariable ist also x_2 und das Tableau T_1 erweitert sich zu

$$\begin{array}{c|cc|c} \tilde{T}_1 & u_1 & x_2 & \\ \hline \bar{x}_1 & \frac{1}{4} & -\frac{1}{2} & \frac{3}{2} \\ u_2 & 0 & (3) & 3 \end{array}$$

Pivotzeile wird die neuen Zeile. Der Umrechnungsschritt mit dem gekennzeichneten Pivot liefert

$$\begin{array}{c|cc|c} T_2 & u_1 & u_2 & \\ \hline \bar{x}_1 & \frac{1}{4} & \frac{1}{6} & 2 \\ x_2 & 0 & \frac{1}{3} & 1 \end{array}$$

3. Schritt: Wir lesen ab

$$\bar{x}_1 = -\frac{1}{4}u_1 - \frac{1}{6}u_2 + 2, \quad x_2 = -\frac{1}{3}u_2 + 1$$

Das ergibt

$$\begin{aligned}\tilde{z}(\tilde{x}) &= -6 \left(-\frac{1}{4}u_1 - \frac{1}{6}u_2 + 2 \right) + 2 \left(-\frac{1}{4}u_1 - \frac{1}{6}u_2 + 2 \right)^2 \\ &\quad - 2 \left(-\frac{1}{4}u_1 - \frac{1}{6}u_2 + 2 \right) \left(-\frac{1}{3}u_2 + 1 \right) + 2 \left(-\frac{1}{3}u_2 + 1 \right)^2\end{aligned}$$

sowie

$$D_{\tilde{x}}\tilde{z}(\tilde{x})_1 = \frac{6}{4} + 2 \cdot 2 \left(-\frac{1}{4}u_1 - \frac{1}{6}u_2 + 2 \right) \left(-\frac{1}{4} \right) - 2 \cdot \left(-\frac{1}{4} \right) \left(-\frac{1}{3}u_2 + 1 \right)$$

und

$$D_{\tilde{x}}\tilde{z}(0)_1 = \frac{6}{4} - \frac{8}{4} + \frac{1}{2} = 0$$

ferner

$$\begin{aligned} D_{\tilde{x}}\tilde{z}(\tilde{x})_2 &= 1 + 2 \cdot 2 \left(-\frac{1}{4}u_1 - \frac{1}{6}u_2 + 2 \right) \left(-\frac{1}{6} \right) \\ &\quad - 2 \left[\left(-\frac{1}{6} \right) \left(-\frac{1}{3}u_2 + 1 \right) + \left(-\frac{1}{4}u_1 - \frac{1}{6}u_2 + 2 \right) \left(-\frac{1}{3} \right) \right] \\ &\quad + 2 \cdot 2 \left(-\frac{1}{3}u_2 + 1 \right) \left(-\frac{1}{3} \right) \end{aligned}$$

und

$$D_{\tilde{x}}\tilde{z}(0)_2 = 1 - \frac{8}{6} - 2 \left(-\frac{1}{6} - \frac{2}{3} \right) - \frac{4}{3} = 1 - \frac{4}{3} + \frac{5}{3} - \frac{4}{3} = 0$$

Damit bricht das Verfahren ab, das letzte Tableau ist optimal.

Endergebnis: Minimierer ist $(x_1, x_2) = (-2, 1)$ mit Zielwert $z = -6$.

3.2.3 Zur praktischen Durchführung des Verfahrens

Aufwendiger Teil bei der Durchführung des Verfahrens von Beale ist die Berechnung der Richtungsableitungen. Diese wollen wir näher untersuchen. Wir gehen dabei analog zum letzten Abschnitt vor.

Berechnung von $D_{\tilde{x}}\tilde{z}(\tilde{x})$

Gegeben sei das anfängliche Tableau des Verfahrens, in das die Zielfunktion wie im letzten Abschnitt besprochen, eingetragen ist:

$$\begin{array}{c|c||c} T & x & \min \\ * & Q & -p \\ y & A & b \end{array}$$

Wollen wir nun zur aktuellen Basislösung x die Ableitung berechnen, so lautet diese $\nabla z(x) = Qx + p$. Wenn dazu nun die Richtungsableitung in Richtung des k -ten Einheitsvektors und damit die im Verfahren genannte künstliche Restriktion berechnet werden soll, so lautet diese

$$u + D_x z(x)_k = 0 \iff u + Q_k x = -p_k$$

Diese Zeile kann unmittelbar als k -te Zeile im oberen Teil des obigen Tableaus abgelesen werden! Und diese elegante Möglichkeit ergibt sich in jedem Schritt des Verfahrens, wenn die Zielfunktion grundsätzlich mit dem Tableau umgerechnet wird!

Man beachte: Die genannte Zeile kann mit zur Quotientenminimierung herangezogen werden und braucht erst bei Bedarf, nämlich wenn diese Zeile Pivotzeile wird, in Kopie *als neue zusätzliche Zeile* (externe Pivotzeile) dem Tableau unten hinzugefügt zu werden.

Damit eröffnen sich grundsätzlich die Möglichkeit einer Tableauform des Verfahrens, die im Folgenden beschrieben werden soll.

Normalform des Verfahrens

Die Tableauform des Verfahrens führt daher fortlaufend das erweiterte Simplextableau

$$\begin{array}{c|c||c} T & x & \min \\ \hline * & Q & -p \\ y & A & b \end{array}$$

Wir nennen sie Normalform des Verfahrens. Es sei das Verfahren von Beale in Normalform (ohne Berücksichtigung von Entartungen) noch einmal geschlossen notiert.

Algorithmus von Beale (Normalform) **Ausgangspunkt** ist das oben dargestellte Tableau für die gestellte Aufgabe.

Phase 1: Rechne den unteren Teil des Tableaus so um, dass er zulässig wird. Rechne bei jedem Schritt den oberen Teil begleitend mit um.

Phase 2: Der untere Teil des Tableaus sei nun zulässig.

(1) Wähle ein *freies* $k \in NBV$ mit $(-p_k) \neq 0$. Ist $(-p_k) < 0$, so multipliziere die k -te Spalte des Tableaus und die k -te Zeile des oberen Teils des Tableaus mit (-1) .

Ist eine solche Pivotspalte nicht wählbar, so wähle ein *eigentliches* $k \in NBV$ mit

$$\begin{aligned} (-p_k) &> 0, & \text{falls } k \text{ vorzeichenbeschränkt} \\ (-p_k) &\neq 0, & \text{falls } k \text{ frei} \end{aligned}$$

Ist im zweiten Fall $(-p_k) < 0$, so multipliziere die k -te Spalte des Tableaus und die k -te Zeile des oberen Teils des Tableaus mit (-1) .

Ist auch eine solche Wahl nicht möglich, **stopp1**: das Tableau ist subminimal.

(2) Wähle eine Pivotzeile wie im Simplexverfahren aus der k -ten Zeile des oberen Tableau-Teils und allen Zeilen des unteren Teils. Ist eine solche Wahl nicht möglich, **stopp2**: Es existiert kein globales Minimum zur Aufgabe.

(3) Ist die k -te Zeile des oberen Teils Pivotzeile geworden, so füge dem unteren Teil des Tableaus eine Kopie der Pivotzeile mit einer neuen freien künstlichen Variablen hinzu und benutze dann diese Zeile als Pivotzeile für den Austauschschritt.

(4) Rechne das gesamte Tableau nach den Simplexregeln um. Füge anschließend die *amputierte negative alte Pivotzeile* als Pivotspalte in das Feld * ein. Rechne den oberen Teil des Tableaus mit der neuen Spalte als Pivotspalte unter Beibehaltung des Pivotelements noch einmal nach den Regeln des Simplexverfahrens um.

(5) Streiche alle Zeilen mit künstlichen freien Kopfvariablen aus dem unteren Teil des Tableaus und gehe zurück zu (1).

Wir rechnen noch einmal unsere Beispiele nach dem neuen Formalismus

Beispiel 3.7 Die Aufgabe lautet:

$$\begin{aligned} \text{minimiere} \quad & z(x_1, x_2) = -6x_1 + 2x_1^2 - 2x_1x_2 + 2x_2^2 \\ \text{s.d.} \quad & x_1 + x_2 \leq 2 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Es ist $D_x z(x) = (-6 + 4x_1 - 2x_2, -2x_1 + 4x_2)$. Daraus bekommen wir das folgende Anfangstableau

$$T_0 \left| \begin{array}{cc|c} x_1 & x_2 & \\ \hline (4) & -2 & 6 \\ -2 & 4 & 0 \\ \hline y & 1 & 1 & 2 \end{array} \right.$$

Man beachte, dass sich die gegebene Zielfunktion mit der Matrix $Q = \begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix}$ beschreibt.

1. Schritt: Im Tableau T_0 signalisiert der Eintrag $6 > 0$, dass die 1. Spalte des Tableaus Pivotspalte wird. Pivotzeile (aus der 1. und der 3. Zeile) wird die erste Zeile. Füge dem Tableau also eine Kopie der 1. Zeile mit neuer künstlicher Variabler unten hinzu und rechne um:

$$T_0 \left| \begin{array}{cc|c} x_1 & x_2 & \\ \hline 4 & -2 & 6 \\ -2 & 4 & 0 \\ \hline y & 1 & 1 & 2 \\ u_1 & (4) & -2 & 6 \end{array} \right. \rightsquigarrow \bar{T}_0 \left| \begin{array}{cc|c} u_1 & x_2 & \\ \hline (-4) & -1 & 0 & 0 \\ 2 & \frac{1}{2} & 3 & 3 \\ \hline y & -\frac{1}{4} & \frac{3}{2} & \frac{1}{2} \\ x_1 & \frac{1}{4} & -\frac{1}{2} & \frac{3}{2} \end{array} \right. \rightsquigarrow \bar{T}_1 \left| \begin{array}{cc|c} u_1 & x_2 & \\ \hline \frac{1}{4} & 0 & 0 \\ 0 & 3 & 3 \\ \hline y & -\frac{1}{4} & \frac{3}{2} & \frac{1}{2} \\ x_1 & \frac{1}{4} & -\frac{1}{2} & \frac{3}{2} \end{array} \right.$$

2. Schritt:

$$\bar{T}_1 \left| \begin{array}{cc|c} u_1 & x_2 & \\ \hline \frac{1}{4} & 0 & 0 \\ 0 & 3 & 3 \\ \hline y & -\frac{1}{4} & \left(\frac{3}{2}\right) & \frac{1}{2} \\ x_1 & \frac{1}{4} & -\frac{1}{2} & \frac{3}{2} \end{array} \right. \rightsquigarrow \bar{T}_1 \left| \begin{array}{cc|c} u_1 & y & \\ \hline \frac{1}{4} & 0 & 0 \\ (-\frac{3}{2}) & \frac{1}{2} & -2 & 2 \\ \hline x_2 & -\frac{1}{6} & \frac{2}{3} & \frac{1}{3} \\ x_1 & \frac{1}{6} & \frac{1}{3} & \frac{2}{3} \end{array} \right. \rightsquigarrow T_2 \left| \begin{array}{cc|c} u_1 & y & \\ \hline \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{4}{3} & -\frac{4}{3} \\ \hline x_2 & -\frac{1}{6} & \frac{2}{3} & \frac{1}{3} \\ x_1 & \frac{1}{6} & \frac{1}{3} & \frac{2}{3} \end{array} \right.$$

3. Schritt:

$$T_2 \left| \begin{array}{cc|c} u_1 & y & \\ \hline \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{4}{3} & -\frac{4}{3} \\ \hline x_2 & -\frac{1}{6} & \frac{2}{3} & \frac{1}{3} \\ x_1 & \frac{1}{6} & \frac{1}{3} & \frac{2}{3} \\ u_2 & \left(\frac{1}{3}\right) & -\frac{1}{3} & \frac{1}{3} \end{array} \right. \rightsquigarrow \bar{T}_2 \left| \begin{array}{cc|c} u_2 & y & \\ \hline (-\frac{1}{3}) & -1 & 0 & 0 \\ \frac{1}{3} & 1 & 1 & -1 \\ \hline x_2 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ x_1 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ u_1 & 3 & -1 & 1 \end{array} \right. \rightsquigarrow T_3 \left| \begin{array}{cc|c} u_2 & y & \\ \hline 3 & 0 & 0 \\ 0 & 1 & -1 \\ \hline x_2 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ x_1 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{array} \right.$$

Nun ist keine Pivotspalte mehr wählbar, das letzte Tableau ist subminimal.

Beispiel 3.8 Gesucht ist ein Minimierer der Funktion $z(x_1, x_2) = 6x_1 + 2x_1^2 + 2x_1x_2 + 2x_2^2$.

Es ist $D_x z(x) = (6 + 4x_1 + 2x_2, 2x_1 + 4x_2)$. Daraus bekommen wir das folgende Anfangstableau (mit freien Variablen x_1, x_2)

$$T_0 \left| \begin{array}{cc|c} x_1 & x_2 & \\ \hline 4 & 2 & -6 \\ 2 & 4 & 0 \end{array} \right.$$

wobei x_1 und x_2 freie (eigentliche) Variablen sind.

1. Schritt: Wegen $-6 \neq 0$ wird die erste Spalte mit der freien Variablen x_1 Pivotspalte. Da allerdings $-6 < 0$, muss die Variable durch ihr Negatives ersetzt werden. Dazu werden die erste Spalte und die erste Zeile (Symmetrie von Q !) mit (-1) multipliziert und jetzt die Pivotzeile gesucht, die sich allein aus der ersten Zeile ermittelt.

$$T_1 \left| \begin{array}{cc|c} (-x_1) & x_2 & \\ \hline (4) & -2 & 6 \\ -2 & 4 & 0 \end{array} \right.$$

Nun wird die erste Zeile als externe Pivotspalte aufgebaut und das Tableau zweimal umgerechnet.

$$\begin{array}{c|cc|c} T_1 & (-x_1) & x_2 & \\ \hline & 4 & -2 & 6 \\ & -2 & 4 & 0 \\ \hline u_1 & (4) & -2 & 6 \end{array} \rightsquigarrow \begin{array}{c|cc|c} \bar{T}_1 & u_1 & x_2 & \\ \hline & (-4) & -1 & 0 \\ & 2 & \frac{1}{2} & 3 \\ \hline & (-x_1) & \frac{1}{4} & -\frac{1}{2} \\ & & \frac{3}{2} & \frac{3}{2} \end{array} \rightsquigarrow \begin{array}{c|cc|c} T_2 & u_1 & x_2 & \\ \hline & \frac{1}{4} & 0 & 0 \\ & 0 & 3 & 3 \\ \hline & (-x_1) & \frac{1}{4} & -\frac{1}{2} \\ & & \frac{3}{2} & \frac{3}{2} \end{array}$$

Die letzte Zeile könnte als Zeile mit freier Kopf-Variablen gestrichen werden. Wir lassen sie aber im Tableau, da es sich um eine eigentliche Variable handelt. Allerdings nimmt diese Zeile an keiner Pivotzeilenwahl teil, so dass die Variable $(-x_1)$ auf immer in der Basis verbleibt.

2. Schritt: Wir suchen erneut eine Pivotspalte. Wegen $3 \neq 0$ und x_2 frei, ist dies die zweite Spalte und weil $3 > 0$ gilt, braucht nicht mit (-1) multipliziert zu werden. Wir können also unmittelbar umrechnen. Da die dritte Zeile nicht an der Quotientenminimierung teilnimmt, wird die Pivotzeile nur aus Zeile

2 bestimmt. Diese wird Pivotzeile und als externe Pivotzeile eingeführt.

T_2	u_1	x_2		T_2	u_1	u_2		T_2	u_1	u_2	
	$\frac{1}{4}$	0	0	0	$\frac{1}{4}$	0	0		$\frac{1}{4}$	0	0
	0	3	3	\rightsquigarrow	(-3)	0	-1	\rightsquigarrow	0	$\frac{1}{3}$	0
(- x_1)	$\frac{1}{4}$	$-\frac{1}{2}$	$\frac{3}{2}$	(- x_1)	$\frac{1}{4}$	$\frac{1}{6}$	2	(- x_1)	$\frac{1}{4}$	$\frac{1}{6}$	2
u_2	0	(3)	3	x_2	0	$\frac{1}{3}$	1	x_2	0	$\frac{1}{3}$	1

3. Schritt: Es ist keine Pivotspalte mehr wählbar. Daher ist das Verfahren über **stopp1** beendet und das Tableau ist subminimal. Optimallösung ist $(x_1^*, x_2^*) = (-2, 1)$ mit Zielwert $z^* = -12 + 8 - 4 + 2 = -6$.

Ein weiteres Beispiel rechnet eine lineare Optimierungsaufgabe. Hier ergibt sich die Besonderheit, dass auf Grund der Linearität der Zielfunktion die Gradienten niemals von den Variablen abhängen können, sondern konstant sind. Neu einzuführende Zeilen im Tableau haben also in allen Nichtbasisvariablen den Koeffizienten null. Dies bedeutet, dass diese Zeilen an keiner Quotientenminimierung zur Pivotzeilenbestimmung teilnehmen, daher nie Pivotzeile werden können und immer sofort wieder gestrichen werden. Deshalb braucht man diese Zeilen abweichend vom normalen Verfahren von Beale gar nicht erst einzuführen. Der Ablauf des Verfahrens stimmt daher mit dem des Simplexverfahrens überein.

Beispiel 3.9 Wir rechnen das Beispiel

$$\begin{aligned} \text{minimiere} \quad & z(x_1, x_2) = -6x_1 + 2x_2 \\ \text{s.d.} \quad & x_1 + x_2 \leq 2 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Wir berechnen die Ableitungen: $D_x z(x) = (-6, 2)$. Daraus erhalten wir als Anfangstableau

T_0	x_1	x_2	
	0	0	6
	0	0	-2
y	(1)	1	2

1. Schritt: Im Tableau T_0 signalisiert der Eintrag $6 > 0$, dass die 1. Spalte des Tableaus Pivotspalte wird. Pivotzeile (aus der 1. und der 3. Zeile) wird

die dritte Zeile.

$$\begin{array}{c|cc|c} T_0 & x_1 & x_2 & \\ \hline & 0 & 0 & 6 \\ & 0 & 0 & -2 \\ \hline y & (1) & 1 & 2 \end{array} \rightsquigarrow \begin{array}{c|cc|c} \bar{T}_0 & y & x_2 & \\ \hline & (-1) & 0 & 6 \\ & -1 & 0 & -2 \\ \hline x_1 & 1 & 1 & 2 \end{array} \rightsquigarrow \begin{array}{c|cc|c} \bar{T}_1 & y & x_2 & \\ \hline & 0 & 0 & -6 \\ & 0 & 0 & -8 \\ \hline x_1 & 1 & 1 & 2 \end{array}$$

Das Tableau ist optimal. O-Lösung ist: $(x_1^*, x_2^*) = (2, 0)$ mit Zielwert $z^* = -12 + 0 = -12$.

Vergleichen wir dies mit dem normalen Simplexverfahren:

$$\begin{array}{c|cc|c} T_0 & x_1 & x_2 & \\ \hline z & 6 & -2 & 0 \\ y & (1) & 1 & 2 \end{array} \rightsquigarrow \begin{array}{c|cc|c} \bar{T}_1 & y & x_2 & \\ \hline z & -6 & -8 & -12 \\ x_1 & 1 & 1 & 2 \end{array}$$

3.2.4 Übungsaufgaben

Aufgabe 1

Erstellen Sie ein MATLAB Programm zum Verfahren von Beale und testen sie es an verschiedenen selbsterzeugten Beispielaufgaben, die auch lineare und unrestringierte Optimierungsaufgaben umfassen

3.3 Nichtlineares Simplexverfahren

Wir wollen uns nun der linear restringierten nichtlinearen Optimierung zuwenden. Vorgegeben sei die folgende Aufgabe in Standardform

$$(SPB) \quad \begin{cases} \text{minimiere} & f(x) \\ \text{s.d.} & Hx = h, x \geq 0 \end{cases}$$

mit nichtlinearer stetig differenzierbarer Funktion $f : \mathbf{R}^n \rightarrow \mathbf{R}$. Hierin sei $H \in \mathbf{R}^{m \times n}$ als von maximalem Zeilenrang vorausgesetzt.

Wir setzen diese Form voraus, da wir bereits darüber gesprochen haben, wie man eine allgemeinere Aufgabe der linear restringierten Optimierung in dieses Standardformat überführt.

3.3.1 Konstruktion von Suchrichtungen

Ausgangspunkt der Überlegungen ist erneut die Definition subminimaler Punkte. Ist x zulässig, aber nicht subminimal zu (SPB) , so gibt es zulässige Abstiegsrichtungen d in x . Wir wissen, dass diese die Bedingungen $Hd = 0$ und $d_j \geq 0$ für $x_j = 0$ erfüllen müssen. Will man die Zielfunktion von zulässigem x aus ein Stück absenken, so muss konkret eine zulässige Abstiegsrichtung gefunden werden.

Das Simplexverfahren der linearen Optimierung verwendet als Suchrichtungen die Kanten des Polyeders. Dies sind pro Ecke nur endlich viele und auch insgesamt wegen der endlichen Eckenanzahl nur endlich viele potentielle Suchrichtungen. Das zu beschreibende nichtlineare Simplexverfahren verwendet auch genau diese Suchrichtungen, mit dem Unterschied, dass sämtliche potentiellen Richtungen an sämtlichen zulässigen Punkten angelegt werden dürfen!

Die konkrete Berechnungsmethode einer Kanten-Suchrichtung beim linearen Simplexverfahren ist die folgende: Man befindet sich an einer Ecke, beschrieben durch eine Basismatrix B . Mit dieser wird das zugehörige Tableau festgelegt. Ist oBdA $H = (B, N)$ die zu B gehörige Zerlegung mit der Nichtbasismatrix N , so lautet im linearen Simplexverfahren das verkürzte Tableau zur Aufgabe $\min \{z = c^T x + q : Hx = h, x \geq 0\}$ wegen $(B, N) \begin{pmatrix} x_B \\ x_N \end{pmatrix} = h \implies$

$$x_B + B^{-1}Nx_N = B^{-1}h$$

$$\begin{array}{c|c} & x_N^T \\ \hline z & - (c_N^T - c_B^T B^{-1}N) \\ \hline x_B & B^{-1}N \end{array} \quad \left\| \begin{array}{c} \min \\ q + c_B^T B^{-1}h \\ B^{-1}h \end{array} \right.$$

Eine Kanten-Suchrichtung d erhält man daraus, indem man eine der Nichtbasisvariablen mit positivem Eintrag in der Zielfunktionszeile gleich 1, alle anderen Nichtbasisvariablen gleich 0 setzt und die zur Basis gehörigen Komponenten an der entsprechenden Spalte negativ abliest. Es gilt dann also:

$$d_B = -B^{-1}Nd_N$$

wobei d_N nach Konstruktion ein Einheitsvektor ist.

Man beachte: es gilt dann

$$\begin{aligned} c^T d &= c_N^T d_N + c_B^T d_B \\ &= c_N^T d_N - c_B^T B^{-1}Nd_N \\ &= (c_N^T - c_B^T B^{-1}N) d_N < 0 \end{aligned}$$

und

$$\begin{aligned} Hd &= Bd_B + Nd_N \\ &= -BB^{-1}Nd_N + Nd_n = 0 \end{aligned}$$

Es ist *hinreichend* dafür, dass die Suchrichtung eine *zulässige Abstiegsrichtung* ist, dass die rechte Seite des Tableaus, also das x_B der zugehörigen Basislösung, *positiv ist in allen Komponenten* ist.

Diese Vorgehensweise kopiert man im nichtlinearen Fall, dies führt zu folgender **Idee:** betrachtet werde ein vorgegebener zulässiger Punkt x und irgendeine Zerlegung von H in $H = (B, N)$ mit invertierbarer Matrix B . Die Rolle von c spielt nun der Gradient $\nabla f(x)$ (lineare Approximation): Man bildet

$$r_N := \nabla f(x)_N^T - \nabla f(x)_B^T B^{-1}N$$

sucht dort eine *negative* Komponente, setzt d_N dort auf den Wert 1, sonst d_N auf null, bildet $d_B = -B^{-1}Nd_N$ und hat so eine Richtung mit $\nabla f(x)^T d < 0$ und $Hd = 0$.

Hinreichend dafür, dass $x + \lambda d$ zulässig ist für kleines $\lambda > 0$ ist wieder, dass die zu B gehörigen Komponenten von x alle positiv sind. Ist dies erfüllt, so liefert die Konstruktion eine *zulässige Abstiegsrichtung*!

Die hinreichende Bedingung kann erfüllt werden, wenn die Ecken des zulässigen Bereichs alle *nicht entartet* sind. Dann nämlich haben alle zulässigen x mindestens m positive Komponenten und die Basen der Ecken eines Randpolyeders, in dem x innerer Punkt ist, haben die geforderte Eigenschaft.

Bleibt die Frage zu klären, ob immer, solange x kein KKT Punkt ist, eine negative Komponente von r_N gefunden werden kann, mit der wie oben beschrieben eine zulässige Abstiegsrichtung konstruiert werden kann. Dies ist in der Tat zweifelhaft, da das zugrundeliegende x keine Ecke mehr sein und so in den Nichtbasisvariablen nicht mehr notwendig null sein muss.

Zu bemerken ist allerdings, dass im Falle einer *positiven* Nichtbasis-Komponente j von x auch ein *positiver* Wert von r_N von Nutzen ist, da dann $d_j = -1$ festgelegt werden könnte und die Variable beim Fortschreiten entlang der zulässigen Abstiegsrichtung d abgesenkt würde.

Wir wollen daher die bisherige *Konstruktion präzisieren*.

Vorgegeben sei weiterhin ein zulässiges x und eine Zerlegung $H = (B, N)$ mit invertierbarer Matrix B so, dass die zu B gehörigen Komponenten von x alle positiv sind. (Aber auch zu N gehörige Komponenten von x können positiv sein). Wir beschreiben mit I_N die Indexmenge der in N befindlichen Variablen und setzen I_B entsprechend fest.

Dann bilden wir den **reduzierten Gradienten** r bzgl B

$$r^T := \nabla f(x)^T - \nabla f(x)_B^T B^{-1} H = \left(0, \nabla f(x)_N^T - \nabla f(x)_B^T B^{-1} N \right)$$

und daraus die Konstanten

$$\alpha := \max \{-r_j : r_j \leq 0\}$$

und

$$\beta := \max \{x_j r_j : r_j \geq 0\}$$

Prüfe dann, ob $\alpha \geq \beta$ oder $\beta \geq \alpha$ und bestimme einen Index $k \in I_N$ wie folgt: k sei ein Index, bei dem α das Maximum annimmt, falls $\alpha \geq \beta$, bzw ein Index, bei dem β sein Maximum annimmt, falls $\beta \geq \alpha$. Ist $\alpha = \beta$, so wähle man eine der beiden Möglichkeiten.

Zu beachten ist: gilt nicht gerade $\alpha = \beta = 0$, so findet man wegen $r_B = 0$ so tatsächlich ein $k \in I_N$ mit $r_k < 0$ im Falle $\alpha \geq \beta$ bzw ein $k \in N$ mit $x_k r_k > 0$, also $r_k > 0$ im Falle $\beta \geq \alpha$. Gilt also nicht $\alpha = \beta = 0$, so liefert die obige Konstruktion eine zulässige Abstiegsrichtung im Punkte x ! Das dies genügt, zeigt

Lemma 3.10 *Seien ein zulässiges x zu (SPB) und $H = (B, N)$ mit $x_B > 0$ gegeben. Genau dann ist x ein KKT Punkt zu (SPB), wenn $\alpha = \beta = 0$ gilt.*

Beweis: Genau dann ist x ein KKT Punkt von (SPB), wenn es Vektoren $v \in \mathbf{R}^m$ und $u \in \mathbf{R}^n$ gibt mit

$$\begin{aligned} 0 &= \nabla f(x)^T + v^T H - u^T \\ Hx &= h, \quad x \geq 0 \\ u^T x &= 0, \quad u \geq 0 \end{aligned}$$

Setzen wir die vorgegebene Zerlegung von H und entsprechend von x und u ein, so ergibt sich

$$\begin{aligned} 0 &= \left(\nabla f(x)_B^T, \nabla f(x)_N^T \right) + v^T (B, N) - (u_B^T, u_N^T) \\ h &= (B, N) \begin{pmatrix} x_B \\ x_N \end{pmatrix}, \quad x_B \geq 0, \quad x_N \geq 0 \\ 0 &= (u_B^T, u_N^T) \begin{pmatrix} x_B \\ x_N \end{pmatrix}, \quad u_B \geq 0, \quad u_N \geq 0 \end{aligned}$$

Daher ist ein *zulässiges* x genau dann KKT Punkt, wenn

$$\begin{aligned} 0 &= \left(\nabla f(x)_B^T, \nabla f(x)_N^T \right) + v^T (B, N) - (u_B^T, u_N^T) \\ 0 &= (u_B^T, u_N^T) \begin{pmatrix} x_B \\ x_N \end{pmatrix}, \quad u_B \geq 0, \quad u_N \geq 0 \end{aligned}$$

Gilt $x_B > 0$, so reduziert sich die zweite Zeile in diesen Bedingungen äquivalent zu

$$u_B = 0, \quad u_N^T x_N = 0, \quad u_N \geq 0$$

Wegen $u_B = 0$ liest man nun aus der ersten Zeile der Bedingungen ab

$$v^T = -\nabla f(x)_B^T B^{-1}$$

Setzt man dies wiederum ein, so reduziert sich die erste Zeile äquivalent zu

$$v^T = -\nabla f(x)_B^T B^{-1}, \quad u_N^T = \nabla f(x)_N^T - \nabla f(x)_B^T B^{-1}N$$

Damit gilt $u_N = r_N$ und das gegebene x ist genau dann KKT Punkt, wenn

$$\begin{aligned} r_N^T x_N &= 0, \quad r_N \geq 0 \quad \text{und} \\ v^T &= -\nabla f(x)_B^T B^{-1}, \quad u_N^T = \nabla f(x)_N^T - \nabla f(x)_B^T B^{-1}N, \quad u_B = 0 \end{aligned}$$

Damit ist das zulässige x genau dann ein KKT Punkt, wenn $r_N^T x_N = 0$, $r_N \geq 0$ gilt. Dies aber ist nach Konstruktion äquivalent mit $\alpha = \beta = 0$.

□

Aus dem Beweis halten wir fest: Ist x KKT Punkt, so können die zugehörigen Lagrange-Multiplikatoren angegeben werden zu

$$v^T = -\nabla f(x)_B^T B^{-1}, \quad u_N^T = \nabla f(x)_N^T - \nabla f(x)_B^T B^{-1}N, \quad u_B^T = 0$$

3.3.2 Das Verfahren

Damit ist die Gesamtkonstruktion nun abgesichert. Wir erhalten genau dann eine Kantensuchrichtung, wenn x (noch) kein KKT Punkt ist. Bleibt zu klären, ob die sukzessive Anwendung dieser Suchrichtungen mit geeigneten Schrittweiten einen KKT Punkt der gegebenen Aufgabe bestimmen kann. Wir formulieren das nun wohldefinierte Verfahren im Zusammenhang.

Nichtlineares Simplexverfahren von Zangwill

- (0) **Start:** Wähle ein zu (SPB) zulässiges x .
- (1) **Auswahl:** Wähle eine Zerlegung von H in $H = (B, N)$ so, dass B invertierbar und die zu B gehörigen m Basisvariablen die *maximal positiven* Werte der Komponenten von x haben. Bilde

$$r^T := \nabla f(x)^T - \nabla f(x)_B^T B^{-1}H = \left(0, \nabla f(x)_N^T - \nabla f(x)_B^T B^{-1}N\right)$$

und daraus die Konstanten

$$\alpha := \max \{-r_j : r_j \leq 0\}$$

und

$$\beta := \max \{x_j r_j : r_j \geq 0\}$$

Ist $\alpha = \beta = 0$, **stopp**, x ist KKT Punkt zu (SPB)

Prüfe dann, ob $\alpha \geq \beta$ oder $\beta \geq \alpha$ und bestimme *einen* Index $k \in I_N$ wie folgt: k sei ein Index, bei dem α das Maximum annimmt, falls $\alpha \geq \beta$ bzw ein Index, bei dem β sein Maximum annimmt, falls $\beta \geq \alpha$.

Setze dann den Vektor d_N wie folgt fest:

im Falle $\alpha \geq \beta$ setze

$$d_j = \begin{cases} 0, & \text{für alle } j \in I_N, j \neq k \\ 1, & \text{für } j = k \end{cases}$$

oder setze im Falle $\beta \geq \alpha$

$$d_j = \begin{cases} 0, & \text{für alle } j \in I_N, j \neq k \\ -1, & \text{für } j = k \end{cases}$$

und bilde anschließend

$$d_B = -B^{-1}Nd_N$$

- (2) **Line search:** Bestimme das Minimum \bar{x} von f entlang des Strahls $x + \lambda d$, $0 \leq \lambda \leq \lambda_{\max}$ mit

$$\lambda_{\max} = \min_{j=1, \dots, n} \left\{ \frac{-x_j}{d_j} : d_j < 0 \right\} \cup \{\infty\}$$

Wähle dann ein zulässiges x mit $f(x) \leq f(\bar{x})$ und gehe zu 1.

Konvergenz des Verfahrens

Wir wollen nun eine Konvergenz des Verfahrens beweisen in dem Sinne, dass wir zeigen, dass jeder Häufungspunkt der Iteriertenfolge ein KKT Punkt zu (SPB) ist. Dazu benötigen wir zunächst ein technisches

Lemma 3.11 *Es sei X eine nichtleere abgeschlossene Teilmenge des \mathbf{R}^n und $f : \mathbf{R}^n \rightarrow \mathbf{R}$ eine stetig differenzierbare Abbildung. Wir betrachten das Problem*

$$\min \{f(x) : x \in X\}$$

Ferner betrachte ein beliebiges Verfahren, das in jedem Schritt, ausgehend von zulässigem $x \in X$ die nächste Iterierte wie folgt berechnet: Wähle eine zulässige Abstiegsrichtung d und bestimme den Minimierer \bar{x} bei Minimumbildung von f entlang der zulässigen Strecke auf dem Strahl $x + \lambda d$, $\lambda \geq 0$. Wähle dann eine beliebige zulässige Stelle \hat{x} mit $f(\hat{x}) \leq f(\bar{x})$ als nächste Iterierte.

Sei $\{x^k\}_{k \in \mathbf{N}}$ eine Iteriertefolge des Verfahrens und $\{d^k\}_{k \in \mathbf{N}}$ die Folge der zugehörigen Suchrichtungen. Dann kann es keine Teilfolge $\{(x^k, d^k)\}_{k \in K}$ geben mit folgenden Eigenschaften:

1. $x^k \rightarrow x$ für $k \in K$
2. $d^k \rightarrow d$ für $k \in K$
3. es gibt ein $\delta > 0$ mit $x^k + \lambda d^k \in X$ für alle $\lambda \in [0, \delta]$ und alle $k \in K$
4. $\nabla f(x)^T d < 0$

Beweis: (vgl. [BSS], Lemma 10.2.6)

Nehmen wir an, es gäbe eine Folge $\{(x^k, d^k)\}_{k \in K}$ mit den genannten Eigenschaften. Dann gibt es nach Bedingung 4. ein $\varepsilon > 0$ mit $\nabla f(x)^T d = -2\varepsilon$. Da die x^k gegen x und die d^k gegen d konvergieren, folgt aus der stetigen Differenzierbarkeit von f , dass ein $\delta' > 0$ existiert so, dass

$$\nabla f(x^k + \lambda d^k)^T d^k < -\varepsilon \quad \text{für alle } \lambda \in [0, \delta'] \text{ und } k \in K \text{ genügend groß} \quad (3.1)$$

Setze nun $\bar{\delta} := \min\{\delta, \delta'\} > 0$ und betrachte $k \in K$ genügend groß. Nach Bedingung 3. und der Konstruktion der nächsten Iterierten ist dann

$$f(x^{k+1}) \leq f(x^k + \bar{\delta} d^k)$$

Gemäß dem Mittelwertsatz gilt

$$f(x^k + \bar{\delta} d^k) = f(x^k) + \bar{\delta} \nabla f(\tilde{x}^k)^T d^k$$

für eine Zwischenstelle \tilde{x}^k zwischen x^k und $x^k + \bar{\delta}d^k$. Nach (3.1) folgt damit

$$f(x^{k+1}) < f(x^k) - \varepsilon\bar{\delta} \quad \text{für genügend große } k \in K \quad (3.2)$$

Da die Folge der $f(x^k)$ monoton abfällt, gilt $\lim_{k \rightarrow \infty} f(x^k) = \lim_{k \rightarrow \infty} f(x^{k+1}) = f(x)$. Bilden wir damit in (3.2) den Grenzwert, so ergibt sich die wegen $\varepsilon, \bar{\delta} > 0$ offensichtlich unmögliche Aussage

$$f(x) \leq f(x) - \varepsilon\bar{\delta}$$

□

Satz 3.12 *Betrachtet werde das Problem (SPB) mit stetig differenzierbarer Funktion $f: \mathbf{R}^n \rightarrow \mathbf{R}$. Die Aufgabe sei so geartet, dass alle Ecken m positive Komponenten haben und jede Wahl von m Spalten von H zu einer invertierbaren Matrix B führt. Dann ist jeder Häufungspunkt einer Iteriertenfolge des nichtlinearen Simplexverfahrens von Zangwill ein KKT Punkt von (SPB).*

Beweis: (vgl. [BBS], Theorem 10.7.2)

Zunächst sei festgestellt, dass die Forderung, dass jede Wahl von m Spalten von H zu einer invertierbaren Matrix B führt, für die Wohldefiniertheit des Verfahrens sorgt, weil damit die zu den ausgewählten m Komponenten gehörige Teilmatrix von H auch invertierbar ist.

Sei $\{x^k\}_{k \in K}$ eine konvergente Teilfolge der Iteriertenfolge mit Grenzwert \hat{x} . Nehmen wir an, \hat{x} sei kein KKT Punkt. Wir wollen unter dieser Annahme eine Teilfolge konstruieren, die es nach Lemma 3.11 gar nicht geben darf.

Es sei $\{d^k\}_{k \in K}$ die Folge der zu x^k gemäß Verfahrensvorschrift gehörigen Suchrichtungen. Wir nutzen nun den Auswahlprozess für die d^k und die Stetigkeit der Schrittweitenbestimmung aus.

Insgesamt gibt es im Auswahlprozess der d^k nur endlich viele Möglichkeiten für die Wahl von d^k . Mindestens eines der d^k muss also in der Folge unendlich oft ausgewählt sein. Durch Übergang zu einer Teilfolge können wir daher oBdA annehmen, dass für alle k die gleiche Menge I der Indizes die m maximal positiven Komponenten vom x^k angibt. Dann ist I auch eine Menge

von Indizes von m maximal positiven Komponenten von \hat{x} . Es sei B die zu I gehörige Basismatrix.

Ferner gilt $\hat{x}_B > 0$. Denn würde für ein $i \in B$ $\lim_{k \in K} x_i^k = 0$ gelten, so wäre nach Konstruktion \hat{x} eine entartete Ecke, deren Existenz ausdrücklich ausgeschlossen ist.

Da \hat{x} nach Annahme kein KKT Punkt ist, muss für \hat{x} nach den Vorschriften des Verfahrens eine Suchrichtung \hat{d} mit $\nabla f(\hat{x})^T \hat{d} < 0$ berechenbar sein, denn es gilt nach Lemma 3.10 für diesen zulässigen Punkt im Auswahlprozess *nicht* $\alpha = \beta = 0$. Wegen der Stetigkeit von ∇f und damit von r, α, β kann oBdA $\hat{d} := \lim_{k \in K} d^k$ diese Richtung sein.

Damit haben wir die Eigenschaften 1., 2. 4. der Teilfolge aus Lemma 3.11 bereits nachgewiesen.

Nun wollen wir auch Eigenschaft 3. nachweisen. Ist $\hat{d} \geq 0$, dann ist $x^k + \lambda d^k$ zulässig für alle $\lambda \geq 0$. Sei also $d^k \not\geq 0$ angenommen. Dann ist $\hat{x} + \lambda \hat{d}$ zulässig für alle $\lambda \in [0, \delta]$ für ein $\delta > 0$. Da die $x^k \rightarrow \hat{x}$ streben und $d^k = \hat{d}$ gilt, ist

$$\min_{j=1, \dots, n} \left\{ \frac{-x_j^k}{d_j^k} : d_j^k < 0 \right\} \geq \delta/2 \quad \text{für genügend große } k \in K$$

Deshalb ist $x^k + \lambda d^k$ zulässig für alle $\lambda \in [0, \delta/2]$

□

Ein Beispiel

Beispiel 3.13 *Wir betrachten noch einmal eine Aufgabe, die wir schon mit dem Verfahren von Beale gerechnet haben. Dies bietet sich insofern an, als das Verfahren von Zangwill im Ablauf große Ähnlichkeit mit dem von Beale hat. Betrachten wir also*

$$\begin{array}{ll} \text{minimiere} & z(x_1, x_2) = -6x_1 + 2x_1^2 - 2x_1x_2 + 2x_2^2 \\ \text{s.d.} & x_1 + x_2 \leq 2 \\ & x_1, x_2 \geq 0 \end{array}$$

Wir wollen die Bearbeitung der Aufgabe in Tableaus durchführen, die am Normaltableau der linearen Optimierung orientiert sind. Wir führen folgendes Tableau für der Stelle \bar{x}

min	x^T
x^T	\bar{x}^T
$\nabla f(x)^T$	$\nabla f(\bar{x})^T$
$B^{-1}H$	$B^{-1}H$
r	$\nabla f(\bar{x})^T - \nabla f(\bar{x})_B^T B^{-1}H$

wobei die linke Spalte die Bezeichnungen für die Werte an der Stelle \bar{x} darstellen, die in der rechten Spalte stehen. Nennen wir also die Schlupfvariable der einzigen Restriktion x_3 , so ergibt sich beim Startpunkt $x^T = (0, 0, 2)$ das folgende Starttableau

min	x_1	x_2	x_3
x^T	0	0	2
$\nabla f(x)^T$	-6	0	0
$B^{-1}H$	1	1	1
r	-6	0	0

wobei zuvor $\nabla f(x_1, x_2, x_3)^T = (-6 + 4x_1 - 2x_2, -2x_1 + 4x_2, 0)$, und daraus $\nabla f(0, 0, 2)^T = (-6, 0, 0)$ berechnet wurde. Es ergab sich aus der zweiten Zeile für die Indexmengen der Nichtbasis und der Basis $I_N = \{1, 2\}$ sowie $I_B = \{3\}$, sodass wegen $H = (1, 1, 1)$ sich $B^{-1} = (1)$ ergab. Damit fand man $B^{-1}H = (1, 1, 1)$ und schließlich

$$r^T = \nabla f(x)^T - \nabla f(x)_B^T B^{-1}H = (-6, 0, 0) - 0 \cdot (1, 1, 1) = (-6, 0, 0)$$

Es ergibt sich dann

$$\alpha = \max\{6, 0, 0\} = 6 \quad \text{und} \quad \beta = \max\{0 \cdot 0, 2 \cdot 0\} = 0$$

und wegen $\alpha > \beta$ damit $k = 1$. Setze daher $d_1 = 1$, $d_2 = 0$ und

$$d_3 = -B^{-1}H_{\cdot k}d_k = -1 \cdot 1 \cdot 1 = -1$$

Damit haben wir $\min\{f(x + \lambda d) : \lambda \geq 0\}$ zu berechnen. Dies tun wir durch Nullsetzen der Ableitung

$$\begin{aligned} 0 &= \frac{\partial}{\partial \lambda} f(x + \lambda d) = \nabla f(x + \lambda d)^T d \\ &= 1 \cdot (-6 + 4\lambda - 2 \cdot 0) + 0 \cdot (-2 \cdot 1 + 4 \cdot 0) + (-1) \cdot 0 \\ &= -6 + 4\lambda, \end{aligned}$$

also gilt $\lambda = \frac{3}{2}$. Wegen $\lambda = \frac{3}{2} < \lambda_{\max} = \left\{ \frac{-x_3}{-1} \right\} = 2$ ergibt sich der nächste Iterationspunkt zu

$$x^T = \left(0 + \frac{3}{2}, 0, 2 - \frac{3}{2} \right) = \left(\frac{3}{2}, 0, \frac{1}{2} \right) \quad \text{mit}$$

$$f(x) = -6 \cdot \frac{3}{2} + 2 \cdot \left(\frac{3}{2} \right)^2 = -\frac{9}{2}$$

Damit müssen wir nun $I_B = \{1\}$ und $I_N = \{2, 3\}$ wählen. Es bleibt wegen $B = (1)$ aber bei $B^{-1}H = (1, 1, 1)$. Ferner gilt

$$\nabla f(x)^T = \left(-6 + 4 \cdot \frac{3}{2} - 2 \cdot 0, -2 \cdot \frac{3}{2} + 4 \cdot 0, 0 \right) = (0, -3, 0)$$

und

$$r^T = (0, -3, 0) - 0 \cdot (1, 1, 1) = (0, -3, 0).$$

Insgesamt hat man als nächstes Tableau also

\min	x_1	x_2	x_3
x^T	$\frac{3}{2}$	0	$\frac{1}{2}$
$\nabla f(x)^T$	0	-3	0
$B^{-1}H$	1	1	1
r	0	-3	0

Daraus berechnen sich $\alpha = 3$ bei $k = 2$, $\beta = 0$. Es ist also $k = 2$ zu wählen und es gilt $d_3 = 0$, $d_2 = 1$ sowie

$$d_1 = -B^{-1}H_{,k}d_k = -1 \cdot 1 \cdot 1 = -1$$

Berechnen wir $\min \{f(x + \lambda d) : \lambda \geq 0\}$.

$$\begin{aligned} 0 &= \nabla f(x + \lambda d)^T d \\ &= (-1) \cdot \left(-6 + 4 \left(\frac{3}{2} - \lambda \right) - 2\lambda \right) + 1 \cdot \left(-2 \left(\frac{3}{2} - \lambda \right) + 4\lambda \right) \\ &= 12\lambda - 3, \end{aligned}$$

Daher gilt $\lambda = \frac{1}{4}$. Auf der anderen Seite ist $\lambda_{\max} = \left\{ \frac{-x_1}{-1} \right\} = \frac{3}{2}$. Also wird das gesuchte Minimum angenommen bei $\lambda = \frac{1}{4}$ und der nächste Iterationspunkt ergibt sich zu $x^T = \left(\frac{5}{4}, \frac{1}{4}, \frac{1}{2} \right)$ mit

$$f(x) = -6 \cdot \frac{3}{2} + 2 \cdot \left(\frac{5}{4} \right)^2 - 2 \cdot \frac{5}{4} \cdot \frac{1}{4} + 2 \cdot \left(\frac{1}{4} \right)^2 = -\frac{51}{8}$$

Nun müssen wir erneut $I_B = \{1\}$ und $I_N = \{2, 3\}$ wählen und es bleibt bei $B^{-1}H = (1, 1, 1)$. Ferner gilt

$$\nabla f(x)^T = \left(-6 + 4 \cdot \frac{5}{4} - 2 \cdot \frac{1}{4}, -2 \cdot \frac{5}{4} + 4 \cdot \frac{1}{4}, 0\right) = \left(-\frac{3}{2}, -\frac{3}{2}, 0\right)$$

und

$$r^T = \left(-\frac{3}{2}, -\frac{3}{2}, 0\right) - \left(-\frac{3}{2}\right) \cdot (1, 1, 1) = \left(0, 0, \frac{3}{2}\right)$$

insgesamt also

<i>min</i>	x_1	x_2	x_3
x^T	$\frac{5}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
$\nabla f(x)^T$	$-\frac{3}{2}$	$-\frac{3}{2}$	0
$B^{-1}H$	1	1	1
r	0	0	$\frac{3}{2}$

Daraus ergibt sich $\alpha = 0$ und $\beta = \frac{3}{4}$ mit $k = 3$ und $d_3 = -1, d_2 = 0$ und $d_1 = -1 \cdot (-1) = 1$. Das unrestringierte Minimum auf dem Strahl ist dann

$$0 = 1 \cdot \left(-6 + 4\left(\frac{5}{4} + \lambda\right)\right) + 0 + (-1) \cdot 0 = -1 + 4\lambda$$

also gilt $\lambda = \frac{1}{4}$, dagegen gilt $\lambda_{\max} = \left\{\frac{-x_3}{-1}\right\} = \frac{1}{2}$. Der kleinere der beiden Werte ist $\lambda = \frac{1}{4}$, also lautet der neue Iterationspunkt $x^T = \left(\frac{5}{4}, \frac{1}{4}, \frac{1}{2}\right) + \frac{1}{4}(1, 0, -1) = \left(\frac{3}{2}, \frac{1}{4}, \frac{1}{4}\right)$. In dieser Weise schreitet das Verfahren im Zick-Zack Kurs heran. Es benötigt insgesamt 7 Iterationen.

Den Weg des Verfahrens zeigt die folgende Skizze:

Wir sehen, die Bearbeitung der Aufgabe mit dem Verfahren von Zangwill liefert dasselbe Ergebnis wie die mit dem Verfahren von Beale, das Verfahren von Zangwill geht aber im Zick-Zack Kurs auf die Optimallösung zu.

Bemerkung: Ein Problem beim Verfahren von Zangwill ist die immer wieder notwendige Berechnung von $B^{-1}N$. Zu beachten ist allerdings, dass die Matrix $B^{-1}N$ im Simplextableau abgelesen werden kann. Führt man also das entsprechende Simplextableau, so muss nur $\nabla f(x)$ durch Einsetzen des aktuellen x neu ermittelt werden, dann können alle Daten des Verfahrens durch einfache Matrizenmultiplikation berechnet werden. Genau so gut kann die Basisinverse geführt und aufdatiert werden. Dies ist dann eine *revidierte Version des Verfahrens*.

Modifikationen

- In Hinblick auf Lemma 3.11 kann das Verfahren von Zangwill auch als *Mantelverfahren* für eine andere, ggf. theoretisch nicht abgesicherte Verfahren dienen. Man kann nämlich nach jedem Schritt des Verfahrens einen beliebigen Zwischenschritt einfügen, sofern dieser die Zielfunktion nur absenkt.

Anders herum ausgedrückt: Wenn man in einem beliebigen anderen Verfahren nach jeweils endlich vielen Schritten, wenn der Zielfunktionswert abgesenkt wurde, einen Schritt des Zangwill Verfahrens anschließt, so ist jeder Häufungspunkt der Teilfolge der Iterationspunkte, an die sich ein Zangwill-Schritt anschloss, ein KKT Punkt der gegebenen Aufgabe. Wir werden später darauf zurückkommen.

- In der bisherigen Version leidet das Verfahren von Zangwill unter einem *Zick-Zack-Laufen*, denn es gibt ja insgesamt nur begrenzt viele Richtungen, entlang derer sich das Verfahren bewegen kann. Diesem kann durch die **Active-Set Strategie** entgegengesteuert werden. Darunter versteht man die Strategie, sich solange wie möglich im aktuellen kleinsten Randpolyeder aufzuhalten. Bewerkstelligt wird dies, indem *zu jeden Zeitpunkt* die Variablen, deren aktueller x -Wert null ist, auf null festgehalten werden. Das Verfahren optimiert dann nur in dem durch die Variablen mit Wert null definierten Randpolyeder. Ist *in diesem Unterraum* das $\max\{\alpha, \beta\}$ kleiner als ein vorgegebenes $\varepsilon > 0$, so werden wieder alle Variablen freigegeben.

Einen gesonderten Konvergenzbeweis brauchen wir hier nicht zu betrachten, wenn wir die Anwendung der Strategie im Sinne des unter 1.

Gesagten sehen: Es werden einfach endlich viele zielfunktionsabsenkende Schritte eingefügt.

Experimentell hat sich diese Strategie als günstig erwiesen. Wir werden später darauf zurückkommen. Zu beachten ist außerdem, dass die Active-set-Strategie im Verfahren von Beale verwendet wird.

- Wir modifizieren nun das Verfahren von Zangwill leicht, indem wir die Konstruktion des Parameters β verändern. Es sei für beliebige Zahl $y \geq 0$ festgelegt: $y^c = \min\{y, 1\}$. y^c kappt also y auf den Wert 1, falls $y > 1$ gilt. Wir verändern dann die Berechnung von β wie folgt

$$\beta := \max \{x_j^c r_j : r_j \geq 0\}$$

Da dies weiterhin ein in x stetiger Berechnungsprozess ist, läuft der Konvergenzbeweis zum Verfahren von Zangwill unverändert durch. Außerdem ist Lemma 3.10 unverändert verwendbar, denn es gilt offensichtlich

$$r_N^T x_N = 0, r_N \geq 0 \iff r_N^T x_N^c = 0, r_N \geq 0$$

- Es gibt eine weitere Modifikation des Verfahrens von Zangwill, die sich von diesem Verfahren nur in der Wahl des Vektors d_N unterscheidet. Dieses **Verfahren der reduzierten Gradienten von Wolfe** [1963], das historisch sogar früher liegt als das von Zangwill [1967], legt d_N wie folgt fest:

$$d_j = \begin{cases} -r_j, & \text{für alle } j \in I_N, r_j \leq 0 \\ -x_j r_j, & \text{für alle } j \in I_N, r_j > 0 \end{cases}$$

ohne dass dazu Konstanten α und β berechnet werden müssen. Für diese Modifikation geht der oben dargestellte Konvergenzbeweis nahezu unverändert durch.

- Für eine Simplex-nahe Implementierung des Zangwill-Verfahrens benutzt man am besten die folgende Modifikation: man wählt ein $\varepsilon > 0$ und benutzt bei der Konstruktion der Abstiegsrichtungen d nicht notwendig die m maximalen Komponenten des aktuellen x , sondern einfach m frei gewählte Komponenten $x_i > \varepsilon$. Geht dies nicht, so wählt man die m maximalen Komponenten. Man kann sich leicht überlegen, dass mit dieser Modifikation der Auswahlregel von d der Konvergenzbeweis auch durchläuft.

Der Vorteil dieser Konstruktion ist, dass die gewählten Basen B nicht von Schritt zu Schritt stark schwanken. Man kann z.B. versuchen, die Basis so wenig wie möglich zu verändern. Dann ist der Aufwand für die Berechnung von $B^{-1}N$ geringer.

- *Freie Variablen* sind für das Verfahren nicht störend. Es kann weiterhin jederzeit eine freie Variable in der Nichtbasis durch ihr Negatives ersetzt werden. Zu beachten ist dabei, dass die entsprechende Komponente des reduzierten Gradienten nach der Kettenregel auch durch ihr Negatives ersetzt werden muss. Zu interpretieren ist das so, dass nach Ersetzen einer Variablen x_i durch $x_i = x_{i1} - x_{i2}$ die ursprüngliche Spalte um ihr Negatives ergänzt wird und dann beide Spalten zur Pivotspaltenwahl zur Verfügung stehen. Wird eine davon Pivotspalte, wird die andere wieder gestrichen. Freie Variablen in der Basis nehmen an der Quotientenminimierung nicht teil.
- Es sei betont, dass *nirgendwo* in den bisherigen Überlegungen benötigt wurde, dass die *Basis nichtleer sein muss*. Dies bedeutet, dass das Verfahren auch *unrestringierte bzw. (teilweise) nur vorzeichenbeschränkte Aufgaben* lösen kann. Das Verfahren vereinfacht sich dann, weil kein Basiswechsel stattfinden muss.
- Man kann auch wie beim Verfahren von Beale jederzeit, aber *nur endlich oft*, zusätzliche Restriktionen mit künstlichen freien Variablen einführen. Dies ist für den Konvergenzbeweis nicht schädlich und kann zu einer *Synthese der Verfahren von Zangwill und Beale* führen. Bei endlich vielen zusätzlichen Restriktionen läuft der Konvergenzbeweis trotzdem durch. Außerdem könnte man Beale-Schritte im Sinne der zuerst genannten Modifikation ansehen.

3.3.3 Spezialfall: Quadratische Optimierung

Ein Schwachpunkt des Verfahrens von Zangwill ist die Notwendigkeit der eindimensionalen Minimierung im Baustein (2) des Verfahrens. Bei beliebiger nichtlinearer Zielfunktion ist dies problematisch, da das Minimum nur numerisch und oft nur schwierig in sehr vielen Unteriterationen berechnet werden kann, andererseits aber das Erreichen des Minimums theoretisch wichtig ist. Im Spezialfall quadratischer Zielfunktion ist dagegen die Bestimmung des

Minimums einfach und das Verfahren kann so wie im obigen Beispiels durchgeführt werden.

Verfahren von Hildreth und D'Esopo

Besondere Erwähnung verdient die quadratische Optimierungsaufgabe, bei der die Restriktionen nur aus den Vorzeichenbeschränkungen bestehen. Wir hatten bereits im Abschnitt über Dualität/Alternative Formulierungen gesehen dass man im Falle streng konvexer Zielfunktion die duale Aufgabe so formulieren kann, dass hier eine konvexe quadratische Zielfunktion nur unter Vorzeichenbeschränkungen optimiert wird. Zur Lösung einer solchen Aufgabe ist das Verfahren von Zangwill sehr gut einsetzbar. Dabei kann das Verfahren auf die Situation zurechtgeschnitten werden. Zur Beschreibung dieser speziellen Variante des Verfahrens orientieren wir uns an obigem Beispiel.

Wir betrachten die Aufgabe

$$(VB) \quad \min \left\{ f(x) = \frac{1}{2}x^T Qx + p^T x : x \geq 0 \right\}.$$

Dann gelten für das Verfahren von Zangwill gemäß der oben gemachten Feststellung, dass die Basismenge I_B auch leer sein darf, die folgenden Vereinfachungen:

Startend bei $x^0 = 0$ ergibt sich für die erste Iteration und nach Konstruktion auch für alle weiteren Iterationsschritte immer als Basisindexmenge $I_B = \emptyset$ und entsprechend $I_N = \{1, \dots, n\}$. Weiterhin ist für alle Iterationsschritte $r = \nabla f(x)$.

Dadurch vereinfacht sich das Verfahren von Zangwill erheblich: d_N ist ja im Verfahren von Zangwill theoretisch immer eine zulässige Abstiegsrichtung. Die Zielfunktion, eingeschränkt auf die Gerade, die durch d_N bestimmt ist, hat höchstens einen Minimierer. Dadurch genügt es, d_N generell als Koordinateneinheitsvektor festzulegen, wobei die Komponente k mit der 1 über α und β bestimmt wird. Der Minimierer von f entlang der Geraden wird dann durch

$$0 = \nabla f(x + \lambda e_k)_k = Q_k \cdot (x + \lambda e_k) + p_k \implies \lambda = -\frac{Q_k \cdot x + p_k}{q_{kk}} = -\frac{r_k}{q_{kk}}$$

bestimmt, wobei x die aktuelle Iterierte ist. Es ist also einfach

$$x_k := \max\{0, x_k + \lambda\}$$

neu festzulegen, das Vorzeichen von λ bestimmt die Abstiegsrichtung. Insgesamt stellt sich das zurechtgeschnittene Verfahren wie folgt dar

Modifiziertes Verfahren von Zangwill zur Lösung von (VB)

(0) **Start:** Vorgegeben sei die Aufgabe (VB). Setze $x = 0$.

(1) **Auswahl:** Berechne $r = \nabla f(x)$ und \hat{r} durch

$$\hat{r}_j = \begin{cases} -r_j, & \text{falls } r_j \leq 0 \\ x_j r_j, & \text{falls } r_j > 0 \end{cases}$$

für $j = 1, \dots, n$ und daraus einen Index k mit

$$\hat{r}_k := \max\{\hat{r}_j : j = 1, \dots, n\}$$

Ist $\hat{r}_k = 0$, **stopp**, x ist stationärer Punkt zu f .

(2) **Line search:** Berechne

$$\lambda_k = -\frac{r_k}{q_{kk}}$$

Setze dann $x_k := \max\{0, x_k + \lambda_k\}$ und gehe zu (1).

Beispiel 3.14 Wir betrachten die Aufgabe

$$\begin{aligned} \text{minimiere} \quad & z(x_1, x_2) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1 - 2x_2 \\ \text{s.d.} \quad & 2x_1 + 3x_2 \leq 6 \\ & x_1 + 4x_2 \leq 5 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Diese Aufgabe beschreibt sich in Matrizenform $\min\{\frac{1}{2}x^T Qx + p^T x : Ax \leq a\}$ mit

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad p = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \quad A = \begin{pmatrix} 2 & 3 \\ 1 & 4 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad a = \begin{pmatrix} 6 \\ 5 \\ 0 \\ 0 \end{pmatrix}$$

Die dazu duale Aufgabe lautet (vgl. Abschnitt Dualität)

$$\max \left\{ -\frac{1}{2} (p + A^T \lambda)^T Q^{-1} (p + A^T \lambda) - a^T \lambda : \lambda \geq 0 \right\}$$

bzw

$$\min \left\{ f(\lambda) = \frac{1}{2} \lambda^T A Q^{-1} A^T \lambda + [A Q^{-1} p + a]^T \lambda : \lambda \geq 0 \right\}$$

wobei die Konstante in der Zielfunktion weggelassen wurde. Man berechnet

$$G := A Q^{-1} A^T = \begin{pmatrix} 13 & 14 & -2 & -3 \\ 14 & 17 & -1 & -4 \\ -2 & -1 & 1 & 0 \\ -3 & -4 & 0 & 1 \end{pmatrix}, \quad g := A Q^{-1} p + a = \begin{pmatrix} -2 \\ -4 \\ 1 \\ 2 \end{pmatrix}$$

Wir wenden nun das obige Verfahren auf die duale Aufgabe an.

1. Durchgang: Es ist $\nabla f(\lambda) = G\lambda + g$, speziell $r = \nabla f(0) = g$, so dass $\hat{r}^T = (2, 4, 0 \cdot 1, 0 \cdot 2) = (2, 4, 0, 0)$, wobei die betragsmaximale Komponente bei $k = 2$ angenommen wird. Es berechnet sich dann $-r_2/g_{22} = 4/17$. Damit berechnet sich λ_2 neu zu $\lambda_2 = 4/17$ und es gilt nun $\lambda^T = (0, 4/17, 0, 0)$.

2. Durchgang: Es gilt

$$r = \nabla f(\lambda) = 4/17 \begin{pmatrix} 14 \\ 17 \\ -1 \\ -4 \end{pmatrix} + \begin{pmatrix} -2 \\ -4 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{22}{17} \\ 0 \\ \frac{13}{17} \\ \frac{18}{17} \end{pmatrix}$$

und daraus $\hat{r}^T = (0 \cdot \frac{22}{17}, 0, 0 \cdot \frac{13}{17}, 0 \cdot \frac{18}{17}) = (0, 0, 0, 0)$. Damit ist das Verfahren bereits beendet. die Optimallösung der dualen Aufgabe lautet: $\lambda^{*T} = (0, 4/17, 0, 0)$.

Berechnen wir hieraus noch die primale Optimallösung über die Formel $x = -Q^{-1} (p + A^T \lambda)$

$$\begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} = - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \left(\begin{pmatrix} -1 \\ -2 \end{pmatrix} + \begin{pmatrix} 2 & 3 \\ 1 & 4 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}^T \begin{pmatrix} 0 \\ \frac{4}{17} \\ 0 \\ 0 \end{pmatrix} \right) = \begin{pmatrix} \frac{13}{17} \\ \frac{18}{17} \end{pmatrix}$$

Zu beachten ist, dass dieses Ergebnis auch an der Bemerkung zu den Lagrange-Multiplikatoren nach Lemma 3.10 und der Tatsache, dass das Dual des Duals die primale Aufgabe darstellt, hätte abgelesen werden können!

Die Idee, eine gegebene quadratische Optimierungsaufgabe mit streng konvexer Zielfunktion über die entsprechende duale Aufgabe zu lösen, stammt wohl ursprünglich von HILDRETH und D'ESOPPO (1957/59). Diese schlugen ein Verfahren vor, das sehr eng verwandt ist mit der soeben durchgeführten Vorgehensweise. Der einzige Unterschied ist nämlich, dass sie keine spezielle Komponente k des Lösungsvektors x bestimmen, um f entlang der entsprechenden Koordinatenrichtung zu minimieren, sondern dass sie zyklisch *alle Koordinatenachsen nacheinander* auswählen und nach n Schritten dabei wieder von vorne anfangen. Sie können die Konvergenz auch dieser Vorgehensweise beweisen.

Wendet man dieses Verfahren auf das obige Beispiel an, so muss man immerhin dreimal sämtlich Komponenten von λ durchgehen, bis man die Optimallösung gefunden hat. In diesem Beispiel erweist sich das modifizierte Verfahren von Zangwill als deutlich überlegen.

3.3.4 Übungsaufgaben

Aufgabe 1

Erstellen Sie ein Matlab Programm zum Verfahren von Zangwill, das auch die Modifikation der Berechnung von β beinhaltet. Testen Sie Ihr Programm im Vergleich zum Verfahren von Beale.

Aufgabe 2

Programmieren Sie die Modifikation des Verfahrens von Zangwill und das Verfahren von Hildreth und D'Esopo zur Lösung von (VB) und testen sie Ihre Programm vergleichend an den Dualen von ausgewählten Testaufgaben.

Aufgabe 3

Erstellen Sie möglichst viele Matlab Programme zu den oben angesprochenen Modifikationen des Verfahrens von Zangwill und testen Sie die Programme an ausgewählten Beispielaufgaben.

Kapitel 4

Quadratische KKT-Simplexverfahren

Das Ziel der bisher betrachteten Verfahren zur Lösung linear restringierter Optimierungsaufgaben war es, einen KKT Punkt der gegebenen Aufgabe zu finden. KKT Punkte sind Punkte, die das System der KKT Bedingungen erfüllen. Diese stellen ein notwendiges Optimalitätskriterium dar.

Bisher haben wir *Abstiegsverfahren* konstruiert, die unmittelbar auf die gegebene Aufgabe angewandt wurden. Eine weitere Idee zur Ermittlung eines KKT Punkts ist, das KKT-System als System von Gleichungen und Ungleichungen *direkt zu lösen*. Im Folgenden stellen wir klassische Verfahren der Quadratischen Optimierung vor, die genau diesen Weg beschreiten. Dabei beschränken wir uns auf eine knappe Darstellung, die technische Beweise auslässt.

Ferner beschränken wir uns im Folgenden auf eine Standardform der quadratischen Optimierungsaufgabe, wohl wissend, dass der Allgemeinfeld darauf zurückgeführt werden kann

$$(QPS) \quad \begin{cases} \text{minimiere} & f(x) = \frac{1}{2}x^T Qx + p^T x \\ \text{s.d.} & Ax \leq a, \quad x \geq 0 \end{cases},$$

wobei wir $A \in \mathbf{R}^{m \times n}$, $Q \in \mathbf{R}^{n \times n}$ und $a \in \mathbf{R}^n$ voraussetzen.

Zu (QPS) gehört das KKT System

$$(QPSKKT) \quad \begin{cases} Qx + p + A^T\lambda - \mu = 0, \lambda, \mu \geq 0 \\ Ax + y = a, x, y \geq 0 \\ 0 = \lambda^T y + \mu^T x \end{cases}$$

Wir beziehen uns auf **Beobachtungen** zu dieser Aufgabe, die wir zum Teil bereits weiter oben notiert haben:

- Die ersten beiden Zeilen des Systems (QPSKKT) beschreiben ein konvexes Polyeder \bar{P} , über dem z.B. die quadratische Zielfunktion $t(x, y, \lambda, \mu) = \lambda^T y + \mu^T x \geq 0$ minimiert werden kann. Eine Optimalösung dieser *primal-dualen* Aufgabe (QPDS)

$$\begin{aligned} \min t &= \lambda^T y + \mu^T x \text{ unter den Nebenbedingungen} \\ Qx + p + A^T\lambda - \mu &= 0, \lambda, \mu \geq 0 \\ Ax + y &= a, x, y \geq 0 \end{aligned}$$

mit Zielwert 0 ist eine Lösung von (QPSKKT). Genau dann besitzt die Aufgabe (QPS) einen KKT Punkt x^* , wenn die Aufgabe (QPDS) einen Minimierer $(x^*, y^*, \lambda^*, \mu^*)$ mit $z(x^*, y^*, \lambda^*, \mu^*) = 0$ besitzt.

- Die Zielfunktion von (QPDS) kann aber nur null sein, wenn in der optimalen Lösung von je zwei zueinander zugehörigen Komponenten von λ und y bzw. μ und x jeweils mindestens eine null ist. Das sind insgesamt mindestens $(n + m)$ Nullen bei insgesamt $2(n + m)$ Variablen. Da die genannten ersten beiden Zeilen von (QPSKKT) genau $(n + m)$ Gleichungen enthalten, bedeutet dies, dass die Lösungen von (QPSKKT) im nichtentarteten Fall durchweg Eckpunkte von \bar{P} sein müssen. Dies zeigt aber, dass, wie in der linearen Optimierung, die Lösungen von (QPSKKT) unter den zulässigen Basislösungen (und nur unter diesen) gesucht werden können.
- Die Zielfunktion von (QPDS) ist nicht konvex, sie wird beschrieben durch

$$t(z_1, z_2) = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}^T \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \quad \text{mit } z_1 = \begin{pmatrix} x \\ y \end{pmatrix} \text{ und } z_2 = \begin{pmatrix} \lambda \\ \mu \end{pmatrix}$$

Erfreulicherweise ist allerdings die Einschränkung der Zielfunktion auf den vom Polyeder \bar{P} aufgespannten affinen Unterraum L konvex, wenn nur Q positiv semidefinit ist. Dies ist wie folgt einzusehen:

Sei z mit $z^T = (x^T, y^T, \lambda^T, \mu^T)$ beliebig aus L . L ist gegeben durch die beiden Gleichungen $Qx + A^T\lambda - \mu = 0$ und $Ax + y = 0$. Dann gilt

$$\begin{aligned} \lambda^T y + \mu^T x &= \lambda^T y + (Qx + A^T\lambda)^T x \\ &= \lambda^T y + x^T Qx + \lambda^T Ax \\ &= \lambda^T y + x^T Qx + \lambda^T (-y) = x^T Qx \geq 0 \end{aligned}$$

Damit ergeben sich mehrere Ansatzpunkte zur Lösung von $(QPSKKT)$.

1. Man wendet Simplexschritte auf den zulässigen Bereich vom $(QPDS)$ an und versucht, durch gezieltes Wandern von Ecke zu Ecke zu erreichen, dass in einer Ecke schließlich die Komplementaritätsbedingung $\lambda^T y + \mu^T x = 0$ erfüllt ist.
2. Man löst die Aufgabe $(QPDS)$ mit einem beliebigen Verfahren der quadratischen Optimierung. Genau dann, wenn sich als Ergebnis der minimale Funktionswert null ergibt, ist die x Komponente der Lösung auch Optimallösung zu (QPS) . Zu beachten ist, dass jeder KKT Punkt von $(QPDS)$ bereits globaler Minimierer ist.
3. Man erweitert die Aufgabe um künstliche Variablen so, dass die Komplementarität bereits anfänglich erfüllt ist. Dann versucht man, unter Aufrechterhaltung der Komplementarität, die künstlichen Variablen wieder zu entfernen, solange bis eine Ecke von \bar{P} erreicht ist.

In diesem Kapitel verfolgen wir zunächst die ersten beiden Ideen (**primal-duale Verfahren**) und zeigen, mit welchen Strategien $\lambda^T y + \mu^T x = 0$ erreicht werden kann. Danach betrachten wir den dritten Ansatz.

4.1 Primal-Duale Verfahren

Wir wollen nun den Ansatz verfolgen, ersatzweise zur Lösung von (QPS) die primal-duale Aufgabe (QPDS)

$$\begin{aligned}\min t(x, y, \mu, \lambda) &= \lambda^T y + \mu^T x \text{ unter den Nebenbedingungen} \\ Qx + p + A^T \lambda - \mu &= 0, \lambda, \mu \geq 0 \\ Ax + y &= a, x, y \geq 0\end{aligned}$$

zu lösen. Diese Aufgabe *sei nichtentartet*.

4.1.1 Das Verfahren von Barankin und Dorfman

Wir wissen bereits, dass eine Optimallösung dieser Aufgabe unter den zulässigen Basislösungen zu suchen ist. Man kann daher wie beim Simplexverfahren bei einer Ecke des Polyeders starten und dann vom Ecke zu Ecke wandern. Das Problem ist, dass die Zielfunktion nichtlinear ist und eine negative Richtungsableitung entlang einer Kantenrichtung keine Garantie dafür ist, dass die nächste anbelaufene Kante einen niedrigeren Zielfunktionswert hat als die letzte Ecke.

In einem von BARANKIN UND DORFMAN 1956 vorgeschlagenen Verfahren wird jeweils der zu erwartende Zielfunktionswert einer Nachbarecke vorab ausgerechnet und danach entschieden, zu welcher Nachbarecke per ATS übergegangen wird. Wir wollen verstehen, wie man das machen kann.

Schlüssel zum Verständnis ist die besonders einfache Zielfunktion der primal-dualen Aufgabe (QPDS). Betrachten wir den Vektor $z^T = (x^T, y^T, \mu^T, \lambda^T) \in \mathbf{R}^{2(n+m)}$, so wollen wir diesem einen **adjungierten Vektor** \tilde{z} zuordnen, der lediglich durch Umsortieren der Variablen entsteht:

$$\tilde{z}^T = (\mu^T, \lambda^T, x^T, y^T)$$

Mit dieser Festlegung schreibt sich die Zielfunktion t wie folgt

$$t(x, y, \mu, \lambda) = \frac{1}{2} z^T \tilde{z}$$

Der Einfachheit halber können wir natürlich die Zielfunktion $T := 2t$ betrachten, um dem lästigen Faktor $\frac{1}{2}$ aus dem Wege zu gehen. Zu beachten

ist, dass sich mit dem adjungierten Vektor leicht rechnen lässt. Es gelten die *Rechengesetze*

$$a^T \tilde{b} = b^T \tilde{a} \quad \widetilde{(a + b)} = \tilde{a} + \tilde{b}, \quad a^T (\lambda \tilde{b}) = \lambda a^T \tilde{b}$$

Damit können wir folgende Betrachtung anstellen: Ist

$$\frac{z_B}{z_N} \left| \begin{array}{c} z_N^T \\ D \end{array} \right\| \frac{\min}{d^0}$$

ein zulässiges (verkürztes) Tableau zu (QPDS) ohne Zielfunktion, so kann daraus die Basislösung z abgelesen werden und die Zielfunktion zu $T(z) = z^T \tilde{z}$ berechnet werden. Gehen wir davon aus, dass von dieser Ecke zur nächsten Ecke übergegangen werden soll, so geschieht das so, dass eine Pivotspalte \bar{d}^k gewählt wird und eine Pivotzeile über das Quotientenminimum

$$\delta_k = \min \left\{ \frac{\bar{d}_j^0}{\bar{d}_j^k} : \bar{d}_j^k > 0 \right\}$$

berechnet wird. Daraus kann wie gewohnt die neue Ecke z_{neu} ermittelt werden und anschließend der neue Zielfunktionswert $T(z_{neu}) = z_{neu}^T \tilde{z}_{neu}$.

Zur einfacheren Durchführung dieser Rechnung ist es sinnvoll, zum **Quasinormaltableau** überzugehen. Bei diesem stehen alle Variablen *in der Originalreihenfolge* in der *Basis*, wenn die triviale Gleichung $z_N - z_N = 0$ mit ins Tableau aufgenommen wird. Aus dem genannten Tableau wird so das Tableau

$$\frac{z}{z_N} \left| \begin{array}{c} D \\ d^0 \end{array} \right\|,$$

das eine negative Einheitsmatrix enthält. Ist hier die gleiche Pivotspalte d^k ausgewählt, so bestimmt sich die Pivotzeile wie zuvor aus

$$\delta_k = \min \left\{ \frac{d_j^0}{d_j^k} : d_j^k > 0 \right\} \geq 0$$

ATS zu diesem Tableau vollziehen sich analog zum Normaltableau: Man normiert die k -te Spalte durch Division durch das *negative* Pivotelement und zieht Vielfache dieser Spalte so von den anderen Spalten ab, dass dort in der Pivotzeile sonst nur Nullen entstehen. Ferner ergibt sich die neue rechte Seite zu $d^0 - \delta_k d^k$.

Der Vorteil dieser Darstellung ist, dass die Basislösung $z = d^0$ unmittelbar an der rechten Seite abgelesen werden und so auch der Zielfunktionswert leicht berechnet werden kann. Außerdem ergibt sich die neue Ecke zu

$$z_{neu} = z - \delta_k d^k = d^0 - \delta_k d^k$$

und damit deren Zielfunktionswert zu

$$\begin{aligned} T(z_{neu}) &= (z - \delta_k d^k)^T (\tilde{z} - \delta_k \tilde{d}^k) \\ &= T(z) - 2\delta_k d^{0T} \tilde{d}^k + \delta_k^2 d^{kT} \tilde{d}^k \end{aligned}$$

Man erkennt unmittelbar, dass $T(z_{neu}) < T(z)$ nur gelten kann, wenn gilt:

$$\gamma_k := 2d^{0T} \tilde{d}^k - \delta_k d^{kT} \tilde{d}^k > 0, \quad k = 1, \dots, (n + m)$$

Deshalb benutzen wir diese Werte als Zielfunktionskoeffizienten in unserem Tableau, die Werte, die die Pivotspalte festlegen: Werte $\gamma_k > 0$ signalisieren, dass der Zielfunktionswert der neuen Basislösung niedriger liegt als der der alten Basislösung. Wir setzen zur Abkürzung noch

$$\begin{aligned} \alpha_k &: = d^{0T} \tilde{d}^k, \quad k = 0, 1, \dots, (n + m) \\ \beta_k &: = d^{kT} \tilde{d}^k, \quad k = 1, \dots, (n + m) \end{aligned}$$

Es gilt dann

$$\gamma_k = 2\alpha_k - \delta_k \beta_k, \quad T(z) = \alpha_0 \quad \text{und} \quad T(z_{neu}) = \alpha_0 - \delta_k \gamma_k$$

Zu bemerken ist, dass $\gamma_k > 0$ wegen $\delta_k > 0$ im Falle $\beta_k \geq 0$ eintreten kann, wenn $\alpha_k > 0$ gilt. Daher ist anzuraten, die Werte α_k, β_k zuerst zu berechnen. Gilt $\alpha_k \leq 0$, $\beta_k \geq 0$ gilt, brauchen δ_k und γ_k nicht berechnet zu werden. Ist Q positiv semidefinit, so gilt sogar immer $\beta_k \geq 0$ wegen der Konvexität von T auf \bar{P} .

Das Verfahren sieht dann wie folgt aus: Pivotspalten k werden ausgesucht über $\gamma_k > 0$, Pivotzeilen wie üblich ermittelt, es folgt ein ATS mit Neuberechnung der $\alpha_k, \beta_k, \gamma_k, \delta_k$. Eine Optimallösung ist genau dann erreicht, wenn $\alpha_0 = 0$ gilt.

Beispiel 4.1 Wir betrachten die konvexe quadratische Optimierungsaufgabe, die wir auch schon mit der Modifikation des Verfahrens von Zangwill gelöst haben:

$$\begin{aligned} \text{minimiere} \quad & z(x_1, x_2) = -x_1 + \frac{1}{2}x_1^2 - 2x_2 + \frac{1}{2}x_2^2 \\ \text{s.d.} \quad & 2x_1 + 3x_2 \leq 6 \\ & x_1 + 4x_2 \leq 5 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Das Anfangstableau zu dieser Aufgabe lautet zunächst

	x_1	x_2	λ_1	λ_2	
μ_1	-1		-2	-1	-1
μ_2		-1	-3	-4	-2
y_1	2	3			6
y_2	1	(4)			5

Diese Tableau ist leider noch nicht zulässig. Deshalb müssen wir es zunächst zulässig machen. Wir tun dies mit der zweiten Phase des DPA. Zunächst wählen wir die Zeile " μ_2 " als Hilfszielfunktionszeile aus und wählen die Spalte x_2 als Pivotspalte. Dann kommt y_2 in die Nichtbasis

	x_1	y_2	λ_1	λ_2	
μ_1	-1		-2	-1	-1
μ_2	$\frac{1}{4}$	$\frac{1}{4}$	(-3)	-4	$-\frac{3}{4}$
y_1	$\frac{5}{4}$	$-\frac{3}{4}$			$\frac{9}{4}$
x_2	$\frac{1}{4}$	$\frac{1}{4}$			$\frac{5}{4}$

Da die zweite Zeile noch nicht zulässig ist, wird sie erneut als Hilfszielfunktion ausgewählt. Pivotspalte wird " λ_1 ", μ_2 kommt in die Nichtbasis

	x_1	y_2	μ_2	λ_2	
μ_1	$-\frac{7}{6}$	$-\frac{1}{6}$	$-\frac{2}{3}$	$\frac{5}{3}$	$-\frac{1}{2}$
λ_1	$-\frac{1}{12}$	$-\frac{1}{12}$	$-\frac{1}{3}$	$\frac{4}{3}$	$\frac{1}{4}$
y_1	$\left(\frac{5}{4}\right)$	$-\frac{3}{4}$			$\frac{9}{4}$
x_2	$\frac{1}{4}$	$\frac{1}{4}$			$\frac{5}{4}$

Hilfszielfunktionszeile wird nun die erste Zeile, Pivotspalte werde die Spalte " x_1 ".

	y_1	y_2	μ_2	λ_2	
μ_1	$\frac{14}{15}$	$-\frac{13}{15}$	$-\frac{2}{3}$	$\frac{5}{3}$	$\frac{8}{15}$
λ_1	$\frac{1}{15}$	$-\frac{2}{15}$	$-\frac{1}{3}$	$\frac{4}{3}$	$\frac{2}{15}$
x_1	$\frac{4}{5}$	$-\frac{3}{5}$			$\frac{9}{5}$
x_2	$-\frac{1}{5}$	$\frac{2}{5}$			$\frac{4}{5}$

Nun ist Zulässigkeit erreicht und wir können das eigentliche Verfahren starten, indem wir dem Tableau eine negative Einheitsmatrix hinzufügen, die Basis ordnen, und schließlich um die Werte $\alpha_k, \beta_k, \delta_k, \gamma_k$ ergänzen

	y_1	y_2	μ_2	λ_2	
x_1	$\frac{4}{5}$	$-\frac{3}{5}$			$\frac{9}{5}$
x_2	$-\frac{1}{5}$	$\frac{2}{5}$			$\frac{4}{5}$
y_1	-1				0
y_2		-1			0
μ_1	$\frac{14}{15}$	$-\frac{13}{15}$	$-\frac{2}{3}$	$\frac{5}{3}$	$\frac{8}{5}$
μ_2			-1		0
λ_1	$\frac{1}{15}$	$-\frac{2}{15}$	$-\frac{1}{3}$	$(\frac{4}{3})$	$\frac{2}{5}$
λ_2				-1	0
α_k	$\frac{64}{25}$	$-\frac{63}{25}$	-2	3	$\frac{144}{25}$
β_k	$\frac{102}{75}$			0	
δ_k	$\frac{12}{7}$			$\frac{3}{10}$	
γ_k	$\frac{488}{175}$			6	

An diesem Tableau lesen wir ab, dass es sich lohnt, die Spalte " λ_2 " als Pivotspalte zu nehmen. Diese Wahl liefert das eingerahmte Pivotelement. Die Prognose für die Entwicklung des Zielfunktionswertes ist

$$T(z_{neu}) = \frac{144}{25} - \frac{3}{10}6 = \frac{99}{25}$$

Weil dieser Wert noch nicht null ist, wird das Folgetableau vollständig berechnet. Der entsprechende ATS liefert

	y_1	y_2	μ_2	λ_1	
x_1	$\frac{4}{5}$	$-\frac{3}{5}$			$\frac{9}{5}$
x_2	$-\frac{1}{5}$	$\frac{2}{5}$			$\frac{4}{5}$
y_1	-1				0
y_2		-1			0
μ_1	$(\frac{51}{60})$	$-\frac{7}{10}$	$-\frac{1}{4}$	$-\frac{5}{4}$	$\frac{11}{10}$
μ_2			-1		0
λ_1				-1	0
λ_2	$\frac{1}{20}$	$-\frac{1}{10}$	$-\frac{1}{4}$	$\frac{3}{4}$	$\frac{3}{10}$
α_k	$\frac{241}{100}$	$-\frac{111}{50}$	$-\frac{3}{4}$	$-\frac{9}{4}$	$\frac{99}{25}$
β_k	$\frac{102}{75}$				
δ_k	$\frac{66}{51}$				
γ_k	$\frac{153}{50}$				

Damit muss die erste Spalte als Pivotspalte gewählt werden, das Pivotelement ist eingerahmt. Die Prognose für den folgenden Zielfunktionswert ist

$$T(z_{neu}) = \frac{99}{25} - \frac{66}{51} \frac{153}{50} = 0$$

Es zeigt sich also, dass der folgende ATS der letzte sein wird, weil das Verfahren sein Ziel erreicht hat. also rechnen wir nur noch die Werte der Original-Variablen aus

$$\begin{aligned} x_1 &= \frac{9}{5} - \frac{66}{51} \frac{4}{5} = \frac{13}{17} \\ x_2 &= \frac{4}{5} + \frac{66}{51} \frac{1}{5} = \frac{18}{17} \end{aligned}$$

■

Zu beachten ist: Die beschriebene Vorgehensweise hat leider *keine Garantie auf Erfolg*, da es durchaus Ecken gibt, deren sämtliche Nachbarecken höhere Zielfunktionswerte haben und die dennoch nicht optimal sind, bei denen also $\alpha_o > 0$ und $\gamma_k \leq 0$ gilt für alle $k = 1, \dots, 2(n + m)$. Das Verfahren scheitert dann.

Skizze:

Zuweilen genügt es, dem Verfahren über eine solche Klippe hinwegzuhelfen, indem man für begrenzte Zeit auch Pivotspalten k mit $\gamma_k \leq 0$ wählt (z.B. immer das maximale γ_k), um wieder zu Ecken mit $\gamma_k > 0$ zu gelangen. Besser ist allerdings, systematisch auch Übergänge zu *nicht unmittelbar benachbarten* Ecken zuzulassen, wenn dann sogar theoretisch ein Fortkommen des Verfahrens garantiert ist. Wir besprechen eine solche Vorgehensweise im folgenden Abschnitt.

4.1.2 Das Verfahren von Frank und Wolfe

Um Abhilfe aus dem Dilemma der ungesicherten Konvergenz des letzten Verfahrens zu erhalten, **setzen** wir Q nun als *positiv semidefinit voraus* und unterstellen, dass *(QPS) einen KKT Punkt besitzt*.

Wir verfolgen die *Idee*, die gegebene Zielfunktion an einer vorgegebenen zulässigen Stelle zu *linearisieren*, um in der so entstehenden linearen Optimierungsaufgabe über das Simplexverfahren von Ecke zu Ecke zu wandern und damit in der Nähe der Startecke nach einer weiteren Ecke mit niedrigerem (Original-)Zielfunktionswert zu suchen. Die Hoffnung ist dabei, dass die linearisierte Zielfunktion noch genügend Aussagekraft hat, um diese neue Ecke finden zu können.

Die Originalzielfunktion lautet $T(z) = z^T \tilde{z}$. Berechnet man deren Gradienten, so erhält man leicht $\nabla T(z) = 2\tilde{z}$. Wir betrachten also als an der zulässigen Stelle w linearisierte Zielfunktion (bei Vernachlässigung vom Konstanten in der linearen Approximation) $l(z) = \tilde{w}^T z$. Dazu stellen wir folgende Überlegung an:

Es sei \hat{z} eine Optimallösung der linearisierten Aufgabe, die die Hilfszielfunktion $l(z) = \tilde{w}^T z$ über dem gesamten zulässigen Bereich minimiert. Ferner sei \bar{z} eine Optimallösung der gegebenen Aufgabe, die die Zielfunktion $T(z) = z^T \tilde{z}$ über dem zulässigen Bereich minimiert. Es sei weiterhin w die Stelle, an der die Linearisierung vorgenommen wurde. Wir dürfen dabei $T(w) > T(\bar{z}) = 0$ annehmen, da sonst nicht mehr zu suchen ist.

Dann können wir

$$\bar{z} = w + p \quad \text{und} \quad \hat{z} = w + q$$

ansetzen. Nehmen wir an, \hat{z} sei nicht gleichzeitig auch Optimallösung der gegebenen Aufgabe, d.h. es gelte $T(\hat{z}) > T(\bar{z}) = 0$.

Nach Konstruktion gilt dann

$$\tilde{w}^T \hat{z} \leq \tilde{w}^T \bar{z} \implies \tilde{w}^T (w + q) \leq \tilde{w}^T (w + p) \implies \tilde{w}^T q \leq \tilde{w}^T p$$

Dann gilt die Abschätzung

$$\tilde{w}^T w + 2\tilde{w}^T q \leq \tilde{w}^T w + 2\tilde{w}^T p \leq \tilde{w}^T w + 2\tilde{w}^T p + \tilde{p}^T p$$

letzteres wegen $\tilde{p}^T p \geq 0$ auf grund der Konvexität von T über \bar{P} .

Nun gilt nach Konstruktion aber $\tilde{w}^T w + 2\tilde{w}^T p + \tilde{p}^T p = (\tilde{w} + \tilde{p})^T (w + p) = T(\tilde{z}) = 0$, so dass $\tilde{w}^T w + 2\tilde{w}^T q \leq 0$ gilt. Daraus schließen wir

$$\begin{aligned} \tilde{w}^T w + 2\tilde{w}^T q &\leq 0 \implies \\ 2\tilde{w}^T w - \tilde{w}^T w + 2\tilde{w}^T q &\leq 0 \implies \\ 2\tilde{w}^T w + 2\tilde{w}^T q &\leq \tilde{w}^T w \implies \\ 2\tilde{w}^T (w + q) &\leq \tilde{w}^T w \implies \\ 2\tilde{w}^T \hat{z} &\leq \tilde{w}^T w = T(w) \end{aligned}$$

Betrachten wir nun die Richtungsableitung von T längs $r := (\hat{z} - w)$ an der Stelle w , $\nabla T(w)^T r$. Es gilt

$$\begin{aligned} \nabla T(w)^T r &= 2\tilde{w}^T (\hat{z} - w) \\ &= 2\tilde{w}^T \hat{z} - 2\tilde{w}^T w \\ &\leq \tilde{w}^T w - 2\tilde{w}^T w \\ &= -\tilde{w}^T w = -T(w) < 0 \end{aligned}$$

Damit ist geklärt, dass die Zielfunktion T von w aus in Richtung r echt abfällt! Man berechnet leicht den Punkt, an dem die Zielfunktion von w aus in Richtung $r = (\hat{z} - w)$ am meisten abgefallen ist. Dazu betrachte $z = w + \tau r$ für ein τ und setze $0 = \nabla T(z)^T r = 2[w + \tau r]^T r$. Es ergibt sich als Scheitelpunkt der *nach unten geöffneten* Parabel:

$$\hat{\tau} = \frac{-\tilde{w}^T r}{\tilde{r}^T r} > 0$$

Es ist dann $\tau = \hat{\tau}$, maximal aber $\tau = 1$. Daher bietet sich folgende nun wohldefinierte Vorgehensweise als Verfahren an, bei dem eine Iterationsfolge w^0, w^1, w^2, \dots von zulässigen (nicht notwendig Basis-) Lösungen konstruiert wird, für die die Folge $\{T(w^k)\}_{k \in \mathbf{N}}$ streng monoton abfällt.

Verfahren von Frank und Wolfe

1. **Start** Vorgelegt sei ein zulässiges Simplextableau zur Aufgabe (QPDS) ohne Zielfunktionszeile mit Basislösung w , so dass $T(w) > 0$ gilt.
2. **Lineare Zielfunktion:** Führe die Komponenten von \tilde{w} als Zielfunktionskoeffizienten ins Tableau ein.

3. **Simplexverfahren:** Führe das lineare Simplexverfahren, startend beim aktuellen Tableau, solange durch, bis für die Basislösung z des Tableaus $z^T \tilde{z} = 0$ oder $2\tilde{w}^T z \leq \tilde{w}^T w$ gilt.
4. **Abbruch:** Ist $z^T \tilde{z} = 0$, **stopp**, (z liefert einen KKT Punkt von (QPS))
5. **Aktualisierung:** Sonst berechne τ mit

$$r = (\hat{z} - w), \quad \tau = \min \left\{ 1, -\frac{\tilde{w}^T r}{\tilde{r}^T r} \right\},$$

setze $w = w + \tau r$ und gehe zu 1.

Zu beachten ist: Frank und Wolfe zeigen in ihrer Arbeit, dass unter den gegebenen Voraussetzungen das Verfahren *nach endlich vielen Schritten stoppt!*

Beispiel 4.2 Wir betrachten wieder die Aufgabe

$$\begin{aligned} \text{minimiere} \quad & z(x_1, x_2) = -x_1 + \frac{1}{2}x_1^2 - 2x_2 + \frac{1}{2}x_2^2 \\ \text{s.d.} \quad & 2x_1 + 3x_2 \leq 6 \\ & x_1 + 4x_2 \leq 5 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Das Anfangstableau zu dieser Aufgabe lautet zunächst

	x_1	x_2	λ_1	λ_2	
μ_1	-1		-2	-1	-1
μ_2		-1	-3	-4	-2
y_1	2	3			6
y_2	1	4			5

Diese Tableau ist leider noch nicht zulässig. Dies erledigen wir mit zwei ATS

	x_1	x_2	μ_1	μ_2	
λ_1	$\frac{4}{5}$	$-\frac{1}{5}$	$-\frac{4}{5}$	$\frac{1}{5}$	$\frac{2}{5}$
λ_2	$-\frac{3}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$-\frac{2}{5}$	$\frac{1}{5}$
y_1	2	3			6
y_2	1	4			5

Hier beginnt nun das eigentliche Verfahren. Die aktuelle Ecke bezeichnen wir mit

$$w = z = (x_1, x_2, y_1, y_2, \mu_1, \mu_2, \lambda_1, \lambda_2) = \left(0, 0, 6, 5, 0, 0, \frac{2}{5}, \frac{1}{5}\right)$$

und berechnen $T(w) = 2 \left(6 \cdot \frac{2}{5} + 5 \cdot \frac{1}{5}\right) = \frac{34}{5} > 0$. Damit ist die Ecke noch nicht optimal und wir starten die Minimierung über das Simplexverfahren mit dem obigen Tableau und der Zielfunktion

$$l(z) = \tilde{w}^T z = \frac{2}{5}y_1 + \frac{1}{5}y_2 + 6\lambda_1 + 5\lambda_2$$

Diese muss allerdings erst auf die Nichtbasis des Tableaus umgerechnet werden. Dies geschieht, indem die Zeilen des Tableaus mit dem Koeffizienten der Zielfunktion multipliziert und die Summe als Zielfunktionszeile eingetragen wird:

	x_1	x_2	μ_1	μ_2	min
l	$\frac{14}{5}$	$\frac{14}{5}$	$-\frac{9}{5}$	$-\frac{4}{5}$	$\frac{34}{5}$
λ_1	$\left(\frac{4}{5}\right)$	$-\frac{1}{5}$	$-\frac{4}{5}$	$\frac{1}{5}$	$\frac{2}{5}$
λ_2	$-\frac{3}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$-\frac{2}{5}$	$\frac{1}{5}$
y_1	2	3			6
y_2	1	4			5

Wir wählen die erste Spalte als Pivotspalte und erhalten das eingeklammerte Pivot. Der ATS liefert

	λ_1	x_2	μ_1	μ_2	min
l	$-\frac{7}{2}$	$\frac{7}{2}$	$\frac{39}{5}$	$-\frac{3}{2}$	$\frac{27}{5}$
x_1	$\frac{5}{4}$	$-\frac{1}{4}$	-1	$\frac{1}{4}$	$\frac{1}{2}$
λ_2	$\frac{3}{4}$	$\frac{1}{4}$		$-\frac{1}{4}$	$\frac{1}{2}$
y_1	$-\frac{5}{2}$	$\frac{7}{2}$	2	$-\frac{1}{2}$	5
y_2	$-\frac{5}{4}$	$\left(\frac{17}{4}\right)$	1	$-\frac{1}{4}$	$\frac{9}{2}$

Da der aktuelle Zielwert noch nicht die Hälfte des anfänglichen Werts, $\frac{17}{4}$, unterschritten hat und auch $T(z) = \frac{9}{2} > 0$ gilt, muss ein weiterer ATS durchgeführt werden. Wir wählen das eingeklammerte Pivotelement und erhalten

	λ_1	y_2	μ_1	μ_2	min
l	$-\frac{42}{17}$	$-\frac{14}{17}$	$\frac{593}{85}$	$-\frac{22}{17}$	$\frac{144}{85}$
x_1	$\frac{20}{17}$	$\frac{1}{17}$	$-\frac{16}{17}$	$\frac{4}{17}$	$\frac{13}{17}$
λ_2	$\frac{14}{17}$	$-\frac{1}{17}$	$-\frac{1}{17}$	$-\frac{4}{17}$	$\frac{4}{17}$
y_1	$-\frac{25}{17}$	$-\frac{14}{17}$	20	$-\frac{5}{17}$	$\frac{22}{17}$
x_2	$-\frac{17}{17}$	$\frac{4}{17}$	$\frac{17}{4}$	$-\frac{1}{17}$	$\frac{17}{18}$

Nun ist mit $l(z) = \frac{144}{85}$ die Hälfte des anfänglichen Wertes, $\frac{17}{4}$, unterschritten, die lineare Optimierung endet zunächst hier. Gleichzeitig gilt aber auch $T(z) = 0$, deshalb braucht keine neue Linearisierung der Zielfunktion mehr durchgeführt werden, das Verfahren ist beendet. Das Endergebnis ist das bekannte.

Bemerkung: Wir haben das Verfahren von Frank und Wolfe als eigenständiges Verfahren formuliert und wissen nun, dass es in endlich vielen Schritten eine Ecke z ermittelt mit $T(z) = 0$. Ausgangspunkt der Überlegungen war aber, eine Hilfskonstruktion für die Situation im Verfahren von Barankin und Dorfman zu finden, in der dieses Verfahren stoppt, weil zu keiner benachbarten Ecke mit kleinerem Zielfunktionswert übergegangen werden kann. Hier kann die Vorgehensweise von Frank und Wolfe herangezogen werden, diese Lücke zu füllen.

Ein **Hybridverfahren** geht wie folgt vor:

1. Man startet mit der Anwendung von Barankin und Dorfman, bis dieses Verfahren bei einer Ecke z^1 stoppt. Gilt $T(z^1) = 0$, so ist das Hybridverfahren beendet. **Stopp.**
2. Gilt $T(z^1) > 0$, so startet man Frank und Wolfe mit $z^1 = w$ und führt dieses Verfahren solange durch, bis eine Ecke z^2 angelaufen wurde, für die $T(z^2) < T(z^1)$. Dies ist nach endlich vielen Schritten der Fall! Gilt nun immer noch $T(z^2) > 0$, so geht man zu 1.

4.1.3 Übungsaufgaben

Aufgabe 1

Man programmiere eine Kombination aus dem Verfahren von Barankin und Dorfman einerseits und dem Verfahren von Frank und Wolfe andererseits, wobei immer, wenn möglich, das erste Verfahren eingesetzt wird, sonst das zweite. Dies soll über Parameter zu steuern sein, so dass auch jeweils nur eines der Verfahren zum Einsatz kommt. Testen Sie dies Verfahren an zufallserzeugten Beispielaufgaben mit konvexer Zielfunktion.

4.2 Die Verfahren von Wolfe

Vorgegeben sei die Aufgabe (QPS), zu der wir das KKT System lösen wollen. Wir wollen nun Verfahren besprechen, bei denen nach Einführung künstlicher Variablen die Komplementarität durchgehend für alle Basislösungen erfüllt ist.

4.2.1 Die kurze Form

Das zuerst zu beschreibende Verfahren von Wolfe ist in der Literatur für die andere Standardform der quadratischen Aufgabe angegeben

$$(QPSG) \quad \begin{cases} \text{minimiere} & \bar{f}(\bar{x}) = \frac{1}{2}\bar{x}^T \bar{Q}\bar{x} + \bar{p}^T \bar{x} \\ \text{s.d.} & \bar{H}\bar{x} = \bar{h}, \bar{x} \geq 0 \end{cases},$$

in die wir die unsere vorgegebene Aufgabe zunächst überführen wollen. Dies gelingt, indem wir Schlupfvariablen einführen und

$$\bar{H} = (A, I), \bar{h} = a, \bar{x}^T = (x^T, y^T), \bar{Q} = \begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix}, \bar{p} = \begin{pmatrix} p \\ 0 \end{pmatrix} \quad \text{und} \quad \bar{f}(\bar{x}) = f(x)$$

setzen. Das führt zu den KKT Bedingungen

$$(QPSGKKT) \quad \begin{cases} \bar{Q}\bar{x} + \bar{p} + \bar{H}^T \mu - \lambda = 0, \lambda \geq 0 \\ \bar{H}\bar{x} = \bar{h}, \bar{x} \geq 0 \\ 0 = \lambda^T \bar{x} \end{cases}$$

Die *Idee von WOLFE* ist nun, das durch die ersten beiden Zeilen dargestellte Polyeder in ein Simplextableau unter Benutzung künstlicher Variabler so einzubringen, dass die Komplementarität anfänglich gegeben ist. Dann wird versucht, durch Simplexschritte die künstlichen Variablen in die Nichtbasis zu bringen, ihnen also in der Basislösung den Wert null zuzuweisen und dabei die Komplementarität $0 = \lambda^T \bar{x}$ im Tableau immer aufrechtzuerhalten.

Gesucht ist also eine Basislösung, bei der die Gleichung $0 = \lambda^T \bar{x}$ erfüllt ist. Es bietet sich an, zunächst alle Variablen in die Nichtbasis zu nehmen. Dann enthält die Basis in ihrem unteren Teil Nullen. Wir füllen den oberen Teil der Basis mit einem künstlichen Vektor z freier Variablen. Nun führt man ATS durch, die den unteren Teil des Tableaus zulässig (Phase 0 und Phase 1) machen, wobei die Zeilen mit den künstlichen Variablen nicht an der

Pivotzeilenwahl teilnehmen. Danach multiplizieren wir die oberen Gleichungen, die unzulässig sind, mit (-1) , so dass das Tableau nun zulässig ist. Die jetzt verbleibenden künstlichen Variablen erklären wir nun zu vorzeichenbeschränkt und bezeichnen sie oBdA wieder alle mit z_1, \dots, z_n . Das Tableau hat dann die Form

$$\begin{aligned} z + \bar{Q}\bar{x}^1 + \bar{H}^T\mu - \lambda &= -\bar{p} \\ \bar{x}^2 + \bar{A}\bar{x}^1 &= \bar{a} \end{aligned} \quad (4.1)$$

Das Tableau erhält dann eine Zielfunktionszeile, die der Minimierung der Hilfszielfunktion

$$w(z) = \sum_{i=1}^n z_i$$

entspricht. Diese lineare Optimierungsaufgabe wird nun mit Simplex austauschschritten, die w absenken, gelöst, wobei nur ATS erlaubt sind, nach deren Ausführung *nicht* für ein j sowohl \bar{x}_j und λ_j beide gleichzeitig in der Basis sind und auch keine ATS, die \bar{x}_j und λ_j austauschen für ein j . Wir nennen dies **eingeschränkte Austauschregel**.

Ein Beispiel

Beispiel 4.3 *Wir betrachten erneut die quadratische Optimierungsaufgabe*

$$\begin{aligned} \text{minimiere} \quad & z(x_1, x_2) = -x_1 - 2x_2 + \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 \\ \text{s.d.} \quad & 2x_1 + 3x_2 \leq 6 \\ & x_1 + 4x_2 \leq 5 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Führen wir diese zunächst in eine Aufgabe vom Format (QPSG) über:

$$\bar{H}^T = \begin{pmatrix} 2 & 1 \\ 3 & 4 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \bar{h} = \begin{pmatrix} 6 \\ 5 \end{pmatrix}, \quad \bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}, \quad \bar{Q} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \bar{p} = \begin{pmatrix} -1 \\ -2 \\ 0 \\ 0 \end{pmatrix}$$

Dies führt zu folgendem (verkürzten) Anfangstableau der kurzen Form

	x_1	x_2	x_3	x_4	μ_1	μ_2	λ_1	λ_2	λ_3	λ_4	
w	1	1			6	6	-1	-1	-1	-1	3
z_1	1				2	1	-1				1
z_2		1			3	4		-1			2
z_3					1				-1		0
z_4						1				-1	0
0	2	3	1								6
0	1	4		1							5

Als erstes bringen wir x_3 und x_4 und dann μ_1 und μ_2 in die Basis und dafür z_3 und z_4 und zwei Nullen in die Nichtbasis, wo alle entsprechenden Spalten sofort gestrichen werden können. Da wir nicht wirklich an den optimalen Werten der μ s interessiert sind und diese freie Variablen sind, können auch die entsprechenden Zeilen anschließend gestrichen werden.

	x_1	x_2	λ_1	λ_2	λ_3	λ_4	
w	1	1	-1	-1	5	5	3
z_1	(1)		-1		2	1	1
z_2		1		-1	3	4	2
x_3	2	3					6
x_4	1	4					5

Wir bringen nun x_1 in die Basis, das Pivotelement ist markiert. Dadurch kommt z_1 in die Nichtbasis, die zugehörige Spalte wird also nach dem ATS sofort gestrichen.

	x_2	λ_1	λ_2	λ_3	λ_4	
w	1	0	-1	3	4	2
x_1		-1		2	1	1
z_2	1		-1	3	4	2
x_3	3	2		-4	-2	4
x_4	(4)	1		-2	-1	4

Nun bringen wir x_2 in die Basis (λ_3 und λ_4 sind nicht erlaubt), dadurch kommt x_4 in die Nichtbasis

	x_4	λ_1	λ_2	λ_3	λ_4	
w	$-\frac{1}{4}$	$-\frac{1}{4}$	-1	$\frac{14}{4}$	$\frac{17}{4}$	1
x_1		-1		2	1	1
z_2	$-\frac{1}{4}$	$-\frac{1}{4}$	-1	$\frac{14}{4}$	$(\frac{17}{4})$	1
x_3	$-\frac{3}{4}$	$\frac{5}{4}$		$-\frac{10}{4}$	$-\frac{5}{4}$	1
x_2	$\frac{1}{4}$	$\frac{1}{4}$		$-\frac{2}{4}$	$-\frac{1}{4}$	1

Als Nächstes bringen wir λ_4 in die Basis, stattdessen geht z_2 aus der Basis, wodurch unsere Hilfszielfunktion auf null gebracht ist. Das nächste Tableau ist also das Endtableau:

	x_4	λ_1	λ_2	λ_3	
x_1	$\frac{1}{17}$	$-\frac{16}{17}$	$\frac{4}{17}$	$\frac{20}{17}$	$\frac{13}{17}$
λ_4	$-\frac{1}{17}$	$-\frac{1}{17}$	$-\frac{4}{17}$	$\frac{14}{17}$	$\frac{4}{17}$
x_3	$-\frac{14}{17}$	$\frac{20}{17}$	$-\frac{5}{17}$	$-\frac{25}{17}$	$\frac{22}{17}$
x_2	$\frac{4}{17}$	$\frac{17}{17}$	$-\frac{1}{17}$	$-\frac{5}{17}$	$\frac{18}{17}$

Die gesuchte Optimallösung lautet damit $x^{*T} = (\frac{13}{17}, \frac{18}{17}, \frac{22}{17}, 0)$, was sehr wohl mit dem uns schon bekannten Ergebnis übereinstimmt.

Konvergenzuntersuchung

Wie wir gesehen haben, hat die angewendete Strategie bei unserem Beispiel zum Ziel geführt. Dies mag an der speziellen Struktur der Aufgabe gelegen haben. Wir wollen untersuchen, wann die Strategie garantiert zum Ziel führt. Dazu verwenden wir das folgende mithilfe der linearen Dualitätstheorie zu beweisende

Lemma 4.4 Vorgegeben sei die lineare Optimierungsaufgabe

$$\begin{aligned}
 & \text{minimiere} && \hat{q}^T z \\
 & \text{s. d.} && \hat{Q}x + \hat{H}^T \mu - \lambda + \hat{D}z = -\hat{p} \\
 & && \hat{H}x = \hat{h} \\
 & && x, \lambda \geq 0, z \geq 0, x_R = 0, \lambda_S = 0
 \end{aligned}$$

bei der $\hat{q}, \hat{Q}, \hat{H}, \hat{D}, \hat{p}, \hat{h}$ in passender Dimension vorgegeben sind, \hat{Q} positiv semidefinit, und R und S zwei vorgegebene disjunkte Indermengen.

Ist dann z^* der z -Teil einer Optimallösung $(\bar{x}^*, \lambda^*, z^*)$ der Aufgabe und geben $x_S^* > 0$, $\lambda_R^* > 0$ genau die positiven Komponenten von x^* bzw λ^* an, so gibt es einen Vektor r mit

$$\hat{Q}r = 0, \hat{H}r = 0, \hat{q}^T z^* = -\hat{p}^T r$$

Beweis: [KKR], Kap 7, 2. und 4.

□

Dieses Lemma wollen wir wie folgt anwenden: Es werde das skizzierte Verfahren angewendet und das Verfahren ende mit einem Abbruchtableau, bei dem $w(z)$ noch nicht den Wert $w(z) = 0$ erreicht hat. Sei dann S die Menge der Indizes i in der Basis, für die die Variable \bar{x}_i in der Basislösung positiv ist und R die Menge der Indizes i in der Basis, für die die Variable λ_i in der Basislösung positiv ist. Nach Konstruktion befinden sich dann die zu S gehörigen Komponenten λ_i und die zu R gehörigen Komponenten von \bar{x} in der Nichtbasis, nehmen also in der Basislösung den Wert null an. Weil das Verfahren an dieser Stelle nicht weitergeht, ist das letzte Tableau (Nichtentartung vorausgesetzt) Optimaltableau zu der im Lemma genannten Aufgabe mit

$$\hat{q}^T = e^T = (1, \dots, 1), \hat{Q} = \bar{Q}, \hat{H} = \bar{H}, \hat{D} = D, \hat{p} = \bar{p}, \hat{h} = \bar{h}.$$

Das Lemma findet Anwendung, es gibt daher einen Vektor r mit $\bar{Q}r = 0$, $\bar{H}r = 0$, $e^T z^* = -\bar{p}^T r$, wobei $e^T z^*$ den aktuellen Wert der Hilfsziel-funktion angibt. Man kann zeigen, dass diese Folgerung auch im entarteten Fall gilt.

Unter folgenden **hinreichenden Bedingungen** an die Aufgabenstellung ist dieser Wert gleich null und das letzte Tableau gibt in seiner Basislösung einen gesuchten KKT Punkt für (QPSGKKT) an:

$$\bar{p} \perp \ker(\bar{Q}) \cap \ker(\bar{H})$$

Insbesondere ist das der Fall, wenn $\bar{p} = 0$ gilt oder wenn $\ker(\bar{Q}) = 0$ gilt, also \bar{Q} positiv definit ist.

Damit ist abgeklärt, warum das Verfahren beim letzten Beispiel funktionier-te, denn dort gilt $\ker(\bar{Q}) \cap \ker(\bar{H}) = 0$.

4.2.2 Die lange Form

Leider enthält die bisherige Betrachtung keine Aussage über die Lösbarkeit des KKT Problems, wenn das obige Verfahren scheitert. Betrachten wir daher eine Modifikation des eben beschriebenen Verfahrens, die generell funktioniert, solange Q nur positiv semidefinit ist.

Die kurze Form ermittelt erfolgreich einen KKT Punkt, wenn $\bar{p} = 0$ gilt. Dies machen wir uns wie folgt zu nutze: wir bringen die Aufgabe in die Form

$$\begin{aligned}\bar{Q}\bar{x} + \bar{H}^T\mu - \lambda + \tau\bar{p} &= 0, \\ \bar{H}\bar{x} &= \bar{h}, \\ 0 = \lambda^T\bar{x}, \lambda \geq 0, \bar{x} \geq 0, \tau &= 1\end{aligned}$$

und lösen zunächst in **einer ersten Phase** das Problem für $\tau = 0$. Für diesen Wert des Parameters τ ermittelt die kurze Form eine Lösung. Wir erhalten dadurch eine Basislösung zur Aufgabe

$$(A) \quad \begin{cases} \text{minimiere} & -\tau \\ \text{s.d.} & \bar{Q}\bar{x} + \bar{H}^T\mu - \lambda + \tau\bar{p} = 0 \\ & \bar{H}\bar{x} = \bar{h}, \\ & \lambda \geq 0, \bar{x} \geq 0, \tau \geq 0 \end{cases}$$

mit $\tau = 0$, die alle geforderten Komplementaritätsbedingungen erfüllt.

Eine erneute Anwendung des Simplexverfahrens *unter der eingeschränkten Austauschregel* auf die Aufgabe (A) kann dann **in einer Phase 2** zu einem KKT Punkt der gegebenen Aufgabe führen, sofern $\tau = 1$ erreicht wird. Fragen wir uns, unter welchen hinreichenden Bedingungen dies der Fall ist.

Nun, Hilfe leistet wieder das obige Lemma. Wenden wir dies an, wenn das Simplexverfahren mit der eingeschränkten Austauschregel, angewendet auf die Aufgabe (A), keine Pivotspalte mehr wählen kann. Setzen wir

$$\hat{q}^T = -1, \hat{Q} = \bar{Q}, \hat{H} = \bar{H}, \hat{D} = \bar{p}, \hat{p} = 0, \hat{h} = \bar{h}, z = \tau$$

so liefert uns das Lemma die Existenz eines Vektors r mit $(-1)\tau^* = 0 \cdot r = 0$. Ein Stocken des Verfahrens kann also nur bei $\tau = 0$ erfolgen! Damit gilt, lediglich unter der Voraussetzung, dass \bar{Q} positiv semidefinit ist:

- entweder die Bearbeitung der Aufgabe (A) mit dem Simplexverfahren mit der eingeschränkten Umrechnungsregel bricht ab, dann bei $\tau = 0$ und es existiert kein KKT Punkt zu (QPSGKKT).

- oder das Simplexverfahren mit der eingeschränkten Umrechnungsregel führt zu einem Abbruch wegen unbeschränkter Zielfunktion. In diesem Fall gibt es natürlich eine Zwischenlösung mit $\tau = 1$ und daher einen KKT Punkt zu (QPSGKKT).

Beispiel 4.5 *Betrachten wir noch einmal die Aufgabe*

$$\begin{aligned} \text{minimiere} \quad & z(x_1, x_2) = -x_1 + \frac{1}{2}x_1^2 - 2x_2 + \frac{1}{2}x_2^2 \\ \text{s.d.} \quad & 2x_1 + 3x_2 \leq 6 \\ & x_1 + 4x_2 \leq 5 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Das Anfangstableau zur langen Form lautet für die erste Phase offensichtlich

	x_1	x_2	x_3	x_4	μ_1	μ_2	λ_1	λ_2	λ_3	λ_4	τ		
w	1	1			6	6	-1	-1	-1	-1	-3	0	
z_1	1				2	1	-1					-1	0
z_2		1			3	4		-1				-2	0
z_3					1				-1				0
z_4						1				-1			0
0	2	3	1										6
0	1	4		1									5

Wir führen nun das Simplexverfahren unter Sperrung der Spalte " τ " mit der eingeschränkten Austauschregel durch. Zunächst mal läuft das Verfahren ab wie in der kurzen Form: als erstes werden die Nullen in die Nichtbasis und die freien Variablen in die Basis gebracht. Das ergibt

	x_1	x_2	λ_1	λ_2	λ_3	λ_4	τ	
w	1	1	-1	-1	5	5	-3	0
z_1	(1)		-1		2	1	-1	0
z_2		1		-1	3	4	-2	0
x_3	2	3						6
x_4	1	4						5

Auch das nächste Pivot kann identisch gewählt werden und führt zu

	x_2	λ_1	λ_2	λ_3	λ_4	τ	
w	1	0	-1	3	4	-2	0
x_1		-1		2	1	-1	0
z_2	(1)		-1	3	4	-2	0
x_3	3	2		-4	-2	2	6
x_4	4	1		-2	-1	1	5

Nun allerdings weicht die Vorgehensweise wegen der starken Entartung ab: als Pivotspalte muss die Spalte "x₂" gewählt werden, der ATS bringt z₂ in die Nichtbasis. Damit ist die Hilfszielfunktion auf null gebracht und wir können nun in Phase 2 erneut unter der eingeschränkten Austauschregel τ maximieren

	λ ₁	λ ₂	λ ₃	λ ₄	τ	
-τ	0	0	0	0	1	0
x ₁	-1		2	1	-1	0
x ₂		-1	3	4	-2	0
x ₃	2	3	-13	-14	8	6
x ₄	1	4	-14	-17	(9)	5

Es muss nun die Spalte "τ" als Pivotspalte, sodass τ in die Basis und dafür x₄ in die Nichtbasis kommen.

	λ ₁	λ ₂	λ ₃	λ ₄	x ₄	
-τ	$-\frac{1}{9}$	$-\frac{4}{9}$	$\frac{14}{9}$	$\frac{17}{9}$	$-\frac{1}{9}$	$-\frac{5}{9}$
x ₁	$-\frac{8}{9}$	$\frac{4}{9}$	$\frac{4}{9}$	$-\frac{8}{9}$	$\frac{1}{9}$	$\frac{5}{9}$
x ₂	$\frac{2}{9}$	$-\frac{1}{9}$	$-\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{10}{9}$
x ₃	$\frac{10}{9}$	$\frac{5}{9}$	$-\frac{5}{9}$	$\left(\frac{10}{9}\right)$	$-\frac{8}{9}$	$\frac{14}{9}$
τ	$\frac{1}{9}$	$\frac{4}{9}$	$-\frac{14}{9}$	$-\frac{17}{9}$	$\frac{1}{9}$	$\frac{5}{9}$

Es kann nun nur die Pivotspalte "λ₄" gewählt werden, Quotientenminimierung liefert einen Steigerungsfaktor von $\frac{14}{10}$ für τ, wodurch dieser über den Wert von τ = 1 hinausläuft. Daher kann nur mehr das $\frac{4}{17}$ fache der Spalte "λ₄" von der rechten Seite abgezogen werden, um τ = 1 zu erreichen. Dies liefert neben λ₄ = $\frac{4}{17}$

$$\begin{aligned}
 x_1 &= \frac{5}{9} + \frac{4}{17} \frac{8}{9} = \frac{13}{17} \\
 x_2 &= \frac{10}{9} - \frac{4}{17} \frac{2}{9} = \frac{18}{17} \\
 x_3 &= \frac{14}{9} - \frac{4}{17} \frac{10}{9} = \frac{22}{17}
 \end{aligned}$$

und damit das erwartete Ergebnis.

4.2.3 Vereinfachte Formen

Betrachten wir, wie die KKT Bedingungen (*QPSGKKT*) explizit aussehen, wenn wir die Daten der Aufgabe (*QPS*) einsetzen:

$$(QPSGKKT) \quad \begin{cases} \begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} A^T \\ I \end{pmatrix} \mu - \begin{pmatrix} \lambda^1 \\ \lambda^2 \end{pmatrix} = - \begin{pmatrix} p \\ 0 \end{pmatrix}, \lambda^1, \lambda^2 \geq 0 \\ (A, I) \begin{pmatrix} x \\ y \end{pmatrix} = a, x, y \geq 0 \\ 0 = \begin{pmatrix} \lambda^1 \\ \lambda^2 \end{pmatrix}^T \begin{pmatrix} x \\ y \end{pmatrix} \end{cases}$$

oder

$$\begin{aligned} -Qx - A^T \mu + \lambda^1 &= p, \lambda^1 \geq 0 \\ -\mu + \lambda^2 &= 0, \lambda^2 \geq 0 \\ Ax + y &= a, x, y \geq 0 \\ 0 &= \lambda^{1T} x + \lambda^{2T} y \end{aligned} \tag{4.2}$$

wobei wir die erste Zeile mit (-1) multipliziert haben. Es sei weiterhin $a \geq 0$ vorausgesetzt.

Vereinfachte kurze Form

Wenn wir nun aus dieser Struktur heraus eine geeignete Anfangsbasis aufstellen möchten, so bietet es sich an, neben dem Vektor y für die dritte Gleichung den Vektor λ^1 für die erste Gleichung und den Vektor λ^2 für die zweite Gleichung zu wählen. Danach werden als erstes die Variablen μ in die Basis und die Variablen λ^2 in die Nichtbasis gebracht. Dabei ändert sich das Tableau nur dadurch, dass diese Variablen ausgetauscht werden, alle anderen Daten bleiben unverändert, da die entsprechenden Zeilen im Tableau jeweils genau einen Eintrag (-1) und sonst nur Nullen haben.

Dann erfüllt die Basislösung des entstandenen Tableaus die Komplementaritätsbedingungen! Außerdem können die Zeilen mit den freien Variablen μ sofort gestrichen werden. Letztlich bedeutet der bisherige Vorgang lediglich, dass die Variablen μ im System durch die Variablen λ^2 ersetzt und dann gestrichen werden.

Dabei hat die Wahl von λ^1 für die Basis leider den Nachteil, dass u.U. keine *zulässige* Basislösung festgelegt ist. Gibt es negative Komponenten in p , so korrigieren wir dies durch Einführung einer künstlichen Variablen z , indem wir dem gesamten System $z \cdot s$ hinzuaddieren, wo s ein Vektor ist, dessen Komponenten $s_i = -1$ sind, falls $p_i < 0$ und $s_i \leq 0$ beliebig sonst. Wir können dann die Zulässigkeit des Tableaus durch einen einzigen ATS erreichen. Dazu wählen wir die künstliche Spalte als Pivotspalte und die Pivotzeile i als eine Zeile, bei der p den am meisten negativen Wert annimmt. Man überzeugt sich leicht, dass das Tableau mit diesem ATS zulässig wird. Die bestehende Komplementarität wird durch den ATS nicht gestört, da die Variable z in die Basis kommt und eine andere Variable aus λ^1 die Basis verlässt.

Zu beachten ist, dass in dem Fall, in dem bereits von vorne herein $p \geq 0$ gilt, also keine Zusatzvariable z eingeführt werden muss, durch das Erfülltsein der Komplementaritätsbedingungen bereits ein KKT Punkt zu (QPS) gefunden wurde!

Wir haben also ein für unser Verfahren geeignetes zulässiges Ausgangstableau gefunden und können nun das Simplexverfahren wie zuvor mit der eingeschränkten Austauschregel durchführen mit dem Ziel, die künstliche Variable z unter Aufrechterhaltung der Komplementarität auf Null zu bringen. Zur Überprüfung der Konvergenz dieser Vorgehensweise bemühen wir wieder unser obiges technisches Lemma mit den Daten

$$\hat{q}^T = (0, \dots, 0, 1), \quad \hat{Q} = \bar{Q}, \quad \hat{H} = \bar{H}, \quad \hat{D} = \bar{s}, \quad \hat{p} = \bar{p}, \quad \hat{h} = \bar{h},$$

wobei \bar{s} der Vektor ist, der aus s mit ergänzten Nulleinträgen entsteht und bei dem das Vorzeichen in den Komponenten, die anfänglich in (4.2) mit (-1) multipliziert wurden, wieder umgedreht wurde. Dabei tun wir so, als seien z und s von vorneherein eingeführt, was offensichtlich keinen Unterschied macht.

Die Aussage ist dann die gleiche wie zuvor: Das Verfahren arbeitet erfolgreich, wenn $\bar{p} \perp \ker(\bar{Q}) \cap \ker(\bar{H})$ gilt, was z.B. zutrifft, wenn $p = 0$ oder $\ker(Q) = 0$ gilt. Im letzteren Fall gilt nämlich für $\bar{r} \in \ker(\bar{Q}) \cap \ker(\bar{H})$

$$\begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{und} \quad (A, I) \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = 0 \implies \\ Qr_1 = 0 \quad \text{und} \quad Ar_1 + r_2 = 0 \implies (r_1, r_2) = (0, 0)$$

Beispiel 4.6 Betrachten wir erneut die Aufgabe

$$\begin{aligned}
 \text{minimiere} \quad & z(x_1, x_2) = -x_1 + \frac{1}{2}x_1^2 - 2x_2 + \frac{1}{2}x_2^2 \\
 \text{s.d.} \quad & 2x_1 + 3x_2 \leq 6 \\
 & x_1 + 4x_2 \leq 5 \\
 & x_1, x_2 \geq 0
 \end{aligned}$$

Anfangstableau ist nun

	x_1	x_2	λ_3	λ_4	z	
λ_1	-1		-2	-1	-1	-1
λ_2		-1	-3	-4	(-1)	-2
x_3	2	3				6
x_4	1	4				5

Ein ATS mit dem gekennzeichneten Pivotelement führt zu

	x_1	x_2	λ_3	λ_4	λ_2	
λ_1	-1	(1)	1	3	-1	1
z		1	3	4	-1	2
x_3	2	3				6
x_4	1	4				5

Das Tableau ist damit zulässig und komplementär. Die nächste Pivotspalte ist eine wählbare mit positivem Eintrag in der Zeile "z". Offensichtlich ist hier nur die Spalte "x₂" möglich. Quotientenminimierung führt auf das markierte Pivot. Der ATS liefert

	x_1	λ_1	λ_3	λ_4	λ_2	
x_2	-1	1	1	3	-1	1
z	1	-1	2	1	0	1
x_3	5	-3	-3	-9	3	3
x_4	(5)	-4	-4	-12	4	1

Nun kommt x_1 in die Basis und x_4 verlässt die Basis.

	x_4	λ_1	λ_3	λ_4	λ_2	
x_2	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{3}{5}$	$-\frac{1}{5}$	$\frac{6}{5}$
z	$-\frac{1}{5}$	$-\frac{1}{5}$	$\frac{14}{5}$	$(\frac{17}{5})$	$-\frac{4}{5}$	$\frac{4}{5}$
x_3	-1	1	1	-9	3	3
x_1	$\frac{1}{5}$	$-\frac{4}{5}$	$-\frac{4}{5}$	$-\frac{12}{5}$	$\frac{4}{5}$	$\frac{1}{5}$

Nun kommt noch λ_4 in die Basis und die Hilfsvariable z kommt in die Nichtbasis und ist damit auf null gebracht.

	x_4	λ_1	λ_3	λ_2	
x_2	$\frac{4}{17}$	$\frac{4}{17}$	$-\frac{5}{17}$	$-\frac{1}{17}$	$\frac{18}{17}$
λ_4	$-\frac{1}{17}$	$-\frac{1}{17}$	$\frac{14}{17}$	$-\frac{4}{17}$	$\frac{4}{17}$
x_3	$-\frac{14}{17}$	$\frac{20}{17}$	$-\frac{25}{17}$	$-\frac{5}{17}$	$\frac{22}{17}$
x_1	$\frac{1}{17}$	$-\frac{16}{17}$	$\frac{20}{17}$	$\frac{4}{17}$	$\frac{13}{17}$

Das Ergebnis ist das gleiche wie zuvor.

Vereinfachte Lange Form

Besonders interessant ist die Übertragung der Ideen der langen Form auf den vereinfachten Fall. Nimmt man zunächst p in die linke Seite des Gleichungssystems auf, so ergibt sich aus den obigen Überlegungen direkt ein *erstes zulässiges und komplementäres* Tableau

	x	λ^2	τ	
λ^1	$-Q$	$-A^T$	$-p$	0
y	A	0	0	a

zum System

$$\begin{aligned} -Qx - A^T \mu + \lambda^1 - \tau p &= 0, \lambda^1, \lambda^2 \geq 0 \\ \mu - \lambda^2 &= 0 \\ Ax + y &= a, x, y \geq 0 \end{aligned}$$

wobei in der Basislösung zunächst $\tau = 0$ gilt. Dies liegt daran, dass die rechte Seite von vorneherein zulässig ist und die Ausgangsüberlegungen zu einem komplementären Tableau führen: $\lambda^{1T}x = 0$ und $\lambda^{2T}y = 0$

Danach wird τ maximiert. Wiederum zeigt das obige Lemma, dass die Maximierung von τ unter diesen Bedingungen bei Verwendung der eingeschränkten Pivotwahl entweder zur Optimallösung mit $\tau = 0$ führt, was in der Regel bereits nach wenigen ATS erkannt wird oder τ beliebig erhöht werden kann. Eine Erhöhung auf den Wert $\tau = 1$ führt zum gesuchten KKT Punkt.

Beispiel 4.7 Betrachten wir noch einmal die Aufgabe

$$\begin{array}{ll}
 \text{minimiere} & z(x_1, x_2) = -x_1 + \frac{1}{2}x_1^2 - 2x_2 + \frac{1}{2}x_2^2 \\
 \text{s.d.} & 2x_1 + 3x_2 \leq 6 \\
 & x_1 + 4x_2 \leq 5 \\
 & x_1, x_2 \geq 0
 \end{array}$$

Anfangstableau ist nun

	x_1	x_2	λ_3	λ_4	τ	
$-\tau$					1	0
λ_1	-1		-2	-1	(1)	0
λ_2		-1	-3	-4	2	0
x_3	2	3				6
x_4	1	4				5

Dieses Tableau ist bereits komplementär und zulässig. Es soll die Zielfunktion $-\tau$ minimiert werden. Der ATS mit dem sich zwangsweise ergebenden markierten Pivot liefert

	x_1	x_2	λ_3	λ_4	λ_1	
$-\tau$	1		2	1	-1	0
τ	-1		-2	-1	1	0
λ_2	(2)	-1	1	-2	-2	0
x_3	2	3				6
x_4	1	4				5

wobei die Hilfszielfunktion $-\tau$ im Folgenden gestrichen werden kann, da τ selbst in der Basis auftritt. τ muss nun maximiert werden, d. h. negative Einträge in der τ Zeile führen zur Pivotspalte. Der nächste ATS bringt daher x_1 in die Basis und λ_2 in die Nichtbasis

	λ_2	x_2	λ_3	λ_4	λ_1	
τ	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{3}{2}$	-2	0	0
x_1	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	-1	-1	0
x_3	-1	4	-1	2	2	6
x_4	$-\frac{1}{2}$	($\frac{9}{2}$)	$-\frac{1}{2}$	1	1	5

Nun muss das eingeklammerte Pivot gewählt werden, was zu folgendem Ta-

bleau führt

	λ_2	x_4	λ_3	λ_4	λ_1	
τ	$\frac{4}{9}$	$\frac{1}{9}$	$-\frac{14}{9}$	$-\frac{17}{9}$	$\frac{1}{9}$	$\frac{5}{9}$
x_1	$\frac{4}{9}$	$\frac{1}{9}$	$\frac{4}{9}$	$-\frac{8}{9}$	$-\frac{8}{9}$	$\frac{2}{9}$
x_3	$-\frac{5}{9}$	$-\frac{8}{9}$	$-\frac{5}{9}$	$(\frac{10}{9})$	$\frac{10}{9}$	$\frac{14}{9}$
x_2	$-\frac{1}{9}$	$\frac{2}{9}$	$-\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{10}{9}$

Es kann nun nur die Pivotspalte " λ_4 " gewählt werden, Quotientenminimierung liefert einen Steigerungsfaktor von $\frac{14}{10}$ für τ , wodurch dieser über den Wert von $\tau = 1$ hinauskäme. Daher kann nur mehr das $\frac{4}{17}$ fache der Spalte " λ_4 " von der rechten Seite abgezogen werden, um $\tau = 1$ zu erreichen. Dies liefert neben $\lambda_4 = \frac{4}{17}$

$$\begin{aligned}
 x_1 &= \frac{5}{9} + \frac{4}{17} \frac{8}{9} = \frac{13}{17} \\
 x_2 &= \frac{10}{9} - \frac{4}{17} \frac{2}{9} = \frac{18}{17} \\
 x_3 &= \frac{14}{9} - \frac{4}{17} \frac{10}{9} = \frac{22}{17}
 \end{aligned}$$

und damit das erwartete Ergebnis.

Bei den beiden vereinfachten Formen ergeben sich die folgende

Beobachtungen:

- Beide Varianten benutzen nach dem ersten ATS eine Hilfszielfunktion, die durch eine Zeile im Tableau gegeben ist.
- Die $2(n + m)$ in Komplementarität zu bringenden Variablen sind ab dem zweiten Tableau so verteilt, dass in der Basis $(n + m - 1)$ und in der Nichtbasis $(n + m + 1)$ Variablen gelistet sind, dass diese Variablen außer der in der Basis befindlichen künstlichen Variablen die einzigen Variablen des Tableaus sind und dass die Komplementarität gegeben ist, weil in der Nichtbasis genau ein Paar der zusammengehörigen Variablen zusammen auftritt. Alle anderen Paare von zusammengehörigen Variablen sind eins zu eins auf Basis und Nichtbasis verteilt.
- Nach der eingeschränkten Austauschregel muss in jedem Schritt eine der Variablen aus dem in der Nichtbasis vorhandenem Variablen-Paar

die Nichtbasis verlassen, weil die entsprechende Spalte Pivotspalte wird. Dieses Paar in der Nichtbasis ist als solches aber erst durch den letzten ATS entstanden. Daher kann diejenige der beiden Variablen, die in letzten Schritt in die Nichtbasis gelangte, nicht sofort wieder die Pivotspalte festlegen, da deren Eintrag in der Zielfunktionszeile genau das falsche Vorzeichen hat. Es legt also genau diejenige der beiden Nichtbasisvariablen die neue Pivotspalte fest, die nicht im letzten ATS in die Nichtbasis gekommen war. Die Pivotspalte ist also in jedem Fall eindeutig bestimmt, sogar unabhängig von der Zielfunktion, wenn das Verfahren unter den oben festgestellten Voraussetzungen mit der eingeschränkten Pivotregel auf jeden Fall zu einem positiven Ende führt.

4.2.4 Übungsaufgaben

Aufgabe 1

Programmieren Sie alle vier Varianten des Verfahrens von Wolfe und testen Sie diese gegeneinander. Berücksichtigen Sie dabei die zuletzt gemachten Beobachtungen.

4.3 Das Verfahren von Lemke

Wir betrachten weiterhin die quadratische Optimierungsaufgabe

$$(QPS) \quad \begin{cases} \text{minimiere} & f(x) = \frac{1}{2}x^T Qx + p^T x \\ \text{s.d.} & Ax \leq a, \quad x \geq 0 \end{cases}$$

wobei wir $A \in \mathbf{R}^{m \times n}$, $Q \in \mathbf{R}^{n \times n}$ und $a \in \mathbf{R}^n$ voraussetzen.

Die KKT-Bedingungen zu (QPS) lauten

$$(QPSKKT) \quad \begin{cases} Qx + p + A^T \lambda - \mu = 0, \quad \lambda, \mu \geq 0 \\ Ax + y = a, \quad x, y \geq 0 \\ 0 = \lambda^T y + \mu^T x \end{cases}$$

In Matrixform schreibt sich dies zu

$$\begin{pmatrix} \mu \\ y \end{pmatrix} - \begin{pmatrix} Q & A^T \\ -A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} p \\ a \end{pmatrix}$$

$$\begin{pmatrix} x \\ \lambda \end{pmatrix}^T \begin{pmatrix} \mu \\ y \end{pmatrix} = 0, \quad \begin{pmatrix} x \\ \lambda \end{pmatrix} \geq 0, \quad \begin{pmatrix} \mu \\ y \end{pmatrix} \geq 0$$

Diese Formulierung ist vom Format eines **Linear Complementary Problem (LCP)**

$$\begin{aligned} w - Mz &= q \\ w \geq 0, \quad z &\geq 0 \\ w^T z &= 0 \end{aligned} \tag{4.3}$$

Dieses Gleichungs-/Ungleichungssystem wollen wir erneut mit einem dem Simplexverfahren der linearen Programmierung ähnlichen Verfahren lösen. Dabei halten wir uns eng an die Ideen und Beobachtungen zur *vereinfachten kurzen Form* des Verfahrens von Wolfe und untersuchen weiterhin Folgerungen aus der besonderen Struktur der KKT Bedingungen für die Lösbarkeit der Aufgabe (QPS).

4.3.1 Die Idee von Lemke

Vorgegeben sei ein LCP (4.3) mit $p \times p$ Matrix M und p -Vektor q . **Gesucht** sind Vektoren w und z , die dieses System erfüllen.

Es ist von vorneherein nicht klar, ob das System (4.3) eine Lösung (w, z) besitzt. Zunächst machen wir die folgende Beobachtung:

Ist $q \geq 0$, so besitzt (4.3) eine Lösung und diese kann unmittelbar angegeben werden. Setzt man nämlich $w := q$, $z = 0$, so erfüllen w und z alle geforderten Bedingungen.

Wirklich interessant ist also nur der Fall $q \not\geq 0$, wir können diese Forderung für das Folgende **voraussetzen**.

Wir erweitern die gegebene Aufgabenstellung zu

$$\begin{aligned} w - Mz - z_0e &= q \\ w \geq 0, z \geq 0, z_0 &\geq 0 \\ w^T z &= 0 \end{aligned} \tag{4.4}$$

wobei wie üblich $e^T = (1, \dots, 1) \in \mathbf{R}^p$ gesetzt ist.

Zu diesem erweiterten LCP kann eine zulässige Lösung unmittelbar angegeben werden: wegen $q \not\geq 0$ ist

$$z_0 := \max \{-q_i : 1 \leq i \leq p\} \tag{4.5}$$

positiv und wir erhalten mit $w := q + z_0e$, $z := 0$ offensichtlich eine zulässige Lösung von (4.4)!

Zu beachten ist, dass die ersten beiden Zeilen von (4.4)

$$w - Mz - z_0e = q, w \geq 0, z \geq 0, z_0 \geq 0 \tag{4.6}$$

ein lineares Gleichungs-/Ungleichungssystem bilden. Die angegebene Lösung kann wie folgt leicht in eine *zulässige* Basislösung zu diesem System überführt werden, bei der z und die Komponente aus w mit dem Index s , bei dem das Maximum in (4.5) angenommen wird, die Nichtbasis- und die übrigen Variablen die Basisvariablen bilden:

Zunächst einmal ist aus der Struktur des zulässigen Bereichs klar, dass eine (unzulässige) Basislösung zu (4.6) existiert, bei der genau die Variablen des Vektors w in der Basis sind. Im zugehörigen Tableau führe dann einen ATS durch, bei dem z_0 in die Basis und w_s in die Nichtbasis gelangt. Die zugehörige Basislösung ist genau die oben angegebene. Dieser ATS ist möglich, da die Spalte zu z_0 nur Einträge (-1) besitzt. Die Zulässigkeit der Basislösung ergibt sich unmittelbar aus der Definition von z_0 .

Zu beachten ist, dass die Komplementaritätsbedingungen $w^T z = 0$ durch diese zulässige Basislösung erfüllt sind. Leider nimmt aber z_0 in dieser Lösung einen in der Regel positiven Wert an, es liegt also keine Lösung der ursprünglichen Systems (4.3) vor.

Im Folgenden wollen wir durch eine Reihe von weiteren Austauschschritten des Simplexverfahrens unter Aufrechterhaltung der Zulässigkeit zu (4.4) die Variable z_0 auf null bringen, wobei darauf zu achten ist, dass niemals einander entsprechende Variablen aus w und z sich beide gleichzeitig in der Basis befinden. Gelingt dies, so ist die Basislösung des letzten Tableaus eine zulässige Lösung von (4.3).

Die Vorgehensweise ist dabei die folgende: die ATS werden solange durchgeführt, bis die Variable z_0 , die im ersten ATS in die Basis gelangte, den Wert null erlangt, z.B. indem sie wieder zurück in die Nichtbasis kommt und damit in der Basislösung den Wert null erhält.

Solange z_0 noch in der Basis verweilt, gibt es immer genau einen Variablenindex $1 \leq s \leq p$, so dass w_s und z_s gemeinsam zur Nichtbasis gehören. Eine dieser beiden Variablen verlässt im nächsten ATS die Nichtbasis wieder und zwar genau diejenige, die nicht im ATS unmittelbar davor in die Nichtbasis gelangt ist. Ein neuer Variablenindex s' gelangt stattdessen in die Nichtbasis, sodass nun s' genau der Variablenindex ist, zu dem die beiden Variablen $w_{s'}$ und $z_{s'}$ in der Nichtbasis sind (sofern der letzte ATS nicht z_0 in die Nichtbasis beförderte).

Wir formulieren diese Idee nun als Verfahren von Lemke. **Zu beachten** ist, dass dieses Verfahren drei Ausgänge hat. Ferner besitzt das Verfahren zu den Indizes $i = 1, \dots, n$ jeweils zwei Variablen, z_i und w_i und zum Index 0 eine Variable, z_0 .

Das Verfahren von Lemke

1. **Start** Vorgelegt sei das kanonische Simplextableau zum zulässigen Bereich (4.6) ohne Zielfunktionszeile. Ist das Tableau zulässig, **stopp0**, ($w := q$, $z := 0$ ist der gesuchte KKT Punkt von (4.3))

Sonst führe einen ATS durch zur Pivotspalte k mit der Variablen z_0 und zur Pivotzeile i , die durch

$$i := \arg \max \{-q_j : 1 \leq j \leq p\}$$

Setze s gleich dem Variablenindex, für den w_s in die Nichtbasis gelangt.

2. **Pivotspalte** Ist $s = 0$, **stopp1**, (die Basislösung zum letzten Tableau ist Lösung zu (4.3))

Sonst wähle als Pivotspalte k die Spalte zu der Variablen mit dem Index s , die nicht im letzten Schritt in die Nichtbasis gelangte.

3. **Pivotzeile** Wähle die Pivotzeile i , dass beim ATS das Tableau zulässig bleibt. Geht dies nicht, **stopp2**, (**Ray-termination**)
4. **Austauschschritt** Führe den ATS durch und setze s gleich dem Variablenindex der Variablen, die die Basis verlässt. Gehe zu 2.

Beispiel 4.8 *Betrachten wir erneut die Aufgabe*

$$\begin{array}{ll}
 \text{minimiere} & z(x_1, x_2) = -x_1 + \frac{1}{2}x_1^2 - 2x_2 + \frac{1}{2}x_2^2 \\
 \text{s.d.} & 2x_1 + 3x_2 \leq 6 \\
 & x_1 + 4x_2 \leq 5 \\
 & x_1, x_2 \geq 0
 \end{array}$$

Anfangstableau ist nun

	x_1	x_2	λ_3	λ_4	z_0	
λ_1	-1		-2	-1	-1	-1
λ_2		-1	-3	-4	(-1)	-2
x_3	2	3			-1	6
x_4	1	4			-1	5

Ein ATS mit dem gekennzeichneten Pivotelement führt zu

	x_1	x_2	λ_3	λ_4	λ_2	
λ_1	-1	(1)	1	3	-1	1
z_0		1	3	4	-1	2
x_3	2	4	3	4	-1	8
x_4	1	5	3	4	-1	7

Das Tableau ist damit zulässig und komplementär. Es ist $s = 2$, die nächste Pivotspalte ist die Spalte " x_2 ". Quotientenminimierung führt auf das mar-

kierte Pivot. Der ATS liefert

	x_1	λ_1	λ_3	λ_4	λ_2	
x_2	-1	1	1	3	-1	1
z_0	1	-1	2	1	0	1
x_3	6	-4	-1	-8	3	4
x_4	(6)	-5	-2	-11	4	2

Es ist $s = 1$ und es kommt x_1 in die Basis, x_4 verlässt die Basis ($s = 4$)

	x_4	λ_1	λ_3	λ_4	λ_2	
x_2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{7}{6}$	$-\frac{1}{3}$	$\frac{4}{3}$
z_0	$-\frac{1}{6}$	$-\frac{1}{6}$	$\frac{7}{3}$	$\left(\frac{17}{6}\right)$	$-\frac{2}{4}$	$\frac{2}{3}$
x_3	-1	1	1	3	-1	2
x_1	$\frac{1}{6}$	$-\frac{5}{6}$	$-\frac{1}{3}$	$-\frac{11}{6}$	$\frac{2}{3}$	$\frac{1}{3}$

Nun kommt noch λ_4 in die Basis und die Hilfsvariable z_0 kommt in die Nichtbasis und ist damit auf null gebracht.

	x_4	λ_1	λ_3	λ_2	
x_2	$\frac{4}{17}$	$\frac{4}{17}$	$-\frac{5}{17}$	$-\frac{1}{17}$	$\frac{18}{17}$
λ_4	$-\frac{1}{17}$	$-\frac{1}{17}$	$\frac{14}{17}$	$-\frac{4}{17}$	$\frac{4}{17}$
x_3	$-\frac{14}{17}$	$\frac{20}{17}$	$-\frac{25}{17}$	$-\frac{5}{17}$	$\frac{22}{17}$
x_1	$\frac{1}{17}$	$-\frac{16}{17}$	$\frac{20}{17}$	$\frac{4}{17}$	$\frac{13}{17}$

Das Ergebnis ist das gleiche wie zuvor.

4.3.2 Konvergenzanalyse

Das Verfahren von Lemke gleicht dem Simplexverfahren darin, da es von Ecke zu Ecke des zulässigen Bereichs (4.6) wandert. Das dies überhaupt sinnvoll ist, liegt an Bemerkung ??, die zusichert, dass die gesuchte Optimallösung eine Ecke des zulässigen Bereichs sein muss.

Das Verfahren verläuft daher automatisch endlich, wenn gesichert ist, dass keine der angelaufenen Basislösungen mehrfach angelaufen wird. Dies kann natürlich nicht garantiert werden, wenn Entartungen auftreten. Der folgende Satz sichert aber die Endlichkeit zu, wenn entartete Basislösungen ausgeschlossen werden.

Satz 4.9 *Es werde unterstellt, dass alle von Verfahren von Lemke angelaufenen Ecken des Bereichs (4.6) nicht entartet sind. Dann sind alle angelaufenen Ecken verschieden und das Verfahren ist endlich.*

Beweis: [BSS], Lemma 11.1.5

□

Verläuft das Verfahren endlich und endet nicht schon bei **stopp0**, so endet es entweder mit **stopp1** oder mit **stopp2**. Im ersten Fall ist das Ergebnis eine Lösung von (4.3), das Verfahren hat seine Aufgabe erfüllt. Endet es bei **stopp2**, so scheitert das Verfahren bei der Suche nach einer Lösung von (4.3) und es ist an dieser Stelle unklar, ob dies an der Unfähigkeit des Verfahrens liegt, eine Lösung von (4.3) zu ermitteln oder ob es unter Umständen gar keine solche gibt.

Um diese Unsicherheit zu beseitigen, müssen wir mehr Informationen über die Matrix M haben.

Definition 4.10 *Es sei M eine $p \times p$ Matrix. Dann heißt M **copositiv**, wenn $z^T M z \geq 0$ gilt für alle $z \geq 0$. Sie heißt **copositiv-plus**, wenn sie copositiv ist und zusätzlich die Bedingung*

$$z \geq 0, z^T M z = 0 \implies (M + M^T) z = 0 \quad \text{für alle } z \in \mathbf{R}^p$$

erfüllt ist.

Beachte: Offensichtlich ist M copositiv, wenn M nur nichtnegative Einträge enthält und M ist copositiv-plus, wenn zusätzlich die Diagonaleinträge alle positiv sind.

Nun können wir den Konvergenzsatz verschärfen und erhalten die Sicherheit, einen KKT Punkt zu finden, wenn ein solcher existiert.

Satz 4.11 *Die Matrix M von (4.3) sei copositiv-plus. Es werde unterstellt, dass alle von Verfahren von Lemke angelaufenen Ecken des Bereichs (4.6) nicht entartet sind. Dann endet das Verfahren nach endlich vielen Schritten.*

Besitzt (4.3) eine zulässige Lösung, so endet das Verfahren mit einer Lösung von (4.3). Andernfalls endet das Verfahren mit der Ray-termination.

Beweis: [BSS], Theorem 11.1.8

□

4.3.3 Spezialfall Quadratische Optimierung

Wir wollen klären, inwieweit man das Verfahren von Lemke zum sicheren Lösen einer quadratischen Optimierungsaufgabe einsetzen kann.

Die Aufgabe (QPS)

Betrachten wir die quadratische Optimierungsaufgabe (QPS), deren KKT System (QPSKKT) sich als LCP schreiben lässt. Wendet man das Verfahren von Lemke auf (QPSKKT) an und endet dieses nach endlich vielen Schritten mit einem KKT Punkt, so hat man eine suboptimale Lösung von (QPS) bestimmt.

Will man allerdings sicher sein, dass das Verfahren erfolgreich ist, so sollten die Voraussetzungen von Satz 4.11 zur Verfügung stehen. Klären wir also ab, unter welchen Voraussetzungen die zu

$$(QPSKKT) \quad \begin{cases} \begin{pmatrix} \mu \\ y \end{pmatrix} - \begin{pmatrix} Q & A^T \\ -A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} p \\ a \end{pmatrix} \\ \begin{pmatrix} x \\ \lambda \end{pmatrix}^T \begin{pmatrix} \mu \\ y \end{pmatrix} = 0, \begin{pmatrix} x \\ \lambda \end{pmatrix} \geq 0, \begin{pmatrix} \mu \\ y \end{pmatrix} \geq 0 \end{cases}$$

gehörige Matrix

$$M = \begin{pmatrix} Q & A^T \\ -A & 0 \end{pmatrix}$$

copositiv-plus ist.

Satz 4.12 *Gilt $y^T Q y \geq 0$ für alle $y \geq 0$, so ist M copositiv. Gilt darüber hinaus, dass die Bedingung*

$$y \geq 0, y^T Q y = 0 \implies Q y = 0 \quad \text{für alle } y \in \mathbf{R}^n \quad (4.7)$$

erfüllt ist, so ist M copositiv-plus.

Insbesondere ist M copositiv-plus, wenn Q positiv semidefinit ist.

Beweis: Zunächst gelte $y^T Q y \geq 0$ für alle $y \geq 0$. Wir zeigen, dass M dann copositiv ist. Sei $z^T := (y^T, x^T) \geq 0$ vorgegeben. Dann gilt

$$z^T M z = (y^T, x^T) \begin{pmatrix} Q & A^T \\ -A & 0 \end{pmatrix} \begin{pmatrix} y \\ x \end{pmatrix} = y^T Q y \geq 0 \quad \text{für } z \geq 0 \quad (4.8)$$

also ist M copositiv.

Gelte nun zusätzlich die Bedingung (4.7). Dann ist zu zeigen, dass M sogar copositiv-plus ist. Unterstellen wir also $z \geq 0$ und $z^T M z = 0$. Dann gilt

$$(M + M^T) z = \begin{pmatrix} 2Qy \\ 0 \end{pmatrix}$$

Gemäß (4.8) gilt $0 = z^T M z = y^T Q y$. Aus (4.7) folgt also $Qy = 0$. Damit gilt $(M + M^T) z = 0$ und M ist copositiv-plus.

Sei nun Q als positiv semidefinit vorausgesetzt. Dann gilt $y^T Q y \geq 0$ sogar unabhängig von $y \geq 0$, also ist M copositiv. Für die Behauptung bleibt noch nachzuweisen, dass Q die Eigenschaft (4.7) hat. Sei daher $y \geq 0$ und $y^T Q y = 0$ vorausgesetzt. Setzen wir $d := Qy$. Dann gilt wegen der positiven Semidefinitheit der Matrix Q für beliebiges $\lambda > 0$

$$0 \leq (y^T - \lambda d) Q (y - \lambda d) = y^T Q y + \lambda^2 d^T Q d - 2\lambda \|d\|^2$$

Da $y^T Q y = 0$ gilt, kann diese Ungleichung durch λ geteilt werden. Lässt man dann λ gegen null gehen, so zeigt sich $0 = d = Qy$.

□

Das Verfahren von Lemke berechnet also für die Aufgabe (QPS) bei konvexer Zielfunktion einen KKT Punkt genau dann, wenn ein solcher existiert. Dies ist nach unserer Kenntnis z. B. der Fall, wenn der zulässige Bereich von (QPS) beschränkt ist oder wenn f streng konvex ist oder wenn Q positiv semidefinit ist und $p = 0$ gilt. Endet das Verfahren dagegen mit der Ray-Termination, so existiert kein KKT Punkt zur Aufgabe.

Zu beachten ist, dass das Verfahren von Lemke bei Anwendung auf die quadratische Optimierungsaufgabe (QPS) mit konvexer Zielfunktion sehr ähnlich verläuft wie die vereinfachte kurze Form von Wolfe. Der einzige Unterschied ist die konkrete Wahl der Spalte bei der Einführung der künstlichen Variablen z_0 . Wählt man diese bei Wolfe genauso, verlaufen die Verfahren identisch. Bei Lemke gewinnt man zusätzlich die Sicherheit, dass kein KKT Punkt existiert, wenn das Verfahren "scheitert".

4.3.4 Übungsaufgaben

Aufgabe 1

Erstellen Sie ein Programm zum Verfahren von Lemke für die Aufgabe (*QPS*) und testen Sie es an zufallserzeugten Beispielen mit konvexer Zielfunktion. Beurteilen Sie abschließend sämtliche quadratischen KKT Verfahren experimentell.

Teil II

Allgemeinere Konzepte

Kapitel 5

Unrestringierte Optimierung

Im Folgenden wollen wir nun *allgemeinere Verfahren* (als Simplexverfahren) der linear restringierten Optimierung konstruieren und dabei grundlegende neuere Konstruktionsprinzipien der nichtlinearen Optimierung besprechen. Startpunkt sollen *Verfahren der unrestringierten Optimierung* bilden, deren Konstruktion wegen fehlender Restriktionen sicherlich am einfachsten ist. Danach versuchen wir, die erarbeiteten Prinzipien auf *linear gleichungsrestringierte Aufgaben* und auf den allgemeinen Fall der *linear restringierten Optimierungsaufgaben* zu übertragen.

Eine **grundsätzliche Vorgehensweise** zur Lösung von (P) , die auch schon bei den besprochenen Simplexverfahren Anwendung fand, ist die folgende:

Man startet bei einer zulässigen Lösung, in diesem Fall also bei einem Punkt x^0 aus dem zulässigen Bereich M , sucht sich eine **zulässige Abstiegsrichtung** d , also eine Richtung, in der die Zielfunktion, ausgehend von x^0 ein Stück weit abfällt und versucht dann, entlang dieser Richtung mit einer geeigneten **Schrittweite** α so weit fortzuschreiten, dass ein neuer zulässiger Punkt $\bar{x}^0 = x^0 + \alpha d$ gefunden wird, in dem f einen erheblich niedrigeren Funktionswert hat. Dann wählt man eine neue Iterierte $x^1 \in M$, für die $f(x^1) \leq f(\bar{x}^0)$ gilt, z. B. $x^1 = \bar{x}^0$. Dies setzt man iteriert fort, solange bis das Verfahren **abgebrochen** werden kann, weil man mit dem erreichten Ergebnis zufrieden ist.

Allgemeines zulässiges Abstiegsverfahrens

- (0) **Start:** Wähle einen Punkt $x^0 \in M$. Setze $k = 0$.
- (1) **Abbruch:** Genügt x^k einem geeigneten Abbruchkriterium, **Stopp**.
- (2) **Suchrichtung:** Wähle in x^k eine zulässige Abstiegsrichtung d^k .
- (3) **Schrittweite:** Wähle eine Schrittweite $\alpha_k > 0$ so, dass mit $\bar{x}^k := x^k + \alpha_k d^k$ gilt:
$$\bar{x}^k \in M \quad \text{und} \quad f(\bar{x}^k) \text{ ist erheblich kleiner als } f(x^k).$$
- (4) **Iteration:** Wähle $x^{k+1} \in M$ mit $f(x^{k+1}) \leq f(\bar{x}^k)$, setze $k := k + 1$ und gehe zu (1).

Zu diesem Konzept sind natürlich vorab einige **Bemerkungen** zu machen:

1. Das Abbruchkriterium unterbricht *ein potentiell unendlich verlaufendes Verfahren*. Man kann also von der letzten Iterierten nur erwarten, dass es sich um eine numerische Lösung, d.h. um ein annäherndes Ergebnis handelt.
2. Die Geschwindigkeit, mit der sich das Verfahren seinem Ziel annähert, ist unter Anderem abhängig von der Wahl *geeigneter guter Suchrichtungen und Schrittweiten*. In diesem Zusammenhang ist zu präzisieren, was unter "erheblich kleiner" verstanden wird.
3. Es ist nicht davon auszugehen, dass das Verfahren gegen einen globalen Minimierer konvergiert. Wenn es denn konvergiert, kann es gegen einen lokalen Minimierer konvergieren. Wenn es nicht konvergiert, so kann trotzdem ein Häufungspunkt der Iteriertenfolge eine gesuchte (lokale) Lösung sein und es ist nützlich, einen solchen zu kennen.

In den folgenden Abschnitten und Kapiteln wird es darauf ankommen, zu dem genannten Prototyp eines Abstiegsverfahrens durch Spezialisierung der Wahlmöglichkeiten, z. B. der Abstiegsrichtung und der Schrittweite, Konvergenzsätze zu beweisen.

5.1 Sequentielle Approximation

Wir kennen bereits Verfahren der unrestringierten Optimierung. Für quadratische Polynome ist speziell das Verfahren von Beale zu nennen, für allgemeine nichtlineare Funktionen Varianten des Verfahrens von Zangwill: zum einen das Verfahren von Wolfe, zum anderen die Spezialisierung des Verfahrens von Zangwill, die grundsätzlich in Richtung der Koordinatenachsen sucht. Diese Verfahren benutzen spezielle Suchrichtungen und Schrittweiten, die eine eindimensionale Minimierung verlangen. Über den numerischen Nachteil dieser Schrittweiten haben wir bereits gesprochen.

Es ist eine grundsätzliche Idee zur Lösung von (P) und zur Ausgestaltung des zulässigen Abstiegsverfahrens, ersatzweise mithilfe einer zulässigen Startlösung x^0 eine geeignete (lineare oder quadratische) Ersatzaufgabe als *Approximation der gegebenen Aufgabe* zu lösen, um zu dieser (näherungsweise) eine Optimallösung \bar{x}^1 zu bestimmen. Entlang der Suchrichtung $(\bar{x}^1 - x^0)$ wird mit geeigneter Schrittweite eine möglichst gute neue zulässige Lösung x^1 gesucht. Mit Hilfe von x^1 wird dann eine verbesserte lineare oder quadratische Approximation der Aufgabenstellung und eine neue gute zulässige Lösung x^2 ermittelt, die wiederum zu einer erneut besseren Approximation führt. Eine iterative Fortführung dieser Vorgehensweise soll eine Folge von Lösungen x^0, x^1, x^2, \dots ergeben, die im Grenzprozess zur Lösung der gegebenen Aufgabe führt.

Dies liefert (hoffentlich) nicht nur gute Suchrichtungen, es unterstreicht auch die *Bedeutung der Linearen und Quadratischen Optimierung*.

5.1.1 Sequentielle lineare Approximation

Betrachtet werde die Aufgabe (P) $\min \{f(x) : x \in X\}$ mit gegebener *stetig differenzierbaren* Funktion $f : \mathbf{R}^n \rightarrow \mathbf{R}$ und abgeschlossener konvexer Teilmenge $X \subset \mathbf{R}^n$.

Zulässige Abstiegsrichtungen in einem Punkte x ergeben sich, wie wir wissen, z.B. als zulässige Lösungen der Aufgabe

$$(ZA) \quad \begin{cases} \text{minimiere} & \nabla f(x)^T d \\ \text{s.d.} & x + d \in M \end{cases} .$$

mit negativem Funktionswert. Wir fassen $\nabla f(x)^T d$ als *lineare Approximation* von f an der Stelle x auf, bei der der konstante Term weggelassen wurde. (ZA) sei also eine *approximierende (lineare) Ersatzaufgabe*, von der oben gesprochen wurde. Dabei kann die Aufgabe (ZA) durch die Forderung $\|d\| \leq 1$ zu (ZA1) ergänzt werden, mit irgendeiner Norm $\|\cdot\|$, damit eine Optimallösung existiert.

Wir wollen für unsere erstes Verfahren aber speziell die *euklidische Norm* auswählen. Im Fall $M = \mathbf{R}^n$ ist aus geometrischen Gründen unmittelbar klar, dass (ZA1) eine eindeutige Lösung hat, nämlich

$$\bar{d} = -\frac{1}{\|\nabla f(x)\|} \nabla f(x)$$

Skizze:

Andererseits ergibt sich das Ergebnis auch direkt aus der Cauchy-Schwarzschen Ungleichung

$$\left| \nabla f(x)^T d \right| \leq \|\nabla f(x)\| \|d\| \leq \|\nabla f(x)\|$$

und

$$\nabla f(x)^T \bar{d} = -\frac{\nabla f(x)^T \nabla f(x)}{\|\nabla f(x)\|} = -\|\nabla f(x)\|$$

Wir wählen diese Suchrichtungen für das zu konstruierende Verfahren, der Einfachheit halber setzen wir $d^k := -\nabla f(x^k)$.

Von den zugehörigen Schrittweiten erwarten wir, dass sie die die Zielfunktion "erheblich absenken". Dieser Terminus ist undefiniert. Die stärkste Absenkung erhalten wir natürlich, wenn wir die Schrittweite α_k wie bisher berechnen:

Minimierungsregel Vorgegeben seien ein Iterationspunkt x^k und eine zulässige Abstiegsrichtung d^k in x^k . Wähle dann als Schrittweite

$$\alpha_k := \arg \min \{ f(x^k + \alpha d^k) : \alpha \geq 0, x^k + \alpha d^k \in M \}$$

Diese Schrittweite ist aber oft nur aufwendig mit ungesichertem Ergebnis zu berechnen. Daher werden wir uns mit einer etwas schwächeren Forderung zufrieden geben. Schrittweiten wählen wir nun gemäß der

Armijo-Regel Vorgewählt seien Zahlen $\sigma, \beta \in (0, 1)$. Für einen Iterationspunkt x^k und eine zulässige Abstiegsrichtung d^k in x^k mit $x^k + d^k \in M$ wähle als Schrittweite die Zahl $\alpha_k := \max \{\beta^l : l = 0, 1, 2, 3, \dots\}$, so dass gilt

$$f(x^k + \alpha_k d^k) \leq f(x^k) + \sigma \alpha_k \nabla f(x^k)^T d^k$$

Skizze:

Es gilt dann zunächst ein Satz, der die Wohldefiniertheit der Konstruktion zeigt:

Satz 5.1 Seien $f : \mathbf{R}^n \rightarrow \mathbf{R}$ stetig differenzierbar und $\sigma, \beta \in (0, 1)$ vorgegeben und $x, d \in \mathbf{R}^n$ mit $\nabla f(x)^T d < 0$. Dann existiert ein endliches $l \in \mathbf{N}$ mit

$$f(x + \beta^l d) \leq f(x) + \sigma \beta^l \nabla f(x)^T d$$

Beweis: Angenommen, für alle $l \in \mathbf{N}$ gelte $f(x + \beta^l d) > f(x) + \sigma \beta^l \nabla f(x)^T d$. Dann gilt auch

$$\frac{f(x + \beta^l d) - f(x)}{\beta^l} > \sigma \nabla f(x)^T d$$

Wegen der vorausgesetzten stetigen Differenzierbarkeit von f gilt für $l \rightarrow \infty$

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

und da $\sigma \in (0, 1)$ gewählt wurde, folgt im Widerspruch zur Voraussetzung $\nabla f(x)^T d \geq 0$.

□

Beispiel 5.2 Wir bestimmen eine Schrittweite nach der Armijo-Regel im Punkte $x_1 = -1$ für die Funktion $f(x) = x^2$ mit der Ableitung $f'(x) = 2x$. Als Richtung wählen wir $d = -f'(-1) = 2$, ferner wählen wir $\sigma = 0.2$ und $\beta = 0.9$.

Es gilt zunächst für $\alpha = \beta^0 = 1$:

$$1 = (-1 + 1 \cdot 2)^2 > (-1)^2 + 0.2 \cdot 1 \cdot 2(-1) \cdot 2 = 0.2$$

Mit $\alpha = \beta^1 = 0.9$ ergibt sich

$$0.64 = (-1 + 0.9 \cdot 2)^2 > (-1)^2 + 0.2 \cdot 0.9 \cdot 2(-1) \cdot 2 = 0.28$$

mit $\alpha = \beta^2 = 0.81$

$$0.3844 = (-1 + 0.81 \cdot 2)^2 > (-1)^2 + 0.2 \cdot 0.81 \cdot 2(-1) \cdot 2 = 0.352$$

und schließlich mit $\alpha = 0.9^3 = 0.729$

$$0.20976 = (-1 + 0.729 \cdot 2)^2 < (-1)^2 + 0.2 \cdot 0.729 \cdot 2(-1) \cdot 2 = 0.4168$$

Skizze:

Nun sind wir soweit, ein *Verfahren der unrestringierten Optimierung* ($M = \mathbf{R}^n$) zu formulieren. Es heißt wegen der Wahl der Suchrichtung **Verfahren des steilsten Abstiegs** oder

Gradientenverfahren

- (0) **Start:** Wähle einen Punkt $x^0 \in \mathbf{R}^n$ und $\varepsilon \geq 0$, $\sigma, \beta \in (0, 1)$. Setze $k = 0$.
- (1) **Abbruch:** Gilt $\|\nabla f(x^k)\| \leq \varepsilon$, **Stopp**.

- (2) **Suchrichtung:** Setze $d^k := -\nabla f(x^k)$
- (3) **Schrittweite:** Setze $\bar{x}^k := x^k + \alpha_k d^k$ mit der Eigenschaft: $\alpha_k := \max \{ \beta^l : l = 0, 1, 2, 3, \dots \}$, so dass

$$f(x^k + \alpha_k d^k) \leq f(x^k) + \sigma \alpha_k \nabla f(x^k)^T d^k$$

- (4) **Iteration:** Wähle $x^{k+1} \in \mathbf{R}^n$ mit $f(x^{k+1}) \leq f(\bar{x}^k)$, setze $k := k + 1$ und gehe zu (1).

Praktikable Beispielwerte für die Parameter des Verfahrens sind:

$$\beta = 0.5, \quad \sigma = 10^{-4}, \quad \varepsilon = 10^{-6}$$

Es sei bemerkt, dass man natürlich immer $x^{k+1} = \bar{x}^k$ wählen kann.

Beispiel 5.3 (vgl. [BSS], chapt. 8) Wir betrachten die Funktion $f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$. Wendet man das Gradientenverfahren auf diese Funktion an, startend bei $(x_1, x_2) = (0, 3)$, so nimmt das Verfahren ungefähr den folgenden Verlauf, um sich der eindeutigen Optimallösung $x^* = (2, 1)$ anzunähern. Wenn der Linesearch genau ausgeführt wird, lautet eine Abbruchlösung z.B. $x = (2.28, 1.15)$ mit $\|\nabla f(x)\| = 0.09$ nach 8 Iterationen:

Skizze:

■

Zum Beweis eines Konvergenzsatzes für das Gradientenverfahren benötigen wir zunächst ein technisches

Lemma 5.4 (vgl. [GK1], Lemma 8.2)

Seien $f : \mathbf{R}^n \rightarrow \mathbf{R}$ stetig differenzierbar, $x, d \in \mathbf{R}^n$, $\{x^k\}, \{d^k\} \subseteq \mathbf{R}^n$ mit $\{x^k\} \rightarrow x$, $\{d^k\} \rightarrow d$ sowie $\{\alpha_k\} \subseteq \mathbf{R}_{++}$ mit $\{\alpha_k\} \rightarrow 0$ konvergente Folgen. Dann ist

$$\lim_{k \rightarrow \infty} \frac{f(x^k + \alpha_k d^k) - f(x^k)}{\alpha_k} = \nabla f(x)^T d$$

Beweis: Über den Mittelwertsatz können wir für jedes k

$$f(x^k + \alpha_k d^k) - f(x^k) = \alpha_k \nabla f(\xi^k)^T d^k$$

mit geeigneten Vektoren ξ^k auf der Strecke zwischen x^k und $x^k + \alpha_k d^k$ schreiben. Dann konvergiert $\{\xi^k\}$ für $k \rightarrow \infty$ gegen x^k und wegen der stetigen Differenzierbarkeit von f auch $\{\nabla f(\xi^k)^T d^k\} \rightarrow \nabla f(x)^T d$. Daraus folgt unmittelbar die Behauptung. □

Nun können wir einen Konvergenzsatz zum Gradientenverfahren formulieren und beweisen.

Satz 5.5 (vgl. [GK1], Satz 8.3) Ist $f : \mathbf{R}^n \rightarrow \mathbf{R}$ stetig differenzierbar und $\varepsilon = 0$, so endet das Verfahren nach endlich vielen Schritten mit einem stationären Punkt von f oder aber jeder Häufungspunkt einer durch das Gradientenverfahren erzeugten unendlichen Iterationsfolge ist ein stationärer Punkt von f .

Beweis: Es ist nur der Fall zu betrachten, dass das Verfahren unendlich verläuft.

Es sei x^* ein Häufungspunkt der Iteriertenfolge $\{x^k\}$ und sei $\{x^k\}_K$ eine gegen x^* konvergente Teilfolge. Es werde angenommen, dass $\nabla f(x^*) \neq 0$ gilt.

Da die Folge $\{f(x^k)\}$ nach Konstruktion monoton abfällt, ist sie gegen $f(x^*)$ konvergent. Dies trifft ebenso zu für die Folge $f(x^0), f(\bar{x}^0), f(x^1), f(\bar{x}^1), \dots$

Daher konvergiert die Folge $f(x^k) - f(\bar{x}^k)$ gegen 0. Nach Konstruktion gilt somit

$$f(\bar{x}^k) - f(x^k) = \alpha_k \nabla f(x^k)^T d^k = -\alpha_k \|\nabla f(x^k)\|^2 \rightarrow 0$$

und wegen $\{\nabla f(x^k)\}_K \rightarrow \nabla f(x^*) \neq 0$ nach Annahme schließt man hieraus auf $\{\alpha_k\}_K \rightarrow 0$. Daher kann die Folge dieser α_k nicht durchgehend gleich $\beta^0 = 1$ sein, für hinreichend großes $k \in K$ muss daher

$$\frac{f(x^k + \beta^{l_k-1} d^k) - f(x^k)}{\beta^{l_k-1}} > \sigma \nabla f(x^k)^T d^k$$

gelten, wobei l^k der eindeutig bestimmte Exponent aus der Armijo-Regel ist. Es gilt dann natürlich auch $\{\beta^{l_k-1}\}_K \rightarrow 0$ und es folgt aus dem letzten Lemma beim Übergang $k \rightarrow \infty, k \in K$

$$\|\nabla f(x^*)^T d^*\| \geq \sigma \|\nabla f(x^*)^T d^*\|$$

wobei $d^k = -\nabla f(x^k) \rightarrow -\nabla f(x^*)^T =: d^*$ gelte. Daher gilt auch

$$-\|\nabla f(x^*)\|^2 \geq -\sigma \|\nabla f(x^*)\|^2$$

was wegen $\nabla f(x^*) \neq 0$ aber $\sigma \in (0, 1)$ widerspricht. □

Man beachte: Aus der Formulierung des Gradientenverfahrens ist unmittelbar klar, dass der Konvergenzsatz auch trägt, wenn anstelle der Armijo-Regel die schärfere Minimierungsregel Anwendung findet (in der Praxis auch, wenn diese nur numerisch erfüllt wird).

5.1.2 Sequentielle quadratische Approximation

Wir wollen nun eine *zweimal stetig differenzierbare* Funktion $f : \mathbf{R}^n \rightarrow \mathbf{R}$ unrestringiert über *sukzessive quadratische Approximation* minimieren.

Bekanntermaßen erfüllt ein Minimierer der Funktion f die KKT Bedingung

$$\nabla f(x) = 0. \tag{5.1}$$

Sei x^0 beliebig aus \mathbf{R}^n vorgewählt, z.B. als Näherungslösung von $\nabla f(x) = 0$. Wir ersetzen die gegebene Funktion durch ihre *quadratische Approximation* an der Stelle x^0

$$q(x) = f(x^0) + \nabla f(x^0)^T (x - x^0) + \frac{1}{2} (x - x^0)^T \nabla^2 f(x^0) (x - x^0),$$

bilden deren Gradienten, der mit der linearen Approximation von $F(x) := \nabla f(x)$ übereinstimmt,

$$\nabla q(x) = \nabla f(x^0) + \nabla^2 f(x^0) (x - x^0) = F(x) + F'(x) (x - x^0) =: T_F(x)$$

und betrachten zur Lösung von (5.1) ersatzweise die Gleichung $\nabla q(x) = T_F(x) = 0$

$$\begin{aligned} 0 &= F(x^0) + F'(x^0) (x - x^0) \\ \implies F'(x^0) x &= F'(x^0) x^0 - F(x^0) \end{aligned}$$

wobei $F'(x)$ die Jacobimatrix an der Stelle x angibt. Ist $F'(x^0)$ invertierbar, so folgt

$$x^1 = x^0 - F'(x^0)^{-1} F(x^0)$$

für die Lösung x^1 der Ersatzgleichung. Oft ist x^1 eine bessere Näherungslösung der Gleichung $F(x) = \nabla f(x) = 0$ als x^0 , daher wiederholt man den beschriebenen Vorgang mit x^1 anstelle von x^0 . Führt man dies iteriert durch, d.h. verwendet man die Rekursionsformel

$$x^{k+1} = x^k - F'(x^k)^{-1} F(x^k) \quad \text{für } k = 0, 1, 2, 3, \dots \quad (5.2)$$

so nennt man diese Vorgehensweise bekanntermaßen **Newtoniteration**.

Newtonverfahren

In der Praxis setzt man $\Delta x^k := x^{k+1} - x^k$ und löst zur Minimierung der Funktion f anstelle der Rekursionsgleichung (5.2) iteriert die Gleichung

$$F'(x^k) \Delta x^k = -F(x^k) \quad (5.3)$$

Der folgende Satz gilt für eine beliebige stetig differenzierbare Funktion $F : \mathbf{R}^n \rightarrow \mathbf{R}^n$, zu der eine Nullstelle bestimmt werden soll, nicht nur für $F = \nabla f$:

Satz 5.6 *Es sei x^* eine Nullstelle von F und die Jacobimatrix $F'(x^*)$ sei regulär. Dann gibt es eine Umgebung U von x^* so, dass für alle Startpunkte $x^0 \in U$ des Newtonverfahrens gilt:*

1. *das Newtonverfahren ist erfolgreich durchführbar, d.h. die Rekursionsformel ist beliebig oft mit Erfolg anwendbar und erzeugt eine gegen x^* konvergente Iterationsfolge.*
2. *Die Konvergenzrate ist superlinear.*
3. *Ist F' auch lokal Lipschitz-stetig, so ist die Konvergenzrate quadratisch.*

Beweis: [GK2], Satz 5.26

Startet man also das Newtonverfahren *in unmittelbarer Nähe* einer Nullstelle von F , so ermittelt es schnell und zuverlässig die Nullstelle von F , sofern die *Jacobimatrix von F an der Nullstelle invertierbar* ist. Beide Annahmen sind allerdings von vorne herein kaum überprüfbar, was die Anwendbarkeit des Verfahrens stark einschränkt.

Zu beachten ist ferner, dass das Verfahren, eingesetzt zur Minimierung einer zweimal differenzierbaren Funktion f , als Ergebnis auch ein lokales Maximum oder einen Sattelpunkt von f bestimmen kann. Außerdem gibt es ja durchaus Minimalstellen, bei denen die Funktion im Minimalpunkt keine invertierbare Hessematrix besitzt: z. B. $f(x, y) = x^2$ mit der Minimalstelle $(x, y) = (0, 0)$. Über das Verhalten des Newtonverfahrens bei solchen Funktionen sagt der Satz leider nichts.

Beispiel 5.7 (vgl. [BSS], example (8.6.4)) *Wir betrachten noch einmal die Funktion $f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$. Wendet man das Newtonverfahren auf diese Funktion an, startend bei $(x_1, x_2) = (0, 3)$, so nimmt das Verfahren ungefähr den folgenden Verlauf, um sich der eindeutigen Optimallösung $x^* = (2, 1)$ anzunähern. Es lautet eine Abbruchlösung z.B. $x = (1, 83, 0, 91)$ mit $\|\nabla f(x)\| = 0.04$ nach 7 Iterationen.*

Globalisierung

Für größere praktische Anwendbarkeit wird gesorgt, wenn man das Newtonverfahren und das Gradientenverfahren kombiniert. Der Vorteil der schnellen lokalen Konvergenz wird dabei dem global konvergenten Linearisierungsverfahren eingepflegt: das Hybridverfahren erbt vom Gradientenverfahren die globale und vom Newtonverfahren schnelle lokale Konvergenz.

Dabei wollen wir ausnutzen, dass das Gradientenverfahren die Freiheit lässt, nach der Bestimmung von \bar{x}^k einen beliebigen Schritt zu tun, an den nur die Forderung gestellt wird, die Zielfunktion nicht zu erhöhen. Dies kann so ausgestaltet werden, dass nach jedem Übergang von x^k zu \bar{x}^k endlich viele Schritte des Newtonverfahrens eingefügt werden, solange sie berechenbar sind und die Zielfunktion nicht ansteigt.

Diese grundsätzliche Vorgehensweise soll an folgendem Beispiel erläutert werden.

Beispiel 5.8 *Wir betrachten als Beispiel die Funktion $f(x) = \sin(x)$ und wollen ein Subminimum dieser Funktion mit dem Newtonverfahren bestimmen. Wir wollen immer dann einen Schritt des Gradientenverfahrens anwenden, wenn der Schritt des Newtonverfahrens nicht durchführbar ist oder die Absenkung von $|f'(x)|$ nicht mindestens 1% beträgt. Dies begrenzt die Anzahl der hintereinander durchgeführten Newtonschritte oder führt dazu, dass unendlich oft hintereinander Newtonschritte ausgeführt werden.*

Wir starten unser Gesamt-Verfahren mit $x^0 = 0$ und versuchen, einen Schritt des Newtonverfahrens. Die Iterationsformel hierzu lautet

$$x^{k+1} = x^k - f'(x^k) / f''(x^k)$$

Wegen $f'(x) = \cos(x)$ und $f''(x) = -\sin(x)$ ist $f''(0) = 0$ und dieser Wert ist nicht invertierbar. Daher ist ein Newtonschritt nicht ausführbar und wir führen stattdessen einen Schritt des klassischen Gradientenverfahrens durch. Es ergibt sich der Gradient von f an der Stelle $x = 0$ zu $f'(x_0) = \cos(0) = 1$. Daher ist $\bar{d} = -1$ und wir haben eine Line Search durchzuführen für f an der Stelle $x = 0$ in Richtung $\bar{d} = -1$. Nun, es ist $\cos(d) > 0$ für $d = x^0 + \lambda\bar{d} = -\lambda \in [0; 1]$, daher ist f strikt monoton abfallend in diese Richtung und wir können $x^1 = -1$ wählen mit $f(x) = \sin(-1) = -0.84147$. Also

setzen wir $x^1 := -1$ und versuchen nun wieder einen Newtonschritt. Es ist $\cos(-1) = 0.5403$, also ergibt sich probeweise

$$x^2 = x^1 - \frac{f'(x^1)}{f''(x^1)} = -1 - \frac{0.5403}{0.84147} = -1.6421$$

mit $|\cos(-1.6421)| = 0.071243 < 0.9 \cdot 0.5403 = 0.9 \cdot |\cos(-1)|$. Also ist dieser Newtonschritt erfolgreich. Daher führen wir einen weiteren Newtonschritt durch. Wegen $-\sin(-1.6421) = 0.99746$ gilt

$$x^3 = x^2 - \frac{f'(x^2)}{f''(x^2)} = -1.6421 - \frac{-0.071243}{0.99746} = -1.5707.$$

Nun ist $\cos(-1.5707) = 0.000096327$ und dieser Newtonschritt ist wieder erfolgreich. Wir brechen ab, da die angestrebte Lösung numerisch erreicht ist ($x^* = -\frac{\pi}{2} = -1.570796\dots$). ■

Die dargestellte Vorgehensweise hat Nachteile: Möglicherweise ist die willkürlich gewählte Parametereinstellung (1%) so, dass fast nur Gradientenschritte durchgeführt werden, was das Verfahren langsam macht. Man kann außerdem der Meinung sein, dass die *Newtonrichtung*, wenn sie denn eine *Abstiegsrichtung* ist, die die Zielfunktion genügend absenken kann, gegenüber der Gradientenrichtung im Sinne des Gesamtverfahrens die bessere Suchrichtung ist, weil sie als "Suchrichtung zweiter Ordnung" die gesuchte Optimallösung direkter ansteuert und so Zick-Zack-Laufen vermeiden hilft. In diesem Falle würde man sie gerne verwenden, auch wenn sie mit der Schrittweite $\alpha = 1$ nicht zu genügender Absenkung führt. Also behandelt man sie im Hybridverfahren wie die Gradientenrichtung, d.h. man verwendet auch im Zusammenhang mit der Newtonrichtung die Armijo-Regel. Dies macht auch deswegen Sinn, weil die Armijo-Regel ja zunächst die Schrittweite $\beta^0 = 1$ abstestet und diese verwendet, wenn die Zielfunktion mit dieser Schrittweite genügend abgesenkt wird.

Diese Idee verfeinert also die oben beschriebene Vorgehensweise und führt zum

Globalisiertes Newtonverfahren

(0) **Start:** Wähle einen Punkt $x^0 \in \mathbf{R}^n$ und $\rho > 0$, $\varepsilon \geq 0$, $\sigma, \beta \in (0, 1)$.
Setze $k = 0$.

(1) **Abbruch:** Gilt $\|\nabla f(x^k)\| \leq \varepsilon$, **Stopp**.

(2) **Suchrichtung:** Berechne eine Lösung d^k der Newtongleichung

$$\nabla^2 f(x^k) d = -\nabla f(x^k)$$

Ist dieses System nicht lösbar oder ist die Bedingung

$$\nabla f(x^k)^T d^k \leq -\rho \|d^k\|^2 \quad (5.4)$$

nicht erfüllt, so setze $d^k := -\nabla f(x^k)$

(3) **Schrittweite:** Setze $\bar{x}^k := x^k + \alpha_k d^k$ mit der Eigenschaft: $\alpha_k := \max\{\beta^l : l = 0, 1, 2, 3, \dots\}$, so dass

$$f(x^k + \alpha_k d^k) \leq f(x^k) + \sigma \alpha_k \nabla f(x^k)^T d^k$$

(4) **Iteration:** Setze $x^{k+1} := \bar{x}^k$, setze $k := k + 1$ und gehe zu (1).

Dabei soll die Zusatzbedingung (5.4) sicherstellen, dass die berechnete Newtonrichtung genügend absenkend ist. Motiviert ist die Bedingung durch folgende Überlegung:

Für die Gradientenrichtung $d^k := -\nabla f(x^k)$ gilt

$$\nabla f(x^k)^T d^k = -\|\nabla f(x^k)\|^2 = -\|d^k\|^2$$

Will man die Newtonrichtung an diesem Maß der Absenkung der Zielfunktion messen, so sollte man von ihr eine Absenkung ähnlicher Größenordnung verlangen, also etwa

$$\nabla f(x^k)^T d^k \leq -\rho \|d^k\|^2 \quad \text{mit vorgewähltem } \rho > 0$$

Nach dem Konvergenzsatz 5.5 ist die Konvergenz dieses angereicherten Gradientenverfahrens klar, sofern man als Anwender sicherstellt, dass jeweils

nur endlich oft hintereinander Newtonschritte ausgeführt werden. Dies kann vom Anwender natürlich einfach über eine vorgewählte Obergrenze für diese Schrittzahl sichergestellt werden.

Andererseits kann man es dem Verfahren ohne Probleme selbst überlassen, ob nur endlich oft hintereinander Newtonschritte verwendet werden oder nicht, wie der folgende Satz zeigt.

Satz 5.9 *Ist f zweimal stetig differenzierbar, so ist jeder Häufungspunkt der Iteriertenfolge des Globalisierten Newtonverfahrens ein stationärer Punkt von f .*

Beweis: (vgl. [GK1], Satz 9.5)

Wir haben nur noch den Fall zu betrachten, bei dem unendlich oft hintereinander die Newtonrichtung als Suchrichtung gewählt wird. Gehen wir also oBdA davon aus, dass im Verfahren nur noch die Newtonrichtung als Suchrichtung gewählt wird und unterstellen wir, dies führe nicht zu einem stationären Häufungspunkt der Iterationsfolge.

Nehmen wir also an, es gäbe eine gegen $x^* \in \mathbf{R}^n$ konvergente Teilfolge $(x^k)_{k \in K}$ mit $\nabla f(x^*) \neq 0$. Dann folgt aus der Newtongleichung für beliebigen Iterationspunkt x^k der Teilfolge

$$\|\nabla f(x^k)\| = \|\nabla^2 f(x^k) d^k\| \leq \|\nabla^2 f(x^k)\| \|d^k\|$$

Daraus können wir schließen

$$\|d^k\| \geq \frac{\|\nabla f(x^k)\|}{\|\nabla^2 f(x^k)\|} \tag{5.5}$$

denn es kann ja nicht $\|\nabla^2 f(x^k)\| = 0$ sein, da wir die Lösbarkeit des Newtonsystems bei jedem Schritt vorausgesetzt haben.

Damit können wir nun nachweisen, dass die $\|d^k\|$ uniform beschränkt und von null wegbeschränkt sind.

Denn einerseits folgt aus dem vorausgesetzten Erfülltsein der Regel (5.4)

$$\begin{aligned} \|d^k\|^2 &\leq \frac{1}{\rho} \left| \nabla f(x^k)^T d^k \right| \leq \frac{1}{\rho} \|\nabla f(x^k)\| \|d^k\| \\ \implies \|d^k\| &\leq \frac{1}{\rho} \|\nabla f(x^k)\| \end{aligned}$$

Da nach Konstruktion die Folge $\|\nabla f(x^k)\|_{k \in K}$ konvergiert, ist sie beschränkt und daher gibt es ein $c_1 > 0$ mit $\|d^k\| \leq c_1$ für alle $k \in K$.

Nehmen wir andererseits an, es gäbe eine Teilfolge $(\|d^k\|)_{k \in \bar{K}}$, die gegen null strebt. Dann würde aus der Abschätzung (5.5) folgen, dass auch die Folge $(\|\nabla f(x^k)\|)_{k \in \bar{K}}$ eine Nullfolge wäre, denn die Folge der $\|\nabla^2 f(x^k)\|$ ist als konvergente Folge natürlich wieder beschränkt. Die Folge $(\|\nabla f(x^k)\|)_{k \in \bar{K}}$ kann aber keine Nullfolge sein, da sie gegen $\|\nabla f(x^*)\| \neq 0$ konvergiert. Also gibt es auch ein $c_2 > 0$ so, dass $\|d^k\| \geq c_2$ für alle $k \in K$.

Nun ist nach Konstruktion die Folge $(f(x^k))$ monoton fallend und besitzt eine konvergente Teilfolge, also ist sie insgesamt konvergent und die Folge der $f(x^{k+1}) - f(x^k)$ bildet eine Nullfolge. Wegen der Armijo Regel muss damit auch für die Folge

$$\alpha_k \nabla f(x^k)^T d^k \rightarrow 0 \quad (5.6)$$

gelten.

Wir zeigen nun, dass darin die Folge der α_k von null wegbeschränkt bleibt. Nehmen wir an, es gäbe eine Teilfolge der α_k , die gegen null strebt, oBdA beschrieben wieder durch $k \in \bar{K}$. Es sei $\alpha_k = \beta^{l_k}$ mit dem nach der Armijo-Regel eindeutig bestimmten $l_k \in \mathbf{N}_0$. Da die $\alpha_k, k \in \bar{K}$, eine Nullfolge bilden, muss die Folge der l_k gegen unendlich streben. Deshalb muss für hinreichend große $k \in \bar{K}$ dann $l_k > 0$ gelten. Für diese k gilt

$$\frac{f(x^k + \beta^{l_k-1} d^k) - f(x^k)}{\beta^{l_k-1}} > \sigma \nabla f(x^k)^T d^k$$

was nach Lemma 5.4

$$\nabla f(x^*)^T d^* \geq \sigma \nabla f(x^*)^T d^*$$

bedeutet, wobei wegen der uniformen Beschränktheit der d^k oBdA davon ausgegangen werden kann, dass die d^k gegen ein $d^* \in \mathbf{R}^n$ konvergieren, notfalls durch Einschränkung auf eine weitere Teilfolge. Weil die d^k sogar von null wegbeschränkt sind, muss dabei sogar $d^* \neq 0$ gelten.

Nun war $\sigma \in (0, 1)$ vorausgesetzt, so dass wir aus der letzten Ungleichung sofort auf $\nabla f(x^*)^T d^* \geq 0$ schließen können. Andererseits liefert die Forderung (5.4) im Grenzprozess

$$\nabla f(x^*)^T d^* \leq -\rho \|d^*\|^2 < 0,$$

was aber der letzten Abschätzung widerspricht.

Also ist die Folge der $\alpha_k, k \in \bar{K}$, von null wegbeschränkt und wir können aus (5.6) für diese k auf

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

schließen. Wiederum aus (5.4) folgt nun, dass $(\|d^k\|_k)_{k \in \bar{K}} \rightarrow 0$ gelten muss, was nun aber $d^* \neq 0$ widerspricht. Damit ist die Ausgangsannahme widerlegt und der Satz bewiesen.

□

Die Vermutung, dass das globalisierte Newtonverfahren die fulminante lokale Konvergenzgeschwindigkeit des klassischen Newtonverfahrens erbt, bestätigt der folgenden

Satz 5.10 *Wählt man im globalisierten Newtonverfahren den Parameter $\sigma \in (0, \frac{1}{2})$, verschärft die Bedingung (5.4) zu*

$$\nabla f(x^k)^T d^k \leq -\rho \|d^k\|^p$$

für ein vorgewähltes $p > 2$ und ist für einen Häufungspunkt x^ der Iteriertenfolge des Verfahrens $\nabla^2 f(x^*)$ positiv definit, so gelten folgende Aussagen:*

1. *Die Gesamtfolge der Iterierten konvergiert gegen x^* und x^* ist eine strikte lokale Minimalstelle.*
2. *Für alle hinreichend großen $k \in \mathbf{N}$ ist die Suchrichtung immer die Newtonrichtung.*
3. *Für alle hinreichend großen $k \in \mathbf{N}$ wird immer die Schrittweite $\alpha_k = 1$ akzeptiert.*
4. *Die Iteriertenfolge konvergiert superlinear gegen x^* .*
5. *Ist $\nabla^2 f$ lokal lipschitz-stetig, so konvergiert die Iteriertenfolge quadratisch gegen x^**

Beweis: [GK1], Satz 9.10

5.1.3 Spezialfall: Quadratische Optimierung

Wird das *Gradientenverfahren* auf eine quadratische Zielfunktion f angewendet, so kann die Armijo-Regel rechenstechnisch problemlos durch die schärfere Minimierungsregel ersetzt werden.

Ist etwa $f(x) = \frac{1}{2}x^T Qx + p^T x$ mit symmetrischer quadratischer Matrix Q und ist d die ermittelte Suchrichtung mit $\nabla f(x)^T d < 0$, so ist das Minimum von f auf dem Strahl $x + \lambda d$, $\lambda \geq 0$, dadurch zu ermitteln, dass man die Ableitung der Zielfunktion auf dem Strahl gleich null setzt:

$$0 = \frac{\partial}{\partial \lambda} f(x + \lambda d) = \nabla f(x + \lambda d)^T d$$

Es ist also die Richtungsableitung von f längs d an der Stelle $x + \lambda d$ gerade gleich null. Wegen $\nabla f(x) = Qx + p$ ergibt sich

$$0 = \left((x + \lambda d)^T Q + p^T \right) d = (x^T Q + p^T) d + \lambda d^T Q d = \nabla f(x)^T d + \lambda d^T Q d$$

Ist nun $d^T Q d = 0$, so gibt es wegen $\nabla f(x)^T d < 0$ keine Lösung der Gleichung, also fällt $f(x + \lambda d)$ für $\lambda \geq 0$ unbegrenzt ab. Ist dagegen $d^T Q d \neq 0$, so lautet der eindeutige Wert $\hat{\lambda}$ für λ , bei dem das Minimum angenommen wird,

$$\hat{\lambda} = -\frac{\nabla f(x)^T d}{d^T Q d}.$$

Man beachte, dass sich die Einschränkung von f auf die Gerade $x + td$, $t \in \mathbf{R}$, $\tilde{f}(t) := f(x + td)$, wie folgt beschreiben lässt:

$$\begin{aligned} \tilde{f}(t) &= f(x + td) = \frac{1}{2}(x + td)^T Q(x + td) + p^T(x + td) \\ &= \frac{1}{2}x^T Qx + tx^T Qd + \frac{1}{2}t^2 d^T Qd + p^T x + tp^T d \\ &= \frac{1}{2}t^2 d^T Qd + tx^T Qd + tp^T d + \text{const.} \\ &= \frac{1}{2}t^2 d^T Qd + t \nabla f(x)^T d + \text{const.} \end{aligned}$$

Dieses quadratische Polynom ist eine Gerade genau dann, wenn $d^T Q d = 0$ gilt und eine nach oben geöffnete Parabel, wenn $d^T Q d > 0$ gilt, eine nach unten geöffnete Parabel, wenn $d^T Q d < 0$ gilt.

Skizze:

Wendet man bei quadratischer Optimierungsaufgabe die Minimierungsregel im Gradientenverfahren an, so gilt folgender Konvergenzsatz

Satz 5.11 Sei $f(x) = \frac{1}{2}x^T Qx + p^T x + q$ mit $Q \in \mathbf{R}^{n \times n}$ symmetrisch und positiv definit, $p \in \mathbf{R}^n$ und $q \in \mathbf{R}$ vorgegeben. Das Gradientenverfahren mit der Minimierungsregel für die Schrittweitenbestimmung konvergiert für jeden Startvektor $x^0 \in \mathbf{R}^n$ gegen den eindeutig bestimmten globalen Minimierer x^* und es gilt

$$f(x^{k+1}) - f(x^*) \leq \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \right)^2 (f(x^k) - f(x^*)) \quad (5.7)$$

wobei λ_{\max} der größte und λ_{\min} der kleinste Eigenwert der Matrix Q seien.

Beweis: [GK1], Satz 8.6

□

Damit zeigt sich, dass das Gradientenverfahren selbst in der verschärften Form wohl sehr langsam konvergiert. Da sich jede zweimal stetig differenzierbare Funktion f in der Nähe eines globalen Minimums durch eine quadratische Funktion approximieren lässt, beschreibt (5.7) in etwa auch das Konvergenzverhalten im allgemeineren Fall. Eine Anreicherung durch das Newtonverfahren ist also sehr sinnvoll.

Bemerkung: Der Verkleinerungsfaktor in der Formel (5.7) kann mithilfe der bekannten **Kondition** der Matrix Q umgeschrieben werden. Setzt man die Kondition $Kond(Q)$ wie folgt fest:

$$\kappa := Kond(Q) := \frac{\lambda_{\max}}{\lambda_{\min}}$$

so erhält man

$$\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} = \frac{\kappa - 1}{\kappa + 1}$$

und es zeigt sich, dass die Konvergenz um so schlechter ist, je größer die Kondition κ ist.

Wird das *Newtonverfahren* auf eine quadratische Funktion $f(x) = \frac{1}{2}x^T Qx + p^T x$ angewendet, so muss Q als zweite Ableitung von f invertierbar sein. Dann zeigt sich die überragende Qualität der Newtonrichtung als Suchrichtung: wegen $Q(x^1 - x^0) = -(Qx^0 + p) \iff Qx^1 + p = 0$ erkennt man, dass das Newtonverfahren, unabhängig vom Startpunkt x^0 , in *einem einzigen* Schritt einen stationären Punkt x^1 ansteuert! Die Newtonrichtung ist also die "optimale Suchrichtung", da sie von jedem beliebigen Startpunkt unmittelbar auf den (eindeutigen) stationären Punkt zeigt.

Gleichzeitig zeigt sich, dass auch für nicht streng konvexe Funktionen f die Verwendung der Newtonformel sinnvoll ist: man hat lediglich die Gleichung $Qx + p = 0$ mit einem geeigneten Verfahren zu lösen und erhält bei Lösbarkeit einen KKT Punkt.

5.1.4 Übungsaufgaben

Aufgabe 1

Programmieren Sie das Gradientenverfahren in Matlab und testen Sie es gegen Versionen des Simplexverfahrens sowohl für quadratische als auch für beliebige nichtlineare Funktionen.

Aufgabe 2

Erstellen Sie ein Matlab Programm zum globalisierten Newtonverfahren. Testen Sie das Programm an zufallserzeugten quadratischen Funktionen und an nichtlinearen Funktionen aus der Literatur.

5.2 Verfahren konjugierter Richtungen

Wir haben im letzten Abschnitt die Effizienz des Gradientenverfahrens daran gemessen, wie gut es eine streng konvexe quadratische Zielfunktion zu minimieren in der Lage ist. Auf der anderen Seite haben wir mit der Newtonrichtung eine Suchrichtung kennengelernt, die der Gradientenrichtung überlegen zu sein scheint, aber nur mit größerem Aufwand zu berechnen ist.

Bei der Suche nach "besseren Suchrichtungen" sollte man also streng konvexe quadratische Zielfunktionen zum Maßstab nehmen. Im Folgenden konstruieren wir Richtungen, die sich bei der Minimierung solcher Zielfunktionen als besonders geeignet erweisen.

Vorgegeben sei die quadratische Zielfunktion $q(x) = \frac{1}{2}x^T Qx + p^T x$ mit *positiv definierter* symmetrischer Matrix Q . q ist also eine streng konvexe Funktion mit eindeutigen Minimierer. Wir wollen diesen Minimierer bestimmen.

Natürlich kann man mit algebraischen Mitteln einfach die Stelle berechnen, an der der Gradient der Funktion q verschwindet

$$0 = \nabla q(x) = Qx + p$$

d.h. man bekommt den Minimierer x^* durch Lösen des Gleichungssystems $Qx = -p$ mit einem freigewählten Gleichungslöser.

Auf der anderen Seite kann man den Minimierer über die folgenden Überlegungen ermitteln, wodurch ein Verfahren entsteht, das in unmittelbare Konkurrenz zum ausgewählten Gleichungslöser tritt.

5.2.1 Das Verfahren konjugierter Gradienten

Lemma 5.12 *Es sei $x^0 \in \mathbf{R}^n$ beliebig und es seien $d^0, \dots, d^{n-1} \in \mathbf{R}^n$ von Vektoren mit*

$$\begin{aligned} (d^i)^T Qd^j &= 0 \quad \text{für alle } i, j = 0, \dots, n-1, i \neq j, \\ (d^i)^T Qd^i &> 0 \quad \text{für alle } i = 0, \dots, n-1, \end{aligned} \quad (5.8)$$

Dann liefert das Verfahren der sukzessiven eindimensionalen Minimierung längs der Richtungen d^0, \dots, d^{n-1} , d.h. die Berechnung der Iterationsfolge

$$x^{k+1} = x^k + t_k d^k \quad \text{mit} \quad t_k = -\frac{\nabla q(x^k)^T d^k}{(d^k)^T Qd^k} \quad \text{für } k = 0, \dots, n-1 \quad (5.9)$$

nach spätestens n Schritten das Minimum $x^* = x^n$ der Zielfunktion q . Es gilt dann ebenfalls

$$\nabla q(x^{k+1})^T d^j = 0 \quad \text{für alle } j = 0, \dots, k \quad (5.10)$$

Beweis: Zunächst ist bekannt, dass die angegebene Schrittweite t_k diejenige ist, die q von x^k entlang d^k minimiert

$$t_k = \arg \min_{t \in \mathbf{R}} q(x^k + td^k)$$

Es gilt dann mit $x^{k+1} = x^k + t_k d^k$ nach Konstruktion für $k = 0, \dots, n-1$

$$\nabla q(x^{k+1})^T d^k = 0$$

Setzt man zur Abkürzung $g^k := \nabla q(x^k)$ (für alle k), so gilt für alle $i, j = 0, \dots, n-1, i \neq j$,

$$(g^{i+1} - g^i)^T d^j = (Qx^{i+1} + p - Qx^i - p)^T d^j = (Qt_i d^i)^T d^j = t_i (d^i)^T Q d^j = 0$$

Also gilt für $j = 0, \dots, k$

$$\nabla q(x^{k+1})^T d^j = (g^{j+1})^T d^j + \sum_{i=j+1}^k (g^{i+1} - g^i)^T d^j = 0$$

Dies ist gerade Behauptung (5.10). Da die n Vektoren nach Voraussetzung bzgl des Skalarprodukts $\langle u, v \rangle = u^T Q v$ orthogonal sind, bilden sie eine Basis des \mathbf{R}^n . Daher folgt aus $\nabla q(x^n)^T d^j = 0$ für $j = 0, \dots, n-1$ gerade

$$\nabla q(x^n) = 0$$

der Vektor x^n ist also der eindeutige Minimierer der Funktion q .

□

Bezeichnung: Man nennt $k \leq n$ Vektoren d^0, \dots, d^{k-1} mit der Eigenschaft (5.8) **Q -konjugiert oder Q -orthogonal**. (Wir hatten einen solchen Begriff bereits im Zusammenhang mit dem Verfahren von BEALE geprägt).

Konstruktion Q -konjugierter Vektoren

Q -konjugierte Richtungen sind nicht eindeutig. Man erhält Q -konjugierte Vektoren bei positiv definitem Q leicht über das *Gram-Schmidtsche Orthogonalisierungsverfahren*, wenn man das Skalarprodukt $\langle u, v \rangle = u^T Q v$ verwendet und dabei von beliebigen linear unabhängigen Vektoren ausgeht. Wir tun dies nun speziell abgestimmt auf die Funktion q .

Man setzt zunächst $d^0 = -\nabla q(x^0)$ und geht davon aus, dass $d^0 \neq 0$ gilt, also x^0 nicht Minimierer von q ist. Dann setzt man

$$t_0 = -\frac{\nabla q(x^0)^T d^0}{(d^0)^T Q d^0}, \quad x^1 = x^0 + t_0 d^0$$

und hat damit $\nabla q(x^1)^T d^0 = 0$.

Gehen wir nun davon aus, dass die $l+1$ Vektoren d^0, \dots, d^l bereits als Q -konjugierte Vektoren und Punkte x^0, \dots, x^{l+1} konstruiert sind, die die Eigenschaften (5.10) und (5.9) haben für $k = 0, \dots, l$. Wir setzen wieder

$$g^k := \nabla q(x^k) \quad \text{für } k \in \mathbf{N}_0$$

Ferner sei $g^{l+1} = \nabla q(x^{l+1}) \neq 0$ angenommen, da sonst mit x^{l+1} bereits der Minimierer der Funktion q gefunden ist. Dann setzt man d^{l+1} wie im Orthogonalisierungsverfahren an als

$$d^{l+1} = -g^{l+1} + \sum_{i=0}^l \beta_i^l d^i. \quad (5.11)$$

wobei zu beachten ist, dass wegen Eigenschaft (5.10) $g^{l+1} = \nabla q(x^{l+1})$ nicht Linearkombination der d^0, \dots, d^l sein kann.

Dann ist die Forderung $(d^{l+1})^T Q d^j = 0$ für $j = 0, \dots, l$ genau dann erfüllt, wenn

$$\beta_j^l = \frac{(g^{l+1})^T Q d^j}{(d^j)^T Q d^j} \quad \text{für } j = 0, \dots, l \quad (5.12)$$

gilt. Man setzt wieder $x^{l+2} = x^{l+1} + t_{l+1} d^{l+1}$ mit $t_{l+1} = -\frac{(g^{l+1})^T d^{l+1}}{(d^{l+1})^T Q d^{l+1}}$ und hat damit $(g^{l+2})^T d^{l+1} = 0$. Wie im Beweis zum vorigen Lemma folgt dann sogar $(g^{l+2})^T d^j = 0$ für $j = 0, \dots, l+1$.

Vereinfachung der Formeln

Mit den folgenden Überlegungen vereinfachen wir Formel (5.11). Zunächst gilt wegen Formeln (5.11) und (5.10)

$$(g^{l+1})^T d^{l+1} = -\|g^{l+1}\|^2 < 0$$

woraus entnommen werden kann, dass d^{l+1} eine Abstiegsrichtung für q in x^{l+1} ist und dass $t_{l+1} > 0$ gilt. Da die Konstruktion von d^k für alle $k = 0, \dots, l+1$ die gleiche ist, gilt ebenso

$$(g^k)^T d^k = -\|g^k\|^2 < 0 \quad \text{und } t_k > 0. \quad (5.13)$$

und wegen (5.10) und (5.11)

$$(g^{l+1})^T g^j = (g^{l+1})^T \left(\sum_{i=0}^{j-1} \beta_i^{j-1} d^i - d^j \right) = 0 \quad \text{für } j = 0, \dots, l$$

Daraus folgt schließlich wegen $g^{j+1} - g^j = Qx^{j+1} - Qx^j = t_j Qd^j$ und $t_j > 0$ für j

$$(g^{l+1})^T Qd^j = \frac{1}{t_j} (g^{l+1})^T (g^{j+1} - g^j) \quad (5.14)$$

so dass also wegen (5.12)

$$\beta_j^l = 0 \quad \text{gilt für } j = 0, \dots, l-1$$

und wegen (5.14), (5.11) und (5.9)

$$\beta_l^l = \frac{1}{t_l} \frac{(g^{l+1})^T g^{l+1}}{(d^l)^T Qd^l} = \frac{(g^{l+1})^T g^{l+1}}{-(g^l)^T d^l} = \frac{\|g^{l+1}\|^2}{\|g^l\|^2} \quad (5.15)$$

Die Formel (5.11) vereinfacht sich daher erheblich wie folgt:

$$d^{l+1} = -g^{l+1} + \beta_l^l d^l. \quad (5.16)$$

Zusammen mit den Formeln (5.9) ist nun ein sehr einfaches rekursives System zur Lösung der Gleichung $Qx = -p$ mit positiv definitem Q etabliert, welches in Konkurrenz zu anderen Lösungsverfahren tritt. Es trägt den Namen: **Verfahren konjugierter Gradienten (CG)**.

Erwähnenswert ist, dass man im Zusammenhang mit dem CG-Verfahren gerne eine Koordinatentransformation zur Verbesserung der Kondition der Matrix Q vornimmt. Man spricht dann vom **Präkonditionierten CG Verfahren**. Aus Einzelheiten wollen wir hier nicht eingehen (vgl. [GK1], Abschnitt 13.1)

Beispiel 5.13 Betrachten wir die Aufgabe, ein Minimum der quadratischen Funktion $q(x) = -12x_2 + 4x_1^2 + 4x_2^2 - 4x_1x_2$ zu suchen. Deren Gradient ist $\nabla q(x_1, x_2) = (8x_1 - 4x_2, -12 + 8x_2 - 4x_1)$, so dass wir

$$Q = \begin{pmatrix} 8 & -4 \\ -4 & 8 \end{pmatrix} \quad \text{und} \quad p = \begin{pmatrix} 0 \\ -12 \end{pmatrix}$$

ablesen.

Zu lösen ist also die Gleichung $Qx = -p$. Starten wir das CG Verfahren bei $x^0 = (0, 0)$. Dann ist $g^{0T} = \nabla q(x^0)^T = (0, -12)$. Nun bilden wir mit $d^0 = -g^0$ und berechnen x^1 nach Formel (5.9)

$$t_0 = -\frac{g^{0T}d^0}{(d^0)^T Q d^0} = -\frac{(0, -12) \begin{pmatrix} 0 \\ 12 \end{pmatrix}}{(0, 12) \begin{pmatrix} 8 & -4 \\ -4 & 8 \end{pmatrix} \begin{pmatrix} 0 \\ 12 \end{pmatrix}} = \frac{144}{12^2 \cdot 8} = \frac{1}{8}$$

so dass sich

$$x^1 = x^0 + \frac{1}{8}d^0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{1}{8} \begin{pmatrix} 0 \\ 12 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{3}{2} \end{pmatrix}$$

ergibt. Nun berechnen wir g^1

$$\begin{aligned} g^1 &= \begin{pmatrix} 8 & -4 \\ -4 & 8 \end{pmatrix} \begin{pmatrix} 0 \\ \frac{3}{2} \end{pmatrix} + \begin{pmatrix} 0 \\ -12 \end{pmatrix} \\ &= \begin{pmatrix} -6 \\ 12 \end{pmatrix} + \begin{pmatrix} 0 \\ -12 \end{pmatrix} = \begin{pmatrix} -6 \\ 0 \end{pmatrix} \end{aligned}$$

und d^1 nach der Formel (5.16) bei $l = 0$:

$$\begin{aligned} \beta_0^0 &= \frac{\|g^1\|^2}{\|g^0\|^2} = \frac{36}{144} = \frac{1}{4} \\ d^1 &= -g^1 + \beta_0^0 d^0 = \begin{pmatrix} 6 \\ 0 \end{pmatrix} + \frac{1}{4} \begin{pmatrix} 0 \\ 12 \end{pmatrix} = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \end{aligned}$$

Damit kann auch x^2 berechnet werden:

$$t_k = -\frac{g^{1T}d^1}{(d^1)^T Q d^1} = -\frac{\begin{pmatrix} -6 \\ 0 \end{pmatrix}^T \begin{pmatrix} 6 \\ 3 \end{pmatrix}}{\begin{pmatrix} 6 \\ 3 \end{pmatrix}^T \begin{pmatrix} 8 & -4 \\ -4 & 8 \end{pmatrix} \begin{pmatrix} 6 \\ 3 \end{pmatrix}} = \frac{36}{\begin{pmatrix} 6 \\ 3 \end{pmatrix}^T \begin{pmatrix} 36 \\ 0 \end{pmatrix}} = \frac{1}{6}$$

$$x^2 = x^1 + \frac{1}{6}d^1 = \begin{pmatrix} 0 \\ \frac{3}{2} \end{pmatrix} + \frac{1}{6} \begin{pmatrix} 6 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Man prüft leicht nach, dass $x^{2T} = (1, 2)$ tatsächlich die Lösung der Minimierungsaufgabe bzw die Lösung von $Qx + p = 0$ ist.

Skizze:

5.2.2 Das Verfahren von Fletcher-Reeves

Damit können wir nun verallgemeinernd das folgende Verfahren betrachten, das nicht nur zur Minimierung der streng konvexen quadratischen Funktion q herangezogen, sondern sogar für beliebige, stetig differenzierbare Funktion $f : \mathbf{R}^n \rightarrow \mathbf{R}$ formuliert werden kann, weil keine Spezifika der quadratischen Zielfunktion im Verfahren vorkommen. Es werden keine näheren Informationen (ausser der Ableitung) zur Funktion f benötigt.

Vorstellungsmäßig werde dabei die gegebene Zielfunktion f an jedem Iterationspunkt x^k durch eine streng konvexe quadratische Zielfunktion angenähert und bzgl dieser die nächste Suchrichtung ausgerechnet. Lediglich von der bisherigen Formel für die Schrittweite wird Abstand genommen. Stattdessen wird eine Berechnungsmethode verwendet, die in endlich vielen Schritten eine sinnvolle Schrittweite berechnet.

Verfahren FR (Fletcher-Reeves)

1. **Start:** Wähle $x^0 \in \mathbf{R}^n$, $\varepsilon \geq 0$, $0 < \sigma < \rho < \frac{1}{2}$, setze $d^0 := -\nabla f(x^0)$ und $k = 0$.
2. **Abbruch:** Ist $\|\nabla f(x^k)\| \leq \varepsilon$, **stopp1**, (ein KKT-Punkt wurde angenähert)
3. **Schrittweite:** Bestimme eine Schrittweite $t_k > 0$ mit

$$\begin{aligned} f(x^k + t_k d^k) &\leq f(x^k) + \sigma t_k \nabla f(x^k)^T d^k && \text{und} \\ \left| \nabla f(x^k + t_k d^k)^T d^k \right| &\leq -\rho \nabla f(x^k)^T d^k \end{aligned}$$

Existiert keine solche Schrittweite, **stopp2**, (die Zielfunktion ist unbeschränkt)

4. **Aktualisierung:** Setze

$$\begin{aligned} x^{k+1} &: = x^k + t_k d^k \\ \beta_k &: = \frac{\|\nabla f(x^{k+1})\|^2}{\|\nabla f(x^k)\|^2} \\ d^{k+1} &: = -\nabla f(x^{k+1}) + \beta_k d^k \end{aligned}$$

und $k = k + 1$. Gehe zu 2.)

In dieser Version des Verfahrens findet die **strenge Wolfe-Powell Regel** zur Schrittweitenbestimmung Anwendung. Diese kann durch folgende **Skizze** erläutert werden: wir betrachten bei gegebenen x^k , d^k , $f(x^k)$, $\nabla f(x^k)$ mit $\nabla f(x^k)^T d^k < 0$ und (von k unabhängigen Parametern) $\sigma \in (0, \frac{1}{2})$, $\rho \in [\sigma, 1)$ die Funktion $\varphi(t) := f(x^k + t d^k)$:

Berechnung der Schrittweite

Die folgende Prozedur liefert für den allgemeinen Fall eine praktische Berechnungsmöglichkeit für t_k bei beliebiger (stetig differenzierbarer) Zielfunktion f .

Es sei wieder f als stetig differenzierbar und $\sigma \in (0, \frac{1}{2})$, $\rho \in [\sigma, 1)$ vorausgesetzt. Zu $x, d \in \mathbf{R}^n$ mit $\nabla f(x)^T d < 0$ setzen wir wieder $\varphi(t) := f(x^k + td^k)$ und $\psi(t) := \varphi(t) - \varphi(0) - t\sigma\varphi'(0)$. Dann lauten die strengen Wolfe-Powell Regeln

$$\psi(t) \leq 0 \quad \text{und} \quad |\varphi'(t)| \leq \rho |\varphi'(0)|$$

Mit $\langle u, v \rangle$ bezeichnen wir adhoc das durch $u, v \in \mathbf{R}$ festgelegte abgeschlossene Intervall (nicht notwendig $v \geq u$).

Prozedur SWP (strenge Wolfe-Powell Regel)

- (A.0) wähle $t_0 > 0$, $\gamma > 1$ und setze $i := 0$
- (A.1) Ist $\psi(t_i) \geq 0$, so setze $a := 0$ und $b := t_i$ und gehe zu (B.0).
Ist $\psi(t_i) < 0$ und $|\psi'(t_i)| \leq (\rho - \sigma)|\varphi'(0)|$, so setze $t := t_i$. **Stopp**
Ist $\psi(t_i) < 0$ und $\psi'(t_i) > 0$, so setze $a := t_i$, $b := 0$ und gehe zu (B.0).
Ist $\psi(t_i) < 0$ und $\psi'(t_i) < 0$, so setze $t_{i+1} := \gamma t_i$, $i := i + 1$ und gehe zu (A.1).
- (B.0) Wähle $\tau_1, \tau_2 \in (0, \frac{1}{2}]$, setze $j := 0$ und übernehme $a_0 := a$, $b_0 := b$ aus Phase A.
- (B.1) Wähle $t_i \in \langle a_j + \tau_1(b_j - a_j), b_j - \tau_2(b_j - a_j) \rangle$.
- (B.2) Ist $\psi(t_j) \geq \psi(a_j)$, so setze $a_{j+1} := a_j$, $b_{j+1} := t_j$, $j := j + 1$ und gehe zu (B.1).
Ist $\psi(t_j) < \psi(a_j)$ und $|\psi'(t_j)| \leq (\rho - \sigma)|\varphi'(0)|$, so setze $t := t_j$. **Stopp**
Ist $\psi(t_j) < \psi(a_j)$ und $(t_j - a_j)\psi'(t_j) < 0$, so setze $a_{j+1} := t_j$, $b_{j+1} := b_j$, $j := j + 1$ und gehe zu (B.1).
Ist $\psi(t_j) < \psi(a_j)$ und $(t_j - a_j)\psi'(t_j) > 0$, so setze $a_{j+1} := t_j$, $b_{j+1} := a_j$, $j := j + 1$ und gehe zu (B.1).

Es gilt dann der

Satz 5.14 Sei $f : \mathbf{R}^n \rightarrow \mathbf{R}$ stetig differenzierbar und seien $\sigma \in (0, \frac{1}{2})$, $\rho \in [\sigma, 1)$ vorgewählt und $x, d \in \mathbf{R}^n$ mit $\nabla f(x)^T d < 0$ gegeben. Bricht die obige Prozedur ab, so ist das letzte t eine für die strenge Wolfe-Powell Regel gültige Schrittweite. Bricht sie nicht ab, so fällt die Funktion f auf der Iteriertenfolge unbeschränkt.

Beweis: leichte Modifikation des Beweises zu [GK1], Satz 6.6.

Bzgl des Verfahrens FR können nun folgende Aussagen getroffen werden:

Satz 5.15 Sei $f : \mathbf{R}^n \rightarrow \mathbf{R}$ stetig differenzierbar. Dann ist das Verfahren FR wohldefiniert, d.h. lässt sich keine Schrittweite nach der strengen Wolfe-Powell Regel berechnen, so ist die Zielfunktion unbeschränkt.

Beweis: leichte Modifikation des Beweises zu [GK1], Satz 13.6

Satz 5.16 Sei $f : \mathbf{R}^n \rightarrow \mathbf{R}$ stetig differenzierbar. Es sei ferner ∇f lipschitzstetig auf der Levelmenge $\mathcal{L}(x^0) = \{x \in \mathbf{R}^n : f(x) \leq f(x^0)\}$. Dann gilt für die durch das Verfahren erzeugte Iteriertenfolge x^k , $k \in \mathbf{N}$:

f ist unbeschränkt auf der Iteriertenfolge oder es gilt

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

Beweis: leichte Modifikation des Beweises zu [GK1], Satz 13.7.

Man beachte dazu: Seien $x^0 \in \mathbf{R}^n$ und $f : \mathbf{R}^n \rightarrow \mathbf{R}$ zweimal stetig differenzierbar. Ist eine der folgenden Bedingungen erfüllt, so ist ∇f lipschitzstetig auf der Levelmenge $\mathcal{L}(x^0) = \{x \in \mathbf{R}^n : f(x) \leq f(x^0)\}$:

1. $\|\nabla^2 f(x^k)\|$ ist beschränkt auf einer konvexen Obermenge der Levelmenge $\mathcal{L}(x^0)$.
2. die Levelmenge $\mathcal{L}(x^0)$ ist kompakt.

(Beweis: [GK1], Bemerkung 5.4)

Bemerkung: In der praktischen Durchführung hat es sich als sinnvoll erwiesen, nach jeweils n Schritten einen *Restart des Verfahrens* durchzuführen, um sich der jeweiligen quadratischen Approximation besser anzupassen.

Beispiel 5.17 *Wir betrachten noch einmal die Funktion $f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$, auf die wir bereits das Gradientenverfahren und das Newtonverfahren angewendet hatten. Wendet man das Verfahren FR auf diese Funktion an, startend bei $(x_1, x_2) = (0, 3)$, so nimmt das Verfahren ungefähr den folgenden Verlauf, um sich der eindeutigen Optimallösung $x^* = (2, 1)$ anzunähern. Wenn dabei der Line search genau ausgeführt und nach jeweils zwei Schritten ein Restart durchgeführt wird, so lautet eine Abbruchlösung $x = (2.185, 1.094)$ mit $\|\nabla f(x)\| = 0.02$ nach 8 Iterationen:*

5.2.3 Quasi Newton Verfahren

Wir haben in unseren bisherigen Untersuchungen zur unrestringierten Optimierung verschiedene sinnvolle Wahlen von Suchrichtungen und Schrittweiten kennengelernt:

- als Suchrichtungen die negative Gradientenrichtung (steilster Abstieg), die Newtonrichtung, aufdatierte und zueinander konjugierte Richtungen
- als Schrittweitenregeln die Armijo (einschließlich der Schrittweitenwahl 1 bei Newtonverfahren), die strenge Wolfe-Powell-Regel und die Minimierungsregel

Zu bemerken ist, dass es auch eine "normale" **Wolfe Powell Regel** gibt, die eine etwas schwächere Forderung stellt als die strenge Wolfe Powell Regel. Sie verlangt

$$\begin{aligned} \text{nur } \nabla f(x^k + t_k d^k)^T d^k &\geq \rho \nabla f(x^k)^T d^k \\ \text{anstelle von } \left| \nabla f(x^k + t_k d^k)^T d^k \right| &\leq -\rho \nabla f(x^k)^T d^k \end{aligned}$$

Skizze:

Die Wohldefiniertheit dieser Regel soll hier nicht besprochen werden (vgl. [GK1], Kapitel 5 und 6).

Allgemeines unrestringiertes Abstiegsverfahren

Sicher ist es richtig, dass für die Konvergenz des allgemeinen zulässigen Abstiegsverfahrens im Falle der unrestringierten Optimierung die Suchrichtungen und die Schrittweiten gewissen Qualitätsanforderungen unterliegen müssen. Erstaunlicherweise kann dies in recht allgemeiner Weise formuliert werden.

Vorgegeben sei die unrestringierte Optimierungsaufgabe mit (stetig) differenzierbarer Funktion f . Wir benötigen zwei Definitionen, die sich auf den Prototyp des zulässigen Abstiegsverfahren beziehen.

Definition 5.18 *Man sagt: Das Verfahren benutzt **winkelbeschränkte Abstiegsrichtungen**, wenn es eine Konstante $0 < \kappa \leq 1$ gibt, so dass für jede Iterierte x^k und der in diesem Punkt gewählte Abstiegsrichtung d^k gilt*

$$\nabla f(x^k)^T d^k \leq -\kappa \|\nabla f(x^k)\| \|d^k\| \quad (\text{Winkelbedingung})$$

Das Verfahren benutzt **effiziente Schrittweiten**, wenn eine Konstante $\gamma > 0$ gibt so, dass für jede Iterierte x^k mit zugehöriger Suchrichtung d^k und Schrittweite α_k gilt

$$f(x^k + \alpha_k d^k) - f(x^k) \leq -\gamma \left(\frac{|\nabla f(x^k)^T d^k|}{\|d^k\|} \right)^2$$

Man beachte, dass die Winkelbedingung dafür sorgt, dass die Winkel zwischen dem negativen Gradienten $(-\nabla f(x^k))$ und den gewählten Suchrichtungen d^k gleichmäßig von 90° wegbeschränkt bleiben! Die Schrittweitenbedingung dagegen sorgt für einen "erheblichen Abstieg der Zielfunktion".

Nun sind solche Definitionen sinnlos, solange nicht geklärt ist, dass die auch erfüllt werden können. Es ist aber unmittelbar einsichtig, dass der steilste Abstieg $d^k = -\nabla f(x^k)$ unabhängig von der Wahl von κ immer eine winkelbeschränkte Abstiegsrichtung ist. Ferner kann nachgewiesen werden, dass die strenge Wolfe-Powell-Regel unter der Voraussetzung, dass der Gradient $\nabla f(x)$ auf der Levelmenge $\mathcal{L}(x^0) = \{x \in \mathbf{R}^n : f(x) \leq f(x^0)\}$ Lipschitzstetig ist, immer zu effizienten Schrittweiten führt. Das trifft auch zu für die Minimierungsregel und die normale Wolfe-Powell Regel. **Man beachte aber**, dass die Armijo-Regel *nicht* generell zu effizienten Schrittweiten führt.

Bezeichnung: Wir wollen das zulässige Abstiegsverfahren, bei dem bei Anwendung in der unrestringierten Optimierung winkelbeschränkte Abstiegsrichtungen und effiziente Schrittweiten eingesetzt werden, als **Allgemeines Unrestringiertes Abstiegsverfahren** bezeichnen.

Der Beweis eines Konvergenzsatzes für dieses Verfahren ist denkbar einfach:

Satz 5.19 *Betrachtet werde das allgemeine unrestringierte Abstiegsverfahren. Dann ist jeder Häufungspunkt der Iteriertenfolge ein stationärer Punkt der Funktion f .*

Beweis: (vgl. [GK1], Satz 4.6)

Betrachten wir eine beliebige Iterierte x^k im Verfahren. Dann gilt

$$f(x^k + \alpha_k d^k) - f(x^k) \leq -\gamma \left(\frac{|\nabla f(x^k)^T d^k|}{\|d^k\|} \right)^2$$

und aus der Winkelbedingung folgt sofort

$$f(x^k + \alpha_k d^k) - f(x^k) \leq -\gamma \kappa^2 \|\nabla f(x^k)\|^2 \leq 0 \quad (5.17)$$

Sei nun x^* der Grenzwert einer Teilfolge $\{x^k\}_{k \in K}$ der Iteriertenfolge. Da die Gesamtfolge der $f(x^k)$ monoton abfällt und auf einer Teilfolge gegen $f(x^*)$

konvergiert, konvergiert die gesamte Folge der $f(x^k)$ gegen $f(x^*)$. Es folgt also

$$f(x^{k+1}) - f(x^k) \rightarrow 0 \quad \text{für } k \rightarrow \infty.$$

Damit gilt auch

$$f(x^k + \alpha_k d^k) - f(x^k) \rightarrow 0 \quad \text{für } k \rightarrow \infty.$$

und aus (5.17) folgt unmittelbar

$$\|\nabla f(x^k)\| \rightarrow 0 \quad \text{für } k \rightarrow \infty.$$

□

Man beachte: Auch diese Version des zulässigen Abstiegsverfahrens kann dazu dienen, Unterverfahren aufzunehmen, deren eigenständige Konvergenz nicht bewiesen ist.

Ein Konstruktionsprinzip für Abstiegsrichtungen

Es gibt noch eine weitere Möglichkeit, sinnvoll Abstiegsrichtungen zu berechnen. Sie geht aus von einem Ansatz, der an die Newtonrichtungen angelehnt ist und diese verallgemeinert: Suchrichtungen d werden angesetzt als $d := -D\nabla f(x)$ mit *symmetrischer positiv definiten* $n \times n$ Matrix D , die in der Vorstellung die Inverse der Hessematrix approximativ ersetzen soll. Zu beachten ist, dass sowohl die Gradientenrichtung wie auch die Newtonrichtung unter diese Konstruktion fallen.

Ausgangspunkt für diesen Ansatz ist die folgende

Beobachtung: Ist $\nabla f(x) \neq 0$ und D eine positiv definite symmetrische $n \times n$ Matrix, so ist $d := -D\nabla f(x)$ eine Abstiegsrichtung.

Begründung: Ist $D = LL^T$ die Cholesky-Zerlegung der Matrix D mit invertierbarer Matrix L , so folgt

$$\nabla f(x)^T d = \nabla f(x)^T (-D\nabla f(x)) = -\nabla f(x)^T LL^T \nabla f(x) = -\|L^T \nabla f(x)\|^2 < 0$$

Man kann also im allgemeinen unrestringierten Abstiegsverfahren grundsätzlich Abstiegsrichtungen der Form $d = -D\nabla f(x)$ verwenden, sollte dabei

allerdings zum Ziel haben, algorithmisch und rechnerisch günstige positiv definite Matrizen D zu benutzen. Dazu gibt es mehrere Ideen. Man kann z. B. die rechnerisch teure Newtonrichtung $D = \nabla^2 f(x)^{-1}$ nur alle $r \in \mathbf{N}$ Schritte neu zu berechnen und zwischendurch einfach das letzte D weiterverwenden, r frei gewählt (**vereinfachtes Newtonverfahren**). Oder man wählt D geeignet als Diagonalmatrix.

Man kann auch versuchen, ähnlich der Suchrichtungen bei Fletcher-Reeves Verfahren, die Matrizen D_k rechnerisch und für das Verfahren vorteilhaft zu D_{k+1} aufzudatieren. Einen solchen Weg beschreibt das folgende Verfahren. Es basiert auf dem

Lemma 5.20 *Seien D eine positiv definite symmetrische $n \times n$ Matrix, x eine Stelle in \mathbf{R}^n mit $\nabla f(x) \neq 0$ und $0 < \rho < 1$. Ferner sei $d := -D\nabla f(x)$ die daraus berechnete Abstiegsrichtung zur Funktion f und $\alpha > 0$ eine Schrittweite, für die mit $\bar{x} := x + \alpha d$ gilt*

$$\nabla f(\bar{x})^T d \geq \rho \nabla f(x)^T d. \quad (5.18)$$

Setzt man dann $p := \alpha d = \bar{x} - x$ sowie $q := \nabla f(\bar{x}) - \nabla f(x)$ und schließlich

$$\bar{D} := D + \frac{pp^T}{p^T q} - \frac{Dqq^T D}{q^T Dq},$$

so ist \bar{D} wohldefiniert und ebenfalls eine positiv definite symmetrische $n \times n$ Matrix.

Beweis: (vgl. [BSS], Lemma 8.8.5)

Zunächst zeigen wir, dass unter den gegebenen Voraussetzungen $p^T q > 0$ gilt.

$$\begin{aligned} p^T q &= (\bar{x} - x)^T (\nabla f(\bar{x}) - \nabla f(x)) \\ &= \alpha d^T (\nabla f(\bar{x}) - \nabla f(x)) \\ &= \alpha d^T \nabla f(\bar{x}) - \alpha d^T \nabla f(x) \\ &\geq \alpha \rho \nabla f(x)^T d - \alpha d^T \nabla f(x) \\ &= \alpha (\rho - 1) \nabla f(x)^T d \\ &= \alpha (1 - \rho) \nabla f(x)^T D \nabla f(x) \\ &> 0 \end{aligned}$$

Damit gilt auch $q \neq 0$ und wegen der positiven Definitheit von D ist $q^T D q > 0$. Daher ist \bar{D} wohldefiniert.

Es sei nun $y \in \mathbf{R}^n$ als beliebiger Vektor, $y \neq 0$, vorgelegt. Dann gilt

$$y^T \bar{D} y = y^T D y + \frac{(y^T p)^2}{p^T q} - \frac{(y^T D q)^2}{q^T D q}$$

Da D positiv definit ist, gibt es eine ebenfalls positiv definite Quadratwurzel $D^{\frac{1}{2}}$ mit $D = D^{\frac{1}{2}} D^{\frac{1}{2}}$. Wir setzen

$$a = D^{\frac{1}{2}} y \quad \text{und} \quad b = D^{\frac{1}{2}} q$$

Dann gilt zunächst $y^T D y = a^T a$ und $q^T D q = b^T b$ sowie $y^T D q = a^T b$ und es folgt wegen $p^T q > 0$ und $b^T b = q^T D q > 0$ nach der Cauchy-Schwarzschen Ungleichung

$$\begin{aligned} y^T \bar{D} y &= a^T a + \frac{(y^T p)^2}{p^T q} - \frac{(a^T b)^2}{b^T b} \\ &= \frac{(a^T a)(b^T b) - (a^T b)^2}{b^T b} + \frac{(y^T p)^2}{p^T q} \geq 0 \end{aligned}$$

Damit ist \bar{D} zumindest als positiv semidefinit nachgewiesen. Nehmen wir nun an, dass $y^T \bar{D} y = 0$ gilt.

In der obigen Abschätzung kann nur dann das Gleichheitszeichen gelten, wenn

$$(a^T a)(b^T b) = (a^T b)^2 \quad \text{und} \quad y^T p = 0 \quad (5.19)$$

gilt. Dabei kann die erste Gleichheit nur eintreten, wenn $a = \lambda b$ für ein $\lambda \in \mathbf{R}$ gilt. Das hieße nach Definition

$$D^{\frac{1}{2}} y = \lambda D^{\frac{1}{2}} q \implies y = \lambda q$$

Wegen $y \neq 0$ folgt daraus $\lambda \neq 0$ und wegen (5.19) $0 = y^T p = \lambda p^T q$, was direkt $p^T q > 0$ widerspricht. Also ist die Annahme falsch und es gilt $y^T \bar{D} y > 0$, wodurch \bar{D} als positiv definit nachgewiesen ist.

□

Zu beachten ist, dass die Zusatzregel (5.18) durch die normale und die strenge Wolfe-Powell-Regel sowie die Minimierungsregel automatisch erfüllt wird.

Bemerkung: Aus dem Beweis ist leicht ersichtlich, dass das Lemma genauso gilt, wenn von $d := -D^{-1}\nabla f(x)$ anstelle von $d := -D\nabla f(x)$ ausgegangen wird, da D genau dann positiv definit ist, wenn dies für D^{-1} gilt. Außerdem kann die Rolle von p und q ausgetauscht werden.

Angeregt durch das Lemma betrachten wir nun die folgende Spezialisierung des allgemeinen unrestringierten Abstiegsverfahrens, dessen Konvergenz im Sinne des Konvergenzsatzes 5.19 durch das Lemma gesichert ist.

Davidon-Fletcher-Powell (DFP) Verfahren

1. **Start:** Wähle $x^0 \in \mathbf{R}^n$, $\varepsilon \geq 0$ und die Parameter $\kappa, \gamma > 0$ zur Überprüfung der Winkelbedingung und der Effizienz der Schrittweiten sowie die Parameter $\sigma \in (0, \frac{1}{2})$ und $\rho \in [\sigma, 1)$ für die strenge Wolfe-Powell-Schrittweitenregel. Setze $D_0 := I_n$ und $k = 0$.
2. **Abbruch:** Ist $\|\nabla f(x^k)\| \leq \varepsilon$, **stopp1**, (ein KKT-Punkt wurde angenähert)
3. **Schritt:** Erfüllt d^k nicht die Winkelbedingung an der Stelle x^k , so setze $D_k := I$.

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

Berechne eine Schrittweite $\alpha_k > 0$ so, dass die strengen Wolfe-Powell Bedingungen

$$\begin{aligned} f(x^k + \alpha_k d^k) &\leq f(x^k) + \sigma \alpha \nabla f(x^k)^T d^k \\ \left| f(x^k + \alpha_k d^k)^T d^k \right| &\leq -\rho \nabla f(x^k)^T d^k \end{aligned}$$

erfüllt sind. Existiert keine solche Schrittweite, **stopp2**, (die Zielfunktion ist unbeschränkt)

4. **Aktualisierung:** Setze $x^{k+1} := x^k + \alpha_k d^k$ und

$$\begin{aligned} D_{k+1} &: = D_k + \frac{p^k (p^k)^T}{(p^k)^T q^k} - \frac{D_k q^k (q^k)^T D_k}{(q^k)^T D_k q^k} \quad \text{mit} \quad (5.20) \\ p^k &: = \alpha_k d^k = x^{k+1} - x^k \\ q^k &: = \nabla f(x^{k+1}) - \nabla f(x^k) \end{aligned}$$

und $k = k + 1$. Gehe zu 1.

Beispiel 5.21 *Wir betrachten wie schon zuvor die Funktion $f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$. Wendet man das Verfahren DFP auf diese Funktion an, startend bei $(x_1, x_2) = (0, 3)$, so nimmt das Verfahren ungefähr den folgenden Verlauf, um sich der eindeutigen Optimallösung $x^* = (2, 1)$ anzunähern. Wenn dabei wieder der Line search genau ausgeführt und nach jeweils zwei Schritten ein Restart (ähnlich dem Verfahren FR) durchgeführt wird, so lautet eine Abbruchlösung $x = (2.115, 1.058)$ mit $\|\nabla f(x)\| = 0.006$ nach 8 Iterationen.*

Bemerkungen: Ersetzt man im Verfahren $d^k := -D_k \nabla f(x^k)$ durch $d^k := -D_k^{-1} \nabla f(x^k)$ und vertauscht die Rolle von p und q , so ist das Verfahren mit den Namen BROYDEN, FLETCHER, GOLDFARB, SHANNO (**BFGS**) verbunden. (vgl. [GK1], Algorithmus 11.34).

Man kann zur Matrix \bar{D} eine Inverse angeben, die ebenfalls aus einer Rekursionsformel entsteht:

$$\bar{B} = B + \frac{(q - Bp) q^T + q(q - Bp)}{q^T p} - \frac{(q - Bp)^T p}{(q^T p)^2} q q^T \quad (5.21)$$

Man rechnet nach: Ist B die Inverse von D , so ist \bar{B} die Inverse von \bar{D} , d.h. $BD = I \implies \bar{B}\bar{D} = I$. Damit kann man sich das Invertieren von D ersparen, wenn man mit dem Ansatz $d^k := -D_k^{-1} \nabla f(x^k)$ arbeitet.

Als Inverse von \bar{D} ist \bar{B} natürlich auch symmetrisch und positiv definit und die Formel (5.21) kann als eigenständige Rekursionsformel ebenfalls eingesetzt werden. Zusammen mit der Vertauschbarkeit von p und q und der von D und D^{-1} im Ansatz der Suchrichtung entstehen mannigfaltige Möglichkeiten der Ausgestaltung des Verfahrens.

Ferner ist zu bemerken, dass die strenge Wolfe-Powell-Regel im Verfahren auch durch die normale Wolfe-Powell-Regel oder durch die Minimierungsregel ersetzt werden kann. Außerdem kann in einer anderen Variante des

DFP-Verfahrens im Sinne des Konvergenzsatzes 5.19 jederzeit immer wieder für eine endliche vorgewählte Anzahl von Schritten $k \in \mathbf{N}$ die Aufdatierung der Matrix D nach der Formel (6.1) von der Frage des Erfülltseins der Winkelbedingung abgekoppelt werden, d.h. es wird einfach für k Schritte die Aufdatierungsformel angewendet und danach (im Sinne eines Restarts) $D = I$ gesetzt.

Es sei darauf hingewiesen, dass es noch viele andere, konkurrierende Aufdatierungsformeln für D gibt (vgl. [GK1], Kapitel 11). Da ihnen allen die Idee unterliegt, mit D die (Inverse der) Hessematrix von f , sofern existent, zu approximieren und sie damit eine Approximation der Newtonrichtung darstellen wollen, nennt man die daraus entstehenden Verfahren **Quasi-Newton-Verfahren**. Sie sind in der Praxis sehr wirkungsvoll. (Bzgl. weiterer Einzelheiten zur Theorie der Quasi-Newton Verfahren siehe [GK1], Kapitel 11.)

5.2.4 Spezialfall: Quadratische Optimierung

Das *Verfahren von Fletcher-Reeves* wurde über die Minimierung einer streng konvexen quadratischen Zielfunktion motiviert, es sollte also bei einem solchen Aufgabentyp auch besonders gut funktionieren: Es ist anschaulich klar, dass bei streng konvexer quadratischer Zielfunktion f der Wert

$$t_k = \arg \min_{t \in \mathbf{R}} f(x^k + td^k)$$

die genannten Bedingungen bei geeigneter Wahl der Parameter σ und ρ erfüllt (formal gilt dies wegen $\nabla f(x^k + t_k d^k)^T d^k = 0$). Deshalb konvergiert das Fletcher-Reeves Verfahren bei einer solchen Zielfunktion und der Wahl der Schrittweite mit der Minimierungsregel gemäß Lemma 5.12 garantiert in höchstens $n - 1$ Schritten.

Das *Davidon-Fletcher-Powell Verfahren* hat zunächst mit der quadratischen Optimierung direkt nichts zu tun. Unsere Motivation bei der Konstruktion bestand nur darin, ein algorithmisch einfaches Verfahren entstehen zu lassen, das irgendwie zwischen Gradientenverfahren und Newtonverfahren anzusiedeln ist. Dass dieser Verfahrenstyp in der Praxis so erfolgreich ist, liegt nun aber doch an seiner Beziehung zur quadratischen Optimierung. Verwendet man nämlich bei streng konvexer quadratischer Zielfunktion f das DFP Verfahren mit der Minimierungsregel als Schrittweitenregel in der Variante, dass

unabhängig von der Winkelbedingung die Aufdatierungsformal jeweils mindestens $k \geq n$ mal hintereinander verwendet wird, so erweist sich das Verfahren als *Verfahren konjugierter Richtungen* und endet nach endlich vielen Schritten!

Satz 5.22 *Vorgegeben sei die streng konvexe quadratische Zielfunktion $f(x) = \frac{1}{2}x^T Qx + p^T x$ mit positiv definiten $n \times n$ Matrix Q . Es werde das DFP Verfahren in der Variante angewendet, dass die Minimierungsregel zur Schrittweiltenwahl herangezogen und die Aufdatierungsregel ohne Berücksichtigung der Winkelbedingung permanent verwendet wird. Dann sind die Suchrichtungen d^k , $k = 0, 1, 2, \dots$ alle zueinander Q -konjugiert. Ferner gilt $D_n = Q^{-1}$ und x^n ist die gesuchte Optimallösung.*

Beweis: (vgl. [BSS], Theorem 8.8.6)

Wir wollen folgende Behauptungen für $0 \leq j < n$ beweisen:

1. d_0, \dots, d_j sind linear unabhängig
2. $(d^i)^T Qd^k = 0$ für $i \neq k$, $0 \leq i, k \leq j$
3. $D_{j+1}Qp^k = p^k$ bzw. $D_{j+1}Qd^k = d^k$ für $0 \leq k \leq j$ und $p^k = \alpha_k d^k$.

Beginnen wir mit $j = 0$. Hier ist nur die dritte Behauptung zu überprüfen.

Generell gilt für jedes k

$$Qp^k = Q(\alpha d^k) = Q(x^{k+1} - x^k) = \nabla f(x^{k+1}) - \nabla f(x^k) = q^k \quad (5.22)$$

insbesondere also $Qp^0 = q^0$. Daher ergibt sich mit der Aufdatierungsformel für D

$$D_1 Qp^0 = \left(D_0 + \frac{p^0 (p^0)^T}{(p^0)^T q^0} - \frac{D_0 q^0 (q^0)^T D_0}{(q^0)^T D_0 q^0} \right) q^0 = p^0$$

und die dritte Behauptung ist für $j = 0$ bewiesen.

Nun nehmen wir an, dass die drei Behauptungen für $j \leq n-2$ gültig sind. Wir zeigen, dass dies auch für $j+1$ zutrifft. Da nun nach Induktionsannahme die Richtungen d^k , $k \leq j$, Q -konjugiert sind, können wir Lemma 5.12 anwenden und erhalten

$$(d^i)^T \nabla f(x^{j+1}) = 0$$

für alle $i \leq j$. Außerdem gilt nach der Induktionsannahme aus dem dritten Teil der Behauptung: $d^i = D_{j+1} Q d^i$ für $i \leq j$. Damit gilt auch:

$$0 = (d^i)^T \nabla f(x^{j+1}) = (d^i)^T Q D_{j+1} \nabla f(x^{j+1}) = - (d^i)^T Q d_{j+1}.$$

Damit ist die Induktionsbehauptung von Teil 2 bewiesen.

Zur Begründung von Teil 3 betrachte nun $k \leq j + 1$. Dann gilt

$$D_{j+2} Q p^k = \left(D_{j+1} + \frac{p^{j+1} (p^{j+1})^T}{(p^{j+1})^T q^{j+1}} - \frac{D_{j+1} q^{j+1} (q^{j+1})^T D_{j+1}}{(q^{j+1})^T D_{j+1} q^{j+1}} \right) Q p^k.$$

Unter Benutzung von (5.22) folgt daraus wie oben $D_{j+2} Q p^k = p^k$, speziell für $k = j + 1$

$$D_{j+2} Q p^{j+1} = p^{j+1}$$

Nun sei $k \leq j$. Dann folgt aus dem bereits bewiesenen Teil 2:

$$(p^{j+1})^T Q p^k = \alpha_k \alpha_{j+1} (d^{j+1})^T Q d^k = 0$$

und der Induktionsannahme für Teil 3 sowie (5.22)

$$(q^{j+1})^T D_{j+1} Q p^k = (q^{j+1})^T p^k = (p^{j+1})^T Q p^k = \alpha_{j+1} \alpha_k (d^{j+1})^T Q d^k = 0$$

Damit ist auch die Induktionsbehauptung für Teil 3 bewiesen.

Bleibt noch Teil 1. Nehmen wir dazu an, dass

$$\sum_{i=0}^{j+1} \beta_i d^i = 0 \quad \text{gelte für gewisse } \beta_i$$

Multipliziert man dies mit $(d^{j+1})^T Q$ so erhält man mit Teil 2 sofort

$$\beta_{j+1} (d^{j+1})^T Q d^{j+1} = 0.$$

Nun ist nach Konstruktion des Verfahrens $\nabla f(x^{j+1}) \neq 0$ und wie wir wissen d_{j+1} positiv definit, deshalb gilt auch $d^{j+1} = -D_{j+1} \nabla f(x^{j+1}) \neq 0$. Damit gilt auch $(d^{j+1})^T Q d^{j+1} \neq 0$, woraus $\beta_{j+1} = 0$ folgt.

Also gilt nun $\sum_{i=0}^j \beta_i d^i = 0$ und es folgt nach Induktionsannahme $\beta_i = 0$ für $i = 0, \dots, j$. Damit ist auch die Induktionsbehauptung für Teil 1 bewiesen.

Nun gilt nach Teil 3: $D_n Q d^k = d^k$ für $k = 0, \dots, n - 1$. Sei dann \bar{D} die aus den Vektoren d^k spaltenweise aufgebaute Matrix. Dann gilt also $D_n Q \bar{D} = \bar{D}$. Wegen der linearen Unabhängigkeit der Spalten von \bar{D} ist diese Matrix invertierbar und es folgt: $D_n Q = I$. In Sicht auf Lemma 5.12 ist damit alles bewiesen.

□

Beide Verfahren, das Verfahren von Fletcher-Reeves und das Verfahren von Davidon-Fletcher-Powell treten bei quadratischer Zielfunktion in unmittelbarer Konkurrenz zum Verfahren von Beale. Alle drei verwenden sie konjugierte Richtungen als Suchrichtungen. Bei positiv definiten Matrix Q haben sie ähnlichen Verlauf, wenn gleich sie sehr unterschiedlich vorgehen. Die Verfahren FR und DFP zeigen, dass das Prinzip der konjugierten Richtungen auf unterschiedliche Weise auf den Allgemeinfall der unrestrictierten Optimierung übertragen werden kann.

5.2.5 Übungsaufgaben

Aufgabe 1

Programmieren und testen Sie je eine Version des CG Verfahrens zur Lösung von $Qx = -p$ mit und ohne Präkonditionierung.

Aufgabe 2

Testen Sie Matlab Versionen des FR und des DFP Verfahrens auf quadratischen und beliebigen nichtlinearen Funktionen. Vergleichen sie die Ergebnisse mit denen der Simplexverfahren.

5.3 Anwendung im allgemeinen Fall

Ist die Minimierung einer gegebenen Zielfunktion über einem restringierten Bereich vorzunehmen, so können die bisher angeführten Verfahren der unrestringierten Optimierung nicht direkt angewendet werden. Gelingt es allerdings, die Beschränkung des zulässigen Bereichs durch die Restriktionen über eine "Verbiegung" der Zielfunktion weitgehend zu ersetzen, so kann die Bearbeitung restringierter Aufgaben auf die Lösung unrestringierter Aufgabe zurückgeführt werden.

Grundsätzliche Ideen dazu sind

- die Zielfunktion im Inneren des zulässigen Bereichs an den Rändern hochzubiegen, so dass Suchrichtungen automatisch vom Rand wegzeigen. (*Barriere-Ansatz*)

Skizze:

- die Zielfunktion unmittelbar ausserhalb des Randes hochzubiegen, um so Suchrichtungen immer in den zulässigen Bereich hinein zu orientieren. (*Penalty-Ansatz*)

Skizze:

Natürlich müssen diese Ansätze iteriert angewendet werden, damit im Grenzprozess die Restringierung exakt nachgebildet wird.

Wir behandeln im Folgenden beide Ansätze etwas allgemeiner als für unsere vorgegebene Aufgabenstellung notwendig, um die Ähnlichkeit der theoretischen Behandlung herauszustellen.

5.3.1 Barriere Ansatz

Vorgeben sei eine Aufgabenstellung von folgenden Typ

$$(UP) \quad \min \{f(x) : g(x) \leq 0, x \in X\}$$

mit stetiger Funktion $g : \mathbf{R}^n \rightarrow \mathbf{R}^m$ und nichtleerer Teilmenge $X \subseteq \mathbf{R}^n$. Eventuell auftretende Gleichungsrestriktionen seien also in der Forderung $x \in X$ enthalten. Der Aufgabe (UP) ordnen wir eine zugehörige **Ersatzaufgabe** zu

$$(BUP) \quad \inf \{\theta(\mu) : \mu > 0\}$$

wobei $\theta(\mu) := \inf \{f(x) + \mu B(x) : g(x) < 0, x \in X\}$ gesetzt wurde mit einer nichtnegativen und stetigen **Barrierefunktion** $B : U \rightarrow \mathbf{R}$, $U := \{x \in \mathbf{R}^n : g(x) < 0\}$. Dabei ist die Vorstellung, dass sich die Barrierefunktion am Rand von U dem Unendlichen annähert, dass also $B(x)$ umso größer ist, je kleiner $\|g(x)\|_\infty$ ist.

Man beachte: Die Bedingungen $g(x) < 0$ bei der Berechnung von $\theta(\mu)$ haben dabei eher formalen Charakter. In der Praxis kann man sie weglassen, wenn das angewendete Verfahren zur Berechnung von $\theta(\mu)$ die Bedingung automatisch sicherstellt. Dies ist z. B. so, wenn $X = \mathbf{R}^n$ gilt und ein zulässiges Abstiegsverfahren der unrestringierten Optimierung angewendet wird.

Formal fordert man von der Barrierefunktion, dass sie von der Form ist

$$B(x) := \sum_{i=1}^m \phi(g_i(x))$$

mit einer stetigen Hilfsfunktion $\phi : \mathbf{R}_{--} \rightarrow \mathbf{R}$ mit den Eigenschaften

$$\begin{aligned} \phi(y) &\geq 0 \quad \text{für } y < 0 \\ \lim_{y \rightarrow 0^-} \phi(y) &= \infty \end{aligned}$$

Klassische Barrierefunktion sind

$$B(x) := - \sum_{i=1}^m \ln [\min \{1, -g_i(x)\}] \quad \text{oder} \quad B(x) := \sum_{i=1}^m \frac{-1}{g_i(x)}$$

Beispiel 5.23 Als Beispiel betrachten wir die Aufgabe (UP1) $\min \{(x - 2)^2 : x \leq 1\}$.
 Zu dieser Aufgabe lautet bei $X = \mathbf{R}$ das Barriereproblem (BUP1) $\inf \{\theta(\mu) : \mu > 0\}$
 mit

$$\theta(\mu) = \inf \{(x - 2)^2 - \mu \ln(1 - x) : x < 1\}$$

Skizze:

Das Beispiel signalisiert, dass mit abfallendem μ das Minimum der Ersatzfunktion und die Minimalstelle sich immer mehr den entsprechenden Daten der gegebenen Aufgabe annähert. Dies wollen wir theoretisch bestätigen.

Wir beginnen mit einem technischen

Lemma 5.24 Es seien f, g_1, \dots, g_m stetige Funktionen auf \mathbf{R}^n und $X \subseteq \mathbf{R}^n$ sei abgeschlossen und nichtleer. Die Menge $\{x \in X : g(x) < 0\}$ sei nichtleer. Ferner besitze für jedes μ jede Folge $\{x_k\}$ von Punkten in X mit $g(x_k) < 0$, für die $f(x_k) + \mu B(x_k)$ gegen $\phi(\mu)$ konvergiert, eine konvergente Teilfolge (das gilt z.B. wenn X kompakt ist). Dann gilt:

1. Für jedes $\mu > 0$ gibt es ein $x_\mu \in X$ mit $g(x_\mu) < 0$ so, dass

$$\theta(\mu) = f(x_\mu) + \mu B(x_\mu) = \inf \{f(x) + \mu B(x) : g(x) < 0, x \in X\}$$

2. $\inf \{f(x) : g(x) \leq 0, x \in X\} \leq \inf \{\theta(\mu) : \mu > 0\}$

3. Für $\mu > 0$ sind die Funktionen $f(x_\mu)$ und $\theta(\mu)$ nichtabfallende und $B(x_\mu)$ nichtansteigende Funktionen.

Beweis: (vgl. [BSS], Lemma 9.4.2)

1. Es sei $\mu > 0$ vorgegeben. Nach der Definition der Funktion θ gibt es eine Folge $\{x_k\}$ von Elementen aus X , die $g(x_k) < 0$ erfüllen mit der Eigenschaft, dass $f(x_k) + \mu B(x_k) \rightarrow \theta(\mu)$. Nach Voraussetzung hat diese Folge eine konvergente Teilfolge $\{x_k\}_{k \in K}$ mit Grenzwert $x_\mu \in X$. Da g stetig ist, gilt auch $g(x_\mu) \leq 0$. Wir zeigen, dass sogar $g(x_\mu) < 0$ gilt. Wäre dem nicht so, dann wäre $g_i(x_\mu) = 0$ für ein i und nach Definition der Barrierefunktion B hätte man $B(x_k) \rightarrow \infty$ für $k \in K$. Damit wäre aber $\theta(\mu) = \infty$, was nicht sein kann, da $\{x \in X : g(x) < 0\}$ als nichtleer angenommen wurde. Also gilt $\theta(\mu) = f(x_\mu) + \mu B(x_\mu)$ mit $x_\mu \in X$ und $g(x_\mu) < 0$.

2. Da $B(x) \geq 0$ für x mit $g(x) < 0$ gilt, hat man für vorgegebenes $\mu \geq 0$

$$\begin{aligned} \theta(\mu) & : = \inf \{f(x) + \mu B(x) : g(x) < 0, x \in X\} \\ & \geq \inf \{f(x) : g(x) < 0, x \in X\} \\ & \geq \inf \{f(x) : g(x) \leq 0, x \in X\} \end{aligned}$$

3. Seien nun $\mu > \lambda > 0$ vorgegeben. Da $B(x) \geq 0$ für x mit $g(x) < 0$ gilt, folgt $f(x) + \mu B(x) \geq f(x) + \lambda B(x)$ für jedes x mit $g(x) < 0$. Also gilt $\theta(\mu) \geq \theta(\lambda)$. Nach Teil 1 gibt es x_μ und x_λ mit

$$\begin{aligned} f(x_\mu) + \mu B(x_\mu) & \leq f(x_\lambda) + \mu B(x_\lambda) \\ f(x_\lambda) + \lambda B(x_\lambda) & \leq f(x_\mu) + \lambda B(x_\mu) \end{aligned} \tag{5.23}$$

Durch Addition folgt dann $(\mu - \lambda)[B(x_\mu) - B(x_\lambda)] \leq 0$ und wegen $(\mu - \lambda) > 0$ auch $B(x_\mu) \leq B(x_\lambda)$. Setzt man dies wiederum in (5.23) ein, erhält man $f(x_\lambda) \leq f(x_\mu)$ und $\theta(\lambda) \leq \theta(\mu)$.

□

Mithilfe des Lemmas können wir einen allgemeinen Konvergenzsatz beweisen.

Satz 5.25 *f, g_1, \dots, g_m seien stetige Funktionen auf \mathbf{R}^n und $X \subseteq \mathbf{R}^n$ sei abgeschlossen und nichtleer. Die Menge $\{x \in X : g(x) < 0\}$ sei nichtleer. Außerdem besitze das Problem (UP) eine Optimallösung \bar{x} mit folgender Eigenschaft:*

(*) Zu jeder Umgebung N um \bar{x} gebe es ein $x \in X \cap N$ mit $g(x) < 0$.

Dann gilt

$$\min \{f(x) : g(x) \leq 0, x \in X\} = \lim_{\mu \rightarrow 0^+} \theta(\mu) = \inf_{\mu > 0} \theta(\mu)$$

Ferner gilt: Ist eine Nullfolge $\{\mu_k\}_{k \in \mathbf{N}}$ mit $\mu_k > 0$ gegeben und zu μ_k ein $x_{\mu_k} \in X$ mit $g(x_{\mu_k}) < 0$ und $\theta(\mu_k) = f(x_{\mu_k}) + \mu_k B(x_{\mu_k})$, so ist jeder Häufungspunkt der Folge $\{x_{\mu_k}\}$ eine Optimallösung von (UP) . Ferner gilt dann $\mu_k B(x_{\mu_k}) \rightarrow 0$.

Beweis: (vgl. [BSS], Theorem 9.4.3)

(a) Sei \bar{x} eine Optimallösung von (UP) mit der Eigenschaft (*) und sei $\varepsilon > 0$. Dann gibt es wegen der Stetigkeit von f ein $\hat{x} \in X$ mit $g(\hat{x}) < 0$ und $f(\bar{x}) + \varepsilon > f(\hat{x})$. Damit gilt für $\mu > 0$

$$f(\bar{x}) + \varepsilon + \mu B(\hat{x}) > f(\hat{x}) + \mu B(\hat{x}) \geq \theta(\mu)$$

und im Grenzprozess $\mu \rightarrow 0^+$ dann $f(\bar{x}) + \varepsilon \geq \lim_{\mu \rightarrow 0^+} \theta(\mu)$. Da dies für beliebiges $\varepsilon > 0$ gilt, folgt

$$f(\bar{x}) \geq \lim_{\mu \rightarrow 0^+} \theta(\mu)$$

Zusammen mit dem zweiten Teil des vorangegangenen Lemmas gilt also $f(\bar{x}) = \lim_{\mu \rightarrow 0^+} \theta(\mu)$.

(b) Sei nun eine Nullfolge $\{\mu_k\}_{k \in \mathbf{N}}$ mit $\mu_k > 0$ gegeben. Es sei zu μ_k ein $x_{\mu_k} \in X$ mit $g(x_{\mu_k}) < 0$ und $\theta(\mu_k) = f(x_{\mu_k}) + \mu_k B(x_{\mu_k})$ gemäß dem ersten Teil des vorangegangenen Lemmas ausgewählt. Dann gilt

$$\theta(\mu_k) = f(x_{\mu_k}) + \mu_k B(x_{\mu_k}) \geq f(x_{\mu_k}) \geq f(\bar{x})$$

und im Grenzprozess $\mu_k \rightarrow 0^+$ wegen (a) $f(x_{\mu_k}) + \mu_k B(x_{\mu_k}) \rightarrow f(\bar{x})$ und $f(x_{\mu_k}) \rightarrow f(\bar{x})$ sowie $\mu_k B(x_{\mu_k}) \rightarrow 0$. Hat ferner die Folge $\{x_{\mu_k}\}$ eine konvergente Teilfolge mit Grenzwert x' , dann gilt auch $f(x') = f(\bar{x})$. Da die x_{μ_k} alle zulässig zu (UP) sind, gilt dies auch für x' , daher ist auch x' eine Optimallösung von (UP) .

□

Der Konvergenzsatz sichert bei kompaktem X die Konvergenz des folgenden Verfahrens in dem Sinne, dass jeder Häufungspunkt der Iteriertenfolge eine Optimallösung von (UP) ist, sofern Bedingung (*) erfüllt ist.

Barriereverfahren

1. **Start:** Wähle $x^0 \in X$ mit $g(x^0) < 0$ sowie $\varepsilon, \mu_0 > 0$ und ein $\beta \in (0, 1)$. Setze $k = 0$.
2. **Barriereproblem:** Bestimme ausgehend von x^k eine Optimallösung $x_{\mu_k} = x^{k+1}$ des **Barriereproblems**
$$(BUP_{\mu_k}) \quad \min \{f(x) + \mu_k B(x) : g(x) < 0, x \in X\}$$
3. **Abbruch:** Ist $\mu_k B(x^{k+1}) \leq \varepsilon$, Stopp, (eine Näherungslösung von (UP) ist erreicht)
4. **Iteration:** Setze $\mu_{k+1} := \beta \mu_k$, $k := k + 1$ und gehe zu (1).

Um die Anwendbarkeit des Barriere-Prinzips auch ausserhalb der linear restringierten Optimierung aufzuzeigen, betrachten wir folgendes

Beispiel 5.26 (vgl. [BSS], Example 9.4.4) *Betrachtet werde die Aufgabe*

$$\min \{(x_1 - 2)^4 + (x_1 - 2x_2)^2 : x_1^2 - x_2 \leq 0\}$$

Hierbei ist $X = \mathbf{R}^2$ angenommen. Wir wenden das Barriereverfahren mit der Barrierefunktion

$$B(x) = -\frac{1}{x_1^2 - x_2}$$

an mit den Startparametern

$$\mu_0 = 10, \beta = 0.1, x^0 = (0.0, 1.0)^T, \varepsilon = 0.02$$

Nach 6 Schritten ist der Punkt $x^6 = (0.94389, 0.89635)^T$ erreicht mit $\mu_5 B(x^6) = 0.0184 < \varepsilon$, wodurch das Verfahren abbricht.

k	μ_{k-1}	(x^k)	$f(x^k)$	$\mu_{k-1}B(x^k)$
1	10	$\begin{pmatrix} 0.708 \\ 1.531 \end{pmatrix}$	8.3338	9.705
2	1	$\begin{pmatrix} 0.828 \\ 1.11 \end{pmatrix}$	3.8214	2.3591
3	0.1	$\begin{pmatrix} 0.8989 \\ 0.9638 \end{pmatrix}$	2.5282	0.6419
4	0.01	$\begin{pmatrix} 0.9294 \\ 0.9162 \end{pmatrix}$	2.1291	0.1908
5	0.001	$\begin{pmatrix} 0,9403 \\ 0.9011 \end{pmatrix}$	2.0039	0.0590
6	0.0001	$\begin{pmatrix} 0.94389 \\ 0.89635 \end{pmatrix}$	1.9645	0.0184

Der Verlauf des Verfahrens kann auch aus folgender **Skizze** abgelesen werden:

Es sei darauf hingewiesen, dass das Verfahren, je nach den Daten der Aufgabenstellung, unter *algorithmischen Schwierigkeiten* leidet. Es ist nicht immer einfach, einen geeigneten Startpunkt zu finden. Ferner leidet die Rechnung unter Ungenauigkeitsproblemen, je näher man dem Rand kommt. Eine zusätzliche Schwierigkeit besteht darin, dass im Verfahren globale Minima der unrestringierten Unterprobleme gefunden werden müssen. Dass dies gelingt, ist in der Regel nicht gesichert.

5.3.2 Penalty Ansatz

Die Entwicklung eines Penaltyverfahrens verläuft analog zu der des Barriereverfahrens. Vorausgesetzt wird eine Aufgabe der Form

$$(NP) \quad \min \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$$

mit stetigen Funktionen $g : \mathbf{R}^n \rightarrow \mathbf{R}^m$, $h : \mathbf{R}^n \rightarrow \mathbf{R}^p$, $f : \mathbf{R}^n \rightarrow \mathbf{R}$ und einer nichtleeren Teilmenge $X \subseteq \mathbf{R}^n$. Es soll nun die Abweichung von der Zulässigkeit bestraft werden, was im Gegensatz zum Barriereansatz zu einer "Verbiegung der Zielfunktion" außerhalb des zulässigen Bereichs führt. Dazu setzt man folgenden **Penaltyterm** fest:

$$P(x) = \sum_{i=1}^m \phi[g_i(x)] + \sum_{i=1}^m \psi[h_i(x)]$$

mit reelwertigen Funktionen ϕ und ψ , die folgende Eigenschaften haben

$$\phi(y) \begin{cases} = 0, & \text{falls } y \leq 0 \\ > 0, & \text{falls } y > 0 \end{cases}$$

$$\psi(y) \begin{cases} = 0, & \text{falls } y = 0 \\ > 0, & \text{falls } y \neq 0 \end{cases}$$

und betrachtet ersatzweise anstelle (NP) für verschiedene $\mu > 0$ das **Penaltyproblem**

$$(PP\mu) \quad \min \{f(x) + \mu P(x) : x \in X\}$$

Bei $X = \mathbf{R}^n$ wird diese Aufgabe also mit einem Verfahren der unrestringierten Optimierung gelöst.

Man verwendet in der Regel die konkrete **Penaltyfunktion**

$$P(x) = \sum_{i=1}^m [\max\{0, g_i(x)\}]^p + \sum_{i=1}^m |h_i(x)|^p$$

mit einem vorgewählten positiven Parameter p .

Beispiel 5.27 Betrachten wir das einfache Problem $\min \{x : x \geq 2\}$. Interpretieren wir dabei $X = \mathbf{R}$, so lautet das Penaltyproblem mit $p = 2$

$$\min \{x + \mu P(x) : x \in \mathbf{R}\}$$

mit

$$P(x) = \begin{cases} 0, & \text{falls } x \geq 2 \\ (-x + 2)^p, & \text{falls } x < 2 \end{cases}$$

Skizze:

Das Beispiel legt wieder nahe, dass eine Konvergenz der Minimierer der Penaltyproblem bei ansteigendem μ gegen den Minimierer der gegebenen Aufgabe zu erwarten ist. Wir klären dies theoretisch. Zunächst sei wieder ein technisches Lemma benannt.

Lemma 5.28 f, g_1, \dots, g_m seien stetige Funktionen auf \mathbf{R}^n und $X \subseteq \mathbf{R}^n$ sei nichtleer. Sei P eine stetige Penaltyfunktion und es werde unterstellt, dass für jedes $\mu > 0$ ein $x_\mu \in X$ existiert mit $\theta(\mu) = f(x_\mu) + \mu P(x_\mu)$. Dann gilt:

1. $\inf \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\} \geq \sup \{\theta(\mu) : \mu \geq 0\}$
2. Für $\mu \geq 0$ sind die Funktionen $f(x_\mu)$ und $\theta(\mu)$ nichtabfallende und $P(x_\mu)$ nichtansteigende Funktionen.

Beweis: [BSS], Lemma 9.2.1 ■

Mithilfe des Lemmas können wir den folgenden Konvergenzsatz beweisen.

Satz 5.29 *Unterstellt werde, dass das Problem (NP) eine zulässige Lösung besitze. Betrachtet wird eine gegen unendlich strebende Folge $\{\mu_k > 0\}$ und es werde unterstellt, dass für jedes μ_k das zugehörige Penaltyproblem $(PP\mu_k)$ eine Optimallösung $x_{\mu_k} \in X$ besitze und dass die Folge $\{x_{\mu_k}\}$ in einer kompakten Teilmenge von X enthalten sei. Dann gilt:*

$$\min \{f(x) + \mu P(x) : g(x) \leq 0, h(x) = 0, x \in X\} = \lim_{\mu \rightarrow \infty} \theta(\mu) = \sup_{\mu \geq 0} \theta(\mu)$$

wobei $\theta(\mu) = f(x_\mu) + \mu P(x_\mu)$ gilt. Ferner ist jeder Häufungspunkt der Folge $\{x_{\mu_k}\}$ eine Optimallösung zu (NP) und es konvergiert dabei $\mu_k P(x_{\mu_k})$ gegen null für $k \rightarrow \infty$.

Beweis: (vgl. [BSS], Theorem 9.2.2) Nach dem zweiten Teil des Lemmas ist die Funktion θ monoton, deshalb ist

$$\sup_{\mu \geq 0} \theta(\mu) = \lim_{\mu \rightarrow \infty} \theta(\mu)$$

Wir zeigen zunächst, dass $P(x_{\mu_k}) \rightarrow 0$ für $k \rightarrow \infty$. Sei dazu y eine zulässige Lösung von (NP) und $\varepsilon > 0$. Sei x_1 eine optimale Lösung des Problems, $f(x) + \mu P(x)$ mit $\mu = 1$ über X zu minimieren. Sei andererseits $\mu_k \geq \frac{1}{\varepsilon} |f(y) - f(x_1)| + 2 > 1$. Dann gilt nach dem zweiten Teil des Lemmas wegen der Monotonie von f die Ungleichung $f(x_{\mu_k}) \geq f(x_1)$.

Nun sei $P(x_\mu) > \varepsilon$ angenommen. Dann gilt nach Teil 1 des Lemmas:

$$\begin{aligned} & \inf \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\} \\ & \geq \theta(\mu_k) = f(x_{\mu_k}) + \mu_k P(x_{\mu_k}) \geq f(x_1) + \mu_k P(x_{\mu_k}) \\ & > f(x_1) + |f(y) - f(x_1)| + 2\varepsilon \geq f(y) + 2\varepsilon > f(y) \end{aligned}$$

Dies aber widerspricht der vorausgesetzten Zulässigkeit von y .

Also gilt $P(x_{\mu_k}) \leq \varepsilon$ für alle $\mu_k \geq \frac{1}{\varepsilon} |f(y) - f(x_1)| + 2$. Weil $\varepsilon > 0$ beliebig war, folgt, dass $P(x_{\mu_k}) \geq 0$ der Null beliebig nahe kommt, sofern nur μ_k groß genug wird. Also gilt $P(x_{\mu_k}) \rightarrow 0$ für $k \rightarrow \infty$.

Nun sei $\{x_{\mu_k}\}_{k \in K}$ eine konvergente Teilfolge der Folge $\{x_{\mu_k}\}_{k \in \mathbf{N}}$ und \bar{x} ihr Grenzwert. Dann gilt

$$\sup_{\mu \geq 0} \theta(\mu) \geq \theta(\mu_k) = f(x_{\mu_k}) + \mu P(x_{\mu_k}) \geq f(x_{\mu_k}).$$

Da $x_{\mu_k} \rightarrow \bar{x}$ für $k \in K$ folgt

$$\sup_{\mu \geq 0} \theta(\mu) \geq f(\bar{x})$$

Ferner folgt aus der Stetigkeit von P , dass $P(\bar{x}) = 0$. Also gilt $g(\bar{x}) \leq 0$ und $h(\bar{x}) = 0$, ferner nach Konstruktion $\bar{x} \in X$, somit ist \bar{x} zulässig zu (NP) . Gemäß Teil 1 des Lemmas gilt nun:

$$f(\bar{x}) = \inf \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\} = \sup \{\theta(\mu) : \mu \geq 0\}$$

und \bar{x} ist Optimallösung von (NP) . Bleibt die Beobachtung: $\mu P(x_{\mu_k}) = \theta(\mu_k) - f(x_{\mu_k})$. Im Grenzprozess $k \rightarrow \infty$, $k \in K$ folgt nun $\mu P(\bar{x}) = f(\bar{x}) - f(\bar{x}) = 0$, also strebt $\mu P(x_{\mu_k})$ gegen null für $k \rightarrow \infty$, $k \in K$.

□

Korollar 5.30 *Ist $P(x_{\mu_k}) = 0$ für ein k , so ist x_{μ_k} Optimallösung von (NP) .*

Beweis: Wie im Satz folgt, dass x_{μ_k} zulässig zu (NP) ist. Wegen

$$\begin{aligned} f(x_{\mu_k}) &\geq \inf \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\} \\ &\geq \theta(\mu_k) = f(x_{\mu_k}) + \mu P(x_{\mu_k}) \geq f(x_{\mu_k}) \end{aligned}$$

folgt auch die Optimalität.

□

Der Konvergenzsatz sichert bei kompaktem X die Konvergenz des folgenden Verfahrens in dem Sinne, dass jeder Häufungspunkt der Iteriertenfolge eine Optimallösung von (UP) ist.

Penaltyverfahren

1. **Start:** Wähle $x^0 \in X$ sowie $\varepsilon, \mu_0 > 0$ und ein $\beta > 1$. Setze $k = 0$.

2. **Penaltyproblem:** Bestimme ausgehend von x^k eine Optimallösung $x_{\mu_k} = x^{k+1}$ des Penaltyproblems

$$(PNP\mu) \quad \min \{f(x) + \mu_k P(x) : x \in X\}$$

3. **Abbruch:** Ist $\mu_k \alpha(x^{k+1}) \leq \varepsilon$, **Stopp**, (eine Näherungslösung von (NP) ist erreicht)

4. **Iteration:** Setze $\mu_{k+1} := \beta \mu_k$, $k := k + 1$ und gehe zu (1).

Wir betrachten ein zum letzten Beispiel analoges

Beispiel 5.31 (vgl. [BSS], Example 9.2.4) Betrachtet werde die Aufgabe

$$\min \{(x_1 - 2)^4 + (x_1 - 2x_2)^2 : x_1^2 - x_2 = 0\}$$

Hierbei ist $X = \mathbf{R}^2$ angenommen. Wir wenden das Penaltyverfahren mit der Penaltyfunktion

$$P(x) = (x_1^2 - x_2)^2$$

an mit den Startparametern

$$\mu_0 = 0.1, \beta = 10, x^0 = (2.0, 1.0)^T, \varepsilon = 0.02$$

Nach 5 Schritten ist der Punkt $x^6 = (0.94611, 0.89344)^T$ erreicht mit $\mu_4 B(x^5) = 0.0028 < \varepsilon$, wodurch das Verfahren abbricht.

k	μ_{k-1}	(x^k)	$f(x^k)$	$\mu_{k-1} B(x^k)$
1	0.1	$\begin{pmatrix} 1.4539 \\ 0.7608 \end{pmatrix}$	0.0935	0.1831
2	1	$\begin{pmatrix} 1.1687 \\ 0.7407 \end{pmatrix}$	0.5753	0.3908
3	10	$\begin{pmatrix} 0.9906 \\ 0.8425 \end{pmatrix}$	1.5203	0.1926
4	100	$\begin{pmatrix} 0.9507 \\ 0.8875 \end{pmatrix}$	1.8917	0.0267
5	1000	$\begin{pmatrix} 0.94611 \\ 0.89344 \end{pmatrix}$	1.9405	0.0028

Zum Abschluss dieses Abschnitts sei bemerkt, dass ähnliche rechnerischen Schwierigkeiten wie beim Barriereansatz auch beim Penaltyansatz auftreten.

5.3.3 Spezialfall: Quadratische Optimierung

Eine günstige Voraussetzung für die Anwendung von Barriere- oder Penalty-Methoden auf die quadratische Optimierungsaufgabe ist das Vorliegen in der Standardform

$$\min \left\{ \frac{1}{2} x^T Q x + p^T x : Ax \leq a \right\}.$$

Hierbei seien wieder mögliche Vorzeichenbeschränkungen in den Restriktionen integriert..

Barriere-Methoden

Gerne verwendet man die "vereinfachte Barrierefunktion"

$$B(x) := - \sum_{i=1}^m \ln [-g_i(x)] = - \sum_{i=1}^m \ln [A_i \cdot x - a_i]$$

Zu lösen ist dann in jeder Iteration das unrestringierte Barriereproblem bei vorgewähltem $\mu > 0$

$$\min \left\{ \frac{1}{2} x^T Q x + p^T x - \mu \sum_{i=1}^m \ln [A_i \cdot x - a_i] \right\}$$

Aber auch die Übertragung von Verfahren der unrestringierten Optimierung auf den Fall affin-linearer Gleichungsrestriktionen ist möglich und sinnvoll. Wir werden im nächsten Abschnitt darauf zurückkommen.

Zu beachten ist, dass die speziellen Vorgaben des Lemmas und des Konvergenzsatzes beim Barriereansatz erfüllt sind, sobald es einen **inneren zulässigen Punkt** x zur Aufgabe gibt, der also $x > 0$ erfüllt. Das Verfahren arbeitet dann als **Innere-Punkte Verfahren**, da es sich nur im Inneren des zulässigen Bereichs bewegt und die Optimallösung von innen annähert.

Penalty-Methoden

Als Penaltyfunktion verwendet man aus Differenzierbarkeitsgründen gerne

$$P(x) = \sum_{i=1}^m [\max \{0, g_i(x)\}]^p + \sum_{i=1}^m |h_i(x)|^p = \sum_{i=1}^m (-x_i)_+^2 + \sum_{i=1}^m (Hx - h)^2$$

mit $(-x_i)_+ := \max\{0, -x_i\}$. Zu lösen ist also jeweils das unrestringierte Penaltyproblem bei vorgewähltem $\mu > 0$

$$\min \left\{ \frac{1}{2} x^T Q x + p^T x + \mu \left(\sum_{i=1}^m (-x_i)_+^2 + \sum_{i=1}^m (Hx - h)^2 \right) \right\}$$

5.3.4 Übungsaufgaben

Aufgabe 1

Lösen Sie Aufgaben der linear restringierten Optimierung in Standardform über den Barriere Ansatz unter Rückgriff auf die Verfahren FR und DFP.

Aufgabe 2

Verfahren Sie ähnlich wie bei Aufgabe 1 mit dem Penalty Ansatz.

Kapitel 6

Linear gleichungsrestringierte Optimierung

Die nächst komplexere Situation als die der unrestringierten Optimierung bildet die Aufgabenstellung der linear gleichungsrestringierten Optimierung. Hier ist der zulässige Bereich im linearen Fall ein affiner Unterraum und es sollten sich bei dieser "schwachen" Restringierung die Methoden der unrestringierten Optimierung auf diesen eingeschränkten Geltungsbereich übertragen lassen.

Unser Ziel ist also, die Aufgabe

$$(NLG) \quad \begin{cases} \text{minimiere} & f(x) \\ \text{s.d.} & Hx = h \end{cases}$$

bei $(m \times n)$ -Matrix H , die *maximalen Zeilenrang* habe, und stetig differenzierbarer Funktion f mit geeignet angepassten Methoden der unrestringierten Optimierung zu lösen.

6.1 Reduzierte Verfahren

Wir untersuchen zunächst allgemein, wie Verfahren der unrestringierten Optimierung auf die neue Situation übertragen werden können und besprechen dann die Übertragung konkreter Verfahren.

6.1.1 Verwendung von Nullraum-Matrizen

Kennt man eine zulässige Lösung x^0 des affinen Raums V , der durch $Hx = h$ festgelegt ist, so lässt sich V darstellen als $V = x^0 + \ker(H)$. Ist Z irgendeine $(n \times l)$ Matrix, deren Spalten $\ker(H)$ aufspannen, so lässt sich offenbar jedes $y \in \ker(H)$ darstellen als $y = Zz$ für ein $z \in \mathbf{R}^l$. Die gegebene Aufgabe schreibt sich daher als

$$(NLG) \quad \begin{cases} \text{minimiere} & f(x) \\ \text{s.d.} & x = x^0 + Zz, \quad z \in \mathbf{R}^l \end{cases}$$

bzw.

$$(NLR) \quad \begin{cases} \text{minimiere} & F(z) := f(x^0 + Zz) \\ \text{s.d.} & z \in \mathbf{R}^l \end{cases}$$

und somit als unrestringierte Optimierungsaufgabe. Besonders interessant ist diese Umformung, wenn $l < n$ gilt, da dann die Anzahl der Variablen reduziert ist. Wir nennen Z eine **Nullraum-Matrix von H** .

Wie kann man geeignete Nullraum-Matrizen bestimmen?

- *Eine erste Möglichkeit* zur Konstruktion einer Nullraum-Matrix kennen wir bereits aus dem Simplexverfahren. Sei etwa $H = (B, N)$ eine Zerlegung der Matrix H in eine invertierbare $(m \times m)$ (Basis-)Matrix B und eine (Nichtbasis-)Matrix N , so lässt sich ein beliebiges $x \in \ker(H)$ nach Zerlegung $x^T =: (x_B^T, x_N^T)$ beschreiben durch

$$x = \begin{pmatrix} -B^{-1}Nx_N \\ x_N \end{pmatrix}.$$

Setzt man also

$$Z := \begin{pmatrix} -B^{-1}N \\ I_{n-m} \end{pmatrix} \quad \text{und} \quad z := x_N$$

so ist damit eine geeignete Nullraum-Matrix festgelegt. Da $z \in \mathbf{R}^{n-m}$ gilt, ist die Anzahl der unabhängigen Variablen von n auf $n - m$ reduziert.

- *Eine zweite Möglichkeit* zur Konstruktion einer Nullraum-Matrix ergibt sich aus folgender Überlegung: Es ist $\mathbf{R}^n = \mathcal{N}(H) \oplus \mathcal{R}(H^T)$

eine direkte Summe, wobei $\mathcal{N}(H) = \ker(H)$ der Nullraum von H und $\mathcal{R}(H^T) = \{x \in \mathbf{R}^n : x = H^T y \text{ für ein } y \in \mathbf{R}^p\}$ der Bildraum von H^T ist. Dies ist leicht einzusehen, da die Vektoren aus $\ker(H)$ definitionsgemäß auf allen Zeilen von H senkrecht stehen, d.h. auf den Spalten von H^T und damit auf $\mathcal{R}(H^T)$. Aus der klassischen Dimensionsformel der linearen Algebra folgt die Behauptung.

Wir bezeichnen dann mit P_H die Projektionsmatrix, die \mathbf{R}^n orthogonal auf $\mathcal{N}(H)$ projiziert. Es ist also $P_H(x) = x$ für $x \in \mathcal{N}(H)$ und $P_H(x) = 0$ für $x \in \mathcal{N}(H)^\perp = \mathcal{R}(H^T)$. Man findet: Es ist

$$P_H = I - H^T (HH^T)^{-1} H,$$

die gewünschte Projektionsmatrix, denn es gilt $P_H = P_H^T$ und $P_H P_H = P_H$ und es folgt aus

$$P_H x = \left(I - H^T (HH^T)^{-1} H \right) x = x - H^T (HH^T)^{-1} Hx.$$

sofort $P_H x = x$, wenn $x \in \ker(H)$ sowie $P_H x = 0$, wenn $x = H^T y$ für ein y , also $x \in \mathcal{R}(H^T)$.

Die positiv semidefinite $(n \times n)$ Matrix P_H ist daher in obigem Sinne eine Nullraum-Matrix. Leider reduziert sie die Anzahl der Variablen nicht!

Man **beachte**: Es gilt $HP_H = 0$ und $P_H H^T = 0$.

- *Eine dritte Möglichkeit* ergibt sich aus der bekannten HOUSEHOLDERSchen QR Zerlegung von H^T : Man berechnet dabei eine unitäre $(n \times n)$ Matrix U und eine obere invertierbare $(m \times m)$ Dreiecksmatrix R mit

$$UH^T = \begin{pmatrix} R \\ 0 \end{pmatrix}$$

und schreibt dann

$$U = \begin{pmatrix} Y^T \\ Z^T \end{pmatrix}$$

wobei $Y^T \in \mathbf{R}^{m \times n}$ die ersten m Zeilen zusammenfasst und $Z^T \in \mathbf{R}^{(n-m) \times n}$ die restlichen $(n-m)$ Zeilen. Da U invertierbar ist, schreibt sich jedes $x \in \mathbf{R}^n$ eindeutig als

$$x = U^T \begin{pmatrix} x_Y \\ x_Z \end{pmatrix} = (Y, Z) \begin{pmatrix} x_Y \\ x_Z \end{pmatrix} = Yx_Y + Zx_Z,$$

Insbesondere gilt für $x \in \ker(H)$

$$0 = Hx = HU^T \begin{pmatrix} x_Y \\ x_Z \end{pmatrix} = (R^T, 0) \begin{pmatrix} x_Y \\ x_Z \end{pmatrix} = R^T x_Y$$

Da R invertierbar ist, gilt:

$$x \in \ker(H) \iff x_Y = 0 \iff x = Zx_Z.$$

Z ist also eine Nullraum-Matrix.

Bemerkung: Ohne Beweis sei benannt, dass für das Z aus der QR Zerlegung gilt: $P_H = ZZ^T$. Ferner gilt $Z^T Z = I_l$.

6.1.2 Konstruktion von Verfahren

Kehren wir zurück zu unserem Wunsch, die Aufgabe (NLG) über die Ersatz-aufgabe (NLR) zu lösen. Löst man die Aufgabe (NLR) mit einem beliebigen Abstiegsverfahren, so erhält man eine (unendliche) Folge von Iterationspunkten z^k . Die Folge $x^k = x^0 + Zz^k$ gibt dann die zugehörige Folge der Originalvariablen an, $f(x^k) = f(x^0 + Zz^k) = F(z^k)$ die zu beiden gehörige Folge von Funktionswerten. Ferner folgt nach der Kettenregel, dass

$$\nabla F(z^k) = \nabla_z f(x^0 + Zz^k) = Z^T \nabla f(x^k) \quad \text{sowie} \quad F''(z^k) = Z^T f''(x^k) Z.$$

Ist d^k eine zulässige Abstiegsrichtung im z^k , so ist Zd^k die zugehörige Suchrichtung im Punkte x^k , sodass die Folgeiterierte $z^{k+1} = z^k + \alpha_k d^k$ im z -Raum der Folgeiterierten

$$x^{k+1} = x^0 + Z(z^k + \alpha_k d^k) = (x^0 + Zz^k) + \alpha_k Z d^k = x^k + \alpha_k Z d^k$$

im x -Raum entspricht.

Es ist abzuklären, ob in der Situation, in der ein KKT Punkt zu (NLR) bestimmt wurde, die Rücktransformation auch einen KKT Punkt zu (NLG) liefert.

Lemma 6.1 *Sei Z eine beliebige Nullraum-Matrix zu H und sei $x = x^0 + Zz$ vorgegeben.*

Genau dann ist z eine KKT Punkt von (NLR), wenn x ein KKT Punkt von (NLG) ist.

Beweis: Seien also x, z mit $x = x^0 + Zz$ vorgegeben.

Sei zunächst z als KKT Punkt von (NLR) vorausgesetzt. Es gilt also $\nabla F(z) = Z^T \nabla f(x) = 0$. Ist dann $d \in \ker(H)$, so gilt

$$\nabla f(x)^T d = \nabla f(x)^T Z \bar{z} = (Z^T \nabla f(x))^T \bar{z} = \nabla F(z)^T \bar{z} = 0$$

für ein $\bar{z} \in \mathbf{R}^l$. Daher ist $\nabla f(x) \in \ker(H)^T = \mathcal{N}(H)^T = \mathcal{R}(H)$. Es gibt also ein $\lambda \in \mathbf{R}^m$ mit $\nabla f(x) = H^T \lambda$.

Sei nun $x = x^0 + Zz$ als KKT Punkt von (NLG) angenommen. Dann gilt $\nabla f(x) = -H^T \lambda$ für ein $\lambda \in \mathbf{R}^m$. Dann ist

$$\nabla F(z) = Z^T \nabla f(x) = -Z^T H^T \lambda = (\lambda^T H Z)^T = 0$$

□

Wir haben damit nach Wahl einer Nullraum-Matrix Z grundsätzlich **zwei Möglichkeiten zur konkreten Lösung von (NLG)** :

1. Wir transformieren explizit die Aufgabe (NLG) in die Aufgabe (NLR) , lösen diese mit einem beliebigen Verfahren der unrestringierten Optimierung und transformieren das Ergebnis zurück.

Von dieser Möglichkeit wird im Prinzip in den Simplexverfahren in Tableauform Gebrauch gemacht, wenn das Verfahren lokal eine Darstellung unter Verwendung der "Nullraum-Matrix" $-B^{-1}N$ verwendet.

2. Wir transformieren nur theoretisch, wählen ein Verfahren der unrestringierten Optimierung, führen dieses aber möglichst in Termen der Originalaufgabe durch.

Im folgenden wollen wir an Beispielen zeigen, wie sich durch die zweite Möglichkeit Verfahren der unrestringierten Optimierung in reduzierte Verfahren übertragen lassen. Dazu sei eine Nullraum-Matrix Z vorgewählt. Wir bezeichnen weiterhin mit V den von $Hx = h$ aufgespannten affinen Unterraum von \mathbf{R}^n , verwenden die o. g. Beziehungen zwischen den Iterierten z^k und x^k und bezeichnen die Suchrichtungen im z -Raum grundsätzlich mit d^k . Die Suchrichtungen im x -Raum ergeben sich dann zu Zd^k .

Übertragung des Globalisierten Newtonverfahrens

Als erstes betrachten wir das Globalisierte Newtonverfahren der unrestringierten Optimierung. Dieses nimmt als Verfahren für (NLG) die folgende Form an:

Reduziertes Globalisiertes Newtonverfahren

(0) **Start:** Wähle einen Punkt $x^0 \in V$ und $\rho > 0$, $\varepsilon \geq 0$, $\sigma, \beta \in (0, 1)$.
Setze $k = 0$.

(1) **Abbruch:** Gilt $\|Z^T \nabla f(x^k)\| \leq \varepsilon$, **Stopp**.

(2) **Suchrichtung:** Berechne eine Lösung d^k der Newtongleichung

$$Z^T \nabla^2 f(x^k) Z d = -Z^T \nabla f(x^k)$$

Ist dieses System nicht lösbar oder ist die Bedingung

$$\nabla f(x^k)^T Z d^k \leq -\rho \|d^k\|^2$$

nicht erfüllt, so setze $d^k := -Z^T \nabla f(x^k)$

(3) **Schrittweite:** Setze $\bar{x}^k := x^k + \alpha_k Z d^k$ mit der Eigenschaft: $\alpha_k := \max \{\beta^l : l = 0, 1, 2, 3, \dots\}$, so dass

$$f(x^k + \alpha_k Z d^k) \leq f(x^k) + \sigma \alpha_k \nabla f(x^k)^T Z d^k$$

(4) **Iteration:** Setze $x^{k+1} := \bar{x}^k$, $k := k + 1$ und gehe zu (1).

Übertragung des Fletcher-Reeves Verfahrens

Für die Übertragung des Fletcher-Reeves Verfahrens betrachten wir diesmal speziell $Z = P_H$. Denn es ist für das Verfahren vorteilhaft, dass bei der Übertragung von Skalarprodukten des Gradienten mit der Suchrichtung d auf die Projektion des Gradienten verzichtet werden kann, da es für den Gradienten g eine direkte Zerlegung $g = g^1 + g^2$ gibt mit $g^1 \in \mathcal{N}(H)$ und $g^2 \in \mathcal{N}(H)^\perp$ und die Suchrichtung d in $\mathcal{N}(H)$ liegt:

$$g^T d = (g^1 + g^2)^T d = (g^1)^T d + (g^2)^T d = (g^1)^T d = (P_H g)^T d$$

Verfahren RFR (Reduziertes Fletcher-Reeves Verfahren)

1. **Start:** Wähle $x^0 \in V$, $\varepsilon \geq 0$, $0 < \sigma < \rho < \frac{1}{2}$, setze $d^0 := -P_H \nabla f(x^0)$ und $k = 0$.
2. **Abbruch:** Ist $\|P_H \nabla f(x^k)\| \leq \varepsilon$, **Stopp1**
3. **Schrittweite:** Bestimme eine Schrittweite $\alpha_k > 0$ mit

$$\begin{aligned} f(x^k + \alpha_k d^k) &\leq f(x^k) + \sigma \alpha_k \nabla f(x^k)^T d^k \quad \text{und} \\ \left| \nabla f(x^k + \alpha_k d^k)^T d^k \right| &\leq -\rho \nabla f(x^k)^T d^k \end{aligned}$$

Existiert keine solche Schrittweite, **Stopp2**, (die Zielfunktion ist unbeschränkt)

4. **Aktualisierung:** Setze

$$\begin{aligned} x^{k+1} &: = x^k + \alpha_k d^k \\ \beta_k &: = \frac{\|P_H \nabla f(x^{k+1})\|^2}{\|P_H \nabla f(x^k)\|^2} \\ d^{k+1} &: = -P_H \nabla f(x^{k+1}) + \beta_k d^k \end{aligned}$$

und $k = k + 1$. Gehe zu 2.)

Übertragung des DFP Verfahrens

Als nächstes betrachten wir das DFP Verfahren. Hier wählt man am besten Z aus der QR Zerlegung, da dann $Z^T Z = I_{n-m}$ gilt, was folgende Konsequenz hat: Seien $z, \bar{z} \in \mathbf{R}^{n-m}$ und $x = x^0 + Zz$, $\bar{x} = x^0 + Z\bar{z}$. Dann gilt

$$z - \bar{z} = Z^T Z (z - \bar{z}) = Z^T [x^0 + Zz - (x^0 + Z\bar{z})] = Z^T (x - \bar{x})$$

Verfahren RDFP Reduziertes Davidon-Fletcher-Powell Verfahren)

1. **Start:** Wähle $x^0 \in V$, $\varepsilon \geq 0$ und die Parameter $\kappa, \gamma > 0$ zur Überprüfung der Winkelbedingung und der Effizienz der Schrittweiten sowie die Parameter $\sigma \in (0, \frac{1}{2})$ und $\rho \in [\sigma, 1)$ für die strenge Wolfe-Powell-Schrittweitenregel. Setze $D_0 := I_{n-m}$ und $k = 0$.

2. **Abbruch:** Ist $\|Z^T \nabla f(x^k)\| \leq \varepsilon$, **stopp1**, (ein KKT-Punkt wurde angenähert)
3. **Schritt:** Setze $d^k := -D_k Z^T \nabla f(x^k)$. (D_k im z -Raum)
Erfüllt d^k nicht die Winkelbedingung an der Stelle z^k

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

so setze $d^k := -Z^T \nabla f(x^k)$ und $D_k := I_{n-m}$.

Berechne eine Schrittweite $\alpha_k > 0$ so, dass die strengen Wolfe-Powell Bedingungen

$$\begin{aligned} f(x^k + \alpha_k Z d^k) &\leq f(x^k) + \sigma \alpha \nabla f(x^k)^T Z d^k \\ \left| f(x^k + \alpha_k Z d^k)^T d^k \right| &\leq -\rho \nabla f(x^k)^T Z d^k \end{aligned}$$

erfüllt sind. Existiert keine solche Schrittweite, **stopp2**, (die Zielfunktion ist unbeschränkt)

4. **Aktualisierung:** Setze $x^{k+1} := x^k + \alpha_k d^k$ und

$$\begin{aligned} D_{k+1} &: = D_k + \frac{p^k (p^k)^T}{(p^k)^T q^k} - \frac{D_k q^k (q^k)^T D_k}{(q^k)^T D_k q^k} \quad \text{mit} \quad (6.1) \\ p^k &: = \alpha_k d^k = Z^T (x^{k+1} - x^k) \\ q^k &: = Z^T (\nabla f(x^{k+1}) - \nabla f(x^k)) \end{aligned}$$

und $k = k + 1$. Gehe zu 1.

Das Lagrange-Newton-Verfahren

Die Übertragung des normalen Newtonverfahrens kann im Prinzip an der des globalisierten Newtonverfahrens abgelesen werden. Betrachten wir nun noch eine alternative Herleitung des Newtonverfahrens für die Aufgabe (NLG). Wir setzen dazu zur Abkürzung $h(x) = Hx - h$. Mit der Lagrange-Funktion

$$L(x, \mu) = f(x) + \mu^T h(x)$$

können die KKT-Bedingungen der Aufgabe (NLG) formuliert

$$\begin{aligned}\nabla_x L(x, \mu) &= \nabla f(x) + \mu^T \nabla h(x) = 0 \\ h(x) &= 0\end{aligned}$$

und mit der Abkürzung

$$\Phi(x, \mu) := \begin{pmatrix} L_x(x, \mu) \\ h(x) \end{pmatrix}$$

auch als $\Phi(x, \mu) = 0$ geschrieben werden. Diese Gleichung kann mit dem Newtonverfahren (für Gleichungssysteme) angegangen werden, sofern entsprechende Voraussetzungen erfüllt sind. Man spricht dann vom **Lagrange-Newton-Verfahren**.

Es gilt z.B. der

Satz 6.2 *Es sei $(x^*, \mu^*) \in \mathbf{R}^{n+p}$ ein KKT-Punkt der Aufgabe (NLG). Sind die beiden folgenden Bedingungen erfüllt*

1. *Die Gradienten $\nabla h_1(x^*), \dots, \nabla h_p(x^*)$ sind linear unabhängig*
2. *Es gilt $d^T L_{xx}^2(x^*, \mu^*) d \neq 0$ für alle $d \neq 0$ mit $\nabla h_j(x^*)^T d = 0$ ($j = 1, \dots, p$)*

so ist die Jacobimatrix $\Phi'(x^, \mu^*)$ zu Φ regulär.*

Beweis: Die Jacobimatrix $\Phi'(x^*, \mu^*)$ lautet:

$$\Phi'(x^*, \mu^*) = \begin{pmatrix} L_{xx}^2(x^*, \mu^*) & h'(x^*)^T \\ h'(x^*) & 0 \end{pmatrix}$$

Nehmen wir an, dass $\Phi'(x^*, \mu^*) q = 0$ für ein $q \in \mathbf{R}^{n+p}$ gelte. Wir wollen zeigen, dass dann zwangsweise $q = 0$ gelten muss, dass also $\Phi'(x^*, \mu^*)$ regulär ist.

Sei etwa $q =: \begin{pmatrix} q^{(1)} \\ q^{(2)} \end{pmatrix}$ gemäß der Struktur von $\Phi'(x^*, \mu^*)$, dann lässt sich die Gleichung $\Phi'(x^*, \mu^*) q = 0$ schreiben als

$$\begin{pmatrix} L_{xx}^2(x^*, \mu^*) & h'(x^*)^T \\ h'(x^*) & 0 \end{pmatrix} \begin{pmatrix} q^{(1)} \\ q^{(2)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Multipliziert man die erste dieser Gleichungen von links mit $q^{(1)T}$, so ergibt sich

$$q^{(1)T} L_{xx}^2(x^*, \mu^*) q^{(1)} = 0$$

Nach Voraussetzung 2. folgt daraus unmittelbar $q^{(1)} = 0$. Damit wiederum folgt aus der ersten Gleichung

$$h'(x)^T q^{(2)} = 0$$

was wegen der 1. Voraussetzung zu $q^{(2)} = 0$ führt. Also gilt wie gewünscht $q = 0$.

□

Der Satz nennt also Voraussetzungen, unter denen gemäß dem obigen Konvergenzsatz das Newtonverfahren erfolgreich zur Lösung der KKT-Systems von (NLG) angewendet werden kann.

Zu beachten ist, dass wir im obigen Beweis die spezielle lineare Struktur der Nebenbedingungen nicht ausgenutzt haben. Der Satz gilt genauso für eine Nebenbedingung $h(x) = 0$ mit beliebiger zweimal stetig differenzierbarer Funktion $h : \mathbf{R}^n \rightarrow \mathbf{R}^p$.

Es gilt für $h(x) = Hx - h : h'(x) = H, L(x, \mu) = f(x) + \mu^T (Hx - h), L_x(x, \mu) = \nabla f(x) + H^T \mu$ und $L_{xx}(x, \mu) = \nabla^2 f(x)$. Die Bedingung im Satz, dass die Gradienten der Restriktionsfunktionen linear unabhängig sein sollen, entspricht der Forderung, dass H maximalen Zeilenrang hat. Die Newtonrichtung berechnet sich also durch

$$\begin{pmatrix} \nabla^2 f(x) & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \mu \end{pmatrix} = - \begin{pmatrix} \nabla f(x) + \mu^T H \\ Hx - h \end{pmatrix}$$

Setzt man $\mu^+ := \mu + \Delta \mu$, so ergibt sich gleichwertig

$$\begin{pmatrix} \nabla^2 f(x) & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \mu^+ \end{pmatrix} = - \begin{pmatrix} \nabla f(x) \\ Hx - h \end{pmatrix}$$

insbesondere bei *zulässigem* x

$$\begin{pmatrix} \nabla^2 f(x) & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \mu^+ \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ 0 \end{pmatrix}$$

6.1.3 Spezialfall: Quadratische Optimierung

Wenden wir uns nun wieder der Quadratischen Optimierung zu. Im Sinne des letzten Abschnitts betrachten wir nun die quadratische Optimierungsaufgabe

$$(QPG) \quad \begin{cases} \text{minimiere} & f(x) = \frac{1}{2}x^T Qx + p^T x \\ \text{s.d.} & Hx = h \end{cases}$$

mit symmetrischer Matrix $Q \in \mathbf{R}^{n \times n}$ und $H \in \mathbf{R}^{n \times p}$ mit *maximalem Zeilenrang*. Die KKT-Bedingungen hierzu lauten

$$(QGKKT) \quad \begin{cases} Qx + p + H^T \mu = 0 \\ Hx - h = 0 \end{cases}$$

bzw in Matrixform

$$\begin{pmatrix} Q & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} x \\ \mu \end{pmatrix} = \begin{pmatrix} -p \\ h \end{pmatrix} \quad (6.2)$$

Anwendung des Newtonverfahrens

Will man dieses Gleichungssystem mit dem Newtonverfahren lösen, so hat man gemäß (5.3) folgendes Newton-System zu lösen

$$\begin{pmatrix} Q & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} d^{(1)} \\ d^{(2)} \end{pmatrix} = - \begin{pmatrix} Qx + p + H^T \mu \\ Hx - h \end{pmatrix},$$

um in der Newtoniteration vom Punkte x zum Punkte $\bar{x} := x + d^{(1)}$, $\bar{\mu} = \mu + d^{(2)}$ zu gelangen. Bringt man nun $(\bar{x}, \bar{\mu})$ in das Gleichungssystem ein, so erhält man mit einfacher Umformung

$$\begin{aligned} \begin{pmatrix} Q & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} \bar{x} - x \\ \bar{\mu} - \mu \end{pmatrix} &= - \begin{pmatrix} Qx + p + H^T \mu \\ Hx - h \end{pmatrix} \Leftrightarrow \\ \begin{pmatrix} Q & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{\mu} \end{pmatrix} &= \begin{pmatrix} -p \\ h \end{pmatrix} \end{aligned}$$

Es ergeben sich also die KKT-Bedingungen für den Punkt $(\bar{x}, \bar{\mu})$! Dies lässt sich wie folgt interpretieren:

Genau dann ist die Newtoniteration für das KKT-System von beliebigem Startpunkt $(x, \mu) \in \mathbf{R}^{n+p}$ aus durchführbar, wenn das KKT-System für (QPG) lösbar ist. In diesem Fall ermittelt die Newtoniteration einen KKT-Punkt *in einem einzigen Schritt*.

Auf der anderen Seite lohnt es sich nicht, das Newtonverfahren anzuwenden, man kann direkt das KKT System (6.2) lösen.

Bleibe also zu klären, wann die Aufgabe (QPG) einen KKT-Punkt besitzt bzw wann die Newtoniteration für das KKT-System von (QPG) durchführbar ist.

Eine Antwort liefert der obige Satz über das Lagrange-Newton-Verfahren. Es gilt für die Lagrangefunktion

$$L(x, \mu) = \frac{1}{2}x^T Qx + p^T x + H^T \mu$$

also gilt $L_{xx}^2(x, \mu) = Q$. Deshalb lautet eine *hinreichende Bedingung* für die Existenz eines KKT-Punktes:

Gilt $d^T Qd \neq 0$ für alle nichttrivialen $d \in \ker(H)$, so ermittelt das Newtonverfahren bei beliebigem Ausgangspunkt x^0 eine Lösung von $(QGKKT)$ in einem einzigen Schritt und das KKT System zu (QPG) besitzt eine zulässige Lösung.

Diese Bedingung ist natürlich erfüllt, falls Q positiv oder negativ definit ist, falls also f streng konvexe oder streng konkave Funktion ist, aber eben auch, wenn Q eingeschränkt auf $\ker(H)$ positiv oder negativ definit ist.

Lösungsverfahren für (QPG)

Das KKT-System von (QPG) kann auf mindestens zwei Weisen numerisch günstig angegangen werden. **Wir unterstellen**, dass H maximalen Zeilenrang hat, dass Q positiv semidefinit ist und dass das System (6.2) eine Lösung (x, μ) besitzt.

Cholesky-Verfahren Das System (6.2)

$$\begin{pmatrix} Q & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} x \\ \mu \end{pmatrix} = \begin{pmatrix} -p \\ h \end{pmatrix}$$

kann eindeutig gelöst werden, wenn Q sogar positiv definit ist und man Q^{-1} kennt.

Zunächst gilt

$$H^T \mu = -p - Qx$$

also nach Multiplikation mit Q^{-1}

$$Q^{-1}H^T \mu = -Q^{-1}p - x$$

und mit H

$$HQ^{-1}H^T \mu = -(HQ^{-1}p + h)$$

Da Q positiv definit ist und H von maximalem Rang, ist auch $HQ^{-1}H^T$ symmetrisch und positiv definit und das Cholesky Verfahren kann zur Berechnung von μ herangezogen werden. Danach berechnet man x aus

$$x = -Q^{-1}(p + H^T \mu)$$

QR Verfahren Man kann auch wie folgt vorgehen (vgl. [A], Verfahren 6.1.1):

1. **QR Zerlegung von H^T** : Berechne eine unitäre $n \times n$ Matrix U und eine obere Dreiecksmatrix $R \in \mathbf{R}^{m \times m}$ mit

$$UH^T = \begin{pmatrix} R \\ 0 \end{pmatrix}$$

Berechne

$$u := -Up = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad B := UQU^T = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

wobei $u_1 \in \mathbf{R}^m$ und B_{11} seine $m \times m$ Matrix ist.

2. **Berechnung von x** : Berechne x_Y durch Lösen des linearen Gleichungssystems

$$R^T x_Y = h$$

und x_Z durch Lösen des linearen Gleichungssystems

$$B_{22}x_Z = u_2 - B_{21}x_Y$$

Berechne

$$x = U^T \begin{pmatrix} x_Y \\ x_Z \end{pmatrix}$$

3. **Berechnung von μ** : Berechne μ durch Lösen des Gleichungssystems

$$R\mu = u_1 - B_{11}x_Y - B_{12}x_Z$$

Erläuterung der Vorgehensweise: Wie zuvor schon schreiben wir

$$U = \begin{pmatrix} Y^T \\ Z^T \end{pmatrix}$$

wobei $Y^T \in \mathbf{R}^{m \times n}$ die ersten m Zeilen zusammenfasst und $Z^T \in \mathbf{R}^{(n-m) \times n}$ die restlichen $(n - m)$ Zeilen. Dann gilt

$$UH^T = \begin{pmatrix} R \\ 0 \end{pmatrix} \quad \text{und} \quad HU^T = (HY, HZ) = (R^T, 0)$$

Es schreibt sich nun jedes $x \in \mathbf{R}^n$, insbesondere eine Lösung x von (6.2), eindeutig als

$$x = U^T \begin{pmatrix} x_Y \\ x_Z \end{pmatrix} = (Y, Z) \begin{pmatrix} x_Y \\ x_Z \end{pmatrix} = Yx_Y + Zx_Z, \quad (6.3)$$

Z ist dabei eine Nullraum-Matrix für H . Also gilt $Zx_Y \in \ker(H)$, d.h. $HZx_Y = 0$ und

$$h = Hx = HYx_Y = R^T x_Y$$

Da R eine invertierbare obere Dreiecksmatrix ist, lässt sich dieses Gleichungssystem leicht nach x_Y auflösen.

Mit $w = Yx_Y$ haben wir eine zulässige Lösung von $Hx = h$ gefunden und wir können mithilfe dieser Lösung und der Nullraum-Matrix Z die Ersatzaufgabe für quadratische Optimierungsaufgabe (QPG) formulieren:

$$\min \left\{ f(w + Zz) = \frac{1}{2} (w + Zz)^T Q (w + Zz) + p^T (w + Zz) : z \in \mathbf{R}^{n-m} \right\}$$

bzw nach Weglassen von Konstanten

$$(QPR) \quad \min \left\{ F(z) = \frac{1}{2} z^T \tilde{Q} z + \tilde{p}^T z : z \in \mathbf{R}^{n-m} \right\}$$

mit $\tilde{Q} = Z^T Q Z$ und $\tilde{p} = Z^T p + Z^T Q w$.

Aufgrund unserer Voraussetzungen und Lemma 6.1 hat die Aufgabe (QPR) eine Lösung z mit $x = w + Zz$. Wegen $x = Yx_Y + Zx_Z = w + Zx_Z$ ist $z := x_Z$ eine solche Lösung. Also ist x_Z auch Lösung des Gleichungssystems

$$\tilde{Q}z = Z^T Q Z z = -\tilde{p} = -(Z^T p + Z^T Q w) = -Z^T p - Z^T Q Y x_Y \quad (6.4)$$

Zur Lösung dieses Gleichungssystems beachte: aus der QR Zerlegung findet man

$$B := UQU^T = \begin{pmatrix} Y^T \\ Z^T \end{pmatrix} Q(Y, Z) = \begin{pmatrix} Y^T Q Y & Y^T Q Z \\ Z^T Q Y & Z^T Q Z \end{pmatrix} =: \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

und

$$u = -Up = - \begin{pmatrix} Y^T \\ Z^T \end{pmatrix} p = \begin{pmatrix} -Y^T p \\ -Z^T p \end{pmatrix} =: \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

Daher löst sich Gleichung (6.4) über

$$B_{22}x_Z = u_2 - B_{21}x_Y$$

Nach unserer Voraussetzung gibt es eine Lösung dieses Gleichungssystems. Ist die symmetrische Matrix B_{22} invertierbar, so kann eine Cholesky Zerlegung von B_{22} verwendet werden, um die Gleichung schnell und numerisch stabil zu lösen.

Mit x_Y und x_Z haben wir $x = Yx_Y + Zx_Z$ als Lösung von (QPG) bestimmt. Bleibt die Bestimmung von μ . μ hat die KKT-Gleichung $Qx + H^T \mu = -p$ zu erfüllen. Mit (6.3) folgt nach Multiplikation dieser Gleichung mit U

$$UQU^T \begin{pmatrix} x_Y \\ x_Z \end{pmatrix} + UH^T \mu = -Up$$

also

$$\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} x_Y \\ x_Z \end{pmatrix} + \begin{pmatrix} R \\ 0 \end{pmatrix} \mu = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

Daraus liest man ab

$$R\mu = u_1 - B_{11}x_Y - B_{12}x_Z$$

woraus μ leicht berechnet werden kann.

6.2 Verfahren gradientenähnlicher Abstiegsrichtungen

Nachdem wir in den letzten Abschnitten Verfahren der unrestrictierten Optimierung auf den Fall der linear gleichungsrestrictierten Optimierung übertragen haben, wollen wir nun die Lösung solcher Aufgaben durch Spezialisierung des Zulässigen Abstiegsverfahrens direkt angehen.

Abstiegsverfahren, deren Suchrichtungen "ähnlich gerichtet" sind wie der negative Gradient können ein konvergentes Verhalten zeigen, wenn deren Schrittweiten genügend absenkend für die Zielfunktion wirken (vgl. Allgemeines Unrestrictiertes Abstiegsverfahren). Wir besprechen ein solches Verfahren nun für die Aufgabe (*NLG*), indem wir das *Gradientenverfahren* der unrestrictierten Optimierung verallgemeinern. (vgl. [GK1], Abschnitt 8.3). Im Hinblick auf eine spätere Anwendung des Verfahrens legen wir dabei aber eine etwas allgemeinere Aufgabenstellung zugrunde, bei der die zu minimierende Funktion nicht mehr auf ganz \mathbf{R}^n , sondern nur noch auf einer offenen Teilmenge definiert ist.

6.2.1 Gradientenähnliche Suchrichtungen

Betrachtet werde die Aufgabe (*NLGO*), die eine leichte Verallgemeinerung der Aufgabe (*NLG*) darstellt. Bei dieser wird die stetig differenzierbare Funktion $f : \mathcal{U} \subseteq \mathbf{R}^n \rightarrow \mathbf{R}$, \mathcal{U} offen, über $V \cap \mathcal{U}$ mit $V = \{x \in \mathbf{R}^n : Hx = h\}$ minimiert. Im Falle $\mathcal{U} = \mathbf{R}^n$ stimmt (*NLGO*) mit (*NLG*) überein.

Man beachte: Die Aufgabe (*NLGO*) kann auch eine unrestrictierte Optimierung beschreiben, etwa im Fall $H = 0$, $h = 0$ (bei $V = \mathcal{U} = \mathbf{R}^n$).

Wir betrachten nun noch einmal das Gradientenverfahren und verändern es wie folgt: als Abstiegsrichtungen an den Iteriertenstellen $x^k \in V \cap \mathcal{U}$ werden nicht die negativen Gradienten $d^k := -\nabla f(x^k)$ genommen, sondern wir lassen ganz allgemein eine beliebige Abstiegsrichtung zu (*NLGO*) zu, also eine Richtung d^k , die lediglich

$$d^k \in \mathcal{N}(H) \quad \text{mit} \quad \nabla f(x^k)^T d^k < 0 \quad (6.5)$$

respektiert. Dabei wird unterstellt, dass die erste Iterierte ein $x^0 \in V \cap \mathcal{U}$ ist. Wegen der Wahl der Abstiegsrichtungen liegen dann natürlich alle weiteren Iterierten auch in V , bei entsprechend kurzer Schrittweite auch in \mathcal{U} .

Zur Schrittweitenbestimmung verwenden wir weiterhin die Armijo Regel, mit der Modifikation, dass bei jedem Aufruf der Funktion zuerst geprüft werden muss, ob für die aktuelle Stelle überhaupt $x^k + \alpha d^k \in \mathcal{U}$ gilt. Gilt dies nicht, so wird α übersprungen und direkt zu $\alpha = \beta\alpha$ übergegangen (**modifizierte Armijo Regel**). Es entsteht so das **Allgemeine Zulässige Abstiegsverfahren für (NLGO)**.

Nun legen wir einen speziellen Typus von Abstiegrichtungen fest:

Definition 6.3 (vgl. [GK1], Definition 8.8) Sei $f : \mathcal{U} \subseteq \mathbf{R}^n \rightarrow \mathbf{R}$ stetig differenzierbar und $(x^k)_{k \in \mathbf{N}}$ eine Folge von Punkten aus V . Eine zugehörige Folge von Suchrichtungen $(d^k)_{k \in \mathbf{N}}$ aus $\mathcal{N}(H)$ heißt **gradientenähnlich bzgl (NLGO) und (x^k)** , wenn für jede gegen einen Nicht-KKT-Punkt vom (NLGO) konvergierende Teilfolge $(x^k)_{k \in K}$ Konstanten $c > 0$ und $\delta > 0$ existieren, so dass

1. $\|d^k\| \leq c$ für alle $k \in K$
2. $\nabla f(x^k)^T d^k \leq -\delta$ für alle hinreichend großen $k \in K$.

Mit dieser Festlegung können wir den Konvergenzsatz zum Gradientenverfahren und seinen Beweis auf den Fall der Lösung von (NLGO) mit dem so modifizierten Gradientenverfahren übertragen.

Satz 6.4 Wird die Folge der Suchrichtungen im so abgewandelten Verfahren gradientenähnlich gewählt, so ist jeder Häufungspunkt $x \in \mathcal{U}$ der Iteriertenfolge (x^k) ein KKT-Punkt von (NLGO).

Beweis: (vgl. [GK1], Satz 8.9)

Es werde das oben beschriebene Verfahren betrachtet, bei dem x^0 und die Suchrichtungen wie vorausgesetzt gewählt wurden. Nach unseren bisherigen Erkenntnissen ist $x \in V \cap \mathcal{U}$ genau dann ein KKT Punkt zu (NLGO), wenn $P_H \nabla f(x^*) = 0$ gilt.

Sei $(x^k)_{k \in K}$ eine Teilfolge der Iteriertenfolge, die gegen ein $x^* \in V \cap \mathcal{U}$ konvergieren möge. Nehmen wir an, es sei $P_H \nabla f(x^*) \neq 0$, $(x^k)_{k \in K}$ ist also eine gegen einen Nicht-KKT-Punkt von (NLGO) konvergierende Teilfolge der Folge $(x^k)_{k \in \mathbf{N}}$.

Da die Zielfunktionswerte der Iteriertenfolge strikt abfallen und die Teilfolge der $(f(x^k))_{k \in K}$ konvergiert, konvergiert die gesamte Folge $f(x^0), f(\bar{x}^0), f(x^1), \dots, f(x^k), f(\bar{x}^k), f(x^{k+1}), \dots$. Es gilt also

$$\alpha_k \nabla f(x^k)^T d^k = f(x^k) - f(\bar{x}^k) \rightarrow 0,$$

und da nach Voraussetzung die Suchrichtungen gradientenähnlich gewählt wurden, gilt $|\nabla f(x^k)^T d^k| \geq \delta > 0$, so dass zwangsläufig $\alpha_k \rightarrow 0$ gelten muss.

Es möge über die modifizierte Armijo-Regel $\alpha_k = \beta^{l_k}$ gelten. Weil nun $x^* \in \mathcal{U}$ vorausgesetzt wurde und \mathcal{U} offen ist, gibt es eine Kugel $B_r(x^*)$ vom Radius $r > 0$, die ganz in \mathcal{U} liegt und es folgt, dass für k groß genug neben $x^k + \beta^{l_k} d^k$ auch $x^k + \beta^{l_k-1} d^k$ in \mathcal{U} liegen muss und daher für diese k

$$f(x^k + \beta^{l_k-1} d^k) > f(x^k) + \sigma \beta^{l_k-1} \nabla f(x^k)^T d^k$$

gilt. Damit gilt auch

$$\frac{f(x^k + \beta^{l_k-1} d^k) - f(x^k)}{\beta^{l_k-1}} > \sigma \nabla f(x^k)^T d^k$$

Weil nun die d^k auf Grund der Gradientenähnlichkeit beschränkt sind, konvergiert eine Teilfolge der d^k gegen ein $d^* \in V$. Gehen wir oBdA davon aus, dass bereits die Folge $(d^k)_{k \in K}$ gegen d^* konvergiert.

Da wegen $\alpha_k = \beta^{l_k} \rightarrow 0$ auch die β^{l_k-1} eine Nullfolge bilden, folgt nun gemäß dem Lemma 5.4 für $k \rightarrow \infty$

$$\nabla f(x^*)^T d^* \geq \sigma \nabla f(x^*)^T d^*$$

Dies widerspricht aber $\nabla f(x^*)^T d^* \leq -\delta < 0$ und $\sigma < 1$.

□

Bemerkung: Wesentlich im obigen Satz ist, dass der Häufungspunkt x^* in \mathcal{U} liegt. Da \mathcal{U} offen ist, ist dies nicht von alleine gegeben. Eine hinreichende Bedingung dafür ist, dass die Levelmenge

$$\mathcal{L}_{x^0} = \{x \in V \cap \mathcal{U} : f(x) \leq f(x^0)\}$$

kompakt ist. Wir wollen dies im Folgenden voraussetzen.

Wir nennen das so festgelegte Verfahren

Allgemeines Gradientenähnliches Verfahren (für (NLGO))

- (0) **Start:** Wähle einen Punkt $x^0 \in V \cap \mathcal{U}$ und $\varepsilon \geq 0$, $\sigma, \beta \in (0, 1)$. Setze $k = 0$.
- (1) **Abbruch:** Gilt $\|P_H \nabla f(x^k)\| \leq \varepsilon$, **Stopp**.
- (2) **Suchrichtung:** Wähle Suchrichtungen $d^k \in \mathcal{N}(H)$ mit $\nabla f(x^k)^T d^k < 0$, gradientenähnlich bzgl. (NLGO)
- (3) **Schrittweite:** Setze $\bar{x}^k := x^k + \alpha_k d^k$ mit der Eigenschaft: $\alpha_k := \max \{\beta^l : l = 0, 1, 2, 3, \dots\}$, so dass
$$x^k + \alpha_k d^k \in \mathcal{U} \text{ und } f(x^k + \alpha_k d^k) \leq f(x^k) + \sigma \alpha_k \nabla f(x^k)^T d^k$$
- (4) **Iteration:** Wähle $x^{k+1} \in V \cap \mathcal{U}$ mit $f(x^{k+1}) \leq f(\bar{x}^k)$, setze $k := k + 1$ und gehe zu (1).

Bleibt die Frage zu klären, wie man in der Praxis die Wahl gradientenähnlicher Abstiegsrichtungen sicherstellt. Die Problematik dabei ist, dass sich im Verfahren die Wahl der Suchrichtungen und der Iterationspunkte gegenseitig beeinflussen müssen, wenn ein bestimmtes Grenzverhalten erreicht werden soll. Die obige Definition gradientenähnlicher Abstiegsrichtungen ist praktisch so nicht zu handhaben.

6.2.2 Gradientenbezogene Suchrichtungen

Wir betrachten eine Folge von Iterierten und Suchrichtungen des Allgemeinen zulässigen Abstiegsverfahrens für (NLGO), die bei einem $x^0 \in V \cap \mathcal{U}$ startet und (6.5) respektiert. Es gelte $\|P_H \nabla f(x^k)\| > 0$ für alle $k \in \mathbf{N}$, d.h. das Verfahren ende nicht nach endlich vielen Schritten.

- **Behauptung1:** Hinreichend für die Gradientenähnlichkeit ist die Beachtung folgender Regeln bei der Wahl der Suchrichtungen: Wähle Konstanten $c_1 > 0$, $c_2 > 0$ so, dass für alle $k \in \mathbf{N}$ gilt

$$\|d^k\| \leq c_1 \|P_H \nabla f(x^k)\| \quad (6.6)$$

$$\nabla f(x^k)^T d^k \leq -c_2 \|P_H \nabla f(x^k)\|^2 \quad (6.7)$$

Wir nennen Suchrichtungen d^k , die diese beiden Bedingungen erfüllen, **(schwach) gradientenbezogen bzgl (NLGO) und der Iteriertenfolge** $(x^k)_{k \in \mathbf{N}}$.

Beweis: Es sei $(x^k)_{k \in K}$ eine gegen einen Nicht-KKT-Punkt x^* konvergente Teilfolge der Iteriertenpunkte aus $V \cap \mathcal{U}$. Es gilt also $\|P_H \nabla f(x^*)\| \neq 0$. Betrachte die Ungleichung $\nabla f(x^k)^T d^k \leq -c_2 \|P_H \nabla f(x^k)\|^2$ für $k \rightarrow \infty, k \in K$. Dann gibt es also ein $C_2 > 0$ so, dass $C_2 = -c_2 \|P_H \nabla f(x^*)\|^2 < 0$ gilt. Für genügend große $k \in K$ gilt also $\nabla f(x^k)^T d^k \leq -\frac{C_2}{2}$.

Entsprechend betrachte die Ungleichung $\|d^k\| \leq c_1 \|P_H \nabla f(x^k)\|$ für $k \rightarrow \infty, k \in K$. Dann gibt es ein $C_1 > 0$ so, dass $C_1 = c_1 \|P_H \nabla f(x^*)\|$ gilt. Für genügend große $k \in K$, etwa $k \geq k_1 > 0$ gilt also $\|d^k\| \leq 2C_1$. Sei ferner C_3 das Maximum der Werte $\|d^k\|$ für $k \leq k_1$. Dann gilt $\|d^k\| \leq \max\{C_3, 2C_1\}$ für alle $k \in K$.

- **Behauptung2:** Notwendig für die Bedingungen (6.6) und (6.7) sind die Bedingungen (mit den gleichen Konstanten, für alle $k \in \mathbf{N}$)

$$\|d^k\| \leq c_1 \|P_H \nabla f(x^k)\| \quad (6.8)$$

$$\nabla f(x^k)^T d^k \leq -\frac{c_2}{c_1} \|P_H \nabla f(x^k)\| \|d^k\| \quad (6.9)$$

Die zweite Bedingung begrenzt dabei den Winkel zwischen $\|d^k\|$ und $\|\nabla f(x^k)\|$ gleichmäßig auf unter 90° für alle Iterierten (*Winkelbedingung*). Damit ist die Bezeichnung "schwach gradientenbezogen" für die Bedingungen (6.6) und (6.7) verständlich: diese begrenzen den Winkel zwischen der Suchrichtung und dem steilsten Abstieg und gleichzeitig auch die Länge der Suchrichtung nach oben, wobei die Länge des steilsten Abstiegs als Referenz dient.

Beweis: Es seien die Bedingungen (6.6) und (6.7) vorgegeben. Dann

gilt für alle k auch (6.8) und

$$\begin{aligned} \nabla f(x^k)^T d^k &\leq -c_2 \|P_H \nabla f(x^k)\|^2 \leq c_2 \|P_H \nabla f(x^k)\| (-\|P_H \nabla f(x^k)\|) \\ &\leq c_2 \|P_H \nabla f(x^k)\| \left(-\frac{1}{c_1} \|d^k\|\right) = -\frac{c_2}{c_1} \|P_H \nabla f(x^k)\| \|d^k\| \end{aligned}$$

- **Behauptung3:** Hinreichend für die Bedingungen (6.6) und (6.7) sind die Bedingungen mit Konstanten $c_1, c_2, c_3 > 0$ (für alle $k \in \mathbf{N}$)

$$c_1 \|P_H \nabla f(x^k)\| \leq \|d^k\| \leq c_2 \|P_H \nabla f(x^k)\| \quad (6.10)$$

$$\nabla f(x^k)^T d^k \leq -c_3 \|P_H \nabla f(x^k)\| \|d^k\| \quad (6.11)$$

wobei die zweite Bedingung wieder den Winkel zwischen $\|d^k\|$ und $\|\nabla f(x^k)\|$ gleichmäßig auf unter 90° für alle Iterierten begrenzt. Wir nennen Suchrichtungen, die diese beiden Bedingungen erfüllen, **stark gradientenbezogen bzgl (NLGO) und der Iteriertenfolge** $(x^k)_{k \in \mathbf{N}}$, weil nun die Länge der Suchrichtung auch nach unten begrenzt ist.

Beweis: Sind die Bedingungen (6.10) und (6.11) vorgegeben, so gilt wieder direkt (6.6) und

$$\nabla f(x^k)^T d^k \leq -c_3 \|P_H \nabla f(x^k)\| \|d^k\| \leq -c_3 \|P_H \nabla f(x^k)\| c_1 \|P_H \nabla f(x^k)\|$$

Aus den Bedingungen (6.10) und (6.11) ist unmittelbar zu entnehmen, dass die Wahl $d^k := -P_H \nabla f(x^k)$ immer zu gradientenbezogenen Suchrichtungen und damit zu gradientenähnlichen Suchrichtungen führt, wenn $c_2 \geq 1$, $0 < c_1, c_3 \leq 1$ gewählt werden.

Bemerkung: Die bisherigen Ausführungen treffen natürlich insbesondere für die *unrestringierte Optimierung* zu! Damit hat sich gezeigt, dass die Einschränkungen an Suchrichtung und Schrittweite, die zur Konvergenz des *Allgemeinen unrestringierten Abstiegsverfahren* geführt haben, bei Aufrechterhaltung der Konvergenz leicht verschoben werden können, indem die Bedingungen an die Wahl der Suchrichtungen verschärft (zur Winkelbedingung kommt eine weitere Bedingung hinzu) und die an die Schrittweiten gelockert werden (zur Armijo Regel).

Wir erhalten damit folgendes

Allgemeines Gradientenbezogenes Verfahren (für (NLGO))

(0) **Start:** Wähle einen Punkt $x^0 \in V \cap \mathcal{U}$ und $\varepsilon \geq 0$, $\sigma, \beta \in (0, 1)$ sowie $0 < c_2 \leq 1 \leq c_1$. Setze $k = 0$.

(1) **Abbruch:** Gilt $\|P_H \nabla f(x^k)\| \leq \varepsilon$, **Stopp**.

(2) **Suchrichtung:** Wähle $d^k \in \mathcal{N}(H)$ mit

$$\begin{aligned} \|d^k\| &\leq c_1 \|P_H \nabla f(x^k)\| \\ \nabla f(x^k)^T d^k &\leq -c_2 \|P_H \nabla f(x^k)\| \|d^k\| \end{aligned}$$

(3) **Schrittweite:** Setze $\bar{x}^k := x^k + \alpha_k d^k$ mit der Eigenschaft: $\alpha_k := \max \{\beta^l : l = 0, 1, 2, 3, \dots\}$, so dass

$$x^k + \alpha_k d^k \in \mathcal{U} \text{ und } f(x^k + \alpha_k d^k) \leq f(x^k) + \sigma \alpha_k \nabla f(x^k)^T d^k$$

(4) **Iteration:** Wähle $x^{k+1} \in V \cap \mathcal{U}$ mit $f(x^{k+1}) \leq f(\bar{x}^k)$, setze $k := k + 1$ und gehe zu (1).

Im Folgenden können wir nun den bisherigen Konvergenzsatz verschärfen für den Fall, dass die Levelmenge

$$\mathcal{L}_{x^0} = \{x \in V \cap \mathcal{U} : f(x) \leq f(x^0)\}$$

kompakt ist. Zunächst notieren wir eine wesentliche Eigenschaft von über die *modifizierte Armijo Regel* bestimmten Schrittweiten in Form eines technischen

Lemma 6.5 *Es sei \mathcal{L}_{x^0} als kompakt vorausgesetzt. Dann gibt es eine Konstante $c > 0$, so dass für die im gradientenbezogenen Verfahren bestimmten Schrittweiten α_k für genügend großes k gilt*

$$\alpha_k \geq c \frac{|\nabla f(x^k)^T d^k|}{\|d^k\|^2}$$

Beweis: (vgl. [M1], Lemma 2.1 und Satz 2.2)

Wir betrachten ein x^k aus der Iteriertenfolge mit k so groß, dass neben $x^k + \alpha_k d^k$ auch $x^k + \frac{\alpha_k}{\beta} d^k$ in \mathcal{U} liegt.

a) Zunächst versuchen wir ein Intervall anzugeben, auf dem f echt unterhalb der Geraden $\eta(\alpha) = f(x^k) + \alpha \sigma \nabla f(x^k)^T d^k$, $\alpha \geq 0$, verläuft.

Da $\mathcal{L}_{x^0} \subseteq \mathcal{U}$ kompakt ist, gibt es nach dem Zwischenwertsatz ein $\bar{\alpha} > 0$ mit der Eigenschaft, dass

$$f(x^k + \bar{\alpha} d^k) = f(x^k) + \sigma \bar{\alpha} \nabla f(x^k)^T d^k$$

und auch

$$\tau := \alpha_{\min} := \arg \min \left\{ f(x^k + \alpha d^k) = f(x^k) + \sigma \alpha \nabla f(x^k)^T d^k : 0 \leq \alpha \leq \bar{\alpha} \right\}$$

ist nach Konstruktion positiv und es verläuft f auf $(0, \tau)$ echt unterhalb der Geraden η und damit auch innerhalb von $\mathcal{L}_{x^0} \subseteq \mathcal{U}$.

Versuchen wir, τ nach unten abzuschätzen:

b) Nun erhalten wir über den Satz von Taylor für ein $\lambda \in [0, 1]$

$$f(x^k + \tau d^k) = f(x^k) + \tau \nabla f(x^k)^T d^k + \frac{1}{2} \tau^2 d^{kT} \nabla^2 f(x^k + \lambda \tau d^k) d^k$$

und daraus die Abschätzung

$$\begin{aligned} \left| f(x^k + \tau d^k) - f(x^k) - \tau \nabla f(x^k)^T d^k \right| &= \frac{1}{2} \tau^2 \left| d^{kT} \nabla^2 f(x^k + \lambda \tau d^k) d^k \right| \\ &\leq \frac{1}{2} \tau^2 M \|d^k\|^2 \end{aligned}$$

wobei wir $M := \sup \{ \|\nabla^2 f(y)\| : y \in \mathcal{L}_{x^0} \} > 0$ gesetzt haben. Da nach Konstruktion und aus Stetigkeitsgründen $f(x^k + \tau d^k) = \nabla f(x^k) + \sigma \tau \nabla f(x^k)^T d^k$ gilt, folgt

$$(1 - \sigma) \tau \left| \nabla f(x^k)^T d^k \right| \leq \frac{1}{2} \tau^2 M \|d^k\|^2$$

Damit folgt $M > 0$ und wegen $\nabla f(x^k)^T d^k < 0$

$$\tau \geq \frac{2(\sigma - 1) \nabla f(x^k)^T d^k}{M \|d^k\|^2}$$

c) Nun zeigen wir die eigentliche Behauptung. Zunächst folgt für $\alpha_k \geq \beta\tau$ mit

$$\alpha_k \geq \frac{2\beta(1-\sigma) \left| \nabla f(x^k)^T d^k \right|}{M \|d^k\|^2}$$

unmittelbar die Behauptung. Betrachten wir also den Fall $\alpha_k < \beta\tau$. Dieser Fall kann aber gar nicht eintreten, da dann $\tilde{\alpha} := \frac{1}{\beta}\alpha_k < \tau$ und damit bereits $f(x^k + \tilde{\alpha}d^k) \leq \eta(\tilde{\alpha})$ gilt, also $\tilde{\alpha}$ anstelle von α_k von der Armijo Regel als Schrittweite ausgewählt worden wäre.

□

Nun können wir den bisherigen Konvergenzsatz verschärfen

Satz 6.6 *Betrachtet werde die Iteriertenfolge $\{x^k\}$ mit den Suchrichtungen $\{d^k\}$ im Allgemeinen gradientenbezogenen Verfahren für (NLGO). Es sei \mathcal{L}_{x^0} ist kompakt und x^* ein Häufungspunkt der Iteriertenfolge. Dann gilt:*

1. *Die Folge der Funktionswerte $f(x^k)$ konvergiert gegen $f(x^*)$ und die Folge der $\|d^k\|$ bildet eine Nullfolge. Insbesondere gilt $P_H \nabla f(x^*) = 0$ und x^* ist ein KKT Punkt zu (NLGO).*
2. *Ist die Menge Ω der zulässigen Punkte x mit $P_H \nabla f(x) = 0$ endlich, so konvergiert bereits die Folge der Iterierten gegen x^* .*

Beweis: (Vgl. [M1], Satz 2.1 mit Korollar)

1. Es folgt für genügend großes k aus der Armijo Regel

$$f(x^k) - f(x^{k+1}) \geq -\sigma\alpha_k \nabla f(x^k)^T d^k$$

ferner nach dem obigen Lemma

$$\alpha_k \geq c_1 \frac{\left| \nabla f(x^k)^T d^k \right|}{\|d^k\|^2}$$

und wegen der Gradientenbezogenheit der Suchrichtung

$$-\nabla f(x^k)^T d^k \geq c_2 \|P_H \nabla f(x^k)\| \|d^k\| > 0$$

Fügen wir diese Aussagen zusammen, so ergibt sich

$$f(x^k) - f(x^{k+1}) \geq \sigma c \frac{|\nabla f(x^k)^T d^k|^2}{\|d^k\|^2} \geq \sigma c_1 c_2^2 \|P_H \nabla f(x^k)\|^2 > 0$$

Da nun die Folge $\{f(x^k)\}$ wegen $\nabla f(x^k)^T d^k < 0$ monoton abfällt und aufgrund der Stetigkeit eine gegen $f(x^*)$ konvergente Teilfolge besitzt, konvergiert die ganze Folge gegen $f(x^*)$. Damit folgt aus der letzten Ungleichung, dass die Folge $\|P_H \nabla f(x^k)\|$ eine Nullfolge ist, woraus aufgrund der stetigen Differenzierbarkeit von f folgt, dass $P_H \nabla f(x^*) = 0$ gilt und x^* ein KKT Punkt zu (NPGO) ist. Wegen der Gradientenbezogenheit ist mit $\|P_H \nabla f(x^k)\|$ auch $\|d^k\|$ eine Nullfolge.

2. Wegen der Beschränktheit der Schrittweiten durch $\beta^0 = 1$ folgt aus der Tatsache, dass $\|d^k\|$ eine Nullfolge bildet, dass

$$\|x^{k+1} - x^k\| = \alpha_k \|d^k\| \rightarrow 0$$

Da nach 1. alle Häufungspunkte von $\{x^k\}_{k \in \mathbf{N}}$ in Ω liegen, ist $\text{dist}(x^k, \Omega)$ ebenfalls eine Nullfolge. Setzen wir

$$\rho := \min \{\|x - y\| : x, y \in \Omega, x \neq y\} > 0$$

Nun gibt es ein $k_0 \in \mathbf{N}$, so dass für alle $k \geq k_0$ gilt

$$\text{dist}(x^k, \Omega) < \frac{\rho}{4}, \quad \|x^{k+1} - x^k\| < \frac{\rho}{4}$$

Ferner gibt es, da x^* Häufungspunkt der Folge $\{x^k\}_{k \in \mathbf{N}}$ ist, ein $k_1 \geq k_0$ mit $\|x^{k_1} - x^*\| \leq \frac{\rho}{4}$. Damit gilt

$$\|x^{k_1+1} - x^*\| \leq \|x^{k_1+1} - x^{k_1}\| + \|x^{k_1} - x^*\| < \frac{\rho}{2}$$

Weil x^* in Ω liegt und $k_1 \geq k_0$ gilt, gilt sogar $\|x^{k_1+1} - x^*\| < \frac{\rho}{4}$ und über ein Induktionsargument folgern wir, dass sogar $\|x^k - x^*\| \leq \frac{\rho}{4}$ für alle $k \geq k_1 \geq k_0$. Damit ist x^* einziger Häufungspunkt der Folge $\{x^k\}_{k \in \mathbf{N}}$.

□

6.2.3 Transformierter steilster Abstieg

Die Wahl der Suchrichtungen im zuletzt genannten Verfahren ist immer noch nicht sehr konkret und anwenderfreundlich. Wir versuchen sie weiter zu spezialisieren, ohne allzuviel von der Allgemeinheit der Konstruktion zu verlieren. Die Konstruktion von Suchrichtungen, die wir im Vorfeld der Besprechung der Quasi-Newton-Richtungen der unrestringierten Optimierung angesprochen hatten, findet dabei eine Erweiterung in der aktuellen Betrachtung.

Satz 6.7 *Sei $(B_k)_{k \in \mathbf{N}}$ eine Folge symmetrischer $(n \times n)$ -Matrizen mit folgenden Eigenschaften (alle $k \in \mathbf{N}$)*

1. $\mathcal{R}(B_k) = \mathcal{N}(H)$ für alle $k \in \mathbf{N}$
2. B_k eingeschränkt auf $\mathcal{N}(H)$ ist positiv definit für alle $k \in \mathbf{N}$
3. Die positiven Eigenwerte aller B_k seien uniform beschränkt und von null wegbeschränkt.

Dann bilden die Richtungen

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

streng gradientenbezogene Suchrichtungen für (NLGO) und die Folge (x^k) .

Beweis: (vgl. [M1], Lemma 2.2)

Aus der ersten Eigenschaft folgt $HB_k = 0$, sodass $d^k \in \mathcal{N}(H)$ gilt. Ferner folgt aus der Symmetrie der B_k und der zweiten Eigenschaft, dass $\mathcal{N}(B_k) = \mathcal{N}(H)^\perp$ gilt. Daher gilt für jedes $r = r_1 + r_2 \in \mathbf{R}^n = \mathcal{N}(H) \oplus \mathcal{N}(H)^\perp$

$$B_k r = B_k r_1 = B_k P_H r$$

Des weiteren dürfen wir $P_H \nabla f(x^k) \neq 0$ voraussetzen, da sonst x^k KKT-Punkt von (NLG) ist. Daher gilt wegen der Eigenschaft 2.

$$\begin{aligned} \nabla f(x^k)^T d^k &= (P_H \nabla f(x^k))^T d^k = - (P_H \nabla f(x^k))^T B_k \nabla f(x^k) \\ &= - (P_H \nabla f(x^k))^T B_k P_H \nabla f(x^k) < 0 \end{aligned}$$

Also ist d_k Abstiegrichtung.

Nun stellen wir $\nabla f(x^k)$ über eine Orthonormalbasis $v_{k,1}, \dots, v_{k,n}$ von Eigenvektoren der symmetrischen Matrix B_k zu den Eigenwerten $\lambda_{k,1}, \dots, \lambda_{k,n}$ dar.

$$\nabla f(x^k) = \sum_{i=1}^n f_{k,i} v_{k,i}$$

Da B_k den Raum $\mathcal{N}(H)$ auf sich selbst abbildet, kann oBdA unterstellt werden, dass die ersten p Eigenvektoren, v_1, \dots, v_p eine Orthonormalbasis von $\mathcal{N}(H)$, die übrigen eine Orthonormalbasis von $\mathcal{N}(H)^\perp$ bilden. Daher sind nach Voraussetzung die ersten p Eigenwerte positiv, die restlichen null.

Dann lässt sich $\|d^k\|$ wie folgt darstellen

$$\|d^k\|^2 = \|B_k \nabla f(x^k)\|^2 = \sum_{i=1}^n (f_{k,i} \lambda_{k,i})^2 = \sum_{i=1}^p (f_{k,i} \lambda_{k,i})^2$$

Sind nun $\lambda_{k,\min}$ und $\lambda_{k,\max}$ minimaler und maximaler positiver Eigenwert der Matrix B_k , so folgt unmittelbar

$$\lambda_{k,\min}^2 \|P_H \nabla f(x^k)\|^2 \leq \|d^k\|^2 \leq \lambda_{k,\max}^2 \|P_H \nabla f(x^k)\|^2$$

Ferner ergibt sich

$$\begin{aligned} \nabla f(x^k)^T d^k &= - (P_H \nabla f(x^k))^T B_k P_H \nabla f(x^k) \\ &= - \sum_{i=1}^p f_{k,i}^2 \lambda_{k,i} \leq -\lambda_{k,\min} \|P_H \nabla f(x^k)\|^2 \\ &\leq -\frac{\lambda_{k,\min}}{\lambda_{k,\max}} \|P_H \nabla f(x^k)\| \|d^k\| \end{aligned}$$

Da nach Voraussetzung alle positiven Eigenwerte uniform beschränkt und von null wegbeschränkt sind, ist alles gezeigt. □

Zur Konstruktion einer im letzten Satz genannten Folge symmetrischer Matrizen B_k dienen die folgenden Überlegungen, bei denen wir eine *nichtsinguläre Transformationsmatrix* $R \in \mathbf{R}^{n \times n}$ einsetzen. P_{HR} bezeichne dabei die orthogonale Projektion auf den Nullraum $\mathcal{N}(HR)$. Dann gilt zunächst folgendes

Lemma 6.8 *Es sei eine nichtsinguläre Matrix $R \in \mathbf{R}^{n \times n}$ gegeben und $B := RP_{HR}R^T$ gesetzt. Dann hat B die Eigenschaften 1. und 2. des letzten Satzes.*

Beweis: Die Symmetrie der Matrix B ist unmittelbar einzusehen. Ferner gilt

$$HB = HRP_{HR}R^T = 0$$

so dass $\mathcal{R}(B) \subseteq \mathcal{N}(H)$. Gleichheit gilt hier, wenn die durch B induzierte lineare Abbildung injektiv auf $\mathcal{N}(H)$ ist. Wir zeigen, dass B sogar positiv definit auf $\mathcal{N}(H)$ ist.

Sei dazu $d \in \mathcal{N}(H)$ vorgegeben. Dann gilt

$$\begin{aligned} d^T B d &= d^T R P_{HR} R^T d \\ &= d^T R P_{HR} P_{HR} R^T d = \|P_{HR} R^T d\|^2 \geq 0 \end{aligned}$$

andererseits gilt

$$\begin{aligned} \|P_{HR} R^T d\|^2 = 0 &\implies \\ R^T d \in \mathcal{R}((HR)^T) &\implies \\ R^T d = R^T H^T y \text{ für ein } y \in \mathbf{R}^p &\implies \\ d \in \mathcal{R}(H^T) = \mathcal{N}(H)^\perp &\implies d = 0 \end{aligned}$$

□

Ferner prägen wir nun vorab den Begriff:

Definition 6.9 *Wir nennen eine Folge von Matrizen $R_k \in \mathbf{R}^{n \times n}$, $k \in \mathbf{N}$, voll singulärwertbeschränkt, wenn es Konstanten $\beta, \gamma > 0$ gibt so, dass für alle $y \in \mathbf{R}^n$ und alle $k \in \mathbf{N}$ gilt*

$$\beta y^T y \leq y^T R_k R_k^T y \leq \gamma y^T y$$

In der Tat sind die Wurzeln der Eigenwerte der symmetrischen Matrix $R_k R_k^T$ bekanntermaßen die Singulärwerte der Matrix R_k , so dass die Definition Sinn macht. Ferner gilt das folgende

Lemma 6.10 Sei $\{B_k\}_{k \in \mathbf{N}}$ eine Folge symmetrischer, positiv definiter Matrizen des $\mathbf{R}^{n \times n}$. Dann sind folgende Aussagen äquivalent:

1. Die Folgen $\{B_k\}_{k \in \mathbf{N}}$ und $\{B_k^{-1}\}_{k \in \mathbf{N}}$ sind beschränkt
2. Es gibt Konstanten $\beta, \gamma > 0$ so, dass für alle $y \in \mathbf{R}^n$ und alle $k \in \mathbf{N}$ gilt $\beta y^T y \leq y^T B_k y \leq \gamma y^T y$
3. Es gibt Konstanten $\beta, \gamma > 0$ so, dass für alle $y \in \mathbf{R}^n$ und alle $k \in \mathbf{N}$ gilt $\frac{1}{\gamma} y^T y \leq y^T B_k^{-1} y \leq \frac{1}{\beta} y^T y$

Beweis: ([GK1], Lemma 12.8) ■

□

Zu beachten ist ferner, dass über den Spektralsatz für jede symmetrische, positiv semidefinite Matrix $B \in \mathbf{R}^{n \times n}$ gilt:

$$\|B\| = \lambda_{\max}$$

sowie

$$\|B^{-1}\| = \frac{1}{\lambda_{\min}},$$

wenn diese positiv definit ist, wobei λ_{\max} der maximale und λ_{\min} der minimale Eigenwert der Matrix B seien. In diesem Sinne folgt auch

$$\lambda_{\min} y^T y \leq y^T B y \leq \lambda_{\max} y^T y$$

Für die Anwendung des Konvergenzsatzes 6.6 entscheidend ist nun der folgende

Satz 6.11 Es sei $(x^k)_{k \in \mathbf{N}}$ eine Folge von zulässigen Punkten von (NLGO), die in einem Kompaktum $K \subset \mathcal{U}$ liegen, und es sei $\{R_k\}_{k \in \mathbf{N}}$ eine Folge beliebiger reeller nichtsingulärer $n \times n$ Matrizen.

a) Sind die R_k voll singulärwert-beschränkt, so bilden die Richtungen

$$d^k := -R_k P_{HR_k} R_k^T \nabla f(x^k)$$

streng gradientenbezogene Abstiegsrichtungen bzgl (NLGO) und (x^k) .

b) Hängt die Matrix R stetig von $x \in \mathcal{U}$ ab und gilt $R_k = R(x^k)$ für alle k , so sind die R_k voll singulärwert beschränkt.

Beweis: (Vgl. [M1], Korollar 2.2 und Beispiele)

a) Nach dem Satz 6.7 und dem letzten Lemma genügt es zu zeigen, dass die nichtverschwindenden Eigenwerte der Matrix

$$B_k := R_k P_{HR_k} R_k^T$$

uniform beschränkt und von null wegbeschränkt sind. Wir wissen bereits, dass die Eigenwerte der Einschränkung von B_k auf $\mathcal{N}(H)$, $B_{k|N(H)} := \bar{B}_k$, alle positiv und die übrigen Eigenwerte von B_k alle null sind.

Schätzen wir als erstes alle Eigenwerte von B_k nach oben ab: Sei dazu $\lambda_{k,\max}$ maximaler Eigenwert von B_k und $y_{k,\max}$ ein zugehöriger Eigenvektor der Länge eins. Dann gilt

$$(y_{k,\max})^T B_k y_{k,\max} = \lambda_{k,\max} (y_{k,\max})^T y_{k,\max} = \lambda_{k,\max}$$

und es folgt

$$\begin{aligned} \lambda_{k,\max} &= (y_{k,\max})^T B_k y_{k,\max} \\ &= (y_{k,\max})^T R_k P_{HR_k} R_k^T y_{k,\max} \\ &= (y_{k,\max})^T R_k P_{HR_k}^T P_{HR_k} R_k^T y_{k,\max} \\ &= \|P_{HR_k} R_k^T y_{k,\max}\|^2 \\ &\leq \|R_k^T y_{k,\max}\|^2 \\ &= (y_{k,\max})^T R_k R_k^T y_{k,\max} \\ &\leq \gamma (y_{k,\max})^T y_{k,\max} = \gamma \end{aligned}$$

Damit sind alle Eigenwerte von \bar{B}_k uniform nach oben beschränkt. Zeigen wir nun, dass sie auch uniform von null wegbeschränkt sind.

Nehmen wir also an, dass die Folge der kleinsten positiven Eigenwerte der \bar{B}_k eine gegen null konvergierende Teilfolge $(\lambda_{k,\min})_{k \in K}$ hat. Seien $y_{k,\max} \in \mathcal{N}(H)$ mit $\|y_{k,\min}\| = 1$ eine zugehörige Folge. Diese besitzt, da sie in einem Kompaktum liegt, eine konvergente Teilfolge. Gehen wir also oBdA davon aus, dass bereits die $(y_{k,\min})_{k \in K}$ konvergieren, sagen wir gegen $\bar{y} \in \mathcal{N}(H)$. Ferner betrachte die zugehörige Folge der $R_k R_k^T$. Da auch diese nach Voraussetzung in einem Kompaktum liegt und deren Eigenwerte uniform beschränkt und von null wegbeschränkt sind, können wir oBdA davon ausgehen, dass die

Folge $\{R_k R_k^T\}_{k \in K}$ gegen eine symmetrische, positiv definite $n \times n$ Matrix E konvergiert.

Insgesamt betrachten wir nun die Folge

$$R_k P_{HR_k} R_k^T = R_k R_k^T \left(I - H^T (H R_k R_k^T H^T)^{-1} H R_k R_k^T \right)$$

für $k \in K$. Da die Matrix $H R_k R_k^T H^T$ grundsätzlich invertierbar ist und gegen HEH^T konvergiert und auch diese Matrix invertierbar ist, strebt die Folge der $H^T (H R_k R_k^T H^T)^{-1}$ gegen $(HEH^T)^{-1}$. Damit strebt die Folge der $R_k P_{HR_k} R_k^T$ gegen $E \left(I - H^T (HEH^T)^{-1} HE \right)$.

Über die Cholesky Zerlegung der Matrix E , $E = \bar{R} \bar{R}^T$, mit nichtsingulärer Matrix \bar{R} erhalten wir damit

$$0 = \lim_{k \in K} R_k P_{HR_k} R_k^T y_{k, \min} = \bar{R} P_{H\bar{R}} \bar{R}^T \bar{y}.$$

Nach Satz 6.14 folgt daraus aber $\bar{y} \in \mathcal{R}(H^T)$. Wegen $\mathbf{R}^n = \mathcal{R}(H^T) \oplus \mathcal{N}(H)$ widerspricht dies aber $\bar{y} \in \mathcal{N}(H)$, $\|\bar{y}\| = 1$. Damit ist nachgewiesen, dass die positiven Eigenwerte der B_k uniform von null wegbeschränkt sind.

b) Die Matrix $B(x) := R(x) R(x)^T$ sind positiv definit und Isomorphismen des \mathbf{R}^n . Auf dem Kompaktum K besitzt daher $\lambda_{k \max}(x)$ eine obere Schranke, wobei $\lambda_{k \max}(x)$ den größten Eigenwert der Matrix $B(x^k) = R(x^k) R(x^k)^T$ bezeichne. Also besitzen alle (positiven) Eigenwerte der Matrizen $B(x^k)$ eine obere Schranke.

Ferner ist $B(x)$ invertierbar mit größtem Eigenwert $\lambda_{\min}^{-1}(x)$, wobei $\lambda_{\min}(x)$ den kleinsten Eigenwert von $B(x)$ beschreibt. Für die $x \in K$ gibt es dafür eine gemeinsame obere Schranke. Daher sind die (positiven) Eigenwerte der Matrizen $B(x^k)$ von null wegbeschränkt.

Da nun die Eigenwerte der Matrizen $B(x^k)$ uniform beschränkt und von null wegbeschränkt sind, sind die Matrizen $\bar{R}(x^k)$ voll singulärwertbeschränkt.

□

Die Suchrichtungen $d := -B \nabla f(x) := -R P_{HR} R^T \nabla f(x)$ entstehen, wenn der Raum der Iterierten zunächst über R in einen Bildraum transformiert

werden, im Bildraum der steilste Abstieg gebildet und diese Richtung in den Originalraum zurücktransformiert wird.

Skizze:

Die Suchrichtungen können auch durch Lösen von Gleichungssystemen gewonnen werden.

Lemma 6.12 *Seien $R \in \mathbf{R}^{n \times n}$ nichtsingulär und $\bar{d} \in \mathbf{R}^n$ beliebig. Dann sind folgende Aussagen äquivalent:*

1. $\bar{d} = -RP_{HR}R^T \nabla f(x)$
2. es existiert ein $\bar{y} \in \mathbf{R}^p$ so, dass

$$\begin{pmatrix} -(RR^T)^{-1} & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} \bar{d} \\ \bar{y} \end{pmatrix} = \begin{pmatrix} \nabla f(x) \\ 0 \end{pmatrix} \quad (6.12)$$

3. für beliebiges $c \in \mathbf{R}^n$ existieren $\bar{y} \in \mathbf{R}^p$ und $\bar{z} \in \mathbf{R}^n$ so, dass

$$\begin{pmatrix} 0 & H^T & I \\ H & 0 & 0 \\ I & 0 & RR^T \end{pmatrix} \begin{pmatrix} \bar{d} \\ \bar{y} \\ \bar{z} \end{pmatrix} = \begin{pmatrix} c \\ 0 \\ RR^T(c - \nabla f(x)) \end{pmatrix} \quad (6.13)$$

4. $\bar{d} = -(RR^T)(\nabla f(x) - H^T \bar{y})$ mit $\bar{y} = (HRR^T H^T)^{-1} HRR^T \nabla f(x)$.

Beweis: [B1], Lemma 1.3.7 ■

Insbesondere berechnet sich \bar{d} praktisch aus dem Gleichungssystem

$$\begin{aligned} HRR^T H^T \bar{y} &= HRR^T \nabla f(x) \\ \bar{z} &= -H^T \bar{y} \\ \bar{d} &= -(RR^T)(\bar{z} + \nabla f(x)) \end{aligned}$$

das man unschwer aus (6.12) herleitet. Zu beachten ist dabei, dass die Matrizen RR^T und HRR^TH^T wegen der Rangvoraussetzungen symmetrisch und positiv definit sind, d.h. dass das zuletzt genannte Gleichungssystem eine eindeutige Lösung besitzt und das numerisch stabile Choleskyverfahren zur Lösung des Gleichungssystems $HRR^TH^T\bar{y} = HRR^T\nabla f(x)$ herangezogen werden kann. Auf der anderen Seite kann man wieder wie im letzten Abschnitt die QR Zerlegung von H^T heranziehen, um d zu dem System (6.12) zu berechnen. (siehe letzter Abschnitt)

Zu beobachten ist aber, dass man bei der Konstruktion fast immer mit der Matrix $B := RR^T$ auskommt und explizit nicht die Transformation R benötigt. Man kann also auch mit der positiv definiten Matrix B arbeiten anstelle mit der Transformationsmatrix R . Außerdem sei darauf hingewiesen, dass die Systemmatrizen aus den Positionen 2. und 3. des Lemmas dünn besetzt sind, wenn H dünn besetzt und R bzw. B z. B. als Diagonalmatrix gewählt wird.

Zu beachten ist ferner, dass für $R = I$ die im Lemma beschriebene Abstiegsrichtung gerade den auf $\mathcal{N}(H)$ projizierten negativen Gradienten beschreibt, welcher sich aus dem System

$$\begin{pmatrix} I & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} \bar{d} \\ \bar{y} \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ 0 \end{pmatrix}$$

oder alternativ aus

$$\begin{pmatrix} 0 & H^T & I \\ H & 0 & 0 \\ I & 0 & I \end{pmatrix} \begin{pmatrix} \bar{d} \\ \bar{y} \\ \bar{z} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -\nabla f(x) \end{pmatrix}$$

berechnet. Das System (??) liefert gerade die bekannte Formel für $\bar{d} = -P_H\nabla f(x)$.

Beispiel 6.13 *Ein gutes Beispiel für die obige Suchrichtung ergibt sich, wenn f zweimal stetig differenzierbar und $\nabla^2 f(x)$ positiv definit ist. In diesem Falle kann man die Cholesky Zerlegung*

$$\nabla^2 f(x)^{-1} = LL^T$$

betrachten und $R := L$ setzen. Dann berechnet sich die Suchrichtung aus

$$\begin{pmatrix} -\nabla^2 f(x) & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} \bar{d} \\ \bar{y} \end{pmatrix} = \begin{pmatrix} \nabla f(x) \\ 0 \end{pmatrix}$$

wodurch sich

$$\bar{d} = -\nabla^2 f(x)^{-1} (\nabla f(x) - H^T \bar{y})$$

mit $\bar{y} = (H\nabla^2 f(x)^{-1} H^T)^{-1} H\nabla^2 f(x)^{-1} \nabla f(x)$ ergibt.

Interessanterweise kann man das Gleichungssystem wiederentdecken, wenn man versucht, das KKT-System der Aufgabe (NLG) mit dem Newtonverfahren zu lösen. Dieses haben wir bereits als Lagrange-Newton Verfahren besprochen. Es sei hier noch einmal ausgeführt.

Das KKT-System von (NLG) lautet

$$\begin{aligned} -\nabla f(x) + H^T y &= 0 \\ Hx - h &= 0 \end{aligned}$$

und die zu (x, y) gehörige Newtonrichtung zur Lösung dieses Gleichungssystems berechnet sich bei zulässigem x aus

$$\begin{pmatrix} -\nabla^2 f(x) & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} \nabla f(x) - H^T y \\ 0 \end{pmatrix}$$

Setzt man nun $\bar{d} := \Delta x$ und $\bar{y} := y + \Delta y$, so erkennt man das obige Gleichungssystem. Das Newtonverfahren, angewendet auf das KKT-System, liefert also eine Abstiegsrichtung für (NLG). Multipliziert man darin die erste Zeile mit (-1) und ersetzt gleichzeitig die Variable y durch $(-y)$, so erkennt man das GLS des Lagrange-Newton Verfahrens wieder.

Variable Metrik

Analog zum Beispiel schließen wir: Die Konstruktion des transformierten steilsten Abstiegs funktioniert mit beliebiger invertierbarer Transformation $R \in \mathbf{R}^{n \times n}$. Es ist dann RR^T symmetrische und positiv definit. Genau so gut kann man die Cholesky-Zerlegung irgend einer positiv definiten $(n \times n)$ Matrix zur Konstruktion einer Suchrichtung für (NLG) verwenden.

Sei etwa $Q^{-1} := RR^T$, so ergibt sich die zugehörige Suchrichtung aus

$$\begin{pmatrix} -Q & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} \bar{d} \\ \bar{y} \end{pmatrix} = \begin{pmatrix} \nabla f(x) \\ 0 \end{pmatrix}$$

woraus ersichtlich ist, dass die Cholesky-Zerlegung nicht wirklich benötigt wird.

Hier wird eine *neue Metrik* verwendet, die der Aufgabenstellung angepasst ist.

Satz 6.14 Sei $R \in \mathbf{R}^{n \times n}$ eine beliebige invertierbare Matrix. Dann ist mit $Q := (RR^T)^{-1}$

$$\bar{d} = \arg \min \left\{ \nabla f(x)^T d : Hd = 0, d^T Q d \leq 1 \right\}$$

positives Vielfaches von

$$-RP_{HR}R^T \nabla f(x)$$

und dieser Vektor ist zulässige Abstiegsrichtung zu (NLGO) im Punkte x genau dann, wenn $\nabla f(x) \notin \mathcal{R}(H^T)$ gilt.

Beweis: [B1], Satz 1.3.3 ■

Im Satz wird der steilste Abstieg in $x \in V$ bzgl der Metrik $\langle u, v \rangle := u^T Q v$ gesucht! Man spricht daher bei Verfahren, die die genannte Konstruktion gradientenbezogener Abstiegsrichtungen benutzen, oft auch von **Verfahren variabler Metrik**.

6.3 Anwendung im allgemeinen Fall

Betrachten wir nun wieder das allgemeine linear restringierte Problem in der Standardform

$$(SPB) \quad \begin{cases} \text{minimiere} & f(x) \\ \text{s.d.} & Hx = h, x \geq 0 \end{cases}$$

und versuchen wir, unseren neuen Erkenntnisse zur Lösung dieser Aufgabe heranzuziehen. Wir haben zwei Möglichkeiten:

- Zum einen können die bereits früher angesprochene **Active-Set Strategie** heranziehen, bei der wir Abstiegsrichtungen immer nur im kleinsten Randpolyeder suchen, das die aktuelle Iterierte x enthält. Um eine solche zu finden, kann z.B. die Zielfunktion auf die affinen Hülle des Randpolyeders einschränken und diese Konstellation einer linear gleichungsrestringierten Optimierungsaufgabe als Ersatzaufgabe betrachten. *Die Idee* dabei ist, auf dem affinen Unterraum irgendeines der oben erarbeiteten reduzierten Verfahren anzuwenden. Will das Verfahren dabei die aktuelle Iterierte über den Rand hinaus führen, so wird diese auf dem Rand festgehalten, so dass die Iterierte in einem Randpolyeder niedriger Dimension liegt, dessen affine Hülle nun betrachtet wird. Diese Dimensionserniedrigung kann nur endlich oft vorkommen. Nach endlich vielen Dimensionserniedrigungen muss ein KKT Punkt der linear gleichungsrestringierten Ersatzaufgabe im zulässigen Bereich der Originalaufgabe liegen. Ist ein solcher numerisch gefunden, wird die Dimension des affinen Teilraums wieder erweitert.
- Zu anderen können wir von einem **Barriere Ansatz** Gebrauch machen, der die gegebene Aufgabe auch auf linear gleichungsrestringierte Ersatzaufgaben reduziert. *Grundidee* ist hierbei, die Vorzeichenbeschränkungen als einzige Ungleichungsrestriktionen über eine Barrierefunktion aus dem System zu entfernen und die verbleibende linear gleichungsrestringierte Ersatzaufgabe mit einem geeigneten Verfahren zu lösen.

Beide Ansätze sind in gewisser Weise komplementär: Während der erste Ansatz die Iterierten wie im Simplexverfahren im Wesentlichen über den *Rand*

des Polyeders führt, bewegen sich die Iterierten im zweiten Ansatz grundsätzlich nur in *Innern*.

Im Folgenden wollen wir beide Ansätze besprechen. Dabei beschränken wir uns auf die Darstellung der grundsätzlichen Vorgehensweise ohne spezielle Konvergenzbetrachtungen. Konkrete Ausgestaltungen nehmen wir nur für den Spezialfall der quadratischen Optimierung vor.

6.3.1 Active-Set Verfahren

Die Active-Set Strategie haben wir bereits so beschrieben, dass im Verfahren immer wieder ein Iterationspunkt erreicht wird, der im Inneren eines Randpolyeders liegt und (numerisch) ein KKT Punkt der Einschränkung der Aufgabenstellung auf die affine Hülle dieses Randpolyeders ist. Es ist zu klären, wie die das Verfahren danach weiter vorgehen kann. Im Prinzip muss nun zu einem höherdimensionalen Randpolyeder übergegangen werden, an dessen Rand der aktuelle Iterationspunkt liegt, um in diesem Randpolyeder einen neuen Iterationspunkt mit geringerem Zielwert zu finden. Dazu reicht es, eine beliebige zulässige Abstiegsrichtung im aktuellen Punkt zu finden, über die sich das Verfahren dann vom aktuellen Punkt wegbewegt und dabei die Zielfunktion echt absinkt.

Wir formulieren ein allgemeines Active-Set Verfahren für die Aufgabe

$$(SP) \quad \begin{cases} \text{minimiere} & f(x) \\ \text{s.d.} & Hx = h, Ax \leq a \end{cases}$$

wobei H maximalen Zeilenrang habe. Wir bezeichnen den zulässigen Bereich von (SP) mit S . Es sei ein Unterverfahren ausgewählt worden, mit dem die auf affine Untervektorräume eingeschränkte Aufgabe gelöst werden soll.

Allgemeines Active-Set Verfahren

- (0) **Start:** Wähle einen Punkt $x^0 \in S$. Setze $k = 0$ und $x = x^0$.
- (1) **Ersatzaufgabe:** Seien $A^1x = a^1$ die zu den aktiven Ungleichungen gehörigen Gleichungen von $Ax \leq a$. Setze

$$M = \begin{pmatrix} A^1 \\ H \end{pmatrix}$$

Hat M nicht maximalen Zeilenrang, so streiche geeignete Zeilen von M , bis maximaler Zeilenrang erreicht ist und berechne die Projektionsmatrix P_M . Betrachte die Ersatzaufgabe

$$(SPG_x) \quad \begin{cases} \text{minimiere} & f(x) \\ \text{s.d.} & Mx = \begin{pmatrix} a^1 \\ h \end{pmatrix} \end{cases}$$

(2) **Unterverfahren:**

- (a) Gilt $P_M \nabla f(x) = 0$, setze $k = k + 1$, $x^k = x$ und gehe zu (5).
- (b) Bestimme eine Abstiegsrichtung d und eine Schrittweite α gemäß dem gewählten Verfahren zur Lösung von (SPG_x) .
- (c) Ist $\bar{x} := x + \alpha d \in S$, setze $x := \bar{x}$ und gehe zu (2).

(3) **Aktivierung:** Sonst berechne eine maximale Schrittweite α_{\max} , mit der $x + \alpha_{\max} d \in S$ gilt. Setze $x := x + \alpha_{\max} d$ und gehe zu (1).

(4) **Deaktivierung:** Bestimme eine zulässige Abstiegsrichtung d zu (SP) im Punkte x und eine Schrittweite $\bar{\alpha}$ als

$$\bar{\alpha} = \arg \min \{f(x + \alpha d) : x + \alpha d \in V\} > 0$$

Geht dies nicht, **stopp** (x^k ist ein subminimaler Punkt zu (SP))

Sonst setze $x = x + \bar{\alpha} d$ und gehe zu (1)

Konvergenz des Verfahrens

Die Konvergenz dieser Vorgehensweise ist unter geringen Voraussetzungen leicht einzusehen.

Satz 6.15 *Die gegebene Aufgabe (SP) habe die Eigenschaft, dass die Einschränkung der Zielfunktion auf die affine Hülle eines jeden Randpolyeders nur jeweils endlich viele Funktionswerte von KKT Punkten besitzt. Ist die gegebene Aufgabe lösbar, dann endet das Verfahren nach endlich vielen Durchläufen.*

Beweis: Die Aufgabe (SP) sei als lösbar vorausgesetzt. Sei $x \in S$ ein beliebiger zulässiger Punkt von (SP) . Dann ist x nach Konstruktion ein innerer Punkt des (Rand-)Polyeders R , dessen affine Hülle den zulässigen Bereich von (SPG_x) darstellt. Ausgehend von x gibt es also eine zulässige Abstiegsrichtung d , die völlig innerhalb des zulässigen Bereichs von (SPG_x) verläuft und damit parallel zum Polyeders R . d ist damit gleichzeitig zulässige Abstiegsrichtung in x zu (SP) . Nach Konstruktion tritt diese Situation permanent auf, da immer wieder ein (Rand-)Polyeder passend gewählt wird, wenn die nächste Iterierte das Innere von R verlässt. Da es nur endlich viele Randpolyeder gibt, befindet sich die aktuelle Iterierte nach endlich vielen Schritten in einem Randpolyeder, in dessen Innerem alle zukünftigen Iterationspunkte liegen, die im aktuellen Unterverfahren (2) angelaufen werden. Das Unterverfahren berechnet also einen KKT Punkt von (SPG_x) , da die gegebene Aufgabe als lösbar vorausgesetzt wurde und daher die Zielfunktion auf dem letzten Randpolyeder nicht unbegrenzt abfallen kann.

Ist ein KKT Punkt der aktuellen Ersatzaufgabe (SPG_x) angelaufen, so folgt ein Schritt nach Baustein (4), der den aktuellen Zielwert echt absenkt. Das Verfahren kann also zu einem späteren Zeitpunkt diesen KKT Punkt bzw diesen Zielwert nicht mehr anlaufen, da die Folge der angelaufenen Zielwerte monoton fallend ist. Da nun nach Voraussetzung insgesamt nur endlich viele solcher anlaufbaren KKT Punkte existieren, endet das Verfahren nach endlich vielen Durchläufen.

□

Man beachte, dass die Voraussetzung des Satzes z. B. für jede Aufgabe mit konvexer oder mit quadratischer Zielfunktion gegeben ist. Ferner beachte man, dass die Unterfahren natürlich in der Regel unendlich verlaufen und daher numerisch abgebrochen werden müssen. Natürlich birgt der numerische Abbruch dann die Gefahr des Zykelns. Weil die anzulaufenden KKT Punkte im Verfahren unter der gemachten Voraussetzung diskret liegen, kann dem entgegen gewirkt werden, indem man die KKT Punkte bei fortschreitendem Verfahren immer schärfer berechnet. In der Praxis funktioniert das Verfahren. Will man es garantiert konvergent machen, so sollte man z.B. die Schritte im Baustein (4) im Rahmen eines konvergenten Mantelverfahrens wählen.

Die Bestimmung einer zulässigen Abstiegsrichtung im Punkte x im Baustein (4) des Verfahrens kann generell durch Lösen der Hilfsaufgabe

$$(ZA_x) \quad \begin{cases} \text{minimiere} & \nabla f(x)^T d \\ \text{s.d.} & A_1 d \leq 0, Hd = 0 \end{cases}$$

angegangen werden. Andererseits kann unter einer Nichtentartungs-Voraussetzung die zulässige Abstiegsrichtung wie folgt ermittelt werden.

Satz 6.16 *Vorgegeben sei eine beliebige Iterierte x des Verfahrens. Die im Baustein (1) des Verfahrens gebildete Matrix M habe ohne Streichen von Zeilen bereits maximalen Zeilenrang. Es sei $P_M = I - M^T (MM^T)^{-1} M$ die (orthogonale) Projektion auf den auf den zulässigen Bereich von (SPG_x) und sei $P_M \nabla f(x) = 0$, so dass x ein KKT Punkt der Aufgabe (SPG_x) darstellt. Es sei ferner $w = -(MM^T)^{-1} M \nabla f(x)$ und gemäß der Struktur von M sei $w^T =: (u^T, v^T)$.*

Ist dann $u \geq 0$, so ist x sogar ein KKT Punkt von (SP) . Gibt es dagegen eine negative Komponente u_j von u , so entferne die zu u_j gehörige Zeile von M , nenne die entstehende Matrix \hat{M} und bilde die Projektion $P_{\hat{M}}$. Dann ist $\hat{d} := -P_{\hat{M}} \nabla f(x)$ eine zulässige Abstiegsrichtung im Punkte x bzgl (SP) .

Beweis: (vgl. [BSS], Theorem 10.5.4) Nach Voraussetzung gilt $P_M \nabla f(x) = 0$. Daher gilt

$$\begin{aligned} 0 &= P_M \nabla f(x) = \left(I - M^T (MM^T)^{-1} M \right) \nabla f(x) \\ &= \nabla f(x) + M^T w = \nabla f(x) + A^{1T} u + H^T v \end{aligned} \quad (6.14)$$

Ist nun $u \geq 0$, so ist x ein KKT Punkt von (SP) , denn die KKT.Bedingungen dieser Aufgabe lauten

$$\begin{aligned} 0 &= \nabla f(x) + A^T \lambda + H^T \mu \\ Hx &= h, Ax \leq a \\ 0 &= \lambda^T (Ax - a), \lambda \geq 0 \end{aligned}$$

und diese sind erfüllt mit $\lambda^T = (u^T, 0)$ und $\mu = v$ in der durch $A^T = (A^{1T}, A^{2T})$ festgelegten Zerlegung.

Sei also $u_j < 0$ vorausgesetzt für einen Index j und sei \hat{M} wie im Satz festgelegt. Nach Voraussetzung ist \hat{M} von maximalem Zeilenrang.

Betrachte dann $P_{\hat{M}} = I - \hat{M}^T (\hat{M} \hat{M}^T)^{-1} \hat{M}$ und $\hat{w} := - (\hat{M} \hat{M}^T)^{-1} \hat{M} \nabla f(x)$. Nehmen wir an, es gelte $P_{\hat{M}} \nabla f(x) = 0$. Dann gilt:

$$0 = P_{\hat{M}} \nabla f(x) = \left(I - \hat{M}^T (\hat{M} \hat{M}^T)^{-1} \hat{M} \right) \nabla f(x) = \nabla f(x) + \hat{M}^T \hat{w} \quad (6.15)$$

Auf der anderen Seite lässt sich $A^{1T}u + H^T v$ schreiben als $\hat{M}^T \bar{w} + a_j^T u_j$, wobei a_j die j -te Zeile von A^1 sei. Damit gilt nach (6.14)

$$0 = \nabla f(x) + \hat{M}^T \bar{w} + a_j^T u_j \quad (6.16)$$

Bildet man die Differenz aus (6.14) und (6.15), so ergibt sich

$$0 = \hat{M}^T (\hat{w} - \bar{w}) + a_j^T u_j$$

und damit wegen $u_j \neq 0$ eine nichttriviale Darstellung der Form $M^T r = 0$, was der Rangvoraussetzung zu M widerspricht.

Also ist $\hat{d} = -P_{\hat{M}} \nabla f(x) \neq 0$ und eine Abstiegsrichtung in x :

$$\begin{aligned} \nabla f(x)^T \hat{d} &= -\nabla f(x)^T P_{\hat{M}} \nabla f(x) \\ &= -\nabla f(x)^T P_{\hat{M}}^T P_{\hat{M}} \nabla f(x) = -\|P_{\hat{M}} \nabla f(x)\|^2 < 0 \end{aligned}$$

Bleibt noch zu zeigen, dass \hat{d} eine zulässige Richtung ist. Dies ist sie genau dann, wenn $A^1 \hat{d} \leq 0$ und $H \hat{d} = 0$ gilt. Nun wissen wir bereits aus der Konstruktion, dass $\hat{A}^1 \hat{d} = 0$ und $H \hat{d} = 0$ gilt, wobei \hat{A}^1 die Matrix A^1 ohne die Zeile a_i beschreibe. Es genügt also nachzuweisen, dass $a_j \hat{d} \leq 0$ gilt. Dazu betrachten wir die mit $a_j P_{\hat{M}}$ vormultiplizierte Gleichung (6.16) und beachten $\hat{M} P_{\hat{M}} = 0$:

$$\begin{aligned} 0 &= a_j P_{\hat{M}} \left(\nabla f(x) + \hat{M}^T \bar{w} + a_j^T u_j \right) \\ &= -a_j \hat{d} + u_j a_j P_{\hat{M}} a_j^T \end{aligned}$$

Da aber $P_{\hat{M}}$ als Projektionsmatrix positiv semidefinit ist und daher wegen $u_j < 0$ auch $u_j a_j P_{\hat{M}} a_j^T \leq 0$ gilt, ist $-a_j \hat{d} \geq 0$. also gilt die Behauptung.

□

Ein klassischer Vertreter des allgemeinen Active-Set Verfahrens ist das **Verfahren projizierter Gradienten von Rosen**. Es verwendet als Unterverfahren das *reduzierte Gradientenverfahren* und als deaktivierende Suchrichtung die des letzten Satzes. Wir verzichten darauf, dieses Verfahren auszuformulieren.

6.3.2 Innere-Punkte Verfahren

Verfolgen wir nun den Ansatz, in der gegebenen Aufgabe (*SPB*) zunächst die Vorzeichenbeschränkungen über einen Barriereterm zu beseitigen und die verbleibende linear gleichungsrestringierte Ausgabe durch eines der erarbeiteten Verfahren zu lösen.

Barriereansatz

Wir benutzen dazu einen differenzierbaren Barriereansatz

$$h_g(x) = \sum_{i=1}^n g_i \varphi(x_i)$$

mit vorgewähltem Parametervektor $g^T = (g_1, \dots, g_n) \in \mathbf{R}_{++}^n$ und einer differenzierbaren Funktion $\varphi : \mathbf{R}_{++} \rightarrow \mathbf{R}_+$, die die Eigenschaft $\lim_{y \rightarrow 0^+} \varphi(y) = \infty$ besitzt. Beispielfunktionen hierzu sind

$$\varphi(t) = \frac{1}{t^p} \quad \text{mit } p > 0.$$

Die Aufgabe nimmt dann die folgende Gestalt an:

$$(BSPBg) \quad \begin{cases} \text{minimiere} & f(x) + \sum_{i=1}^n g_i \varphi(x_i) \\ \text{s.d.} & Hx = h \end{cases}$$

Diese Aufgabe kann mit jedem der oben besprochenen Verfahren bearbeitet werden! Zu beachten ist allerdings, dass die Zielfunktion nur auf $\mathcal{U} := \mathbf{R}_{++}^n$ definiert ist.

Wählt man nun eine Nullfolge $\{\mu_k\}_{k \in \mathbb{N}}$ mit $\mu_k > 0$ und setzt voraus, dass es zu (SPB) eine zulässiges x gibt mit $g(x) < 0$ (**Innere-Punkte-Annahme**), so sind die Voraussetzungen des Konvergenzsatzes 5.25 erfüllt und das Barriereverfahren, das sukzessive die Ersatzaufgaben $(BSPB\mu_k g)$ mit Zielfunktion

$$f(x) - \mu_k \sum_{i=1}^n g_i \varphi(x_i)$$

löst, liefert eine Iterationsfolge, deren jeglicher Häufungspunkt ein KKT Punkt von (SPB) ist.

Zu beachten ist allerdings, dass für jedes μ_k zur Ersatzaufgabe $(BSPB\mu_k g)$ ein *globaler Minimierer* bestimmt werden muss. Dies ist realistischerweise allenfalls für konvexe Zielfunktion erfüllbar und beinhaltet einen großen Rechenaufwand.

Verfahren mit transformiertem steilsten Abstieg

Wir entfernen uns nun vom eben beschriebenen "reinen" Barriere-Ansatz und verwenden als Hilfsfunktion φ die Logarithmusfunktion $\varphi = \ln$, die nicht mehr die Eigenschaft hat, dass alle ihre Funktionswerte nichtnegativ sind. Die Gesamtkonstruktion des Barriere-Verfahrens bleibt aber unverändert. In diesem Fall muss die Konvergenz des resultierenden Verfahrens erneut bewiesen werden.

Zu beachten ist, dass es bei der Barrierefunktion φ im Wesentlichen auf ihre "Überhöhungseigenschaft" in der Nähe des Nullpunktes ankommt, sodass die Funktion dort nichtnegativ sein sollte. Wenn Sie "weiter weg" negative Werte annimmt, sollte dies nicht weiter störend sein. In diesem Sinne kann $\varphi = \ln$ verwendet werden, wenngleich man diese Funktion für Argumente > 1 gleich null setzen und den Übergang differenzierbar gestalten könnte.

Erfreulicherweise ist die mit \ln verbogene Zielfunktion f tatsächlich konvex, sofern dies für die Ausgangsfunktion zutrifft. Denn die Ableitungen der Funktion $f_g(x) = f(x) - \sum_{i=1}^n g_i \ln x_i$ lauten

$$\nabla f_g(x) = \nabla f(x) - X^{-1}g \text{ und } \nabla^2 f_g(x) = \nabla^2 f(x) + X^{-2}g$$

wobei X^{-1} für die $(n \times n)$ Diagonalmatrix steht, deren Diagonalelemente $x_1^{-1}, \dots, x_n^{-1}$ sind, X^{-2} entsprechend. Die zweite Ableitung ist für $x > 0$ positiv definite Matrix, wenn nur $\nabla^2 f(x)$ positiv semidefinit ist.

Zur Vereinfachung des Gesamtverfahrens wollen wir überlegen, inwieweit es genügt, die Ersatzaufgaben nur lokal und ggf. nur unscharf zu lösen.

Die KKT Bedingungen zur Aufgabe (*BSPBg*) lauten

$$\begin{aligned}\nabla f(x) - X^{-1}g + H^T y &= 0 \\ Hx &= h\end{aligned}$$

mit Lagrangemultiplikatoren $y \in \mathbf{R}^m$.

Setzt man hierin zur Abkürzung $z := X^{-1}g > 0$, so schreibt sich das KKT System nun als

$$(Kg) \quad \begin{cases} \nabla f(x) + H^T y - z = 0 \\ Hx = h \\ Xz = g \end{cases}$$

Es fällt auf, dass dieses System große Ähnlichkeit zu dem KKT System von (*SPB*) aufweist.

$$(KKTSPB) \quad \begin{cases} \nabla f(x) + H^T y - z = 0, \quad z \geq 0 \\ Hx = h, \quad x \geq 0 \\ Xz = 0 \end{cases}$$

Ist $\|g\|$ sehr klein, so fehlen lediglich die Vorzeichenbeschränkungen $x \geq 0$ und $z \geq 0$, damit beide Systeme zumindest numerisch übereinstimmen.

Zu bemerken ist allerdings, dass bei (*Kg*) implizit $x > 0$ gilt. Aus der Komplementaritätsforderung folgt damit aber auch $z > 0$. Damit ist der Unterschied tatsächlich nur numerisch und wir erhalten bei sehr kleinem $\|g\|$ durch Lösen von (*Kg*) näherungsweise einen KKT Punkt von (*SPB*). Wir setzen

$$\phi_x(x, y, z) := \|\nabla f(x) + H^T y - z\|, \quad \phi_y(x, y, z) := \|h - Hx\|, \quad \phi_z(x, y, z) := \|Xz\|$$

als Maß für den Erfülltheitsgrad der KKT Bedingungen fest. Diese Werte müssen klein sein, wenn wir (x, y, z) als numerische Lösung von (*KKTSPB*) ansehen wollen.

Wir wollen nun das Barriereproblem (*BSPBg*) mit dem Verfahren des transformierten steilsten Abstiegs lösen und dabei eine Transformation spezieller Struktur verwenden: Sei x eine beliebige Iterierte. Wir wählen eine $n \times n$ Matrix R als Transformationsmatrix mit

$$RR^T = \left(Q + (SS^T)^{-1}\right)^{-1}$$

wobei S eine nichtsinguläre $n \times n$ Matrix und Q eine positiv semidefinite $n \times n$ Matrix seien In diesem Fall ist mit SS^T auch $(SS^T)^{-1}$ positiv definit und damit auch $(Q + (SS^T)^{-1})^{-1}$ positiv definit. Die Choleskyzerlegung dieser Matrix liefert uns die entsprechende Struktur RR^T .

Die Abstiegsrichtung Δx berechnet sich dann aus dem Gleichungssystem

$$\begin{pmatrix} -(RR^T)^{-1} & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \hat{y} \end{pmatrix} = \begin{pmatrix} \nabla f_g(x) \\ 0 \end{pmatrix}$$

also

$$\begin{pmatrix} -(Q + (SS^T)^{-1}) & H^T \\ H & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \hat{y} \end{pmatrix} = \begin{pmatrix} \nabla f(x) - X^{-1}g \\ 0 \end{pmatrix}$$

Setzen wir zur Abkürzung

$$\hat{z} := X^{-1}g - (SS^T)^{-1} \Delta x$$

so ergibt eine Umformung des Systems

$$\begin{pmatrix} -Q & H^T & I \\ H & 0 & 0 \\ I & 0 & SS^T \end{pmatrix} \begin{pmatrix} \Delta x \\ \hat{y} \\ \hat{z} \end{pmatrix} = \begin{pmatrix} \nabla f(x) \\ 0 \\ SS^T X^{-1}g \end{pmatrix}$$

Eine Lösung $(\Delta x, \hat{y}, \hat{z})$ dieses Systems erfüllt dann mit $Hx = h$, $\bar{y} = -\hat{y}$ das folgende System

$$\begin{aligned} \nabla f(x) + H^T \bar{y} - \hat{z} &= -Q\Delta x \\ h - Hx &= 0 \\ X\hat{z} &= G \left(e - G^{-1}X(SS^T)^{-1} \Delta x \right) \end{aligned}$$

(x, \bar{y}, \hat{z}) kann also als numerischer KKT Punkt von (SPB) gelten, wenn

$$\|Q\Delta x\| \text{ und } \left\| G \left(e - G^{-1}X(SS^T)^{-1} \Delta x \right) \right\|$$

klein sind und wenn $x > 0$ und $\hat{z} > 0$ gelten. Setzen wir (bei festem g) die Levelmenge

$$\mathcal{L}_{x_0}^g = \{x : Hx = h, f_g(x) \leq f_g(x^0)\}$$

als kompakt voraus, so sichert uns z. B. eine stetige Abhängigkeit der Matrizen S und Q von $x \in \mathbf{R}_+^n$ zu, dass

$$\|Q\Delta x\| \text{ nahe an null}$$

und

$$\left\| G \left(e - G^{-1}X (SS^T)^{-1} \Delta x \right) \right\| \text{ nahe an } \|g\|$$

liegt, sobald $\|\Delta x\|$ klein genug ist. Um $x > 0$ und $\hat{z} > 0$ sicherzustellen, sollte allerdings auch

$$G \left(e - G^{-1}X (SS^T)^{-1} \Delta x \right) > 0$$

gelten. $x > 0$ ist nach Konstruktion automatisch gegeben, $z > 0$ folgt dann im Nachhinein.

Konkret ist es vorteilhaft,

$$\left\| G^{-1}X (SS^T)^{-1} \Delta x \right\| \leq \delta < 1 \quad (6.17)$$

für ein vorgegebenes $\delta \geq 0$ zu erreichen. In diesem Fall hat man $x, z > 0$ und die offensichtlichen Abschätzungen

$$0 < X\hat{z} \leq g \quad (6.18)$$

sowie

$$\|X\hat{z} - g\| = \left\| GG^{-1}X (SS^T)^{-1} \Delta x \right\| \leq \delta \|g\| \leq \|g\|$$

Dies führt uns zu folgendem Verfahren

Allgemeines Innere-Punkte Verfahren für (SPB)

- (0) **Start:** Wähle einen Punkt $x \in V \cap \mathbf{R}_{++}^n$ und $\varepsilon_1, \varepsilon_2 \geq 0$, $\delta, \beta \in (0, 1)$, $\gamma > 0$ sowie $g \in \mathbf{R}_{++}^n$.
- (1) **Treibersschritt:** Wähle eine beliebige Suchrichtung und eine beliebige Schrittweite, die x zu einer neuen zulässigen Iterierten überführt.
- (2) **Suchrichtung:** Wähle $Q, S \in \mathbf{R}^{n \times n}$, Q positiv semidefinit, S nichtsingulär und berechne $(\Delta x, \hat{y}, \hat{z})$ über

$$\begin{pmatrix} -Q & H^T & I \\ H & 0 & 0 \\ I & 0 & SS^T \end{pmatrix} \begin{pmatrix} \Delta x \\ \hat{y} \\ \hat{z} \end{pmatrix} = \begin{pmatrix} \nabla f(x) \\ 0 \\ SS^T X^{-1}g \end{pmatrix} \quad (6.19)$$

(3) **Ab-/Umbruch:**

a) Gilt $x > 0$ und $\hat{z} > 0$ und $x_i \hat{z}_i \leq \varepsilon_1$ für $i = 1, \dots, n$ sowie

$$\|\nabla f(x) + H^T \bar{y} - \hat{z}\| \leq \varepsilon_2,$$

Stopp (numerischer KKT Punkt)

b) Gilt $x > 0$ und $\hat{z} > 0$ und $\|X\hat{z} - g\| \leq \gamma \|g\|$ sowie

$$\|\nabla f(x) + H^T \bar{y} - \hat{z}\| \leq \gamma \|g\|,$$

setze $g = \beta g$. Gehe zu (1).

(4) **Schrittweite:** Setze $x := x + \alpha \Delta x$ mit der Eigenschaft:

$$\alpha := \max \{ \beta^l : l = 0, 1, 2, 3, \dots \},$$

so dass

$$x + \alpha \Delta x > 0 \text{ und } \tilde{f}(x + \alpha \Delta x) \leq \tilde{f}(x) + \sigma \alpha \nabla \tilde{f}(x)^T \Delta x$$

und gehe zu (2).

Bemerkung: Der Treiberschritt gibt dem Verfahren mehr praktische Gestaltungsfreiheit. Hier kann ein beliebiger Schritt durchgeführt werden, der Rest des Verfahrens gibt als Mantelverfahren Konvergenzsicherheit.

Konvergenzbetrachtung

Um zu erreichen, dass das Verfahren nach endlich vielen Schritten mit einem numerischen KKT von (SPB) stoppt, sollte es nach jeweils endlich vielen Schritten zu einer Reduzierung von g gelangen: $g = \beta g$. Dazu muss erreicht werden, dass nach jeweils endlich vielen Schritten die in (2) b) genannten Voraussetzungen erfüllt sind. Gilt dabei gleichzeitig $\gamma \|g\| \leq \varepsilon_2$ und $g_i \leq \varepsilon_1$ für alle $i = 1, \dots, n$, so sind auch die Voraussetzungen für (2) a) erfüllt und das Verfahren stoppt.

Um zu klären, dass bei festem g nach endlich vielen Schritten die Bedingungen von (2) b) greifen, machen wir uns zunächst klar, dass der einzelne Treiberschritt kein Hindernis darstellt, wenn diese Aussage unabhängig von

der jeweiligen Ausgangsiterierten ist. Um dies abzuklären, bemühen wir Satz 6.11. Sind dessen Voraussetzungen erfüllt, ist insbesondere nach dem Treibersschritt \mathcal{L}_x^g (x die Iterierte nach dem Treibersschritt) kompakt, so wählt das Verfahren in dieser Phase streng gradientenbezogene Suchrichtungen. Dann besagt Satz 6.6, dass die Folge der Suchrichtungen eine Nullfolge bildet. Dies bedeutet, dass

$$\left\| G^{-1}X (SS^T)^{-1} \Delta x \right\| \leq \delta < 1 \quad \text{und damit} \quad \|X\hat{z} - g\| \leq \gamma \|g\|$$

sowie

$$\|Q\Delta x\| \leq \gamma \|g\|$$

erreicht werden kann, wenn nur $\left\| G^{-1}X (SS^T)^{-1} \right\|$ und $\|Q\|$ beschränkt bleiben.

Betrachten wir also die Voraussetzungen von Satz 6.11. Damit dieser Satz Anwendung findet, müssen bei Kompaktheit von \mathcal{L}_x^g einzig und allein die Transformationen voll singularwert-beschränkt gewählt werden.

Für die verwendete Transformation hatten wir $(RR^T)^{-1} = Q + (SS^T)^{-1}$ angesetzt, Q positiv semidefinit und S nichtsingulär. Setzt man hierin die Folge der S als voll singularwert-beschränkt und die der Q als beschränkt voraus, so gilt wegen

$$y^T \left(Q + (SS^T)^{-1} \right) y = y^T Q y + y^T (SS^T)^{-1} y$$

für alle $y \in \mathbf{R}^n$ mit $\|y\| = 1$, dass auch die Folge der $(RR^T)^{-1}$ und damit die der RR^T uniform beschränkt und von null wegbeschränkt ist.

Bleibt zu klären, wann die Menge \mathcal{L}_x^g kompakt ist. Nun, hierzu genügt es, den zulässigen Bereich M der Aufgabe (SPB) als beschränkt bzw. kompakt zu fordern. Dann nämlich ist die Zielfunktion f auf M beschränkt und aus der Konstruktion der Barrierefunktion bleibt die abgeschlossene Menge $\mathcal{L}_x^g \subset M$ vom Rand wegbeschränkt und ist damit unabhängig von der Wahl von g kompakt.

Damit haben wir geklärt:

Satz 6.17 *Das allgemeine Innere-Punkte Verfahren verläuft unter folgenden Voraussetzungen endlich:*

1. der zulässigen Bereich der Aufgabe (SPB) ist kompakt
2. die Folge der nichtsingulären Matrizen S ist voll singulärwert-beschränkt
3. die Folge der Matrizen Q ist beschränkt

Zur Wahl von Q und S

- Eine besonders einfache Wahl von Q ist $Q = 0$. In diesem Fall gilt automatisch $\|\nabla f(x) + H^T \bar{y} - \hat{z}\| = 0$ und es kann im Verfahren auf die Kontrolle von $\|\nabla f(x) + H^T \bar{y} - \hat{z}\| \leq \gamma \|g\|$ verzichtet werden.
- Ist S eine Diagonalmatrix, so ist eine weitere gute Wahlmöglichkeit für Q als Diagonalmatrix

$$Q = G_0^{-1} X (SS^T)^{-1}$$

wobei G_0 die Diagonalmatrix mit dem anfänglichen $g = g^0$ auf der Diagonalen ist. Die Beschränktheit dieser positiv definiten Matrizen Q auf der Iteriertenfolge ergibt sich aus Beschränktheit von x auf dem Kompaktum \mathcal{L}_x^g und der Beschränktheit der voll singulärwert-beschränkten Matrizen S .

Dann gilt nach der k -ten Anpassung von g :

$$\begin{aligned} \beta^{-k} \|Q\Delta x\| &= \left\| \beta^{-k} G_0^{-1} X (SS^T)^{-1} \Delta x \right\| \leq \delta < 1 \\ \iff \|Q\Delta x\| &\leq \delta \beta^k < \beta^k \end{aligned}$$

und auch in diesem Falle kann im Verfahren auf die Kontrolle von $\|\nabla f(x) + H^T \bar{y} - \hat{z}\| \leq \gamma \|g\|$ verzichtet werden.

- Ist f konvex, so bietet es sich an, $Q = \nabla^2 f(x)$ zu wählen. Q ist dann positiv semidefinit und beschränkt, da $\nabla^2 f$ stetig von x auf einem Kompaktum abhängt.

Diese Wahl ist motiviert durch das Newtonverfahren angewendet auf das KKT System (Kg) in der Form

$$\begin{aligned} H^T y + z - \nabla f(x) &= 0 \\ Hx - h &= 0 \\ z - X^{-1}g &= 0 \end{aligned}$$

Zu diesem System wird die Newton Richtung berechnet aus

$$\begin{pmatrix} -\nabla^2 f(x) & H^T & I \\ H & 0 & 0 \\ GX^{-2} & 0 & I \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \begin{pmatrix} \nabla f(x) - H^T z \\ h - Hx \\ X^{-1}g - z \end{pmatrix}$$

Setzt man hierin $\bar{y} := y + \Delta y$ und $\hat{z} := z + \Delta z$, so ergibt sich

$$\begin{pmatrix} -\nabla^2 f(x) & H^T & I \\ H & 0 & 0 \\ I & 0 & G^{-1}X^2 \end{pmatrix} \begin{pmatrix} \Delta x \\ \bar{y} \\ \hat{z} \end{pmatrix} = \begin{pmatrix} \nabla f(x) \\ h - Hx \\ x \end{pmatrix}$$

Dies stimmt für $S = G^{-\frac{1}{2}}X$ und zulässigem x mit dem System (6.19) überein. Zu beachten ist, dass Q bei dieser Wahl i.a. nicht dünn besetzt ist.

- Im nichtkonvexen Fall kann Q ähnlich zu $\nabla^2 f(x)$ gewählt werden, etwa wie bei den Quasi-Newton Verfahren.

Es ist generell anzuraten, aus Gründen der einfachen Berechnungsmöglichkeiten S als Diagonalmatrix zu wählen. Eine gute Wahl ist die aus dem Newtonverfahren

$$S = G^{-\frac{1}{2}}X$$

Zu beachten ist, dass diese Wahl von S bei festem G voll singulärwert-beschränkt ist, da $x \in \mathcal{L}_{x^0}^g$ gilt. In diesem Fall ist

$$\begin{aligned} \left\| G^{-1}X (SS^T)^{-1} \Delta x \right\| &\leq \delta < 1 \iff \\ \left\| X^{-1} \Delta x \right\| &\leq \delta < 1 \end{aligned}$$

Auch der Ansatz $S = \mu I$ mit kleinem μ lässt sich vertreten. Dann ist $(RR^T)^{-1} = Q + (SS^T)^{-1} = Q + \mu^{-2}I$ vom LEVENBERG-MARQUART Typ.

6.3.3 Spezialfall: Quadratische Optimierung

Wir notieren ein oft (auch in kommerziellen Programmen) verwendetes *Active-Set Verfahren* der streng konvexen quadratischen Optimierung.

Um das oben notierte Allgemeine Active-Set Verfahren zu konkretisieren und zu illustrieren, betrachten wir die Aufgabe

$$(QP) \quad \begin{cases} \text{minimiere} & f(x) = \frac{1}{2}x^T Qx + p^T x \\ \text{s.d.} & Ax \leq a, Hx = h \end{cases}$$

mit *streng konvexer* quadratischer Zielfunktion. Bei streng konvexer Zielfunktion f konvergiert das Lagrange-Newtonverfahren zur Minimierung der Funktion unter affin-linearen Nebenbedingungen ohne Ungleichungsrestriktionen in *einem* Schritt. Es ist also angeraten, ausgehend von einem zulässigen Punkt x die Zielfunktion über der affinen Hülle H des niedrigdimensionalsten Randpolyeders \bar{M} , das x enthält, zu minimieren. Die Newtonrichtung Δx ergibt dann über $\bar{x} := x + \Delta x$ unmittelbar das Minimum von f über H .

Es bietet sich also an, als Unterverfahren im allgemeinen active-set Verfahren im Falle streng konvexer quadratischer Zielfunktion das Lagrange Newton Verfahren einzusetzen.

Die folgende Formulierung des Verfahrens lehnt sich an [GK2], Algorithmus 5.3 an.

Active-Set Algorithmus für (QP)

- (0) **Start:** Vorgegeben sei ein zulässiges $x \in \mathbf{R}^n$ zu (QP). Bezeichne I die Menge der in x aktiven Indizes. Setze den Zähler $k := 0$. Sei ferner $r \in \mathbf{N}_0$ vorgegeben.
- (1) **Suchrichtung:** Bestimme $(\Delta x, \lambda_a, \mu)$ aus

$$\begin{pmatrix} Q & (A_a)^T & H^T \\ A_a & 0 & 0 \\ H & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \lambda_a \\ \mu \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ 0 \\ 0 \end{pmatrix} \quad (6.20)$$

und setze $\lambda_j := 0$ für alle $j \notin I$

- (2) **Aktualisierung:**

- (a) Ist $\Delta x = 0$ und $\lambda_i = 0$, **stopp**, sonst bilde den Index $i := \arg \min \{\lambda_j : 1 \leq j \leq m\}$ und entferne die Restriktion i aus I .

- (b) Ist $\Delta x \neq 0$ und $x + \Delta x$ zulässig zu (QP) , so setze $x := x + \Delta x$ und $k := 0$.

Sonst bilde den Index

$$i = \arg \min \left\{ \frac{a_j - A_j \cdot x}{A_j \cdot \Delta x} : j \notin I \text{ mit } A_j \cdot \Delta x > 0 \right\},$$

füge den Index i zu I hinzu und setze

$$\begin{aligned} \alpha & : = \frac{a_i - A_i \cdot x}{A_i \cdot \Delta x} \\ x & : = x + \alpha \Delta x \end{aligned}$$

Ist $\alpha = 0$, so setze $k := k + 1$, andernfalls setze $k := 0$.

- (3) **Sprung:** Ist $k > r$, so berechne ein zulässiges \bar{x} mit $f(\bar{x}) < f(x)$ und setze $x := \bar{x}$ sowie $k := 0$.

Gehe zu (1).

Zu bemerken ist, dass das Gleichungssystem (6.20) gemäß dem Satz zum Lagrange-Newton Verfahren immer eindeutig lösbar ist, wenn die Matrix $\begin{pmatrix} A_a \\ H \end{pmatrix}$ maximalen Zeilenrang hat. Dies ist natürlich immer durch Weglassen von Zeilen (und Spalten) zu erreichen.

Der Zähler k hat die Aufgabe zu zählen, wie oft *hintereinander* ein entarteter Schritt gemacht wird. Wird diese Anzahl zu groß, so ist auf andere Art und Weise eine zulässige Abstiegsrichtung zu berechnen. Dabei genügt es, wenn ein einziger Schritt dieses Verfahrens gemacht wird. Es kommt nur darauf an, dass das Gesamtverfahren sich von der Stelle bewegt.

Es gilt dann der

Satz 6.18 *Der Algorithmus endet nach endlich vielen Schritten mit einer Optimallösung von (QP) .*

Inhaltsverzeichnis

1	Einleitung	2
2	Theorie linear restringierter Optimierung	7
2.1	Konvexität	7
2.1.1	Charakterisierung konvexer Funktionen	8
2.1.2	Spezialfall: Quadratische Optimierung	10
2.1.3	Übungsaufgaben	11
2.2	Optimalität	12
2.2.1	Subminimale Punkte	12
2.2.2	KKT-Punkte	15
2.2.3	Fritz John Bedingungen	18
2.2.4	Spezialfall: Quadratische Optimierung	20
2.3	Dualität	21
2.3.1	Lagrange Dual	21
2.3.2	Dualitätssätze	23
2.3.3	Spezialfall: Quadratische Optimierung	28

I	Simplexverfahren	36
3	Dreiphasenalgorithmen	37
3.1	Lineare Techniken	37
3.1.1	Transformation auf Standardformat	38
3.1.2	Linearer Dreiphasenalgorithmus	40
3.1.3	Spezialfall: quadratische Optimierung	48
3.1.4	Übungsaufgaben	53
3.2	Quadratisches Simplexverfahren	54
3.2.1	Beschreibung des Verfahrens	55
3.2.2	Beispiele	63
3.2.3	Zur praktischen Durchführung des Verfahrens	69
3.2.4	Übungsaufgaben	75
3.3	Nichtlineares Simplexverfahren	76
3.3.1	Konstruktion von Suchrichtungen	76
3.3.2	Das Verfahren	80
3.3.3	Spezialfall: Quadratische Optimierung	90
3.3.4	Übungsaufgaben	94
4	Quadratische KKT-Simplexverfahren	95
4.1	Primal-Duale Verfahren	98
4.1.1	Das Verfahren von Barankin und Dorfman	98
4.1.2	Das Verfahren von Frank und Wolfe	104
4.1.3	Übungsaufgaben	108
4.2	Die Verfahren von Wolfe	109
4.2.1	Die kurze Form	109
4.2.2	Die lange Form	114
4.2.3	Vereinfachte Formen	117
4.2.4	Übungsaufgaben	123

4.3	Das Verfahren von Lemke	124
4.3.1	Die Idee von Lemke	124
4.3.2	Konvergenzanalyse	128
4.3.3	Spezialfall Quadratische Optimierung	130
4.3.4	Übungsaufgaben	132

II Allgemeinere Konzepte 133

5 Unrestringierte Optimierung 134

5.1	Sequentielle Approximation	136
5.1.1	Sequentielle lineare Approximation	136
5.1.2	Sequentielle quadratische Approximation	142
5.1.3	Spezialfall: Quadratische Optimierung	151
5.1.4	Übungsaufgaben	153
5.2	Verfahren konjugierter Richtungen	154
5.2.1	Das Verfahren konjugierter Gradienten	154
5.2.2	Das Verfahren von Fletcher-Reeves	159
5.2.3	Quasi Newton Verfahren	163
5.2.4	Spezialfall: Quadratische Optimierung	171
5.2.5	Übungsaufgaben	174
5.3	Anwendung im allgemeinen Fall	175
5.3.1	Barriere Ansatz	176
5.3.2	Penalty Ansatz	182
5.3.3	Spezialfall: Quadratische Optimierung	187
5.3.4	Übungsaufgaben	188

6	Linear gleichungsrestringierte Optimierung	189
6.1	Reduzierte Verfahren	189
6.1.1	Verwendung von Nullraum-Matrizen	190
6.1.2	Konstruktion von Verfahren	192
6.1.3	Spezialfall: Quadratische Optimierung	199
6.2	Verfahren gradientenähnlicher Abstiegsrichtungen	205
6.2.1	Gradientenähnliche Suchrichtungen	205
6.2.2	Gradientenbezogene Suchrichtungen	208
6.2.3	Transformierter steilster Abstieg	215
6.3	Anwendung im allgemeinen Fall	225
6.3.1	Active-Set Verfahren	226
6.3.2	Innere-Punkte Verfahren	231
6.3.3	Spezialfall: Quadratische Optimierung	239