Die Penalty-Methode von Pantoja und Mayne

Rolf Felkel[1] [2]

Zusammenfassung: Es wird eine global konvergente Methode zur Minimierung einer nichtlinearen Funktion unter gleichzeitiger Einhaltung nichtlinearer Restriktionen vorgestellt. Der dargestellte Ansatz eignet sich aufgrund seiner theoretischen und numerischen Besonderheiten zur Lösung großer Optimierungsprobleme. Die Theorie der nichtlinearen Optimierung und des Verfahrens werden angegeben und auf die Implementierung sowie auf Resultate wird eingegangen.

1 Einleitung

In der Numerik nichtlinearer Optimierungsprobleme richtet sich das Augenmerk der Forschung in den letzten Jahren immer mehr auf große, schwach besetzte Probleme. Diese entstehen zum Beispiel aus der Diskretisierung von Controling Problemen oder aus der Modellierung von Netzwerken. Bei diesen Problemen hängen die Restriktionen nur von sehr wenigen Variablen ab, was den Einsatz spezieller Speichermethoden erlaubt. Aus diesem Grunde sind die Verfahren, die zur Lösung kleiner und mittlerer Probleme geeignet sind für große Probleme auszuschließen. Der, bei exakten Matrixzerlegungen (z.B. QR), auftretende “fill in” steigert den Speicher- und Rechenaufwand erheblich. Deshalb müssen für große, schwach besetzte Optimierungsprobleme neue Wege gefunden werden.

Der Ansatz, der in dieser Arbeit beschrieben wird, bedient sich einer Methode, die unabhängig von der Dimension des Problems zu sehen ist. Die Penalty-Methode von Pantoja und Mayne [11] reiht sich ein in eine Klasse von Penalty-Methoden, die die Eigenschaft haben jede lokale Lösung eines Optimierungsproblems eindeutig als lokale Minimalstelle einer Penalty-Funktion wiederzugeben [16] [7]. Diesen Vorteil erkauft man sich mit der Tatsache, daß die Penalty-Funktion nicht differenzierbar ist. Aus diesem Grund können die Minimierungsverfahren für unrestringierte Probleme nicht angewendet werden.

[1] TU Darmstadt, Fachbereich Mathematik, Schloßgartenstraße 7, 64289 Darmstadt
[2] e-mail: felkel@mathematik.tu-darmstadt.de

Das Konzept der sequentiellen quadratischen Optimierung (SQP) sieht vor Abstiegsrichtungen durch lokale Approximation des nichtlinearen Optimierungsproblems mit einen quadratischen Unterproblem zu bestimmen. Bei der Lösung dieser quadratischen Optimierungsprobleme ist der Einsatz spezieller Verfahren für große Probleme nötig [4].

Im Folgenden wird zunächst auf die Theorie der nichtlinearen Optimierung eingegangen bevor die Methode von Pantoja und Mayne [11] genauer untersucht wird. Ferner werden Einblicke in das Lösungsverfahren für die quadratischen Unterproblemen gegeben und Resultate werden präsentiert.

2 Theorie der nichtlinearen Optimierung

In diesem Abschnitt soll kurz auf die theoretischen Grundlagen der nichtlinearen Optimierung eingegagen werden. Die angegebenen Definitionen und Sätze können in [16] und [5] nachgeschlagen werden.

Definition 2.1 *Unter einem nichtlinearen Optimierungsproblem (NLO) versteht man die Aufgabe der Minimierung einer nichtlinearen Objektfunktion unter gleichzeitiger Einhaltung nichtlinearer Restriktionen:*

$$\text{Minimiere } f(x)$$

$$\text{unter} \quad \begin{cases} g_i(x) & \geq & 0 & i \in \{1,\ldots,m\} \\ h_j(x) & = & 0 & j \in \{1,\ldots,p\} \end{cases} \tag{NLO}$$

Dabei sind $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ und $h : \mathbb{R}^n \to \mathbb{R}^p$. ■

Bezüglich der Restriktionen unterscheidet man in zulässige und unzulässige Punkte.

Definition 2.2 *Die Menge*

$$\Omega \stackrel{\text{def}}{=} \left\{ x \in \mathbb{R}^n : \begin{array}{ccc} g(x) & \geq & 0 \\ h(x) & = & 0 \end{array} \right\}. \tag{2}$$

ist die zulässige Menge und $x \in \Omega$ heißt zulässiger Punkt. ■

Als Grundvoraussetzungen sollen gelten

1. $\Omega \neq \emptyset$.
2. Ω abgeschlossen in $\mathbb{R}^n$.
3. $f, g, h \in C^2(\mathbb{R}^n)$.

Die Definition der Lösung ergibt sich auf natürliche Weise.

Definition 2.3 *Ein Punkt $x^* \in \Omega$ heißt lokale Lösung von (NLO), wenn eine Umgebung $\mathcal{U}(x^*)$ existiert, so daß für alle $x \in \Omega \cap \mathcal{U}(x^*)$ $f(x) \geq f(x^*)$ gilt und x^* ist eine strenge lokale Lösung, wenn $f(x) > f(x^*)$ erfüllt ist für alle $x \in (\Omega \cap \mathcal{U}(x^*)) \setminus \{x^*\}$.* ■

Beispiel HS16 ($n=2, m=5, p=0$)

$$f(x) = 100(x_2-(x_1)^2)^2+(1-x_1)^2 \quad \text{und} \quad g(x) = \begin{pmatrix} x_1+(x_2)^2 \\ (x_1)^2+x_2 \\ -x_1+0.5 \\ x_1+0.5 \\ -x_2+1 \end{pmatrix} \geq 0$$

Lösungen $x_l^* = \binom{-0.5}{\sqrt{0.5}}$ (lokal) und $x_g^* = \binom{0.5}{0.25}$ (global). Siehe Bild 1.

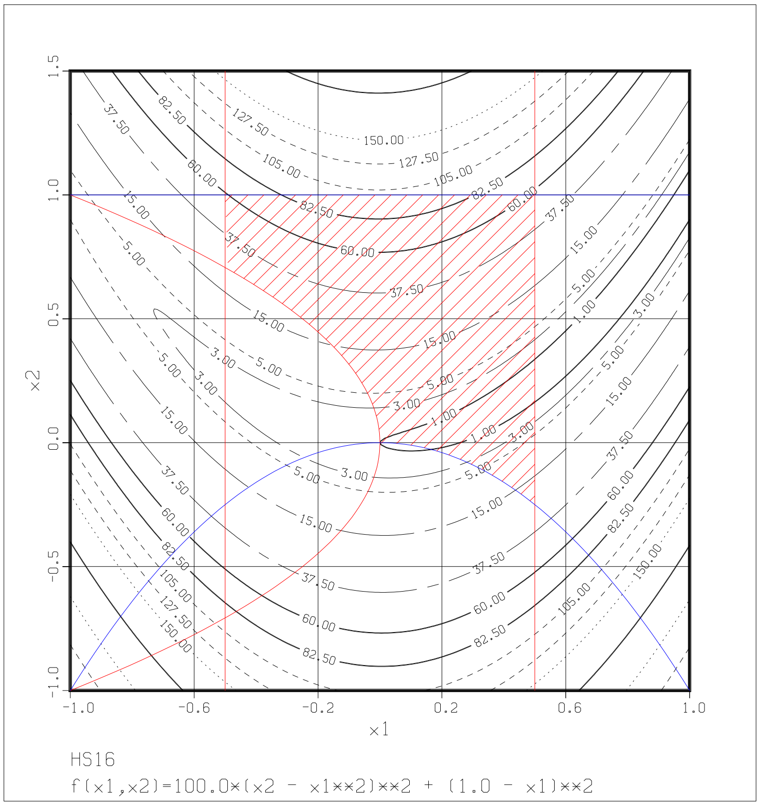

Bild 1 : Zulässige Menge von HS16.

2.1 Notwendige Bedingung erster Ordnung

Eine der wichtigsten Optimalitätsbedingungen ist die notwendige Bedingung erster Ordnung. Sie wurde von Karush bzw. Kuhn und Tucker unabhängig voneinander formuliert. Voraussetzung für diese Bedingung ist eine Einschränkung an die Restriktionen. Nur wenn die Restriktionsqualifikation von Guignard erfüllt ist, kann folgender Satz beweisen werden.

Satz 2.4 *Sei x^* eine lokale Lösung von (NLO) und in x^* gilt die Restriktionsqualifikation von Guignard, dann existieren $\lambda \in \mathbb{R}^m_+$ und $\mu \in \mathbb{R}^p$ mit*

$$\begin{aligned} \nabla f(x^*) - \nabla g(x^*)\lambda^* - \nabla h(x^*)\mu^* &= 0 \\ g(x^*)^T\lambda^* &= 0. \end{aligned} \qquad \text{(KKTB)}$$

Beweis Zum Beweis siehe ([16] Seite 24) ■

Die Vektoren λ und μ werden Lagrangemultiplikatoren genannt und die Funktion

$$L(x;\lambda,\mu) \stackrel{\text{def}}{=} f(x) - g(x)^T\lambda - h(x)^T\mu \tag{4}$$

heißt Lagrangefunktion. Mit $L(x;\lambda,\mu)$ läßt sich die erste Zeile von (KKTB) schreiben als

$$\nabla_x L(x^*;\lambda^*,\mu^*) = 0.$$

2.2 Restriktionsqualifikationen

Im Folgenden ist eine Unterscheidung der Indizies der Ungleichungsrestriktionen sehr hilfreich.

$$\begin{aligned}
\mathcal{A}(x) &\stackrel{\text{def}}{=} \{i \in \{1,\dots,m\} : g_i(x) = 0\} \\
\mathcal{B}(x) &\stackrel{\text{def}}{=} \{i \in \{1,\dots,m\} : g_i(x) > 0\} \\
\mathcal{V}(x) &\stackrel{\text{def}}{=} \{i \in \{1,\dots,m\} : g_i(x) < 0\}.
\end{aligned}$$

Dabei ist $\mathcal{A}(x)$ die Menge der aktiven Restriktionen und $\mathcal{V}(x)$ die Menge der verletzten.

Definition 2.5 *In $x^* \in \Omega$ seien die beiden Mengen*

$$T(\Omega,x^*) \stackrel{\text{def}}{=} \left\{ z \in \mathbb{R}^n : \; \textit{ex.} \; \begin{pmatrix} \alpha_k > 0, \; \textit{und} \\ \Omega \ni x^k \to x^* \end{pmatrix} \; \textit{mit } z = \lim_{k\to\infty} \alpha_k (x^k - x^*) \right\}$$

und

$$Z(x^*) \stackrel{\text{def}}{=} \left\{ z \in \mathbb{R}^n : \; \begin{array}{lll} z^T \nabla g_i(x) & \geq 0 & i \in \mathcal{A}(x) \\ z^T \nabla h_j(x) & = 0 & j \in \{1,\dots,p\} \end{array} \right\}$$

gegeben. Ferner sei zu einer Menge M der positive Normalkegel von M definiert

$$M^N \stackrel{\text{def}}{=} \{z \in \mathbb{R}^n : z^T x \geq 0 \textit{ für alle } x \in M\}.$$

Die Restriktrionsqualifikation von Guignard ist in x^* erfüllt, falls

$$Z^N(x^*) = T^N(\Omega,x^*) \tag{GCQ}$$

gilt. ∎

(GCQ) ist die schwächste Bedingung unter der (KKTB) notwendig aus der Optimalität von x^* folgt. Leider ist (GCQ) sehr schwer zu verifizieren und deshalb ersetzen wir es hier durch die für (GCQ) hinreichende Bedingung von Mangasarian und Fromowitz. Die Mangasarian-Fromowitz-Restriktionsqualifikation (MFCQ) hat auch als Verfahrensvoraussetzung eine große Bedeutung, wie wir noch sehen werden.

Definition 2.6 *In einem Punkt $x \in \Omega$ gilt (MFCQ), falls $\nabla h(x)$ spaltenregulär ist und $z \in \mathbb{R}^n$ existiert mit*

$$\begin{aligned}
&\nabla h(x)^T z = 0 \\
&\nabla g_i(x)^T z > 0 \qquad i \in \mathcal{A}(x).
\end{aligned}$$

∎

Neben der notwendigen Bedingung erster Ordnung existiert noch eine notwendige Bedingung zweiter Ordnung und eine hinreichende Bedingung zweiter Ordnung. Diese werden hier nicht präsentiert. Näheres ist in [16] und [5] nachzulesen. Wichtig für das Verständnis des weiteren Textes ist lediglich die Bedeutung der Multiplikatoren λ und μ, sowie die Verknüpfung der Restriktionen mit der Objektfunktion über die Lagrangefunktion.

3 Die Penalty-Methode von Pantoja und Mayne

Die grundlegende Idee aller Penalty-Methoden ist es das restringierte Problem mittels einer Penalty-Funktion umzuwandeln in ein unrestringiertes Optimierungsproblem. Eine Penalty-Funktion hat die Gestalt

$$\text{Minimiere } \phi(x;\gamma) = f(x) + \gamma s(x). \tag{6}$$

Der Faktor γ wird Penalty-Parameter genannt und für den Penalty-Term $s(x)$ gilt

$$s(x) \begin{cases} = 0 & x \in \Omega \\ > 0 & x \notin \Omega \end{cases} . \tag{7}$$

3.1 Die klassiche Penalty-Methode

Die Penalty-Funktion

$$\phi_K(x;\gamma) = f(x) + \gamma \left(\sum_{j=1}^{p} (h(x))^2 + \sum_{i=1}^{m} (\min(g_i(x), 0))^2 \right) = f(x) + \gamma \left\| \begin{array}{c} h(x) \\ g^-(x) \end{array} \right\|_2^2$$

folgt dem naheliegenden Gedanken die Funktion $f(x)$ über Ω hinaus stetig differenzierbar fortzusetzen. Deshalb können zur Lösung von (6) bekannte Methoden der unrestringierten Optimierung eingesetzt werden.

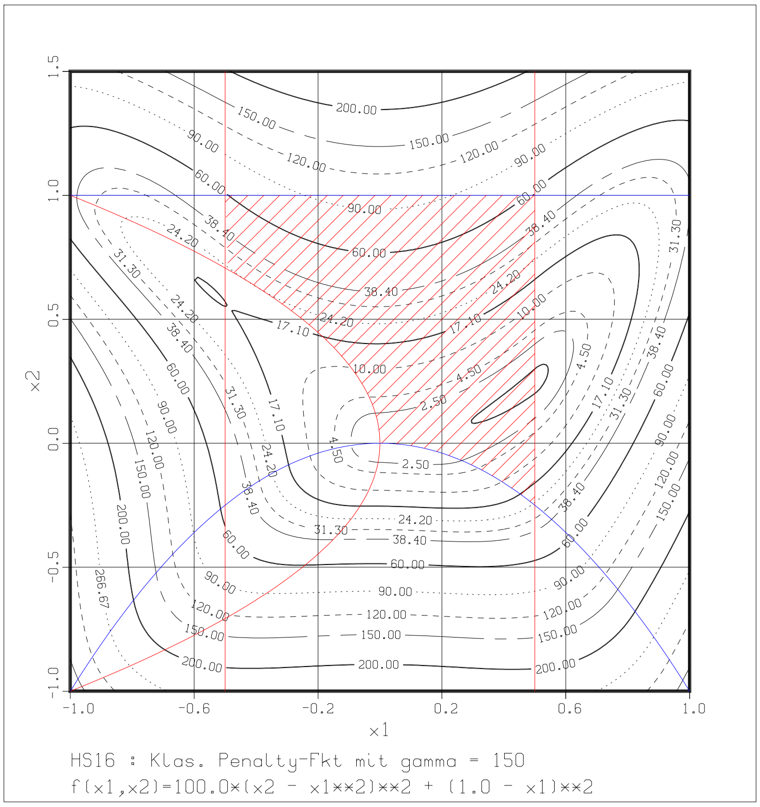

Bild 2 : Klassische Penalty-Funktion mit $\gamma = 150$ angewandt auf HS16

Der Nachteil dieser "inexakten" Methode liegt auf der Hand. Die Lösung x^* von (NLO) kann nur approximiert werden wenn $\gamma \to \infty$. Mit γ steigt aber auch die Konditionszahl der

Hessematrix von $\phi_K(x;\gamma)$ ins Unendliche. Die so auftretenden numerischen Schwierigkeiten machen das Verfahren ineffizient.

Abhilfe können Penalty-Funktionen schaffen, bei denen, für endliches γ, jede lokale Lösung von (NLO) durch eine lokale Minimalstelle der Penalty-Funktion wiedergegeben wird. Wegen (7) kann eine solche "exakte" Penalty-Funktion aber nicht differenzierbar sein.

3.2 Die ∞-Norm-Penalty-Funktion

Die Penalty-Funktion

$$\phi_{PM}(x;\gamma) = f(x) + \gamma \left\| \begin{array}{c} h(x) \\ g^-(x) \end{array} \right\|_\infty \tag{8}$$

erüllt die Forderung der Exaktheit und ist nicht differenzierbar. Die Bilder 3 und 4 verdeutlichen dies. In Bild 3 wird die globale Lösung von HS16 bei $\binom{0.5}{0.25}$ als Minimalstelle von ϕ_{PM} wiedergegeben und in Bild 4 ist $\gamma = 50$ groß genug um auch $\binom{-0.5}{\sqrt(0.5)}$ als Lösung zu liefern.

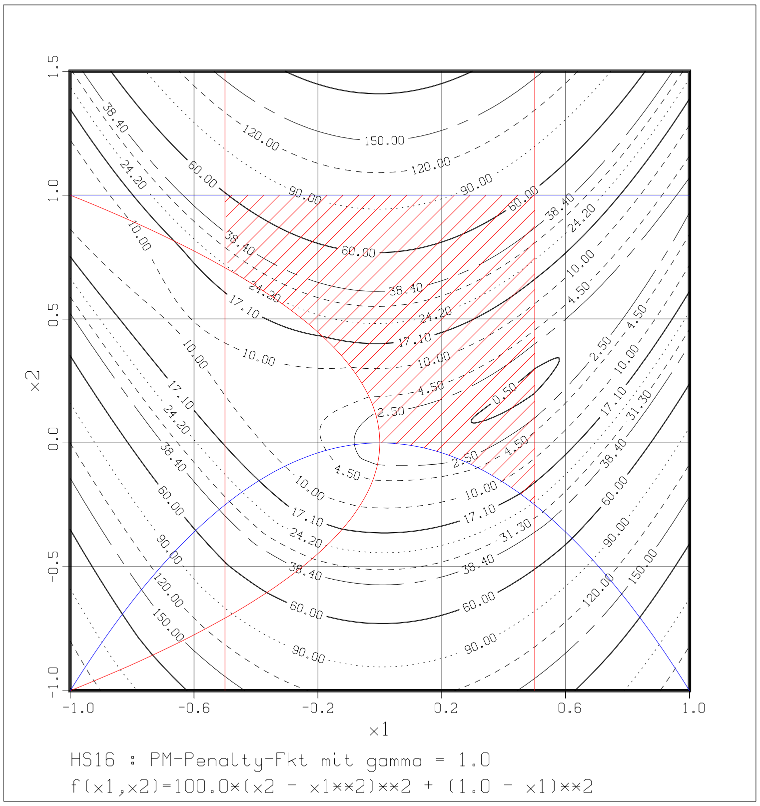

Bild 3 : Penalty-Funktion vom Pantoja & Mayne mit $\gamma = 1$ angewandt auf HS16

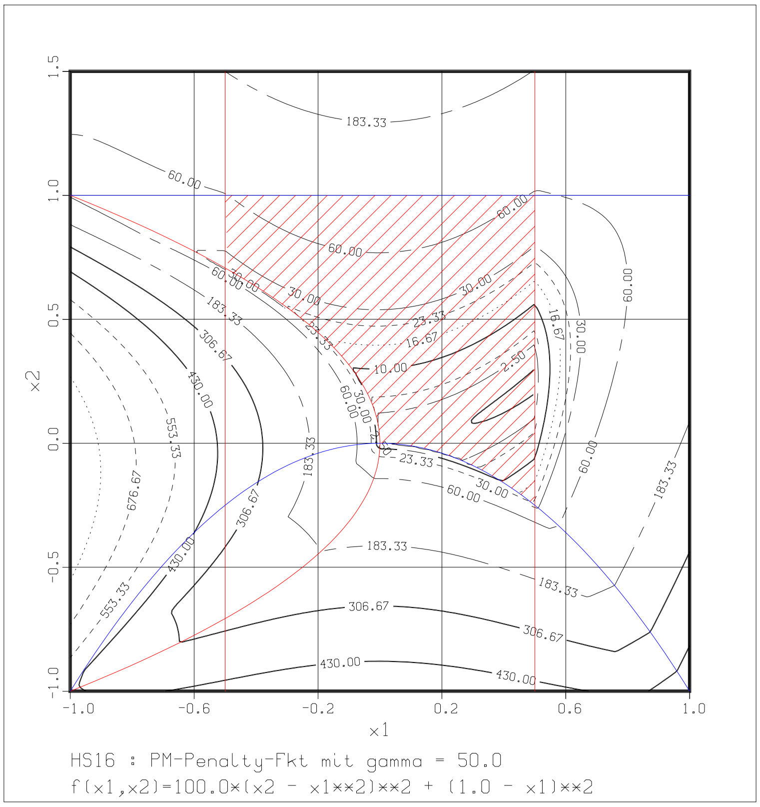

Bild 4 : Penalty-Funktion von Pantoja & Mayne mit $\gamma = 50$ angewandt auf HS16

(8) ist nur ein Repräsentant aus einer ganzen Klasse von nicht differenzierbaren exakten Penalty-Funktionen für den sich die folgenden Sätze beweisen lassen.

Satz 3.1 (Exaktheit der Penalty-Funktion I) *Sei die Menge*

$$\Omega(\tau_0) \stackrel{\text{def}}{=} \left\{ x \in \mathbb{R}^n : \begin{array}{rcl} g(x) & \geq & -\tau_0 \\ |h(x)| & \leq & \tau_0 \end{array} \right\} \tag{9}$$

kompakt für $\tau_0 > 0$ und in jedem Punkt $x \in \Omega(\tau_0)$ gelte die verallgemeinerte Mangasarian-Fromowitz-Bedingung (gMFCQ): $\nabla h(x)$ ist spaltenregulär und $(\exists z \in \mathbb{R}^n)$ mit

$$\begin{array}{rcll} \nabla h(x)^T z & = & -h(x) & \\ \nabla g_i(x)^T z & > & 0 & \qquad i \in \mathcal{A}(x) \cup \mathcal{V}(x). \end{array}$$

Dann gibt es ein $\underline{\gamma} > 0$ so, daß für alle $\gamma > \underline{\gamma}$ jedes lokale Minimum von $\phi_{PM}(x;\gamma)$ auf $\Omega(\tau_0)$ schon in Ω liegt und somit eine Lösung von (NLO) liefert.

Beweis Zum Beweis siehe [16] Satz 3.6.4 Seite 466. ∎

Satz 3.2 (Exaktheit der Penalty-Funktion II) *Sei x^* eine strenge lokale Minimalstelle von (NLO) und es gelte (MFCQ). Dann ist x^* strenge lokale Minimalstelle von $\phi_{PM}(x;\gamma)$ für γ hinreichend groß.*

Beweis Zum Beweis siehe [7] Satz 4.4. ∎

Es existiert sogar eine quantitative Aussage über $\underline{\gamma}$.

Satz 3.3 *Falls in x^* die hinreichende Bedingung zweiter Ordnung erfüllt ist, dann ist x^* strenge lokale Minimalstelle von $\phi_{PM}(x;\gamma)$ für alle*

$$\gamma > \underline{\gamma} = \left\| \begin{array}{c} \mu^* \\ \lambda^* \end{array} \right\|_1 . \tag{10}$$

Dabei sind μ^* und λ^* die Lagrange-Multiplikatoren aus (KKTB).

Beweis Diese Aussage ist eine Folgerung aus dem Satz 4.6. aus [7]. ■

Allgemein läßt sich zeigen, daß die Norm mit der der Penalty-Term definiert wird (hier $\|\cdot\|_\infty$) und die Norm die eine hinreichende Bedingung für die Exaktheit liefert (10) immer duale Normen sind. Siehe [16].

3.3 Bestimmung der Abstiegsrichtungen

Mit diesen drei Sätzen ist sichergestellt, daß durch die Suche nach einem lokalen Minimum von $\phi_{PM}(x;\gamma)$ auf der erweiterten zulässigen Menge $\Omega(\tau_0)$ eine Lösung von (NLO) gefunden werden kann. Es stellt sich aber die Frage **wie** ein solches Minimum zu finden ist. Da die Penalty-Funktion nicht differenzierbar ist kann keine der bekannten Methoden der unrestringierten Minimierung verwendet werden.

Eine Lösung liefert die Methode der sequentiellen quadratischen Optimierung (SQP), bei der in einem gegebene Punkt x die Objektfunktion f im Punkt $x + d$ quadratisch und die Restriktionen linear "approximiert" werden. Es entsteht ein sogenanntes "quadratisches" Optimierungsproblem.

$$\begin{array}{c} \text{Minimiere } \frac{1}{2} d^T B(x,\lambda,\mu) d + \nabla f(x)^T d + \gamma\omega \\ \text{unter } \left\{ \begin{array}{rcll} g_i(x) + \nabla g_i(x)^T d & \geq & -\omega & i \in \{1,\dots,m\} \\ h_j(x) + \nabla h_j(x)^T d & \geq & -\omega & j \in \{1,\dots,p\} \\ -h_j(x) - \nabla h_j(x)^T d & \geq & -\omega & j \in \{1,\dots,p\} \\ \omega & \geq & 0 & \end{array} \right. \end{array} \tag{QP(x)}$$

Die Schlupfvariable ω und die Aufspaltung der Gleichungsrestriktionen soll verhindern, daß die zulässige Menge Ω_{QP} von (QP(x)) leer ist.

Die Matrix $B(x,\lambda,\mu)$ wird dabei in Anlehnung an die Hessematrix der (erweiterten) Lagrangefunktion gewählt. Allerdings muß B positiv definit sein. Pantoja und Mayne schlagen deshalb ein BFGS-Update der Hessematrix der erweiterten Lagrangefunktion vor (Siehe [11]).

Um diese Wahl von B zu motivieren betrachte man (QP(x)) für konstantes $\omega = 0$, $B = \nabla^2 L(x,\lambda,\mu)$ und beschränke sich auf eine vorgegebene Menge aktiver Ungleichungsrestriktionen. Man erhält

$$\begin{array}{c} \text{Minimiere } \frac{1}{2} d^T \nabla^2 L(x,\lambda,\mu) d + \nabla f(x)^T d \\ \text{unter } \left\{ \begin{array}{rcll} g_i(x) + \nabla g_i(x)^T d & = & 0 & i \in \mathcal{A}(x) \\ h_j(x) + \nabla h_j(x)^T d & = & 0 & j \in \{1,\dots,p\} \end{array} \right. . \end{array} \tag{12}$$

Schreibt man (KKTB) für (12) hin so ergibt sich zusammen mit $\lambda_{QP} = \lambda_{\mathcal{A}(x)} - \Delta\lambda_{\mathcal{A}(x)}$ und $\mu_{QP} = \mu - \Delta\mu$ das System

$$\begin{pmatrix} \nabla^2 L(x,\lambda,\mu) & \nabla g_{\mathcal{A}(x)}(x) & \nabla h(x) \\ \nabla g^T_{\mathcal{A}(x)}(x) & 0 & 0 \\ \nabla h(x)^T & 0 & 0 \end{pmatrix} \begin{pmatrix} d \\ \Delta\lambda_{\mathcal{A}(x)} \\ \Delta\mu \end{pmatrix} = \begin{pmatrix} -\nabla L(x,\lambda,\mu) \\ -g_{\mathcal{A}(x)}(x) \\ -h(x) \end{pmatrix}.$$

Die Lösung dieses Systems liefert aber genau einen Korrekturschritt $(d, -\Delta\lambda_{\mathcal{A}(x)}, -\Delta\mu)^T$ des Newton-Verfahrens für das nichtlineare Gleichungssystem

$$\begin{pmatrix} \nabla L(x,\lambda,\mu) \\ g_{\mathcal{A}(x)}(x) \\ h(x) \end{pmatrix} = 0.$$

Die Lösung von (QP(x)) ist also eine Approximation der Newton-Richtung, die die Unzulässigkeit und die Norm des Gradienten der Lagrangefunktion verringert.

Entscheidend für den Erfolg der sequentiellen quadratischen Minimierung ist es, daß die Schlupfvariable ω verschwindet. Pantoja und Mayne zeigen in [11]

Satz 3.4 *Zu einem quadratischen Unterproblem (QP(x)) definieren wir die "erweitertet" zulässige Menge*

$$\Omega(\omega) \stackrel{\text{def}}{=} \left\{ d \in \mathbb{R}^n : \begin{array}{rcll} g_i(x) + \nabla g_i(x)^T d & \geq & -\omega & i \in \{1,\dots,m\} \\ h_j(x) + \nabla h_j(x)^T d & \geq & -\omega & j \in \{1,\dots,p\} \\ -h_j(x) - \nabla h_j(x)^T d & \geq & -\omega & j \in \{1,\dots,p\} \end{array} \right\}.$$

Ferner sei $\lambda^*_{QP} \in \mathbb{R}^{m+2p+1}_+$ der zu (d^*, ω^*) gehörige Lagrangemultiplikator. Falls $\Omega(0) \neq \emptyset$ dann ist

$$\gamma \geq \sum_{i=1}^{m+2p} (\lambda^*_{QP})_i \tag{13}$$

hinreichend für $\omega^* = 0$.

Beweis Zum Beweis siehe [11]. ■

Berücksichtigt man, daß der Lagrangemultiplikator μ der Gleichungsrestriktionen des Ausgangsproblems (NLO) durch die Differenz $\mu \approx (\lambda_{QP})_{i=m+1}^{i=m+p} - (\lambda_{QP})_{i=m+p+1}^{i=m+2p}$ geschätzt werden, so zeigt sich, daß in der Lösung $x^*(\Rightarrow d^* = 0)$ (13) hinreichend ist für (10). Diese Tatsache macht es möglich mit dem Penalty-Parameter γ auch den Schlupf im QP-Problem zu steuern.

Die Komibation aus Penalty-Funktion und sequentieller quadratischer Optimierung erlaubt auch eine andere Interpretation. Bei einem Ansatz mit einer nicht differenzierbaren Penalty-Funktion rückt das eigenliche Minimum der Funktion $\phi_{PM}(x;\gamma)$ mangels Optimalitätsbedingung in den Hindergrund. $\phi_{PM}(x;\gamma)$ kann lediglich verwendet werden um in der Richtung d auf einen Abstieg zu testen. In der Literatur spricht man deshalb auch von SQP-Methoden und nicht von Penalty-Methoden. Ferner bezeichnet man die Penalty-Funktion oft als Abstiegsfunktion ("merit function").

Bild 5 zeigt $\Omega(0)$ und Lösung d^* des QP-Problems zu HS16 im Punkt $x = (1,1)$. $\gamma = 1$ ist dabei hinreichend für $\omega^* = 0$. In Bild 6 ist die Situation aus Bild 3 dargestellt. Zusätzlich ist die Lösung d^* von QP((1,1)) eingefügt. Der nächste Punkt der Iteration wäre (0.5,0.5). Siehe Bild 6.

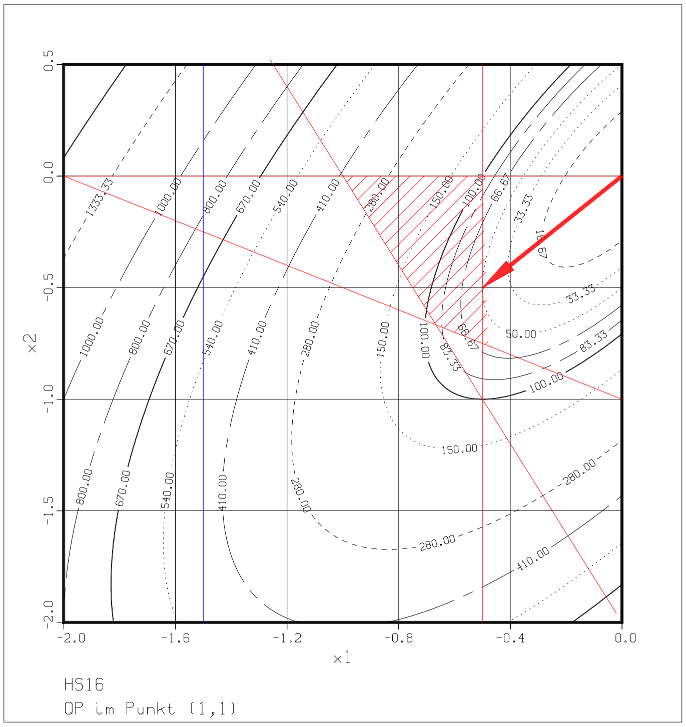

Bild 5 : HS16 : $\Omega(0)$ und d^* von QP((1,1)).

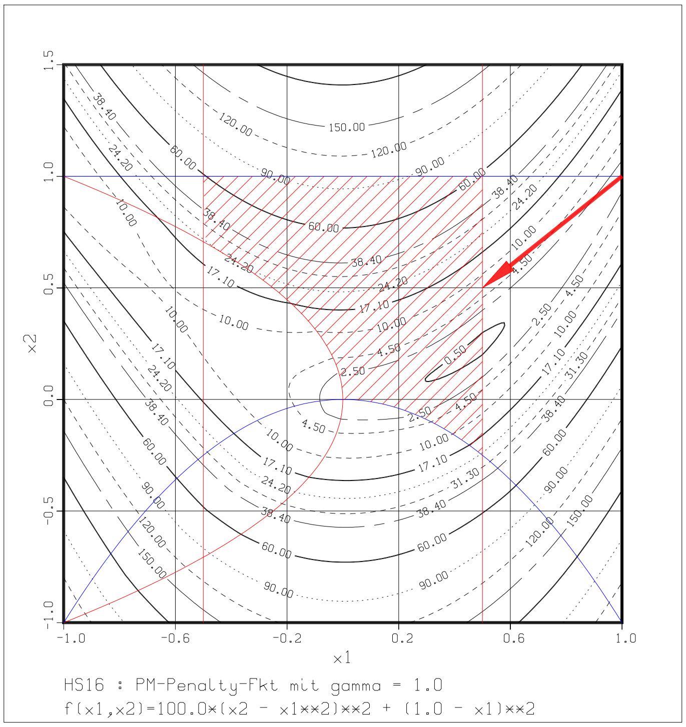

Bild 6 : Penalty-Funktion vom Pantoja & Mayne mit $\gamma = 1$ angewandt auf HS16 und Lösung von QP((1,1)) $d^* = (-0.5, -0.5)$.

3.4 Bestimmung der Schrittweite

Die Abstiegsrichtung d die das Lösungsverfahren für (QP(x)) errechnet liefert nicht unbedingt auch einen Abstieg in $\phi_{PM}(x;\gamma)$. Selbst dann nicht, wenn γ hinreichend groß ist und die Lösung x^* noch nicht erreicht ist. Grund hierfür ist die Nichtlinearität des Ausgangsproblems (NLO). Um mit d einen vernünftigen Abstieg zu erreichen wird der nächste Punkt x_{neu} mit einer Schrittweite σ bestimmt.

Seien $\beta < 1 (= 0.5)$ und $\delta < 1 (= 0.0001)$. Man wählt

$$\sigma = \max\{1, \beta, \beta^2, \beta^3, \ldots\}$$

so, daß

$$\phi_{PM}(x;\gamma) - \phi_{PM}(x + \sigma d;\gamma) \geq \sigma\delta|\theta(x,d;\gamma)|. \tag{14}$$

Dabei ist $\theta(x,d;\gamma) = \nabla f(x)^T d + \gamma\left(\tilde{s}(x,d) - s(x)\right)$ und

$$\tilde{s}(x,d) = \max\left\{\begin{array}{cc} g_i(x) + \nabla g_i(x)^T d\ , & i \in \{1,\ldots,m\} \\ h_j(x) + \nabla h_j(x)^T d\ , & j \in \{1,\ldots,p\} \\ -h_j(x) - \nabla h_j(x)^T d\ , & j \in \{1,\ldots,p\} \\ 0 & \end{array}\right\} \text{ bzw. } s(x) = \left\| \begin{array}{c} h(x) \\ g^-(x) \end{array} \right\|_\infty .$$

$\theta(x,d;\gamma)$ ist eine konvexe Approximation von $\phi_{PM}(x+d;\gamma) - \phi_{PM}(x;\gamma)$.

3.5 Abstiegsrichtungen zweiter Ordnung

Bei starken Nichtlinearitäten in den Restriktionen kann es vorkommen, daß die Richtungen d nur für sehr kleine Schrittweiten σ einen Abstieg liefern. Grund dafür ist, daß in (QP(x)) die Restriktionen nur von erster Ordnung approximiert werden. Eine Approximation von zweiter Ordnung würde allerdings auf quadratische Restriktionen führen und eine Behandlung quadratischer Restriktionen kommt dem Ausgangsproblem (NLO) gleich.

Um diese Schwierigkeiten zu umgehen schlägt Fletcher in [6] vor nach der Berechnung von d^* aus (QP(x)) im Falle $\omega^* = 0$ ein weiteres quadratisches Optimierungsproblem zu lösen. Sei d^* die Lösung von (QP(x)). Dann erfüllt d^* unter anderem die Restriktionen

$$h_j(x) + \nabla h_j(x)^T d^* \left\{\begin{array}{c}\geq \\ \leq\end{array}\right\} 0 \quad , \quad j \in \{1,\ldots,p\}.$$

Gewünscht wäre aber eine quadratische Approximation

$$h_j(x) + \nabla h_j(x)^T d + \frac{1}{2} d^T \nabla^2 h_j(x) d \left\{\begin{array}{c}\geq \\ \leq\end{array}\right\} 0 \quad , \quad j \in \{1,\ldots,p\}. \tag{15}$$

Fletcher schlägt vor $h_j(x+d^*)$ quadratisch zu approximieren

$$h_j(x+d^*) \approx h_j(x) + \nabla h_j(x)^T d^* + \frac{1}{2}(d^*)^T \nabla^2 h_j(x)(d^*) \quad , \quad j \in \{1,\ldots,p\}$$

und damit den quadratischen Anteil von (15) zu schätzen. Für $d \approx d^*$ erhält man mit

$$\left(h_j(x+d^*) - \nabla h_j(x)^T d^*\right) + \nabla h_j(x)^T d \left\{\begin{array}{c}\geq \\ \leq\end{array}\right\} 0 \quad , \quad j \in \{1,\ldots,p\}$$

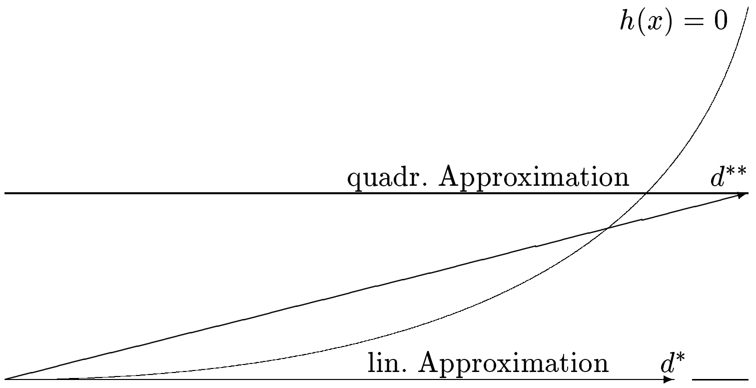

Bild 7 : Abstiegsrichtung zweiter Ordnung bei einer nichtlinearen Restriktion.

eine gute Approximation von (15). Analog kann man mit den aktiven Ungleichungsrestriktionen $g_i(x) + \nabla g_i(x) d^* \geq 0$ verfahren. Das neuen quadratische Optimierungsproblem liefert dann eine Lösung d^{**}. Siehe Bild 7. In diese Richtung wird, wenn sie sich von d^* nicht zu sehr unterscheidet, die nächste Iterierte $x + \sigma d^{**}$ berechnet. Dabei genügt σ (14).

4 Implementierung

In diesem Abschnitt sollen einige Bemerkungen zur Implementierung der beschriebenen Methode gemacht werden.

4.1 Problemstellung

Die wohl wesentlichste Änderung gegenüber der Darstellung in den bisherigen Abschnitten ist die Formulierung des Optimierungsproblems. Da bei (NLO) keine weiteren Informationen über die Form der Restriktionen verfügbar sind gibt es noch weitere Möglichkeiten die Problemstellung zu formulieren. Eine davon ist die in [2] bzw. [3] vorgeschlagene.

$$\text{Minimiere } f(x)$$
$$\text{unter} \quad \left\{ \begin{array}{ccccl} & & c_i(x) & = & 0 \qquad i \in \mathcal{E} \\ (c_l)_i & \leq & c_i(x) & \leq & (c_u)_i \qquad i \in \mathcal{I} \\ (b_l)_i & \leq & x_i & \leq & (b_u)_i \qquad i \in \{1, \ldots, n\} \end{array} \right. \tag{16}$$

Dabei wird der eventuellen Existenz nach unten **und** oben beschränkter Restriktionen und der einfachen Form eingeschränkter Variablen Rechnung getragen. Das drückt sich auch in der Definition der Lagrangefunktion aus

$$L(x, \lambda) \stackrel{\text{def}}{=} f(x) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x). \tag{17}$$

(KKTB) dann hat die Form

$$
\begin{aligned}
i \in \mathcal{I} \text{ und } (c_l)_i < c_i(x) < (c_u)_i &\Rightarrow \lambda_i^* = 0,\\
i \in \mathcal{I} \text{ und } (c_l)_i = c_i(x) < (c_u)_i &\Rightarrow \lambda_i^* \leq 0,\\
i \in \mathcal{I} \text{ und } (c_l)_i < c_i(x) = (c_u)_i &\Rightarrow \lambda_i^* \geq 0
\end{aligned}
$$

$$
P\left(\nabla_x L(x^*, \lambda^*)\right) = 0.
$$

Wobei

$$
P\left(\nabla_x L(x, \lambda)\right)_i \stackrel{\text{def}}{=} \begin{cases} \min\{0, (\nabla_x L(x,\lambda))_i\} & \text{falls } (b_l)_i = x_i < (b_u)_i \\ (\nabla_x L(x,\lambda))_i & \text{falls } (b_l)_i < x_i < (b_u)_i \\ \max\{0, (\nabla_x L(x,\lambda))_i\} & \text{falls } (b_l)_i < x_i = (b_u)_i \\ 0 & \text{falls } (b_l)_i = x_i = (b_u)_i. \end{cases} \tag{18}
$$

Diese Formulierung reduziert den Speicherbedarf unter Umständen erheblich, erschwert aber die Handhabung des Problems.

4.2 Skalierung

Die Penalty-Funktion (8) birgt neben ihrer Nichtdifferenzierbarkeit ein weiteres Problem in sich. Die Maximum-Norm ist extrem anfällig für schlecht skalierte Probleme, da i.A. nur eine Restriktion den Wert des Penalty-Terms bestimmt. Ob eine Restriktion mit dem Faktor 1000 multipliziert wird macht für die zulässige Menge keinen Unterschied. Der Wert der Penalty-Funktion kann sich dadurch aber erheblich ändern. Aus diesem Grunde werden alle Restriktionen und die Objektfunktion skaliert. Die neue Penalty-Funktion lautet dann

$$
\phi_{PM}(x;\gamma,\alpha) \stackrel{\text{def}}{=} \alpha_0 f(x) + \gamma \max \left\{ \begin{array}{ll} |\alpha_i c_i(x)| & i \in \mathcal{E} \\ |\alpha_i \left[c_i(x) - \mathrm{M}((c_l)_i, c_i(x), (c_u)_i)\right]| & i \in \mathcal{I} \\ |x_i - \mathrm{M}(l_i, x_i, u_i)| & i \in \{1,\ldots,n\} \end{array} \right\} \tag{19}
$$

mit $M(a,b,c) = \text{Median}(a,b,c)$. Die Skalierung wird so gesteuert, daß $\|\nabla f(x)\|$ bzw. $\|\nabla c_i(x)\|$ von der Größenordnung 1 sind.

Das zu einem Punkt x gehörende quadratische Unterproblem (QP(x)) hat ebenfalls eine neue Gestalt

$$
\begin{array}{ll}
\text{Minimiere} & \frac{1}{2}\begin{pmatrix} d \\ \omega \end{pmatrix}^T \begin{pmatrix} B(x,\lambda;\alpha) & 0 \\ 0 & \delta \end{pmatrix} \begin{pmatrix} d \\ \omega \end{pmatrix} + \begin{pmatrix} \alpha_0 \nabla f(x) \\ \gamma \end{pmatrix}^T \begin{pmatrix} d \\ \omega \end{pmatrix} \\
\text{unter} & \begin{pmatrix} \left(\alpha_i \nabla c_i(c)^T\right)_{i\in\mathcal{E}} & 1 \\ \left(\alpha_i \nabla c_i(c)^T\right)_{i\in\mathcal{I}} & 1 \\ \left(-\alpha_i \nabla c_i(c)^T\right)_{i\in\mathcal{E}} & 1 \\ \left(-\alpha_i \nabla c_i(c)^T\right)_{i\in\mathcal{I}} & 1 \\ I_n & 0 \\ -I_n & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d \\ \omega \end{pmatrix} \geq \begin{pmatrix} (-\alpha_i c_i(x))_{i\in\mathcal{E}} \\ (-\alpha_i \left(c_i(x) - (c_l)_i\right))_{i\in\mathcal{I}} \\ (\alpha_i c_i(x))_{i\in\mathcal{E}} \\ (\alpha_i \left(c_i(x) - (c_u)_i\right))_{i\in\mathcal{I}} \\ (l_i - x_i)_{i\in\{1,\ldots,n\}} \\ (x_i - u_i)_{i\in\{1,\ldots,n\}} \\ 0 \end{pmatrix}.
\end{array} \tag{20}
$$

Im Gegensatz zu (QP(x)) wird hier für die Schranken-Restriktionen kein Schlupf benötigt, da der Punkt x konstruktiv die “Box” nicht verläßt. Weiterhin wurde auf der Diagonalen der Hessematrix der Objektfunktion von (20) ein $\delta (1 \gg \delta > 0)$ eingeführt um die positive Definitheit von B auf die Gesamtmatrix zu übertragen.

4.3 Der QP-Löser

Das Lösungsverfahren für die quadratischen Unterprobleme (20) ist der entscheidende Teil der Implementierung. Nur wenn diese Unterprobleme schnell gelöst werden können, kann das gesamte Verfahren effizient arbeiten. Die Struktur der linearen Restriktionen erlaubt es (20) mittels einer Barriere-Funktion in ein “unrestringiertes” Problem umzuwandeln. Um die Notation zu erleichtern wird (20) in diesem Anschnitt auf die Form

$$\begin{aligned} &\text{Minimiere } \tfrac{1}{2} d^T C d + p^T d \\ &\text{unter } r(d) = A^T d - b \geq 0. \end{aligned} \tag{21}$$

reduziert, mit $A \in \mathbb{R}^{n \times m}$ und positiv definiter Matrix $C \in \mathbb{R}^{n \times n}$.

Bei der Barriere-Funktion handelt es sich um eine Modifikation der “shifted barrier function” die in [13] beschrieben ist.

$$\Phi(d, w, s) \overset{\text{def}}{=} \frac{1}{2} d^T C d + p^T d - \sum_{i=1}^{m} w_i \text{bar}(r_i(d); s, \eta) \tag{22}$$

Der Barriere-Term bar ist eine Kombination aus einem logarithmischen Term und einem kubischen Polynom

$$\text{bar}(r_i(d); s, \eta) \overset{\text{def}}{=} \begin{cases} \ln(r_i(d) + s) & \text{falls } r_i(d) > -s\eta \\ \alpha_0 + \alpha_1 r_i(d) + \alpha_2 (r_i(d))^2 + \alpha_3 (r_i(d))^3 & \text{falls } r_i(d) \leq -s\eta. \end{cases} \tag{23}$$

Dabei ist $\eta < 1$ und die Koeffizienten $\{\alpha_0, \ldots, \alpha_3\}$ des Polynoms werden so bestimmt, daß $\text{bar}(d; s, \eta) \in C^3(\mathbb{R})$.

Die so entstehenden nichtlinearen Optimierungsprobleme sind konvex.

Durch unrestringierte Minimierung einer Sequenz solcher Barriere-Funktionen und entsprechender Wahl der Gewichte $w_i > 0$ und des “Shifts” $s > 0$ gelang man zur Lösung von (21) (Siehe [4], [13] bzw. [1]). Gradient und Hessematrix von (22) sind gegeben durch

$$\begin{aligned} \nabla_d \Phi(d, w, s) &= Cd + p - A \left(w_i \text{bar}'(r_i(d); s, \eta) \right)_{i=1}^{m} = Cd + p - Au(d, w, s) \\ \nabla^2_{dd} \Phi(d, w, s) &= C + A \text{ diag} \left(w_i \text{bar}''(r_i(d); s, \eta) \right)_{i=1}^{m} \ A^T. \end{aligned} \tag{24}$$

Wobei die Funktion

$$u(d, w, s) \overset{\text{def}}{=} \left(w_i \text{bar}'(r_i(d); s, \eta) \right)_{i=1}^{m}$$

die Lagrangemultiplikatoren des Problems (21) annähert.

Um den Gradienten $\nabla_d \Phi(d, w, s)$ verschwinden zu lassen (was bei einem konvexen Optimierungsproblem hinreichend und notwendig für ein Minimum ist) wird das Truncated-Newton-Verfahren verwendet. Für gegebenes w und s wird die Gleichung

$$\left(C + A \text{ diag} \left(w_i \text{bar}''(r_i(d); s, \eta) \right)_{i=1}^{m} \ A^T \right) \Delta d = - \left(Cd + p - Au(d, w, s) \right)$$

nur näherungsweise gelöst. Dies wird mit einem abgebrochenen Lanczos-Verfahren erreicht. Die Hessematrix von $\Phi(d, w, s)$ wird dabei mit einer gegebene maximalen Zahl $\nu \ll n$ von Lanczos-Schritten zerlegt und mit der Zerlegung die Inverse der Hessematrix approximiert.

$$\left(\nabla^2_{dd}\Phi(d,w,s)\right)^{-1} \approx Q_\nu T_\nu^{-1} Q_\nu^T$$

Die direkte Berechnung der Lanczos-Zerlegung steht in Konkurenz zum Verfahren von Nash (Siehe [10]). Der die Lanczos-Zerlegung implizit verwendet. Der Nachteil dabei ist, daß die Lanczos-Vektoren, die die Orthogonalität verlieren, nicht nachorthogonalisiert werden können. Im Gegensatz dazu erlaubt die explizite Berechnung von $Q_\nu T_\nu Q_\nu^T$ den Einsatz von Reorthogonalisierungsstrategien wie in [14] oder [12] beschrieben.

Nachdem eine neue Iterierte d durch mehrere Truncated-Newton-Iterationen errechnet wurde, werden als Vorbereitung der nächsten Iteration die Gewichte und der Shift neu gewählt

$$s_{\text{neu}} = s_{\text{alt}}\delta \quad (\delta < 1) \quad \text{und} \quad w_i^{\text{neu}} = u_i(d, w^{\text{alt}}, s_{\text{alt}})s_{\text{neu}}.$$

5 Resultate

Die folgende Tabelle gibt eine Zusammenfassungen der Iterationsprotokolle einiger nichtlinearer Optimierungsprobleme. Die Berechnungen wurden auf einer HP9000/715/100-Workstation mit 96 MB Hauptspeicher durchgeführt. Der FORTRAN-Quellcode wurde mit f77 -K +OP4 compiliert. Die angegebenen Daten sind

1. name : Name des Optimierungsproblems aus der Problemsammlung CUTE. Siehe [2]
2. st : Status zum Ende der Berechnung.
   - st= 1 : Iteration erfolgreich. Genaugigkeitsforderungen erfüllt.
   - st= 2 : Die berechneten Schritte σd waren in den letzten $\min(n,5)$ zu klein. Es ist kein Fortschritt mehr zu erwarten.
   - st= 3 : Bei $\omega = 0$ konnte durch den QP-Löser keine Abstiegsrichtung für den Penalty-Term und die Penalty-Funktion mehr gefunden werden. Dies kommt oft in der Nähe der Lösung vor.
   - st= -1 : Abbruch da die Anzahl der Iterationen 500 überschritten hat.
   - st= -2 : Abbruch wegen Fehler im QP-Löser.
   - st= -3 : Abbruch da keine Schrittweite σ bestimmt werden konnte.
3. n : Anzahl der unabhängigen Variablen $x_1, \ldots, x_n$.
4. m : Anzahl der Restriktionen $c_1(x), \ldots, c_m(x)$.
5. #f : Anzahl der benötigten Auswertungen der Objektfunktion.
6. #gf : Anzahl der benötigten Auswertungen des Gradienten der Objektfunktion.
7. #ci : Anzahl der benötigten Auswertungen einzelner Restriktionsfunktionen $c_i(x)$.
8. #gci : Anzahl der benötigten Auswertungen der Gradienten einzelner Restriktionsfunktionen $\nabla c_i(x)$.
9. #hl : Anzahl der benötigten Berechnungen der Hessematrix der Lagrangefunktion (17).

10. it : Anzahl der Iterationen.

11. ktc : Verhältnis zwischen der Norm des projizierten Gradienten der Lagrangefunktion (18) und $\max(1, \|\nabla f(x)\|_2)$ im letzten Punkt der Iteration. Dieser Wert ist bei großen Problemen und solchen mit steilen Gradienten ein besseres Maß für die Güte der Lösung.

12. vc : Wert des unskalierten Penalty-Terms im letzten Punkt der Iteration.

13. fmin : Wert der Objektfunktion im letzten Punkt der Iteration.

14. cpu : Benötigte CPU-Zeit in Sekunden.

name	st	n	m	#f	#gf	#ci	#gci	#hl	it	ktc	vc	fmin	cpu
HS1	1	2	0	38	19	0	0	17	17	0.87e−04	0.00e+00	0.65e−10	.90e−01
HS2	3	2	0	16	8	0	0	7	7	0.61e+00	0.00e+00	0.49e+01	.80e−01
HS3	1	2	0	3	3	0	0	1	1	0.67e−04	0.00e+00	0.12e−03	.10e−01
HS3MOD	3	2	0	7	4	0	0	3	3	0.25e−02	0.00e+00	0.16e−05	.40e−01
HS4	1	2	0	3	3	0	0	1	1	0.00e+00	0.00e+00	0.27e+01	.10e−01
HS5	1	2	0	14	8	0	0	6	6	0.49e−04	0.00e+00	-.19e+01	.50e−01
HS6	1	2	1	24	10	24	10	8	8	0.55e−07	0.14e−08	0.46e−16	.18e+00
HS7	1	2	1	32	9	32	9	5	7	0.14e−04	0.70e−04	-.17e+01	.21e+00
HS8	1	2	2	13	6	26	12	2	4	0.23e−07	0.43e−09	-.10e+01	.15e+00
HS9	1	2	1	15	5	15	2	3	3	0.41e−04	0.69e−05	-.50e+00	.12e+00
HS10	1	2	1	41	14	41	14	9	14	0.18e−06	0.00e+00	-.10e+01	.25e+00
HS11	2	2	1	25	11	25	11	7	11	0.10e+01	0.00e+00	-.85e+01	.25e+00
HS12	1	2	1	40	12	40	12	10	10	0.55e−07	0.82e−03	-.30e+02	.26e+00
HS13	1	2	1	43	25	43	25	23	23	0.52e−06	0.55e−04	0.93e+00	.28e+01
HS14	1	2	2	12	8	24	10	5	6	0.38e−04	0.79e−04	0.14e+01	.20e+00
HS15	1	2	2	16	6	32	12	3	4	0.35e−04	0.00e+00	0.31e+03	.27e+00
HS16	3	2	2	28	7	56	14	6	6	0.20e−03	0.00e+00	0.23e+02	.24e+00
HS17	3	2	2	46	14	92	28	13	13	0.52e−01	0.00e+00	0.10e+01	.52e+00
HS18	3	2	2	33	15	66	30	13	14	0.69e−03	0.00e+00	0.50e+01	.28e+00
HS19	1	2	2	105	31	210	62	28	31	0.53e−08	0.76e−05	-.70e+04	.82e+00
HS20	1	2	3	20	6	60	18	5	5	0.11e−03	0.29e−04	0.40e+02	.32e+00
HS21	1	2	1	3	3	3	2	1	1	0.19e−13	0.00e+00	-.10e+03	.11e+00
HS21MOD	3	7	1	24	5	24	2	4	4	0.43e−02	0.00e+00	-.96e+02	.23e+00
HS22	1	2	2	15	8	30	10	5	6	0.22e−05	0.11e−04	0.10e+01	.19e+00
HS23	1	2	5	25	9	125	38	8	8	0.97e−04	0.43e−04	0.20e+01	.48e+00
HS24	1	2	3	5	5	15	6	3	3	0.41e−04	0.40e−05	-.10e+01	.18e+00
HS25	1	3	0	1	2	0	0	0	1	0.20e−07	0.00e+00	0.33e+02	.20e−01
HS25MOD	1	3	0	24	12	0	0	10	10	0.15e−03	0.00e+00	0.96e−08	.24e+00
HS26	1	3	1	95	28	95	28	26	26	0.14e−03	0.10e−03	0.72e−06	.27e+00
HS27	1	3	1	204	40	204	40	35	38	0.29e−04	0.17e−04	0.40e−01	.34e+00
HS28	1	3	1	4	3	4	2	1	1	0.85e−07	0.39e−13	0.11e−14	.11e+00
HS29	3	3	1	54	10	54	10	9	9	0.35e−03	0.00e+00	-.23e+02	.27e+00
HS30	3	3	1	17	6	17	6	5	5	0.37e−03	0.00e+00	0.10e+01	.19e+00
HS31	1	3	1	20	7	20	7	6	11	0.22e−04	0.00e+00	0.60e+01	.72e+00
HS32	3	3	2	16	4	32	6	3	3	0.25e−02	0.42e−06	0.10e+01	.22e+00
HS33	1	3	2	13	6	26	12	5	5	0.44e−04	0.00e+00	-.46e+01	.19e+00
HS34	3	3	2	23	8	46	16	7	7	0.43e−01	0.00e+00	-.83e+00	.37e+00
HS35	1	3	1	3	3	3	2	1	1	0.30e−04	0.00e+00	0.11e+00	.15e+00
HS35MOD	1	3	1	3	3	3	2	1	1	0.90e−06	0.00e+00	0.25e+00	.13e+00
HS36	1	3	1	4	4	4	2	2	2	0.15e−03	0.15e−04	-.33e+04	.16e+00
HS37	1	3	2	16	7	32	4	5	5	0.32e−04	0.00e+00	-.35e+04	.18e+00
HS38	1	4	0	60	35	0	0	33	33	0.10e−03	0.00e+00	0.65e−11	.25e+00
HS39	1	4	2	42	12	84	24	6	12	0.12e−05	0.24e−04	-.10e+01	.66e+00
HS40	1	4	3	14	7	42	21	5	5	0.41e−04	0.12e−04	-.25e+00	.23e+00
HS41	1	4	1	5	4	5	2	2	3	0.19e−04	0.61e−04	0.19e+01	.19e+00
HS42	1	4	2	9	6	18	8	4	4	0.46e−04	0.83e−04	0.14e+02	.17e+00
HS43	3	4	3	29	11	87	33	10	10	0.42e−02	0.33e−04	-.44e+02	.72e+00
HS44	3	4	6	14	8	84	12	7	7	0.49e−02	0.00e+00	-.15e+02	.16e+01
HS44NEW	3	4	6	26	9	156	12	8	8	0.33e−02	0.74e−05	-.15e+02	.64e+00
HS45	1	5	0	8	4	0	0	2	2	0.00e+00	0.00e+00	0.10e+01	.60e−01
HS46	1	5	2	332	73	664	146	71	71	0.22e−03	0.13e−04	0.36e−06	.70e+00
HS47	1	5	3	108	33	324	99	31	31	0.14e−03	0.69e−05	0.18e−06	.68e+00
HS48	1	5	2	4	3	8	4	1	1	0.17e−12	0.38e−10	0.99e−30	.12e+00
HS49	1	5	2	19	8	38	4	6	6	0.11e−03	0.28e−08	0.30e−05	.20e+00
HS50	1	5	3	16	8	48	6	6	6	0.12e−03	0.16e−10	0.39e−08	.21e+00
HS51	1	5	3	4	3	12	6	1	1	0.21e−12	0.37e−12	0.22e−24	.12e+00
HS52	1	5	3	4	4	12	6	1	2	0.67e−04	0.21e−04	0.53e+01	.13e+00
HS53	1	5	3	7	5	21	6	3	4	0.21e−05	0.93e−04	0.41e+01	.19e+00
HS54	1	6	1	14	9	14	2	6	7	0.20e−03	0.11e−04	-.16e+00	.23e+00
HS55	3	6	6	10	5	60	12	5	5	0.25e−02	0.91e−04	0.67e+01	.53e+00
HS56	1	7	4	38	10	152	40	8	9	0.10e−04	0.61e−05	-.35e+01	.73e+00
HS57	3	2	1	18	7	18	7	6	6	0.16e−01	0.00e+00	0.28e−01	.35e+00
HS59	1	2	3	15	7	45	21	3	5	0.90e−04	0.00e+00	-.68e+01	.18e+00
HS60	3	3	1	21	9	21	9	6	8	0.51e−02	0.14e−04	0.33e−01	.21e+00

HS61	1	3	2	15	8	30	16	6	6	0.13e−03	0.13e−05	-.14e+03	.19e+00
HS62	3	3	1	9	6	9	2	5	5	0.56e−03	0.30e−06	-.26e+05	.21e+00
HS63	1	3	2	30	13	60	15	10	11	0.77e−04	0.21e−03	0.96e+03	.31e+00
HS64	1	3	1	58	33	58	33	29	31	0.11e−03	0.13e−06	0.63e+04	.33e+01
HS65	3	3	1	19	8	19	8	7	7	0.91e−02	0.00e+00	0.95e+00	.25e+00
HS66	3	3	2	18	6	36	12	5	5	0.14e−02	0.00e+00	0.52e+00	.25e+00
HS67	1	3	14	9	6	126	68	4	4	0.79e−06	0.59e−08	-.12e+04	.33e+00
HS68	1	4	2	358	144	716	288	141	143	0.20e−03	0.11e−04	-.92e+00	.24e+02
HS69	3	4	2	72	22	144	44	21	22	0.11e−02	0.67e−06	-.96e+03	.13e+01
HS70	1	4	1	32	16	32	16	14	14	0.15e−03	0.00e+00	0.75e−02	.59e+00
HS71	3	4	2	53	18	106	36	16	17	0.51e−02	0.19e−03	0.17e+02	.89e+00
HS72	1	4	2	44	14	88	28	10	18	0.63e−07	0.26e−08	0.73e+03	.76e+00
HS73	1	4	3	6	5	18	9	2	3	0.60e−05	0.29e−04	0.30e+02	.22e+00
HS74	1	4	5	28	12	140	40	6	17	0.63e−04	0.18e−03	0.51e+04	.20e+01
HS75	1	4	5	23	9	115	31	3	14	0.69e−04	0.15e−04	0.52e+04	.17e+01
HS76	1	4	3	3	3	9	6	1	1	0.25e−04	0.00e+00	-.47e+01	.17e+00
HS77	1	5	2	32	15	64	30	10	13	0.81e−05	0.49e−04	0.24e+00	.31e+00
HS78	1	5	3	11	6	33	18	3	4	0.13e−04	0.51e−04	-.29e+01	.20e+00
HS79	1	5	3	12	6	36	18	3	4	0.12e−03	0.20e−03	0.79e−01	.25e+00
HS80	1	5	3	13	6	39	18	3	4	0.93e−05	0.60e−04	0.54e−01	.20e+00
HS81	1	5	3	38	20	114	60	17	18	0.14e−03	0.30e−03	0.54e−01	.18e+01
HS83	1	5	3	13	8	39	24	5	6	0.25e−04	0.15e−05	-.31e+05	.49e+00
HS84	3	5	3	41	9	123	27	8	8	0.42e−03	0.00e+00	-.52e+07	.10e+01
HS85	3	5	21	148	69	3108	1382	68	68	0.55e−02	0.00e+00	-.22e+01	.19e+02
HS86	3	5	10	6	4	60	20	3	3	0.11e−02	0.00e+00	-.32e+02	.26e+00
HS87	-3	6	4	605	40	2420	160	31	44	0.19e+00	0.44e−03	0.92e+04	.50e+01
HS88	3	2	1	53	23	53	23	21	23	0.34e−03	0.00e+00	0.14e+01	.23e+01
HS89	1	3	1	41	20	41	20	18	18	0.14e−07	0.12e−05	0.14e+01	.39e+01
HS90	3	4	1	45	20	45	20	18	19	0.20e−02	0.26e−07	0.14e+01	.44e+01
HS91	3	5	1	43	19	43	19	18	18	0.19e−02	0.63e−07	0.14e+01	.62e+01
HS92	3	6	1	42	22	42	22	20	21	0.47e−01	0.00e+00	0.14e+01	.89e+01
HS93	3	6	2	18	8	36	16	7	7	0.83e+00	0.00e+00	0.14e+03	.64e+00
HS95	3	6	4	12	6	48	24	4	5	0.12e−02	0.00e+00	0.16e−01	.74e+00
HS96	3	6	4	12	6	48	24	4	5	0.73e−03	0.00e+00	0.17e−01	.72e+00
HS97	3	6	4	22	8	88	32	4	9	0.21e−01	0.00e+00	0.43e+01	.32e+01
HS98	3	6	4	80	30	320	120	27	31	0.28e+00	0.84e−03	0.45e+01	.32e+02
HS99	3	7	2	22	10	44	20	9	9	0.38e−05	0.88e+00	-.83e+09	.90e+00
HS100	1	7	4	37	9	148	36	8	8	0.23e−05	0.13e−03	0.68e+03	.88e+00
HS100LNP	1	7	2	40	13	80	26	10	11	0.11e−05	0.37e−03	0.68e+03	.36e+00
HS100MOD	3	7	4	36	14	144	56	13	13	0.21e+01	0.67e−03	0.68e+03	.14e+01
HS101	3	7	5	34	15	170	75	12	14	0.11e−02	0.00e+00	0.18e+04	.19e+01
HS102	3	7	5	1724	159	8620	795	155	158	0.10e−01	0.54e−04	0.91e+03	.64e+02
HS103	3	7	5	57	25	285	125	22	24	0.28e−01	0.00e+00	0.54e+03	.85e+01
HS104	3	8	5	27	10	135	50	9	9	0.13e+01	0.00e+00	0.40e+01	.25e+01
HS104LNP	1	8	4	28	11	112	44	8	9	0.19e−03	0.38e−03	0.40e+01	.68e+00
HS105	3	8	1	19	10	19	2	9	9	0.68e−03	0.00e+00	0.11e+04	.20e+01
HS106	-1	8	6	4602	501	27612	1509	498	500	0.10e+01	0.10e+02	0.90e+04	.24e+02
HS107	1	9	6	11	7	66	42	5	5	0.11e−03	0.84e−04	0.51e+04	.90e+00
HS108	-3	9	13	160	7	2080	91	6	9	0.12e+03	0.30e+01	0.98e+00	.87e+00
HS109	3	9	10	88	49	880	302	39	53	0.65e−04	0.71e−02	0.54e+04	.50e+02
HS111	1	10	3	4027	447	12081	1341	445	445	0.26e−03	0.95e−04	-.48e+02	.16e+02
HS111LNP	1	10	3	4108	452	12324	1356	450	450	0.26e−03	0.96e−04	-.48e+02	.16e+02
HS112	2	10	3	28	15	84	6	13	19	0.19e−01	0.38e−06	-.48e+02	.28e+01
HS113	3	10	8	38	15	304	81	14	14	0.11e−02	0.40e−05	0.24e+02	.28e+01
HS114	3	10	11	126	35	1386	220	33	34	0.46e−02	0.45e−04	-.16e+04	.42e+01
HS116	-3	13	14	581	21	8134	218	17	26	0.14e+01	0.16e+00	0.25e+03	.16e+02
HS117	3	15	5	66	23	330	115	22	22	0.11e−02	0.00e+00	0.32e+02	.10e+02
HS118	1	15	17	3	3	51	34	1	1	0.87e−06	0.43e−04	0.66e+03	.36e+00
HS119	3	16	8	11	7	88	16	6	6	0.78e−02	0.30e−04	0.24e+03	.82e+00
HS268	3	5	5	4	3	20	10	2	2	0.18e−01	0.00e+00	0.44e−01	.17e+00
GOULDQP2	3	699	349	7	2	2443	698	1	5	0.32e−02	0.95e−07	0.25e−03	.76e+02
GOULDQP3	3	699	349	10	6	3490	698	5	5	0.13e−01	0.15e−03	0.20e+01	.30e+03
EIGMAXA	1	101	101	12	6	1212	606	5	5	0.70e−06	0.73e−04	-.10e+01	.50e+01
EIGMINA	1	101	101	17	8	1717	808	6	6	0.49e−06	0.24e−04	0.10e+01	.73e+01
EIGMAXB	2	101	101	49	25	4949	2525	23	25	0.43e−02	0.11e−04	-.96e+00	.69e+02
EIGMINB	1	101	101	77	25	7777	2525	23	23	0.10e−03	0.11e−04	0.97e−03	.61e+02
EIGMAXC	1	102	102	21	9	2142	918	7	7	0.31e−04	0.57e−04	-.10e+01	.99e+01
EIGMINC	1	102	102	52	16	5304	1632	14	14	0.97e−03	0.68e−04	0.10e+01	.51e+02
ZAMB2-10	1	270	96	17	10	1632	960	6	8	0.28e−03	0.91e−03	-.16e+01	.43e+03
ZAMB2-11	1	270	96	19	11	1824	1056	8	9	0.83e−03	0.41e−03	-.11e+01	.39e+03
ZAMB2-8	1	138	48	98	30	4704	1440	27	28	0.15e−03	0.76e−03	-.15e+00	.32e+03
ZAMB2-9	1	138	48	9	5	432	240	2	3	0.91e−03	0.65e−03	-.35e+00	.15e+03
ZAMB2	1	1326	480	18	10	8640	4800	5	8	0.15e−02	0.58e−03	-.41e+01	.37e+04
HAGER1	3	1001	500	26	14	13000	1000	12	13	0.30e−03	0.25e−02	0.88e+00	.15e+04
HAGER2	3	1001	500	22	12	11000	1000	10	11	0.26e−03	0.43e−02	0.43e+00	.87e+03
HAGER3	3	1001	500	21	11	10500	1000	9	10	0.32e−03	0.11e−01	0.14e+00	.68e+03
HAGER4	3	1001	500	75	37	37500	1000	35	36	0.20e−03	0.13e−02	0.28e+01	.81e+04
HYDROELL	3	1009	1008	5	3	5040	2016	2	2	0.20e+00	0.23e−12	-.34e+07	.15e+04
HYDROELM	3	505	504	4	3	2016	1008	2	2	0.20e+00	0.00e+00	-.34e+07	.88e+02
HYDROELS	3	169	168	9	5	1512	336	4	4	0.20e+00	0.00e+00	-.34e+07	.34e+02
MSQRTA	1	529	529	30	14	15870	7406	9	12	0.11e−02	0.38e−03	0.00e+00	.10e+05

MSQRTB	1	529	529	32	15	16928	7935	10	13	0.19e−02	0.69e−03	0.00e+00	.13e+05
NET1	1	48	57	41	17	2337	414	14	15	0.39e−03	0.82e−06	0.94e+06	.47e+02
NET2	1	144	160	127	38	20320	2624	31	38	0.67e−03	0.85e−03	0.12e+07	.34e+04
NET3	-2	464	521	66	30	34386	7090	17	30	0.14e−02	0.33e+01	0.57e+07	.19e+05
QPCBLEND	1	83	74	15	9	1110	148	7	7	0.11e−04	0.84e−03	-.72e−02	.55e+02
QPCBOEI1	3	384	351	68	18	23868	702	5	21	0.75e+00	0.48e−02	0.12e+08	.25e+04
QPCBOEI2	-2	143	166	56	11	9296	332	0	15	0.98e+00	0.12e+03	0.72e+07	.11e+04
QPCSTAIR	1	467	356	544	304	193664	712	298	307	0.28e−04	0.97e−03	0.62e+07	.11e+06
QPNBLEND	3	83	74	45	31	3330	148	30	30	0.69e−01	0.92e−03	-.75e−02	.14e+04
QPNBOEI1	3	384	351	93	19	32643	702	1	22	0.56e−01	0.11e−01	0.69e+07	.57e+04
QPNBOEI2	-2	143	166	95	26	15770	332	0	29	0.96e+00	0.38e+02	0.18e+07	.33e+04
QPNSTAIR	3	467	356	175	14	62300	712	1	18	0.44e−01	0.13e−04	0.52e+07	.33e+04
STEENBRA	1	432	108	17	9	1836	216	4	7	0.16e−03	0.71e−03	0.17e+05	.65e+03
STEENBRB	-2	468	108	254	34	27432	216	0	53	0.10e+01	0.10e+04	0.27e+12	.18e+05
STEENBRC	-2	540	126	102	13	12852	252	0	13	0.10e+01	0.96e+04	0.59e+09	.53e+04
STEENBRD	3	468	108	233	41	25164	216	4	53	0.10e+01	0.68e−04	0.27e+08	.14e+05
STEENBRE	-2	540	126	89	12	11214	252	0	12	0.10e+01	0.96e+04	0.55e+09	.43e+04
STEENBRF	-2	468	108	254	34	27432	216	0	53	0.10e+01	0.10e+04	0.27e+12	.18e+05
STEENBRG	-2	540	126	102	13	12852	252	0	13	0.10e+01	0.96e+04	0.59e+09	.53e+04
DTOC1L	1	1495	990	12	8	11880	1980	6	6	0.66e−04	0.85e−04	0.12e+02	.21e+03
DTOC1NA	1	1495	990	11	8	10890	7920	6	6	0.60e−04	0.98e−04	0.13e+02	.30e+03
DTOC1NB	1	1495	990	11	7	10890	6930	5	5	0.15e−03	0.12e−03	0.16e+02	.30e+03
DTOC1NC	3	1495	990	205	39	202950	38610	39	39	0.17e+00	0.63e−04	0.25e+02	.64e+04
DTOC1ND	3	1495	990	95	28	94050	27720	27	29	0.14e+00	0.22e−03	0.27e+02	.89e+04
DTOC2	1	598	396	42	16	16632	6336	14	14	0.11e−02	0.77e−03	0.49e+00	.66e+03
DTOC3	1	1499	998	83	42	82834	1996	38	40	0.35e−02	0.41e−04	0.24e+03	.49e+05
DTOC4	1	1499	998	28	15	27944	8483	13	13	0.11e−02	0.49e−03	0.25e+01	.33e+04
DTOC5	1	999	499	21	12	10479	5988	9	10	0.22e−02	0.25e−03	0.15e+01	.10e+04
DTOC6	1	1001	500	59	32	29500	16000	30	30	0.28e−03	0.16e−05	0.68e+04	.25e+04
EIGENA2	1	1260	630	8	6	5040	3780	4	4	0.59e−03	0.10e−03	0.27e−14	.20e+03
EIGENACO	3	1260	630	18	7	11340	4410	6	7	0.87e+00	0.26e−03	0.45e−05	.78e+03
EIGENC2	1	992	496	69	45	34224	22320	43	43	0.23e−02	0.29e−03	0.16e−06	.79e+04

6 Ausblick

Die angegebenen Resultate zeigen erhebliche Unterschiede. Teilweise ist die Genauigkeit und die Geschindigkeit des Verfahrens durchaus brauchbar. Allerdings bleiben im wesentlichen drei Probleme.

Zum Einen muß man bei der Steuerung des Penalty-Parameters äußerst vorsichtig sein. Ein zu früh zu großes γ stellt den QP-Löser vor erhebliche Probleme, wenn das QP-Problem nicht "kompatibel" ist (und somit ein $\omega > 0$ benötigt wird). Es kann dann sehr lange dauern, bis die Unzulässigkeit im QP-Löser reduziert wird. Eine defensive Steuerungsstrategie ist also notwendig. Auf der anderen Seite bewirken langsame Steigerung und ein niedriger Initialwert für γ langsame Konvergenzraten bei Problemen mit quadratischer Zielfunktion und linearen Restriktionen. Diese, eigentlich quadratischen Probleme, sind mit γ hinreichend groß in einem Schritt lösbar.

Zweitens ist die Anzahl der Truncated-Newton-Iterationen im QP-Löser in manchen Iterationen sehr groß. Dieses Problem hängt eng zusammen mit dem Begriff der zentralen Trajektorie. Die zentrale Trajektorie ist die Kurve der Punkte, welche die Barriere-Funktionen exakt minimieren. Weicht die aktuelle Iterierte zu weit von der idealen Lösung ab, so vervielfacht sich der numerische Aufwand.

Das dritte und größte Problem, steckt in der Wahl der Matrix $B(x, \lambda; \alpha)$. Da diese Matrix positiv definit sein soll, ist eine Regularisierung unabdingbar. Es zeigt sich aber, daß schon kleinste Regularisierungen mit einem kleinen Vielfachen der Einheitsmatrix zu sehr "schlechten" Richtungen d führen. Hier sollte die Umstellung des QP-Lösers auf ein Verfahren für indefinite Probleme Besserung bringen. In diesem Fall kann $B(\lambda; \alpha)$ als die Hessematrix der Lagrangefunktion gewählt werden und der Einsatz von Richtungen negativer Krümmung wäre möglich.

Literatur

[1] Auslender, A. , Cominetti, R. , Haddou , M.: *Asymptotic analysis for penalty and barrier methods in convex and linear programming* ; Mathematics of Operations Research 22 (1997) : 43-62

[2] Bongartz, I. , Conn, A.R. , Gould, N. , Toint, Ph.L.: *CUTE: Constrained an unconstrained testing environment* ; (1997)

[3] Conn, A.R. , Gould, N. , Toint, Ph.L.: *LANCELOT: A Fortran package for large-scale nonlinear optimization* ; Springer (1992)

[4] Felkel, R.: *An interior-point method for large-scale QP problems* ; Tech. Report No. 1850 TH Darmstadt (1996)

[5] Fiacco, A.V. , McCormickG.P.: *Nonlinear programming: sequential unconstrained minimization techniques* ; Wiley (1968)

[6] Fletcher, R.: *Practical methods of optimization* ; Wiley (1987)

[7] Han, S.-P. , Mangasarian O.L.: *Exact penalty functions in nonlinear programming* ; Mathematical Programming 17 (1979) : 251-269

[8] Heinz, J. , Spellucci, P.: *A successful implementation of the Pantoja-Mayne SQP method* ; Optimization Methods and Software 4 (1994) : 1-28

[9] Hock, W. , Schittkowski, K.: *Test examples for nonlinear programming codes* ; Springer (1981)

[10] Nash, S.G.: *Newton-type minimization via the Lanczos method* SIAM J. Numerical Analysis 21 (1991) : 93-113

[11] Pantoja, J.F.A. de O. , Mayne D.Q.: *Exact Penalty Function algorithm with simple updating of the Penalty Parameter* ; J.O.T.A 69 (1991) : 441-467

[12] Parlett, B.N.: *The symmetric eigenvalue problem* ; Prectice-Hall Int. (1980)

[13] Polyak, R.: *Modified barrier functions (theory and methods)* ; Mathematical Programming 54 (1992) : 177-222

[14] Simon, H.D.: *The Lanczos algorithm with partial reorthogonalization* ; Math. of Computing 42 (1984) : 115-142

[15] Spellucci, P. , Törnig, W.: *Eigenwertberechnung in den Ingenieurwissenschaften* ; Teubner (1985)

[16] Spellucci, P.: *Numerische Verfahren der nichtlinearen Optimierung* ; Birkhäuser (1993)