



Vergleich zwischen halgblattem Newton-Verfahren und SQP-Verfahren

Bachelorarbeit

am Fachgebiet Optimierung bei partiellen Differentialgleichungen

Institut für Mathematik

Fakultät II Mathematik und Naturwissenschaften

Technische Universität Berlin

vorgelegt von

Vicky Hartanto Tanzil

Betreuer: Prof. Dr. Fredi Tröltzsch

Vicky Hartanto Tanzil

Matr.-Nr.: 308789

Studiengang: Technomathematik

Lehrter Straße 68, 10557 Berlin

Eidesstattliche Erklärung

Die selbständige und eigenhändige Anfertigung versichert an Eides statt
Berlin, den

.....

Unterschrift

Inhaltsverzeichnis

Einleitung	1
1 Grundlagen	2
1.1 Allgemeine Optimierungsprobleme	2
1.2 Optimierungsprobleme ohne Restriktionen	5
1.3 Optimierungsprobleme mit linearen Restriktionen	9
1.4 Optimierungsprobleme mit nichtlinearen Restriktionen	15
2 SQP-Verfahren	17
2.1 Einführung	17
2.2 Formulierung	18
2.3 Aktive-Mengen-Verfahren	20
2.4 Nullraum-Verfahren	24
2.5 Lagrange-Newton-SQP-Verfahren	28
3 Das halbglatte Newton-Verfahren	30
3.1 Grundlagen	30
3.2 Formulierung	32
3.3 Aktive-Mengen-Strategie	34
4 Vergleich	38
4.1 Testprobleme	38
4.2 Numerische Ergebnisse	49
5 Fazit und Ausblick	54
Danksagung	55
Literaturverzeichnis	56

Einleitung

Die nichtlineare Optimierung ist ein bedeutendes Gebiet der Mathematik. Sie findet immer wieder Anwendungen in schwierigen Problemen der Technik und der Wirtschaft. Viele Verfahren wurden entwickelt, um nichtlineare Optimierungsprobleme zu lösen. In dieser Arbeit werden zwei Verfahren verglichen, nämlich das halbglatte Newton-Verfahren und das SQP-Verfahren. Es liegt nahe, zu untersuchen, welches von den beiden Verfahren besser ist.

Das SQP-Verfahren gehört zu den bekanntesten Verfahren der nichtlinearen Optimierung. In seiner Doktorarbeit hat Wilson 1963 das erste SQP-Verfahren entwickelt [Wil63]. Es wird bis heute in vielen Optimierungsproblemen als Standardwerkzeug angewendet und wurde inzwischen sehr viel weiterentwickelt.

Das halbglatte Newton-Verfahren ist nicht so bekannt wie das SQP-Verfahren. Es basiert aber auf dem bekannten Newton-Verfahren. Über dieses Verfahren hat Ulbrich in [Ulbr11] geschrieben: „It was not predictable in 2000 that ten years later semismooth Newton methods would be one of the most important approaches for solving inequality constrained optimization problems in function spaces.“

Die Gliederung der Arbeit sieht wie folgt aus. Zuerst werden die wichtigsten Grundlagen der Optimierungstheorie eingeführt. In Kapitel 2 wird das SQP-Verfahren betrachtet und im darauf folgenden Kapitel das halbglatte Newton-Verfahren. Anschließend werden in Kapitel 4 die numerische Ergebnisse der beiden Verfahren für verschiedene Testprobleme aufgeführt und verglichen.

Kapitel 1

Grundlagen

In diesem Kapitel werden zunächst die grundlegenden Definitionen und die wichtigsten theoretischen Ergebnisse der Optimierungstheorie betrachtet. Wir beschäftigen uns hier mit Optimierungsproblemen, die durch differenzierbare Funktionen charakterisiert werden. Dabei werden verschiedene Optimalitätsbedingungen von Optimierungsproblemen gezeigt, die als Grundlage der numerischen Verfahren dienen werden. Dieses Kapitel orientiert sich an [Alt11, Trö11].

1.1 Allgemeine Optimierungsprobleme

Definition 1.1 Allgemein ist die Aufgabenstellung der Optimierung wie folgt definiert:

$$\min_{x \in \mathcal{F}} f(x). \quad (\text{P})$$

Die Funktion $f: D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ ist die sogenannte Zielfunktion. D sei der Definitionsbereich von f . \mathcal{F} sei eine nichtleere Teilmenge von D , die man als Lösungsmenge bezeichnet. Alle Elemente von \mathcal{F} werden als zulässige Punkte bezeichnet. \mathcal{F} wird durch die sogenannten Nebenbedingungen oder Restriktionen definiert.

Falls $\mathcal{F} = D$ gilt, bezeichnet man das Optimierungsproblem als unrestringiert. Es besitzt also keine Restriktionen. Ansonsten heißt es restringiertes Optimierungsproblem.

In der Regel definiert man ein Optimierungsproblem als Minimierungsproblem, weil ein Maximierungsproblem $\max g(x)$ zum Minimierungsproblem $\min f(x) := -g(x)$ äquivalent ist.

Definition 1.2 (Globale und lokale Lösung, vgl. Definition 1.1.4 in [Alt11, S. 2 f.]

Ein Punkt $x^* \in \mathcal{F}$ heißt globale Lösung des Problems (P) oder globales Minimum, wenn

$$f(x^*) \leq f(x) \quad \forall x \in \mathcal{F} \quad (1.1)$$

gilt. Ein Punkt $x^* \in \mathcal{F}$ heißt *strikte globale Lösung des Problems (P)* oder *striktes globales Minimum*, wenn

$$f(x^*) < f(x) \quad \forall x \in \mathcal{F} \setminus \{x^*\} \quad (1.2)$$

gilt. Ein Punkt $x^* \in \mathcal{F}$ heißt *lokale Lösung des Problems (P)* oder *lokales Minimum*, wenn für eine Umgebung¹ U von x^*

$$f(x^*) \leq f(x) \quad \forall x \in U(x^*) \cap \mathcal{F} \quad (1.3)$$

gilt. Ein Punkt $x^* \in \mathcal{F}$ heißt *strikte lokale Lösung des Problems (P)* oder *striktes lokales Minimum*, wenn für eine Umgebung U von x^*

$$f(x^*) < f(x) \quad \forall x \in U(x^*) \cap \mathcal{F} \setminus \{x^*\} \quad (1.4)$$

gilt.

Von diesen Definitionen kommt der Begriff *globale Optimierung* her. Bei der globalen Optimierung versucht man, eine globale Lösung zu finden. Viele Verfahren versuchen nur lokale Lösungen zu bestimmen, weil globale Lösungen nicht so einfach zu bestimmen sind.

Beispiel 1.1 (Vgl. Beispiel 1.1.2 in [Alt11, S. 1])

Ein Beispiel eines unrestringierten Optimierungsproblems ist

$$\min_{x \in \mathbb{R}} f(x) := (2x - 2)^2(3x + 3)^2 + 10x. \quad (1.5)$$

Dieses Problem hat eine strikte globale Lösung an der Stelle $x^* = -1$ und eine strikte lokale Lösung an der Stelle $x = 1$ (siehe Abbildung 1.1).

Definition 1.3 (Nichtlineare Optimierungsprobleme) Die Aufgabenstellung bei der nichtlinearen Optimierung mit Gleichungs- und Ungleichungsrestriktionen kann man spezifischer wie folgt definieren:

$$\begin{aligned} & \min_{x \in D} f(x) && \text{(PN)} \\ & \text{bei } g(x) \leq 0 \\ & h(x) = 0. \end{aligned}$$

Die Zielfunktion ist wieder die Funktion $f: D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$. Die Nebenbedingungen sind von den Funktionen $g: D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^p$ und $h: D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ abhängig, d. h., die Menge \mathcal{F} sieht hier so aus: $\mathcal{F} = \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}$.

¹Für $\epsilon > 0$ und $x \in \mathbb{R}^n$ heißt $U_\epsilon(x) := \{u \in \mathbb{R}^n \mid \|u - x\| < \epsilon\}$ die ϵ -Umgebung von x . Eine Menge U heißt eine Umgebung von x , wenn es $\epsilon > 0$ gibt, sodass $U_\epsilon(x) \subset U$.

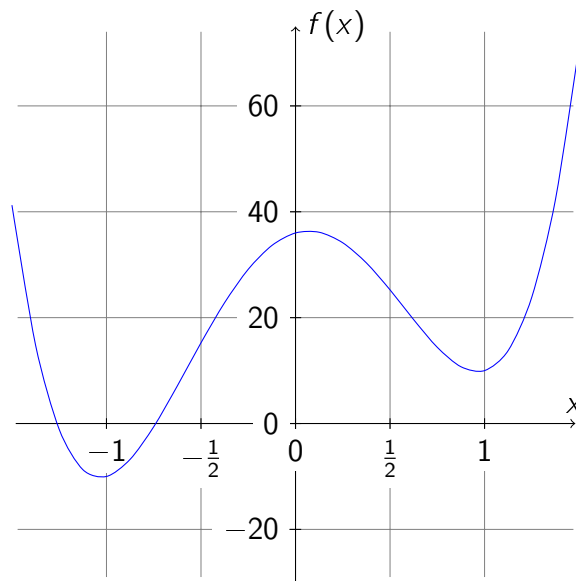


Abbildung 1.1: Die Funktion von Beispiel 1.1

Die Operatoren \leq und $=$ vergleichen hier die Vektoren miteinander elementenweise. Die Zahl 0 ist hier, je nach dem in welchem Kontext, ein Element von \mathbb{R} , \mathbb{R}^p oder \mathbb{R}^m .

Man kann das Problem PN ausführlicher schreiben, indem man die Funktionen g und h in skalare Funktionen $g_1, \dots, g_p: \mathbb{R}^n \rightarrow \mathbb{R}$ und $h_1, \dots, h_m: \mathbb{R}^n \rightarrow \mathbb{R}$ zerlegt, sodass

$$g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_p(x) \end{pmatrix} \quad \text{und} \quad h(x) = \begin{pmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{pmatrix}.$$

Dann bekommt man das Problem

$$\begin{aligned} & \min_{x \in D} f(x) \\ & \text{bei} \quad g_i(x) \leq 0 \text{ für } i = 1, \dots, p \\ & \quad \quad h_j(x) = 0 \text{ für } j = 1, \dots, m. \end{aligned}$$

Es ist hierbei der Unterschied zwischen der linearen Optimierung und der nichtlinearen Optimierung gut zu erkennen. Bei der linearen Optimierung muss die Zielfunktion linear sein (d. h., die Zielfunktion muss in der Form $f(x) = c^T x = \sum_{i=1}^n c_i x_i$ mit $c \in \mathbb{R}^n$ sein) und die Nebenbedingungen sollen durch lineare Gleichungen bzw. Ungleichungen definiert werden. Bei der nichtlinearen Optimierung gibt es dagegen keine Einschränkung, wie die Zielfunktion und die Nebenbedingungen aussehen sollen.

Beispiel 1.2 (Vgl. Problem 1.2 in [NW06, S. 3])

Ein Beispiel eines restringierten Optimierungsproblems ist

$$\begin{aligned} \min_{x \in \mathbb{R}^2} & (x_1 - 2)^2 + (x_2 - 1)^2 \\ \text{bei} & \quad x_1 + x_2 \leq 2 \\ & \quad x_1^2 \leq x_2. \end{aligned} \tag{1.6}$$

Die Abbildung 1.2 zeigt die zulässige Menge \mathcal{F} , die Höhenlinien der Zielfunktion und die optimale Lösung x^* .

Eine Höhenlinie ist eine Menge von Punkten, wo die Zielfunktion einen konstanten Wert annimmt. In Abbildung 1.2 sind die Höhenlinien in Form von Kreisen zu sehen und je größer der Kreis ist, desto größer ist der Wert der Zielfunktion.

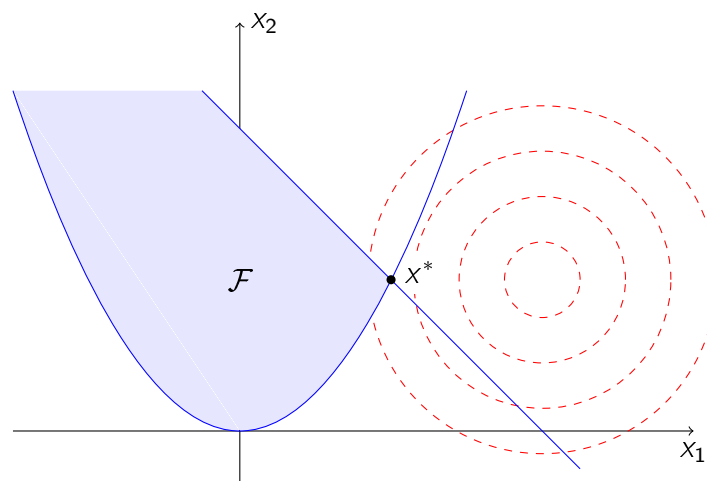


Abbildung 1.2: Geometrische Darstellung des Beispiels 1.2

1.2 Optimierungsprobleme ohne Restriktionen

Wir beginnen nun zuerst mit Optimierungsproblemen ohne Restriktionen und stellen die notwendigen und hinreichenden Optimalitätsbedingungen bereit. Zwei numerische Verfahren hierfür werden auch vorgestellt.

Definition 1.4 (Unrestringierte Optimierungsprobleme)

$$\min_{x \in \mathbb{R}^n} f(x). \tag{PU}$$

Wir nehmen hier der Einfachheit halber an, dass der Definitionsbereich D gleich \mathbb{R}^n sei.

Beispiel 1.3 (Lineare Regression, vgl. Beispiel 1.1.6 in [Alt11, S. 4 f.])

Ein praktisches Beispiel ist die Aufgabe der linearen Regression. Gegeben seien m Messwerte $(\xi_1, \eta_1), \dots, (\xi_m, \eta_m)$. Gesucht ist eine Gerade

$$\eta(\xi) := x_1 \xi + x_2, \quad (1.7)$$

die „optimal“ zu den Messwerten passt. D. h., der Parameter $x = (x_1, x_2)^T \in \mathbb{R}^2$ soll so bestimmt werden, dass die Summe der Fehlerquadrate in den Messpunkten minimiert wird. Somit ist ein unrestringiertes Optimierungsproblem

$$\min_{x \in \mathbb{R}^2} f(x) := \sum_{i=1}^m (\eta(\xi_i) - \eta_i)^2 = \sum_{i=1}^m (x_1 \xi_i + x_2 - \eta_i)^2 \quad (1.8)$$

zu lösen.

Beispiel 1.4 (Nichtlineare Regression, vgl. Abschnitt 2.3.2 in [Alt11, S. 30 f.])

Neben linearen Regressionsaufgaben sind auch oft nichtlineare Regressionsaufgaben zu lösen. Gesucht ist der funktionale Zusammenhang $\eta(\xi)$ zwischen den ξ - und den η -Werten, beispielsweise

$$\eta(\xi) := x_1(\xi - x_2)^2 + x_3 \quad (1.9)$$

oder

$$\eta(\xi) := x_1 e^{\xi x_2}. \quad (1.10)$$

Dabei ist $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ ein Parametervektor, der „optimal“ bestimmt werden soll.

Satz 1.1 (Notwendige Bedingung erster Ordnung, vgl. Satz 3.1.2 in [Alt11, S. 42])

Sei x^* eine lokale Lösung des Problems (PU) und sei f einmal stetig differenzierbar in einer Umgebung von x^* , dann gilt

$$\nabla f(x^*) = 0. \quad (1.11)$$

Die Bedingung (1.11) gilt aber nicht nur für lokale Minima sondern auch für lokale Maxima von f .

Definition 1.5 (Stationärer Punkt, vgl. Definition 3.1.4 in [Alt11, S. 42])

Die Funktion f sei in x^* differenzierbar. Der Punkt x^* heißt stationärer Punkt von f , wenn x^* die notwendige Bedingung (1.11) erfüllt.

Viele Optimierungsverfahren suchen in der Regel nach einem stationären Punkt von f . Aber ein stationärer Punkt muss kein globales oder lokales Minimum sein. Durch folgende notwendige Bedingung kann man zwischen einem lokalen Minimum und einem lokalen Maximum unterscheiden.

Satz 1.2 (Notwendige Bedingung zweiter Ordnung, vgl. Satz 3.1.6 in [Alt11, S. 43])

Sei x^* eine lokale Lösung des Problems (PU) und sei f zweimal stetig differenzierbar in einer Umgebung von x^* , dann gilt (1.11) und

$$x^T f''(x^*)x \geq 0 \quad \forall x \in \mathbb{R}^n. \quad (1.12)$$

$f''(x^*)$ ist also positiv semidefinit.

Satz 1.3 (Hinreichende Bedingung zweiter Ordnung, vgl. Satz 3.1.11 in [Alt11, S. 44])

Sei f zweimal stetig differenzierbar in einer Umgebung von x^* . Die notwendige Bedingung (1.11) sei erfüllt und $f''(x^*)$ sei positiv definit, d. h.

$$x^T f''(x^*)x > 0 \quad \forall x \in \mathbb{R}^n. \quad (1.13)$$

Dann ist x^* eine strikte Lösung des Problems (PU).

Diese hinreichende Bedingung benutzt man in der Regel erst dann, wenn man einen stationären Punkt findet.

Eine wichtige Grundlage für einige Verfahren ist die Definition der Abstiegsrichtung.

Definition 1.6 (Abstiegsrichtung, vgl. Definition 4.1.2 in [Alt11, S. 68])

Die Funktion f sei differenzierbar in x . Ein Vektor $d \in \mathbb{R}^n \setminus \{0\}$ heißt Abstiegsrichtung von f in x , wenn

$$\nabla f(x)^T d < 0 \quad (1.14)$$

gilt.

Sei $x \in \mathbb{R}^n$ mit $\nabla f(x) \neq 0$, dann ist beispielsweise $d = -\nabla f(x)$ eine Abstiegsrichtung von f in x .

Satz 1.4 (Vgl. Lemma 4.1.1 in [Alt11, S. 68])

Seien $x \in \mathbb{R}^n$, f differenzierbar in x und d eine Abstiegsrichtung von f in x . Dann gibt es ein $\hat{\sigma} > 0$ mit

$$f(x + \sigma d) < f(x) \quad \forall \sigma \in]0, \hat{\sigma}[. \quad (1.15)$$

Die meisten Optimierungsverfahren sind iterativ. Sie fangen also mit einem Anfangspunkt x^0 an und versuchen dann weitere Punkte $(x^1, x^2, \dots, x^k, \dots)$ zu finden, die besser als die vorherigen sind. Viele iterative Verfahren zur Bestimmung einer lokalen Lösung sind häufig Abstiegsverfahren. In der k -ten Iteration bestimmen sie zu einem Punkt x^k eine Abstiegsrichtung d^k und eine Schrittweite σ_k so, dass für $x^{k+1} := x^k + \sigma_k d^k$

$$f(x^{k+1}) < f(x^k) \quad (1.16)$$

gilt.

1.2.1 Gradientenverfahren

Das Gradientenverfahren ist ein einfaches Abstiegsverfahren, welches die negativen Gradienten als Abstiegsrichtungen verwendet.

Verfahren 1.1 (Gradientenverfahren, vgl. Verfahren 4.2.39 in [Alt11, S. 98])

1. Wähle einen Startpunkt x^0 und setze $k := 0$.
2. Setze $d^k := -\nabla f(x^k)$.
3. Ist $d^k = 0 \Rightarrow STOP$.
4. Bestimme Schrittweite σ_k so, dass

$$f(x^k + \sigma_k d^k) < f(x^k + \sigma d^k) \quad \forall \sigma \geq 0. \quad (1.17)$$

5. Setze $x^{k+1} := x^k + \sigma_k d^k$ und $k := k + 1. \Rightarrow$ Gehe zu Schritt 2.

Es ist zu beachten, dass die Abbruchbedingung in Schritt 3 nur theoretisch zu verstehen ist. Praktisch gibt man eine Abbruchschranke $\varepsilon > 0$ vor und führt beispielsweise den Test

$$\max \{|d_1^k|, \dots, |d_n^k|\} < \varepsilon \quad (1.18)$$

durch. Es wird damit überprüft, ob alle Elemente von d^k klein genug sind.

Zur Bestimmung der Schrittweite σ_k kann man das Schrittweitenverfahren von Armijo oder das Schrittweitenverfahren von Wolfe-Powell verwenden (vgl. Abschnitt 4.2.7 in [Alt11, S. 90 ff.]).

1.2.2 Newton-Verfahren

Ein bekanntes Verfahren der numerischen Mathematik, um eine nichtlineare Gleichung zu lösen, ist das Newton-Verfahren. Man berechnet mit dem Newton-Verfahren eine Nullstelle von einer gegebenen Abbildung $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$, d. h. eine Lösung $x^* \in \mathbb{R}^n$ der nichtlinearen Gleichung $F(x) = 0$. Für die unrestringierten Optimierungsprobleme kann das Newton-Verfahren angewendet werden, um die Lösung der nichtlinearen Gleichung (1.11), $\nabla f(x) = 0$, zu finden. Es wird dabei vorausgesetzt, dass die Funktion f zweimal differenzierbar ist.

Verfahren 1.2 (Newton-Verfahren, vgl. Verfahren 4.3.1 in [Alt11, S. 107])

1. Wähle einen Startpunkt x^0 und setze $k := 0$.
2. Ist $\nabla f(x^k) = 0 \Rightarrow STOP$.

3. Berechne die Lösung d des linearen Gleichungssystems

$$f''(x^k)d = -\nabla f(x^k). \quad (1.19)$$

Setze $d^k := d$.

4. Setze $x^{k+1} := x^k + d^k$ und $k := k + 1 \Rightarrow$ Gehe zu Schritt 2.

Der Punkt x^{k+1} in jedem Iterationsschritt ist eigentlich die Lösung des Minimierungsproblems, welches durch die quadratische Approximation von f in Punkt x^k definiert ist. In der Umgebung von x^k kann die Funktion f wie folgt approximiert werden:

$$f(x) \approx f(x^k) + \nabla f(x^k)^T(x - x^k) + \frac{1}{2}(x - x^k)^T f''(x^k)(x - x^k). \quad (1.20)$$

Die Ableitung der rechten Seite ist

$$\nabla f(x^k) + f''(x^k)x - f''(x^k)x^k. \quad (1.21)$$

Setzt man diese gleich null, dann bekommt man

$$f''(x^k)(x - x^k) = -\nabla f(x^k) \quad (1.22)$$

$$x - x^k = -[f''(x^k)]^{-1}\nabla f(x^k) \quad (1.23)$$

$$x = x^k - \underbrace{[f''(x^k)]^{-1}\nabla f(x^k)}_d. \quad (1.24)$$

Das ist genau unser Punkt x^{k+1} . D. h., es wird in jedem Iterationsschritt eigentlich das Problem

$$\min_{x \in \mathbb{R}^n} f(x^k) + \nabla f(x^k)^T(x - x^k) + \frac{1}{2}(x - x^k)^T f''(x^k)(x - x^k) \quad (1.25)$$

gelöst, wobei die Konstante $f(x^k)$ weggelassen werden kann.

Es ist später zu sehen, dass das SQP-Verfahren diese Idee auch gebrauchen wird.

Eine Schrittweitensteuerung wie bei dem Gradientenverfahren kann auch durchgeführt werden, d. h., man bestimmt ein $\sigma_k \in \mathbb{R}$ und definiert $x^{k+1} := x^k + \sigma_k d^k$. Dann bekommt man ein Abstiegsverfahren, welches als das gedämpfte Newton-Verfahren bezeichnet wird (vgl. Abschnitt 4.3.2 in [Alt11, S. 111 ff.]).

1.3 Optimierungsprobleme mit linearen Restriktionen

Nun ist die Theorie der restringierten Optimierungsprobleme zu betrachten. Wir konzentrieren uns vorerst auf lineare Restriktionen, d. h., die Nebenbedingungen sind durch lineare Gleichungen und Ungleichungen definiert.

1.3.1 Optimierungsprobleme mit linearen Gleichungsnebenbedingungen

Definition 1.7 (Optimierungsprobleme mit linearen Gleichungsnebenbedingungen)

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) \\ \text{bei } Ax = b. \end{aligned} \quad (\text{PLG})$$

A sei eine $(m \times n)$ -Matrix und b sei ein Vektor mit m Elementen.

D. h., die Menge \mathcal{F} sieht hier so aus: $\mathcal{F} = \{x \in \mathbb{R}^n \mid Ax = b\}$.

Beispiel 1.5 Ein einfaches Beispiel eines Optimierungsproblems (PLG) ist

$$\begin{aligned} \min_{x \in \mathbb{R}^2} x_1^2 + x_2^2 \\ \text{bei } x_1 + x_2 = 1. \end{aligned} \quad (1.26)$$

Die zulässigen Punkte dieses Problems liegen auf einer Gerade (siehe Abbildung 1.3).

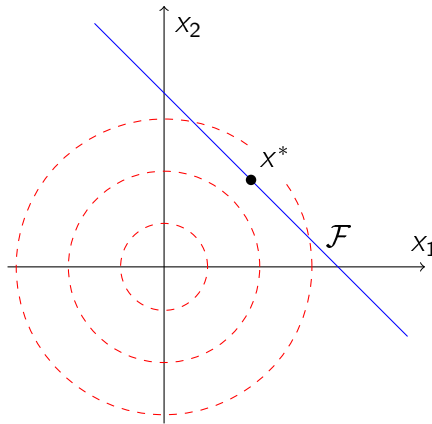


Abbildung 1.3: Geometrische Darstellung des Beispiels 1.5

Satz 1.5 (Notwendige Bedingung, vgl. Satz 5.3.2 in [Alt11, S. 174])

Sei x^* eine lokale Lösung des Problems (PLG) und f sei in x^* differenzierbar. Dann gibt es ein $\lambda \in \mathbb{R}^m$ mit

$$\nabla f(x^*) + A^T \lambda = 0. \quad (1.27)$$

Hat A einen vollen Rang, dann ist λ eindeutig bestimmt.

Diese Bedingung heit die Multiplikatoren-Regel von Lagrange und man bezeichnet λ als Lagrange-Multiplikator.

Satz 1.6 (Hinreichende Bedingung, vgl. Satz 5.3.8 in [Alt11, S. 177])

Sei f in x^* zweimal stetig differenzierbar. Die notwendige Bedingung in Satz 1.5 sei erfllt. Es existiere eine Konstante $\alpha > 0$ mit

$$d^T f''(x) d \geq \alpha \|d\|^2 \quad \forall d \in \text{Kern } A. \quad (1.28)$$

Dann ist x^* eine strikte Lsung des Problems (PLG).

Eine wichtige Grundlage zum Lsen des Problems (PLG) als unrestringiertes Problem ist die Definition der Nullraum-Matrix (vgl. [Alt11, S. 182]).

Definition 1.8 (Nullraum-Matrix)

Eine $(n \times s)$ -Matrix Z heit Nullraum-Matrix von A , wenn fr $d \in \mathbb{R}^n$ gilt

$$d \in \text{Kern } A \Leftrightarrow d = Zz \text{ fr ein } z \in \mathbb{R}^s. \quad (1.29)$$

D. h. $\text{Im } Z = \text{Kern } A$.

Sei w eine Lsung der Gleichung $Ax = b$. Man kann nun die Menge \mathcal{F} des Problems (PLG) so schreiben:

$$\mathcal{F} = w + \text{Kern } A = w + \text{Im } Z = w + \{Zz \mid z \in \mathbb{R}^s\}. \quad (1.30)$$

D. h., jedes Element $x \in \mathcal{F}$ ist mit $w + Zz$, $z \in \mathbb{R}^s$, zu ersetzen. Das Problem (PLG) ist dann quivalent zu

$$\min_{z \in \mathbb{R}^s} F(z) := f(w + Zz). \quad (1.31)$$

Dieses Problem hat keine Nebenbedingung mehr, ist also unrestringiert. Man kann folglich Verfahren fr unrestringierte Probleme anwenden.

1.3.2 Optimierungsprobleme mit linearen Ungleichungsnebenbedingungen

Definition 1.9 (Optimierungsprobleme mit linearen Ungleichungsnebenbedingungen)

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} f(x) && (\text{PLU}) \\ & \text{bei } Ax = b \\ & \quad Gx \leq r. \end{aligned}$$

A sei eine $(m \times n)$ -Matrix mit $m \leq n$ und $b \in \mathbb{R}^m$. G sei eine $(p \times n)$ -Matrix und $r \in \mathbb{R}^p$.

D. h., die Menge \mathcal{F} sieht hier so aus: $\mathcal{F} = \{x \in \mathbb{R}^n \mid Ax = b, Gx \leq r\}$.

Seien $a_k \in \mathbb{R}^n$, $k = 1, \dots, m$, bzw. $g_j \in \mathbb{R}^n$, $j = 1, \dots, p$, Vektoren in der Matrix A bzw. G , sodass

$$A = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix} \quad \text{und} \quad G = \begin{pmatrix} g_1^T \\ \vdots \\ g_p^T \end{pmatrix}. \quad (1.32)$$

Seien $b_k \in \mathbb{R}$, $k = 1, \dots, m$, bzw. $r_j \in \mathbb{R}$, $j = 1, \dots, p$, die Elemente von b bzw. r . Dann kann das Problem (PLU) wie folgt geschrieben werden.

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} f(x) \\ & \text{bei} \quad \langle a_k, x \rangle = b_k \text{ für } k = 1, \dots, m \\ & \quad \quad \langle g_j, x \rangle \leq r_j \text{ für } j = 1, \dots, p. \end{aligned}$$

Beispiel 1.6 (Vgl. Beispiel 16.4 in [NW06, S. 475])

Ein einfaches zweidimensionales Beispiel eines Optimierungsproblems (PLU) ist

$$\begin{aligned} & \min_{x \in \mathbb{R}^2} (x_1 - 1)^2 + (x_2 - 2.5)^2 \\ & \text{bei} \quad -x_1 + 2x_2 \leq 2 \\ & \quad \quad x_1 + 2x_2 \leq 6 \\ & \quad \quad x_1 - 2x_2 \leq 2 \\ & \quad \quad x_1, x_2 \geq 0. \end{aligned} \quad (1.33)$$

Die Abbildung 1.4 zeigt die zulässige Menge \mathcal{F} , die Höhenlinien der Zielfunktion und die optimale Lösung x^* .

Beispiel 1.7 (Vgl. Beispiel 13.2 in [AL07, S. 415 f.])

Gegeben seien zwei Dreiecke R und S in der Abbildung 1.5. Gesucht ist der kürzeste Abstand zwischen diesen Dreiecken und die zugehörigen Punkte $r^* \in R$ und $s^* \in S$, die diesen kürzesten Abstand bilden.

Seien $r = (x_1, x_2)^T \in R$ und $s = (x_3, x_4)^T \in S$. Der quadratische Abstand zwischen r und s ist durch

$$\|r - s\|^2 = (x_1 - x_3)^2 + (x_2 - x_4)^2 = x^T H x \quad (1.34)$$

gegeben, wobei

$$H := \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}. \quad (1.35)$$

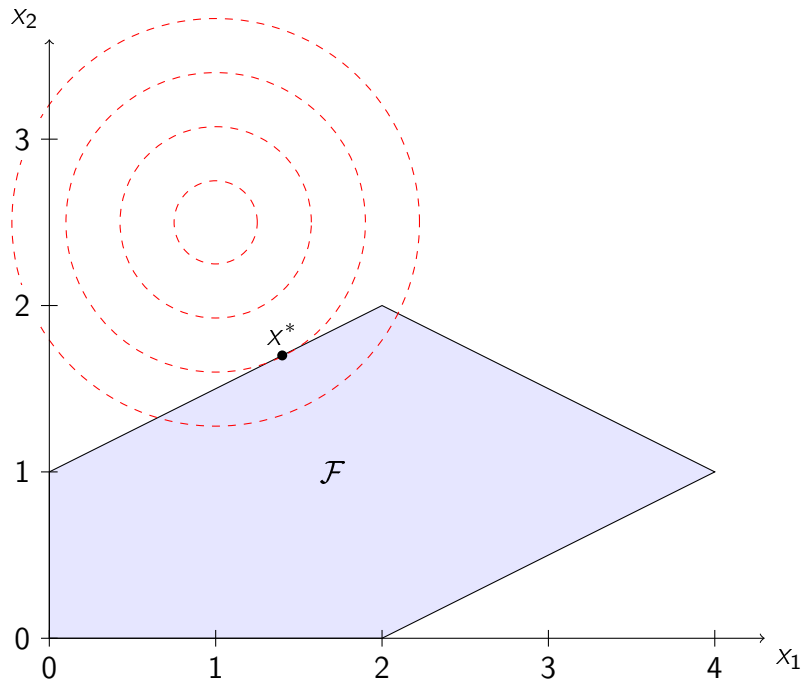


Abbildung 1.4: Geometrische Darstellung des Beispiels 1.6

Die Bedingungen $r \in R$ und $s \in S$ sind durch die Ungleichungen

$$\begin{aligned}
 x_1, x_2 &\geq 0 \\
 x_1 + 2x_2 &\leq 2 \\
 x_4 &\geq 2 \\
 x_3 + x_4 &\geq 3 \\
 x_3 + 2x_4 &\leq 6
 \end{aligned} \tag{1.36}$$

gegeben. D. h., das Problem kann als Optimierungsproblem mit einer quadratischen Zielfunktion und linearen Ungleichungsnebenbedingungen dargestellt werden.

Der folgende Satz hat sich als besonders hilfreich erwiesen, um eine lokale Lösung des Problems (PLU) zu charakterisieren.

Satz 1.7 (Karush-Kuhn-Tucker-Satz, vgl. Satz 5.4.7 in [Alt11, S. 193])

Sei x^* lokale Lösung des Problems (PLU) und f sei in x^* differenzierbar. Dann existieren die Vektoren $\lambda \in \mathbb{R}^m$ und $\mu \in \mathbb{R}^p$ zu x^* mit

$$\nabla f(x^*) + A^T \lambda + G^T \mu = 0 \tag{1.37}$$

$$\mu_j (\langle g_j, x^* \rangle - r_j) = 0 \text{ für } j = 1, \dots, p \tag{1.38}$$

$$\mu \geq 0 \tag{1.39}$$

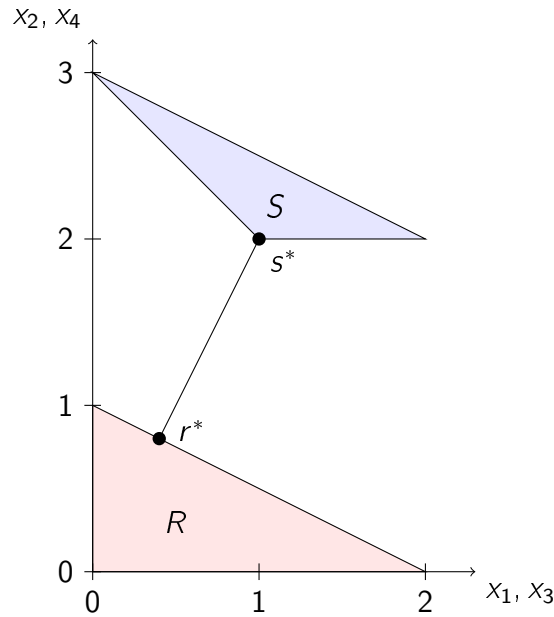


Abbildung 1.5: Geometrische Darstellung des Beispiels 1.7

Die Vektoren λ und μ heißen Lagrange-Multiplikatoren zu x^* .

Sei $x \in \mathcal{F}$. Wir bezeichnen mit

$$J(x) := \{1 \leq j \leq p \mid \langle g_j, x \rangle = r_j\} \quad (1.40)$$

die Indexmenge der in x aktiven Ungleichungsrestriktionen.

Satz 1.8 (Hinreichende Optimalitätsbedingung, vgl. Satz 5.4.13 in [Alt11, S. 198])

Sei f in $x^* \in \mathcal{F}$ zweimal stetig differenzierbar. Die notwendigen Bedingungen von Satz 1.7 seien erfüllt und es gelte mit $\alpha > 0$

$$d^T f''(x^*) d \geq \alpha \|d\|^2 \quad \forall d \in \mathbb{R}^n : \begin{cases} Ad = 0, \\ \langle g_j, d \rangle \leq 0 & \text{für } j \in J(x^*) \text{ mit } \mu_j = 0, \\ \langle g_j, d \rangle = 0 & \text{für } j \in J(x^*) \text{ mit } \mu_j > 0. \end{cases} \quad (1.41)$$

Dann ist x^* eine strikte lokale Lösung des Problems (PLU).

Ein Spezialfall des Problems (PLU) ist das Optimierungsproblem mit unteren und oberen Schranken für die Variablen.

Definition 1.10 (Optimierungsprobleme mit Variablenbeschränkungen)

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} f(x) && \text{(PVB)} \\ & \text{bei } a \leq x \leq b \end{aligned}$$

$a, b \in \mathbb{R}^n$ mit $a \leq b$.

Die Nebenbedingung ist äquivalent zu $Gx \leq r$ mit $G := \begin{pmatrix} -I_n \\ I_n \end{pmatrix}$ und $r := \begin{pmatrix} -a \\ b \end{pmatrix}$.

1.4 Optimierungsprobleme mit nichtlinearen Restriktionen

Abschließend kommt nun das allgemeine nichtlineare Optimierungsproblem. Das Optimierungsproblem (PN) ist

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} f(x) \\ & \text{bei } g_i(x) \leq 0 \text{ für } i = 1, \dots, p \\ & \quad h_j(x) = 0 \text{ für } j = 1, \dots, m. \end{aligned}$$

Um die Optimalitätsbedingung des Problems (PN) zu bekommen, definiert man eine sogenannte *Regularitätsbedingung* (vgl. Definition 7.2.14 in [Alt11, S. 261]), die hier nicht näher betrachtet wird. Die folgende Definition ist eine spezielle Bedingung, aus der die Regularitätsbedingung folgt. Sie kann aber noch aus anderen Bedingungen hergeleitet werden.

Definition 1.11 (Mangasarian-Fromowitz-Regularitätsbedingung, vgl. Satz 7.2.24 in [Alt11, S. 270]) Sei $x \in \mathcal{F}$. Mit $\mathcal{I}(x) := \{i \in \{1, \dots, p\} \mid g_i(x) = 0\}$ wird die Indexmenge der in x aktiven Ungleichungsrestriktionen bezeichnet. Seien g_i und h_j differenzierbar in x für $i = 1, \dots, p$ und $j = 1, \dots, m$. Der Punkt x erfüllt die Mangasarian-Fromowitz-Regularitätsbedingung, wenn die Gradienten

$$\nabla h_j(x), \quad j = 1, \dots, m, \quad \text{linear unabhängig} \tag{1.42}$$

sind und ein Vektor $d \in \mathbb{R}^n$ mit

$$\nabla g_i(x)^T d < 0, \quad i \in \mathcal{I}(x) \quad \text{und} \quad \nabla h_j(x)^T d = 0, \quad j = 1, \dots, m \tag{1.43}$$

existiert.

Der folgende Satz gilt eigentlich mit der allgemeinen Regularitätsbedingung. Es wird aber hier nur die Mangasarian-Fromowitz-Regularitätsbedingung verwendet.

Satz 1.9 (Notwendige Bedingung, vgl. Satz 7.2.11 in [Alt11, S. 260])

Sei x^ eine lokale Lösung des Problems (PN) und x^* erfülle die Mangasarian-Fromowitz-Regularitätsbedingung. Sei f in x^* differenzierbar. Dann existieren Vektoren $\lambda \in \mathbb{R}^m$ und $\mu \in \mathbb{R}^p$, sodass*

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) = 0 \quad (1.44)$$

$$\mu_j \geq 0, \quad \mu_j g_j(x^*) = 0, \quad j = 1, \dots, p \quad (1.45)$$

Für die hinreichende Bedingung zweiter Ordnung verweisen wir auf Satz 7.3.1 in [Alt11, S. 273].

Kapitel 2

SQP-Verfahren

In diesem Kapitel werden zwei SQP-Verfahren vorgestellt. Zuerst wird das SQP-Verfahren für Optimierungsprobleme mit linearen Restriktionen betrachtet. Danach kommt das SQP-Verfahren für allgemeine nichtlineare Optimierungsprobleme. Dazwischen werden zwei Hilfsverfahren zur Lösung quadratischer Optimierungsprobleme mit linearen Restriktionen vorgestellt, nämlich das Aktive-Mengen-Verfahren und das Nullraum-Verfahren. Die Beschreibungen in diesem Kapitel folgen der Herleitung in [Alt11, Trö11].

2.1 Einführung

Das SQP-Verfahren ist ein wichtiges Verfahren, um restringierte nichtlineare Optimierungsprobleme zu lösen. SQP ist eine Abkürzung für sequentielle quadratische Programmierung¹. Die Hauptidee von SQP ist nämlich, iterativ Teilprobleme in Form quadratischer Optimierungsprobleme zu formulieren und zu lösen.

Wir haben im Abschnitt 1.2.2 über das Newton-Verfahren gesehen, dass dieses Verfahren in jedem Iterationsschritt eigentlich ein unrestringiertes quadratisches Problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}(x - x^k)^T f''(x^k)(x - x^k) + \nabla f(x^k)^T(x - x^k) \quad (2.1)$$

löst. Diese Idee wird bei SQP auch verwendet. Aber es wird nicht mehr ein unrestringiertes Problem sein, weil die Nebenbedingungen auch berücksichtigt werden, d. h.

$$\min \frac{1}{2}(x - x^k)^T f''(x^k)(x - x^k) + \nabla f(x^k)^T(x - x^k) \quad (2.2)$$

bei $x \in \mathcal{F}$.

Es wird auch vorausgesetzt, dass $x^k \in \mathcal{F}$ ist.

¹englisch: Sequential Quadratic Programming.

2.2 Formulierung

Zunächst werden in diesem Abschnitt Optimierungsprobleme mit linearen Gleichungs- und Ungleichungsnebenbedingungen (PLU) betrachtet.

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} f(x) \\ & \text{bei } Ax = b \\ & \quad Gx \leq r \end{aligned}$$

Das Teilproblem nach (2.2) lautet dann

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} \frac{1}{2} (x - x^k)^T f''(x^k) (x - x^k) + \nabla f(x^k)^T (x - x^k) \\ & \text{bei } Ax = b \\ & \quad Gx \leq r. \end{aligned} \tag{2.3}$$

Setze $d := x - x^k$. Folglich ist $x = x^k + d$ und das Teilproblem lässt sich schreiben als

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} d^T f''(x^k) d + \nabla f(x^k)^T d \tag{2.4}$$

$$\text{bei } Ax^k + Ad = b \tag{2.5}$$

$$Gx^k + Gd \leq r \tag{2.6}$$

Das ist ein quadratisches Optimierungsproblem mit linearen Restriktionen. Im nächsten Abschnitt wird ein Verfahren für dieses Problem vorgestellt.

Da $x^k \in \mathcal{F}$ ist, d. h., es gilt $Ax^k = b$, kann als Gleichungsnebenbedingung an der Stelle (2.5) die Gleichung $Ad = 0$ geschrieben werden.

Verfahren 2.1 (SQP-Verfahren für (PLU), vgl. Verfahren 6.4.1 in [Alt11, S. 237])

1. Wähle einen zulässigen Startpunkt x^0 und setze $k := 0$.
2. Berechne die Lösung d des Problems

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} \frac{1}{2} d^T f''(x^k) d + \nabla f(x^k)^T d \\ & \text{bei } Ad = 0 \\ & \quad Gx^k + Gd \leq r \end{aligned} \tag{QLU}_k$$

Setze $d^k := d$.

3. Ist $d^k = 0 \Rightarrow STOP$.

4. Setze $x^{k+1} := x^k + d^k$ und $k := k + 1. \Rightarrow$ Gehe zu Schritt 2.

Es ist zu bemerken, dass der Punkt x^k die notwendige Optimalitätsbedingung des Problems (PLU) nach Satz 1.7 erfüllt, falls $d^k = 0$ gilt.

Sei d eine Lösung des Teilproblems (QLU_k). Dann gibt es nach Satz 1.7 die Vektoren λ und μ , sodass

$$f''(x^k)d + \nabla f(x^k) + A^T \lambda + G^T \mu = 0 \quad (2.7)$$

$$\mu_j \geq 0, \quad \mu_j(\langle g_j, x^k \rangle + \langle g_j, d \rangle - r_j) = 0, \quad j = 1, \dots, p. \quad (2.8)$$

Falls $d = 0$ ist, gilt

$$\nabla f(x^k) + A^T \lambda + G^T \mu = 0 \quad (2.9)$$

$$\mu_j \geq 0, \quad \mu_j(\langle g_j, x^k \rangle - r_j) = 0, \quad j = 1, \dots, p, \quad (2.10)$$

d. h., der Punkt x^k erfüllt die notwendige Optimalitätsbedingung.

Der folgende Satz macht eine Aussage zur Konvergenzgeschwindigkeit des Verfahrens.

Satz 2.1 (Vgl. Satz 6.4.4 in [Alt11, S. 243])

Sei x^* lokale Lösung von (PLU). Sei f auf D zweimal stetig differenzierbar und es gelte die Bedingung (1.28). Dann gibt es ein $\delta > 0$, sodass das Verfahren 2.1 für jeden Startpunkt $x^0 \in U_\delta(x^*)$ eine Folge $\{x^k\}$ definiert, die superlinear gegen x^* konvergiert, d. h., es existiert eine Nullfolge $\{r_k\} \subset \mathbb{R}_+$ mit

$$\|x^{k+1} - x^*\| \leq r_k \|x^k - x^*\| \quad \forall k \geq 0. \quad (2.11)$$

Ist zusätzlich f'' auf D Lipschitz-stetig², dann konvergiert die Folge $\{x^k\}$ für jeden Startpunkt $x^0 \in U_\delta(x^*)$ quadratisch gegen x^* , d. h., es existieren $c > 0$ und $K \in \mathbb{N}$ mit

$$\|x^{k+1} - x^*\| \leq c \|x^k - x^*\|^2 \quad \forall k \geq K. \quad (2.12)$$

Es kann gezeigt werden, dass mit den Voraussetzungen des Satzes die Lösbarkeit des Problems (QLU_k) gesichert werden kann. Der Satz zeigt, dass das Verfahren 2.1 leider nur lokal konvergiert. Um ein global konvergentes Verfahren zu bekommen, muss eine geeignete Schrittweite gewählt werden.

² f'' ist auf D Lipschitz-stetig, falls es eine Konstante $L > 0$ mit $\|f''(x) - f''(y)\| \leq L\|x - y\|$ für alle $x, y \in D$ gibt.

2.3 Aktive-Mengen-Verfahren

Es ist jetzt aufzuklären, wie die im SQP-Verfahren vorkommenden quadratischen Teilprobleme mit dem Aktive-Mengen-Verfahren gelöst werden können.

Definition 2.1 (Quadratische Optimierungsprobleme mit linearen Restriktionen)

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) &:= \frac{1}{2} \langle Qx, x \rangle + \langle q, x \rangle & (\text{QLU}) \\ \text{bei } Ax &= b \\ Gx &\leq r \end{aligned}$$

Dabei sei Q eine symmetrische $(n \times n)$ -Matrix und $q \in \mathbb{R}^n$. A sei eine $(m \times n)$ -Matrix mit $m \leq n$ und $b \in \mathbb{R}^m$. G sei eine $(p \times n)$ -Matrix und $r \in \mathbb{R}^p$.

Die Idee des Aktive-Mengen-Verfahrens ist, iterativ durch die aktiven Nebenbedingungen ein Problem mit nur Gleichungsnebenbedingungen zu formulieren und zu lösen.

Sei $x^k \in \mathbb{R}^n$ ein zulässiger Punkt des Problems (QLU). Angenommen, eine Suchrichtung $d \in \mathbb{R}^n$ kann gefunden werden, sodass der Punkt $x^{k+1} := x^k + d$ ein besserer zulässiger Punkt ist. Dieser Vektor d lässt sich finden, indem das Problem

$$\begin{aligned} \min_{d \in \mathbb{R}^n} & \frac{1}{2} \langle Q(x^k + d), x^k + d \rangle + \langle q, x^k + d \rangle & (2.13) \\ \text{bei } A(x^k + d) &= b \\ G(x^k + d) &\leq r \end{aligned}$$

gelöst wird. Dies ist eigentlich das Problem (QLU), wobei die Variable x mit $x^k + d$ ersetzt wurde und das Problem über die Variable d minimiert werden soll.

Da die Konstanten in der Zielfunktion vernachlässigt werden können, ergibt sich als Zielfunktion

$$\frac{1}{2} \langle Qd, d \rangle + \langle Qx^k + q, d \rangle.$$

Weil x^k zulässig ist, d. h., es gilt $Ax^k = b$, vereinfacht sich die Gleichungsnebenbedingung zu $Ad = 0$.

Seien $a_1, \dots, a_m \in \mathbb{R}^n$ bzw. $g_1, \dots, g_p \in \mathbb{R}^n$ wieder die Transponierten der Zeilenvektoren in A bzw. G wie in (1.32) und $J(x^k) := \{1 \leq j \leq p \mid \langle g_j, x^k \rangle = r_j\}$ die Indexmenge der aktiven Ungleichungsrestriktionen zu x^k wie in (1.40).

Ist eine Ungleichungsrestriktion zu x^k mit Index j aktiv, so kann diese Ungleichungsrestriktion auch zu $x^k + d$ aktiv bleiben, wenn $\langle g_j, d \rangle = 0$ gilt.

Seien $J_k := J(x^k)$, $p_k := |J_k|$ und $j_1, \dots, j_{p_k} \in J_k$. Weiter seien

$$G_k = \begin{pmatrix} g_{j_1}^T \\ \vdots \\ g_{j_{p_k}}^T \end{pmatrix} \quad \text{und} \quad B_k = \begin{pmatrix} A \\ G_k \end{pmatrix}. \quad (2.14)$$

Definiere dann folgendes Teilproblem mit nur linearen Gleichungsnebenbedingungen:

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \frac{1}{2} \langle Qd, d \rangle + \langle Qx^k + q, d \rangle \\ \text{bei} \quad & B_k d = 0. \end{aligned} \quad (\text{QLG}_k)$$

Wir setzen hier folgende Annahmen voraus:

$$\begin{aligned} & B_k \text{ hat vollen Rang,} \\ & \text{d. h., } a_1, \dots, a_m, g_{j_1}, \dots, g_{j_{p_k}} \text{ sind linear unabhängig,} \end{aligned} \quad (2.15)$$

und

$$Q \text{ ist positiv definit auf Kern } B_k. \quad (2.16)$$

Sei d eine Lösung des Teilproblems (QLG_k) . Dann existiert nach Satz 1.5 ein Lagrange-Multiplikator $\nu = \begin{pmatrix} \lambda \\ \mu \end{pmatrix}$ mit $\lambda \in \mathbb{R}^m$ und $\mu \in \mathbb{R}^{p_k}$, sodass

$$Qd + Qx^k + q + A^T \lambda + G_k^T \mu = 0. \quad (2.17)$$

Sei $d \neq 0$. Dann ist d eine Abstiegsrichtung (siehe Definition 1.6).

Für den Gradient der Zielfunktion des Problems (QLU) an der Stelle x^k gilt wegen (2.17)

$$\nabla f(x^k) = Qx^k + q = -Qd - A^T \lambda - G_k^T \mu \quad (2.18)$$

und somit

$$\nabla f(x^k)^T d = -d^T Qd - \lambda^T Ad - \mu^T G_k d = -d^T Qd, \quad (2.19)$$

weil $Ad = 0$ und $G_k d = 0$. Wegen der Bedingung (2.16) gilt $d^T Qd > 0$ und folglich

$$\nabla f(x^k)^T d < 0, \quad (2.20)$$

d. h., d ist eine Abstiegsrichtung.

Daher kann nach Satz 1.4 eine Schrittweite σ berechnet werden, sodass

$$f(x^{k+1} := x^k + \sigma d) < f(x^k) \quad (2.21)$$

gilt.

Sei t eine Schrittweite. Es gilt

$$f(x^k + td) = \frac{1}{2}(x^k + td)^T Q(x^k + td) + q^T(x^k + td) \quad (2.22)$$

$$= \frac{1}{2}(x^k)^T Qx^k + t(x^k)^T Qd + \frac{1}{2}t^2 d^T Qd + q^T x^k + tq^T d \quad (2.23)$$

$$= f(x^k) + t(x^k)^T Qd + \frac{1}{2}t^2 d^T Qd + tq^T d \quad (2.24)$$

$$= f(x^k) + \frac{1}{2}t^2 d^T Qd + t(Qx^k + q)^T d \quad (2.25)$$

$$= f(x^k) + \frac{1}{2}t^2 d^T Qd + t\nabla f(x^k)^T d \quad (2.26)$$

$$= f(x^k) + \frac{1}{2}t^2 d^T Qd - td^T Qd \quad (2.27)$$

$$= f(x^k) + (\frac{1}{2}t^2 - t)d^T Qd \quad (2.28)$$

Die Funktion $\frac{1}{2}t^2 - t$ hat ein globales Minimum an der Stelle $t = 1$ mit dem Optimalwert $-\frac{1}{2}$. Man erhält also mit $t = 1$ den größten Abstieg.

Es ist noch zu beachten, dass $x^{k+1} := x^k + td$ nicht unbedingt ein zulässiger Punkt ist. Durch die Bedingungen $Ad = 0$ und $G_k d = 0$ bekommt man zwar

$$Ax^{k+1} = Ax^k + tAd = Ax^k \leq 0 \quad (2.29)$$

und

$$\langle g_j, x^{k+1} \rangle = \langle g_j, x^k \rangle + t\langle g_j, d \rangle = \langle g_j, x^k \rangle = r_j \text{ für } j \in J_k, \quad (2.30)$$

aber es kann sein, dass eine Ungleichung

$$\langle g_j, x^{k+1} \rangle \leq r_j \text{ für } j \in \{1, \dots, p\} \setminus J_k \quad (2.31)$$

verletzt wurde.

Sei nun $j \in \{1, \dots, p\} \setminus J_k$. Falls $\langle g_j, d \rangle < 0$ gilt, dann erfüllt x^{k+1}

$$\langle g_j, x^{k+1} \rangle = \langle g_j, x^k \rangle + t\langle g_j, d \rangle < \langle g_j, x^k \rangle \leq r_j. \quad (2.32)$$

Für den Fall $\langle g_j, d \rangle > 0$ gilt

$$\langle g_j, x^{k+1} \rangle = \langle g_j, x^k \rangle + t\langle g_j, d \rangle \leq r_j \quad (2.33)$$

nur dann, wenn t die Ungleichung

$$t \leq \frac{r_j - \langle g_j, x^k \rangle}{\langle g_j, d \rangle} \quad (2.34)$$

erfüllt. Sei

$$I_k := \{ j \in \{1, \dots, p\} \setminus J_k \mid \langle g_j, d \rangle > 0 \}. \quad (2.35)$$

Dann muss t die Ungleichung (2.34) für alle $j \in I_k$ erfüllen.

Zusammengefasst ist die Schrittweite σ_k wie folgt zu definieren.

$$\sigma_k := \min\{1, \tau_k\}, \quad (2.36)$$

wobei

$$\tau_k := \begin{cases} \min \left\{ \frac{r_j - \langle g_j, x^k \rangle}{\langle g_j, d \rangle} \mid j \in I_k \right\}, & \text{falls } I_k \neq \emptyset \\ \infty, & \text{sonst.} \end{cases} \quad (2.37)$$

Es ist nun der Fall $d = 0$ zu betrachten. In diesem Fall lautet die Gleichung (2.17)

$$Qx^k + q + A^T \lambda + G_k^T \mu = 0. \quad (2.38)$$

Sei $\mu \geq 0$. Setze

$$\lambda^* := \lambda \quad \text{und} \quad \mu^* = (\mu_j^*)_{j=1, \dots, p} \quad \text{mit} \quad \mu_j^* := \begin{cases} \mu_j, & \text{falls } j \in J_k \\ 0, & \text{sonst.} \end{cases} \quad (2.39)$$

Dann erfüllt x^k die notwendigen Bedingungen von Satz 1.7

$$\nabla f(x^k) + A^T \lambda^* + G^T \mu^* = Qx^k + q + A^T \lambda^* + G_k^T \mu^* = 0 \quad (2.40)$$

$$\mu_j^* (\langle g_j, x^k \rangle - r_j) = 0 \quad \text{für } j = 1, \dots, p \quad (2.41)$$

$$\mu^* \geq 0 \quad (2.42)$$

und wegen Voraussetzung (2.16) auch die hinreichende Optimalitätsbedingung (1.8). Daher ist x^k eine optimale Lösung von (QLU) und das Verfahren kann somit gestoppt werden.

Falls $\mu \not\geq 0$, dann wähle den Index $j \in J_k$ mit

$$\mu_j = \min\{\mu_i \mid i \in J_k\} < 0 \quad (2.43)$$

und setze

$$\tilde{J}_k := J_k \setminus \{j\} \quad \text{und} \quad \tilde{p}_k := p_k - 1. \quad (2.44)$$

Mit \tilde{J}_k bekommt man eine neue Matrix \tilde{B}_k und ein neues Problem (QLG_k). Die j -te Ungleichung wird somit nicht mehr als aktiv angesehen.

Sei \tilde{d} die Lösung des neuen Problems mit den zugehörigen Multiplikatoren $\tilde{\lambda}$ und $\tilde{\mu}$. Es gilt die notwendige Bedingung

$$Q\tilde{d} + Qx^k + q + A^T \tilde{\lambda} + G_k^T \tilde{\mu} = 0. \quad (2.45)$$

Angenommen es gilt $\tilde{d} = 0$. Dann lautet die notwendige Bedingung

$$Qx^k + q + A^T \tilde{\lambda} + G_k^T \tilde{\mu} = 0. \quad (2.46)$$

Fasst man diese mit der Gleichung (2.38) zusammen, so bekommt man

$$-A^T \lambda - G_k^T \mu + A^T \tilde{\lambda} + \tilde{G}_k^T \tilde{\mu} = 0 \quad (2.47)$$

$$\sum_{i=1}^m (\tilde{\lambda}_i - \lambda_i) a_i + \sum_{i=1}^{p_k-1} (\tilde{\mu}_i - \mu_i) g_{j_i} = \mu_j g_j \quad (2.48)$$

Für die letzte Gleichung wird o. B. d. A. angenommen, dass $j = p_k$.

Wegen $\mu_j < 0$ heißt es, dass g_j durch die Linearkombination der Vektoren a_1, \dots, a_m und $g_{j_1}, \dots, g_{j_{p_k}}$ darstellbar ist. Das ist ein Widerspruch zu der Voraussetzung (2.15). Daher gilt $\tilde{d} \neq 0$ und dies wäre wie zuvor beschrieben eine Abstiegsrichtung.

Verfahren 2.2 (Aktive-Mengen-Verfahren für (QLU), vgl. Verfahren 6.2.1 in [Alt11, S. 213])

1. Wähle einen Startpunkt x^0 und setze $k := 0$.
2. Löse das Problem $(\text{QLG}_k) \Rightarrow$ Erhalte d^k, λ^k und μ^k .
3. Falls $d^k = 0$
 - 3a. Falls $\mu^k \geq 0 \Rightarrow \text{STOP}$.
 - 3b. Falls $\mu^k \not\geq 0$
 - i. Bestimme $j \in J_k$, sodass $\mu_j^k = \min\{\mu_i^k \mid i \in J_k\}$.
 - ii. Setze $\tilde{J}_k := J_k \setminus \{j\}$ und bestimme \tilde{G}_k .
 - iii. Löse das neue Problem $(\text{QLG}_k) \Rightarrow$ Erhalte $d^k \neq 0, \lambda^k$ und μ^k .
4. Berechne die Schrittweite σ_k nach (2.36).
5. Setze $x^{k+1} := x^k + \sigma_k d^k$ und $k := k + 1. \Rightarrow$ Gehe zu Schritt 2.

Es kann gezeigt werden, dass das Verfahren 2.2 in endlich vielen Schritten die Lösung des Problems (QLU) berechnet (siehe Satz 6.2.3 in [Alt11, S. 214]).

2.4 Nullraum-Verfahren

So wie das SQP-Verfahren die Hilfe von Aktive-Mengen-Verfahren braucht, braucht das Aktive-Mengen-Verfahren die Hilfe von Nullraum-Verfahren, um die entstehenden quadratischen Teilprobleme mit nur linearen Gleichungsnebenbedingungen zu lösen.

Definition 2.2 (Quadratische Optimierungsprobleme mit linearen Gleichungsnebenbedingungen)

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) &:= \frac{1}{2} \langle Qx, x \rangle + \langle q, x \rangle & (\text{QLG}) \\ \text{bei } Ax &= b \end{aligned}$$

Dabei sei Q eine symmetrische $(n \times n)$ -Matrix und $q \in \mathbb{R}^n$. A sei eine $(m \times n)$ -Matrix mit $m \leq n$ und $b \in \mathbb{R}^m$.

Das Problem (QLG) wird hier auf ein unrestringiertes Verfahren mit Hilfe einer Nullraum-Matrix von A (siehe Definition 1.8) reduziert werden. Daher stammt der Name *Nullraum-Verfahren*. Es folgen zunächst sämtliche Schritte des Verfahrens und anschließend die Erklärungen der einzelnen Schritte.

Verfahren 2.3 (Nullraum-Verfahren für (QLG), vgl. Verfahren 6.1.1 in [Alt11, S. 208 f.])

1. Finde mit Hilfe der QR-Zerlegung eine unitäre $(n \times n)$ -Matrix H und eine obere $(m \times m)$ -Dreiecksmatrix R mit

$$HA^T = \begin{pmatrix} R \\ 0 \end{pmatrix}. \quad (2.49)$$

2. Berechne den Vektor $y \in \mathbb{R}^m$ als Lösung der Gleichung

$$R^T y = b. \quad (2.50)$$

3. Berechne

$$h := -Hq = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \quad \text{und} \quad B := HQH^T = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \quad (2.51)$$

wobei $h_1 \in \mathbb{R}^m$ und $B_{11} \in \mathbb{R}^{m,m}$.

4. Berechne den Vektor $z \in \mathbb{R}^{n-m}$ als Lösung der Gleichung

$$B_{22}z = h_2 - B_{21}y. \quad (2.52)$$

5. Berechne x^* mit

$$x^* := H^T \begin{pmatrix} y \\ z \end{pmatrix}. \quad (2.53)$$

6. Der Multiplikator λ ist die Lösung der Gleichung

$$R\lambda = h_1 - B_{11}y - B_{12}z. \quad (2.54)$$

In Schritt 1 bekommt man die Matrizen H und R . Die Matrix H kann so zerlegt werden, dass

$$H^T = \begin{pmatrix} Y & Z \end{pmatrix} \quad (2.55)$$

mit $Y \in \mathbb{R}^{n,m}$ und $Z \in \mathbb{R}^{n,n-m}$ gilt. Wegen der Orthogonalität von H können die Spaltenvektoren in Y und Z eine Basis von \mathbb{R}^n bilden. D. h., für jedes Element $x \in \mathbb{R}^n$ gibt es eindeutig bestimmte Vektoren $x_y \in \mathbb{R}^m$ und $x_z \in \mathbb{R}^{n-m}$, sodass

$$x = Yx_y + Zx_z = H^T \begin{pmatrix} x_y \\ x_z \end{pmatrix} \quad (2.56)$$

gilt. Sei $d \in \text{Kern } A \subseteq \mathbb{R}^n$. Dann gibt es $d_y \in \mathbb{R}^m$ und $d_z \in \mathbb{R}^{n-m}$ mit $d = Yd_y + Zd_z$. Es gilt folglich

$$Ad = AYd_y + AZd_z = AH^T \begin{pmatrix} d_y \\ d_z \end{pmatrix} = \begin{pmatrix} R^T & 0 \end{pmatrix} \begin{pmatrix} d_y \\ d_z \end{pmatrix} = R^T d_y \quad (2.57)$$

und wegen der Invertierbarkeit der Matrix R

$$Ad = 0 \Leftrightarrow d_y = 0 \Leftrightarrow d = Zd_z. \quad (2.58)$$

Daraus folgt

$$\text{Kern } A = \{d \in \mathbb{R}^n \mid Ad = 0\} = \{d \in \mathbb{R}^n \mid \exists d_z \in \mathbb{R}^{n-m} : d = Zd_z\} = \text{Im } Z. \quad (2.59)$$

Deshalb ist diese Matrix Z die Nullraum-Matrix von A .

Sei x^* die optimale Lösung. Dann gibt es y^* und z^* mit

$$x^* = Yy^* + Zz^* = H^T \begin{pmatrix} y^* \\ z^* \end{pmatrix}. \quad (2.60)$$

Es gilt

$$b = Ax^* = AH^T \begin{pmatrix} y^* \\ z^* \end{pmatrix} = \begin{pmatrix} R^T & 0 \end{pmatrix} \begin{pmatrix} y^* \\ z^* \end{pmatrix} = R^T y^* \quad (2.61)$$

Daher ist in Schritt 2 eigentlich der Vektor y^* zu bestimmen.

Da Z eine Nullraum-Matrix ist, gilt $Zz^* \in \text{Kern } A$, d. h. $AZz^* = 0$. Dann erfüllt $w := Yy^*$

$$Aw = Aw + AZz^* = A(w + Zz^*) = A(Yy^* + Zz^*) = Ax^* = b \quad (2.62)$$

Der Vektor w ist somit eine spezielle Lösung der Gleichung $Ax = b$.

Nun ergibt sich für unser Problem (QLG) wie in (1.31) folgende Umformulierung:

$$\min_{z \in \mathbb{R}^{n-m}} f(w + Zz) =: F(z). \quad (2.63)$$

Die Lösung dieses Problems erfüllt nach Satz 1.1 die Optimalitätsbedingung

$$\nabla F(z^*) = 0. \quad (2.64)$$

Es ist

$$F(z) = f(w + Zz) = \frac{1}{2}(w + Zz)^T Q(w + Zz) + q^T(w + Zz) \quad (2.65)$$

$$= \frac{1}{2}w^T Qw + \frac{1}{2}w^T QZz + \frac{1}{2}z^T ZQw + \frac{1}{2}z^T Z^T QZz + q^T w + q^T Zz \quad (2.66)$$

$$= \frac{1}{2}z^T Z^T QZz + w^T QZz + q^T Zz + f(w) \quad (2.67)$$

$$= \frac{1}{2}z^T Z^T QZz + (Z^T Qw + Z^T q)^T z + f(w) \quad (2.68)$$

$$= \frac{1}{2}z^T Z^T QZz + (Z^T QYy^* + Z^T q)^T z + f(Yy^*). \quad (2.69)$$

Für den Gradienten von F gilt dann

$$\nabla F(z) = Z^T QZz + Z^T QYy^* + Z^T q. \quad (2.70)$$

In Schritt 3 sind

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = -Hq = -\begin{pmatrix} Y^T \\ Z^T \end{pmatrix} q = -\begin{pmatrix} Y^T q \\ Z^T q \end{pmatrix} \quad (2.71)$$

und

$$\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = HQH^T = \begin{pmatrix} Y^T \\ Z^T \end{pmatrix} Q \begin{pmatrix} Y & Z \end{pmatrix} = \begin{pmatrix} Y^T QY & Y^T QZ \\ Z^T QY & Z^T QZ \end{pmatrix}. \quad (2.72)$$

Die Bedingung $\nabla F(z^*) = 0$ kann daher als

$$B_{22}z^* + B_{21}y^* - h_2 = 0 \quad (2.73)$$

geschrieben werden. Mit Hilfe dieser Gleichung bestimmt man in Schritt 4 den Vektor z^* und folglich in Schritt 5 die Lösung x^* nach (2.60).

Zum Abschluss wird in Schritt 6 der Multiplikator λ berechnet. Nach Satz 1.5 lautet die Optimalitätsbedingung des Problems (QLG)

$$Qx^* + q + A^T \lambda = 0. \quad (2.74)$$

Multipliziere diese mit H , so ergibt sich

$$0 = HQx^* + Hq + HA^T \lambda = \begin{pmatrix} Y^T \\ Z^T \end{pmatrix} Qx^* - \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \begin{pmatrix} R \\ 0 \end{pmatrix} \lambda. \quad (2.75)$$

Aus dem oberen Teil dieses Gleichungssystems lässt sich die Gleichung

$$0 = Y^T Q(Yy^* + Zz^*) - h_1 + R^T \lambda = B_{11}y^* + B_{12}z^* - h_1 + R^T \lambda \quad (2.76)$$

zur Berechnung von λ finden.

2.5 Lagrange-Newton-SQP-Verfahren

Nun kommt das Verfahren für die allgemeinen nichtlinearen Optimierungsprobleme (PN).

Definition 2.3 Die Lagrange-Funktion $\mathcal{L} : D \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ des Problems (PN) sei durch

$$\mathcal{L}(x, \lambda, \mu) := f(x) + \lambda^T h(x) + \mu^T g(x) \quad (2.77)$$

definiert.

Verfahren 2.4 (SQP-Verfahren für (PN), vgl. Verfahren 8.1.4 in [Alt11, S. 294])

1. Wähle einen Startpunkt $z^0 := (x^0, \lambda^0, \mu^0)$ und setze $k := 0$.
2. Löse das Problem

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} d^T \mathcal{L}_{xx}(x^k, \lambda^k, \mu^k) d + \nabla f(x^k)^T d \quad (2.78)$$

$$\text{bei } h(x^k) + h'(x^k)d = 0 \quad (2.79)$$

$$g(x^k) + g'(x^k)d \leq 0. \quad (2.80)$$

\Rightarrow Erhalte d, λ und μ .

3. Ist $d = 0 \Rightarrow \text{STOP}$.

4. Setze $x^{k+1} := x^k + d$, $\lambda^{k+1} := \lambda$, $\mu^{k+1} := \mu$ und $k := k + 1$. \Rightarrow Gehe zu Schritt 2.

Im Verfahren 2.4 wird also iterativ die Suchrichtung über eine quadratische Näherung der Lagrange-Funktion des Problems berechnet, wobei die Nebenbedingungen linearisiert werden.

Eine ausführliche Erklärung und ein Konvergenzresultat zum Verfahren 2.4 sind in Abschnitt 8.1.2 in [Alt11, S. 290 ff.] zu finden.

Kapitel 3

Das halbglatte Newton-Verfahren

In diesem Kapitel wird das halbglatte¹ Newton-Verfahren vorgestellt. Dieses Verfahren hat eigentlich den gleichen Zweck wie das Newton-Verfahren, und zwar ein Gleichungssystem zu lösen. Man kann aber zeigen, dass es zum Lösen einer restringierten Optimierungsaufgabe verwendet werden kann. Dieses Kapitel basiert auf [Ul11, IK08].

3.1 Grundlagen

Vor der Vorstellung des Verfahrens wird die Definition einer halbglatten Funktion benötigt.

Definition 3.1 (Vgl. Definition 2.1 in [Ul11, S. 25])

Sei $V \subset \mathbb{R}^n$ eine nichtleere offene Menge. Sei $f : V \rightarrow \mathbb{R}^m$ Lipschitz-stetig in der Nähe von $x \in V$, d. h., es gibt eine offene Menge $V(x) \subset V$, wo f Lipschitz-stetig ist. Sei $D_f \subset V$ die Menge aller Punkte $v \in V$, wo f eine Ableitung $f'(v) \in \mathbb{R}^{m,n}$ besitzt. Die Menge

$$\partial_B f(x) := \{M \in \mathbb{R}^{m \times n} \mid \exists \{x^k\} \subset D_f : x^k \rightarrow x, f'(x^k) \rightarrow M\} \quad (3.1)$$

heißt *B-Sub-Ableitung*² von f an der Stelle x . Ferner ist nach Clarke die allgemeine Jacobi-Matrix von f an der Stelle x die konvexe Hülle³

$$\partial f(x) := \text{co } \partial_B f(x). \quad (3.2)$$

¹englisch: semismooth.

²englisch: B-subdifferential, „B“ für Bouligand.

³Die konvexe Hülle einer Menge M ist

$$\text{co } M := \left\{ \sum_{i=1}^k \lambda_i v_i \mid k \in \mathbb{N}, v_i \in M, \lambda_i \geq 0 \text{ für } i = 1, \dots, k, \sum_{i=1}^k \lambda_i = 1 \right\}.$$

Definition 3.2 (Vgl. Definition 2.5 in [Ul11, S. 27])

Sei $V \subset \mathbb{R}^n$ nichtleer und offen. Die Funktion $f: V \rightarrow \mathbb{R}^m$ heißt *halbglatte* in $x \in V$, falls f lokal Lipschitz-stetig in x ist und

$$\lim_{M \in \partial f(x+\tau d), d \rightarrow s, \tau \rightarrow 0^+} M d \quad (3.3)$$

für alle $s \in \mathbb{R}^m$ existiert.

f heißt *halbglatte* (in V), falls f halbglatte in allen $x \in V$ ist.

Beispiel 3.1 (Vgl. Abschnitt 2.5.1 in [Ul11, S. 31 f.])

Die Funktion

$$f: \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \|x\| = \sqrt{x^T x} \quad (3.4)$$

ist eine halbglatte Funktion mit

$$\partial f(x) = \partial_B f(x) = \left\{ \frac{x^T}{\|x\|} \right\} \text{ für } x \neq 0, \quad (3.5)$$

$$\partial_B f(0) = \{v^T \mid v \in \mathbb{R}^n, \|v\| = 1\} \quad \text{und} \quad \partial f(0) = \{v^T \mid v \in \mathbb{R}^n, \|v\| \leq 1\}. \quad (3.6)$$

Beispiel 3.2 (Vgl. Abschnitt 2.5.3 in [Ul11, S. 32 ff.])

Stückweise differenzierbare Funktionen sind halbglatte. Ein Beispiel ist die Funktion

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \min\{x_1, x_2\}. \quad (3.7)$$

Jetzt kann das halbglatte Newton-Verfahren zum Lösen der Gleichung

$$f(x) = 0 \quad (3.8)$$

formuliert werden, wobei $f: V \rightarrow \mathbb{R}^n$, $V \subset \mathbb{R}^n$ offen, halbglatte an der Stelle der Lösung $x^* \in V$ ist.

Verfahren 3.1 (Das halbglatte Newton-Verfahren, vgl. Algorithmus 2.11 in [Ul11, S. 29])

1. Wähle einen Startpunkt x^0 und setze $k := 0$.
2. Ist $f(x^k) = 0 \Rightarrow \text{STOP}$.
3. Wähle $M_k \in \partial f(x^k)$.
4. Berechne die Lösung d des linearen Gleichungssystems

$$M_k d = -f(x^k). \quad (3.9)$$

Setze $d^k := d$.

5. Setze $x^{k+1} := x^k + d^k$ und $k := k + 1$. \Rightarrow Gehe zu Schritt 2.

Das halbglatte Newton-Verfahren funktioniert prinzipiell wie das Newton-Verfahren. Der Unterschied liegt darin, dass hier die Funktion nicht unbedingt glatt ist.

Satz 3.1 (Vgl. Satz 2.12 in [Ul11, S. 29 f.])

Sei $f: V \rightarrow \mathbb{R}^n$ auf der offenen Menge $V \subset \mathbb{R}^n$ definiert. Sei x^* eine Lösung des Problems (3.8). Es gelte:

- (i) Die Funktion f sei halbglatte in x^* .
- (ii) Es existiere eine Konstante $C > 0$, sodass die Matrizen M_k für alle k nichtsingulär mit $\|M_k^{-1}\| \leq C$ sind.

Dann existiert $\delta > 0$, sodass für alle $x^0 \in U_\delta(x^*)$ das Verfahren 3.1 entweder mit $x^k = x^*$ terminiert oder eine Folge $\{x^k\}$ erzeugt, die superlinear gegen x^* konvergiert.

3.2 Formulierung

Es wird nun das halbglatte Newton-Verfahren zur Lösung nichtlinearer Optimierungsprobleme mit linearen Restriktionen formuliert.

Für eine Lösung x^* des Problems (PLU)

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{bei} \quad & Ax = b \\ & Gx \leq r \end{aligned} \tag{PLU}$$

gelten nach Satz 1.7 die Optimalitätsbedingungen

$$\nabla f(x^*) + A^T \lambda + G^T \mu = 0 \tag{3.10}$$

$$Ax^* = b \tag{3.11}$$

$$\mu_j \geq 0, \quad \langle g_j, x^* \rangle \leq r_j, \quad \mu_j(r_j - \langle g_j, x^* \rangle) = 0 \quad \text{für } j = 1, \dots, p \tag{3.12}$$

mit Vektoren $\lambda \in \mathbb{R}^m$ und $\mu \in \mathbb{R}^p$.

Gegeben seien $x_1, x_2 \in \mathbb{R}$. Es gilt

$$\begin{aligned} x_1, x_2 &\geq 0 \\ x_1 x_2 &= 0 \end{aligned} \Leftrightarrow \min\{x_1, x_2\} = 0. \tag{3.13}$$

Die Bedingungen (3.12) sind somit äquivalent zu

$$\min \{\mu_j, r_j - \langle g_j, x^* \rangle\} = 0 \quad \text{für } j = 1, \dots, p, \quad (3.14)$$

oder in Matrixschreibweise:

$$\min \{\mu, r - Gx^*\} = 0, \quad (3.15)$$

wobei der Operator min hier dann elementenweise arbeitet.

Auf diese Weise können die Optimalitätsbedingungen wie folgt geschrieben werden.

$$\nabla f(x^*) + A^T \lambda + G^T \mu = 0 \quad (3.16)$$

$$Ax^* = b \quad (3.17)$$

$$\min \{\mu, r - Gx^*\} = 0. \quad (3.18)$$

Diese Neufassung der Optimalitätsbedingungen motiviert die Definition der halbglatten Funktion

$$F(x, \lambda, \mu) := \begin{pmatrix} \nabla f(x) + A^T \lambda + G^T \mu \\ Ax - b \\ \min \{\mu, r - Gx\} \end{pmatrix} \quad (3.19)$$

und den Einsatz des halbglatten Newton-Verfahrens zur Lösung des Gleichungssystems

$$F(x, \lambda, \mu) = 0. \quad (3.20)$$

Seien die Indexmengen

$$\mathcal{A} := \{j \in \{1, \dots, p\} \mid r_j - \langle g_j, x \rangle < \mu_j\}, \quad (3.21)$$

$$\mathcal{I} := \{j \in \{1, \dots, p\} \mid r_j - \langle g_j, x \rangle \geq \mu_j\}. \quad (3.22)$$

Mit

$$\chi_K(k) := \begin{cases} 1 & \text{für } k \in K, \\ 0 & \text{sonst} \end{cases} \quad (3.23)$$

gilt

$$\min \{\mu, r - Gx\} = \begin{pmatrix} \chi_{\mathcal{I}}(1) \mu_1 + \chi_{\mathcal{A}}(1) (r_1 - \langle g_1, x \rangle) \\ \vdots \\ \chi_{\mathcal{I}}(p) \mu_p + \chi_{\mathcal{A}}(p) (r_p - \langle g_p, x \rangle) \end{pmatrix}. \quad (3.24)$$

Dann ist

$$F'(x, \lambda, \mu) = \begin{pmatrix} f''(x) & A^T & G^T \\ A & 0 & 0 \\ -\chi_{\mathcal{A}}(1) g_1^T & 0 & \chi_{\mathcal{I}}(1) e_1^T \\ \vdots & \vdots & \vdots \\ -\chi_{\mathcal{A}}(p) g_p^T & 0 & \chi_{\mathcal{I}}(p) e_p^T \end{pmatrix} \quad (3.25)$$

ein Element von $\partial F(x, \lambda, \mu)$.

Verfahren 3.2 (Das halbglatte Newton-Verfahren für (PLU))

1. Wähle x^0, λ^0, μ^0 und setze $k := 0$.

2. Ist $F(x^k, \lambda^k, \mu^k) = 0 \Rightarrow \text{STOP}$.

3. Berechne die Lösung $d = \begin{pmatrix} d_x \\ d_\lambda \\ d_\mu \end{pmatrix}$ des linearen Gleichungssystems

$$F'(x^k, \lambda^k, \mu^k)d = -F(x^k, \lambda^k, \mu^k). \quad (3.26)$$

Setze $d_x^k := d_x, d_\lambda^k := d_\lambda$ und $d_\mu^k := d_\mu$.

4. Setze

$$x^{k+1} := x^k + d_x^k, \quad \lambda^{k+1} := \lambda^k + d_\lambda^k, \quad \mu^{k+1} := \mu^k + d_\mu^k \quad (3.27)$$

und $k := k + 1. \Rightarrow$ Gehe zu Schritt 2.

3.3 Aktive-Mengen-Strategie

In diesem Abschnitt wird die Aktive-Mengen-Strategie betrachtet. Dieses Verfahren ist interessant, weil man es in der Praxis als die Implementierung des halbglatten Newton-Verfahrens nehmen kann. Es ist mit dem Aktiven-Mengen-Verfahren in Abschnitt 2.3 verwandt.

Als Literatur über dieses Verfahren sei auf [IK08, Kapitel 7, S. 189 ff.] verwiesen. Es wurde in [IK08, Abschnitt 8.4, S. 240 ff.] gezeigt, dass es als halbglatte Newton-Verfahren interpretiert werden kann.

Es ist hier das Problem (PLU)

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} f(x) && \text{(PLU)} \\ & \text{bei } Ax = b \\ & \quad Gx \leq r \end{aligned}$$

zu betrachten.

Eine der Optimalitätsbedingungen dieses Problems ist nach Satz 1.7

$$\nabla f(x) + A^T \lambda + G^T \mu = 0. \quad (3.28)$$

Dies ist ein nichtlineares Gleichungssystem, wenn f nichtlinear ist.

Sei x^k ein Punkt in der Nähe der Lösung des Problems. Die Approximation der Funktion f in der Umgebung von x^k lautet

$$f(x) \approx f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2} (x - x^k)^T f''(x^k) (x - x^k). \quad (3.29)$$

Die Approximation des Gradienten von f ist folglich

$$\nabla f(x) \approx \nabla f(x^k) + f''(x^k)x - f''(x^k)x^k. \quad (3.30)$$

Ersetzt man $\nabla f(x)$ in der Bedingung (3.28) mit seiner Approximation, so lässt sie sich schreiben als

$$\nabla f(x^k) + f''(x^k)x - f''(x^k)x^k + A^T \lambda + G^T \mu = 0 \quad (3.31)$$

$$\Leftrightarrow f''(x^k)x + A^T \lambda + G^T \mu = f''(x^k)x^k - \nabla f(x^k). \quad (3.32)$$

Dieses lineare Gleichungssystem wird in der Aktive-Mengen-Strategie verwendet.

Verfahren 3.3 (Aktive-Mengen-Strategie für (PLU))

1. Wähle x^0 und setze $\lambda^0 := 0$, $\mu^0 := 0$, $k := 0$.

2. Falls

$$\nabla f(x^k) + A^T \lambda^k + G^T \mu^k = 0$$

$$Ax^k = b$$

$$\min\{\mu^k, r - Gx^k\} = 0$$

\Rightarrow STOP.

3. Bestimme

$$\mathcal{A}_k := \{i \in \{1, \dots, p\} \mid r_i - \langle g_i, x^k \rangle < \mu_i^k\}, \quad (3.33)$$

$$\mathcal{I}_k := \{i \in \{1, \dots, p\} \mid r_i - \langle g_i, x^k \rangle \geq \mu_i^k\}. \quad (3.34)$$

4. Löse das Problem

$$f''(x^k)x + A^T \lambda + G^T \mu = f''(x^k)x^k - \nabla f(x^k) \quad (3.35)$$

$$Ax = b \quad (3.36)$$

$$\langle g_i, x \rangle = r_i \quad \text{für } i \in \mathcal{A}_k \quad (3.37)$$

$$\mu_i = 0 \quad \text{für } i \in \mathcal{I}_k. \quad (3.38)$$

\Rightarrow Erhalte x , λ und μ .

5. Setze $x^{k+1} := x$, $\mu^{k+1} := \mu$ und $k := k + 1 \Rightarrow$ Gehe zu Schritt 2.

Das Verfahren 3.3 kann als halbglattes Newton-Verfahren (Verfahren 3.2) interpretiert werden. Sei k irgendein Iterationsschritt in beiden Verfahren. Bei dem halbglatten Newton-Verfahren ist dann in Schritt 3 folgendes Gleichungssystem zu lösen:

$$\begin{pmatrix} f''(x) & A^T & G^T \\ A & 0 & 0 \\ -\chi_{\mathcal{A}}(1) g_1^T & 0 & \chi_{\mathcal{I}}(1) e_1^T \\ \vdots & \vdots & \vdots \\ -\chi_{\mathcal{A}}(p) g_p^T & 0 & \chi_{\mathcal{I}}(p) e_p^T \end{pmatrix} \begin{pmatrix} d_x \\ d_\lambda \\ d_\mu \end{pmatrix} = - \begin{pmatrix} \nabla f(x^k) + A^T \lambda^k + G^T \mu^k \\ Ax^k - b \\ \chi_{\mathcal{I}}(1) \mu_1^k + \chi_{\mathcal{A}}(1) (r_1 - \langle g_1, x^k \rangle) \\ \vdots \\ \chi_{\mathcal{I}}(p) \mu_p^k + \chi_{\mathcal{A}}(p) (r_p - \langle g_p, x^k \rangle) \end{pmatrix},$$

d. h.

$$f''(x^k) d_x + A^T d_\lambda + G^T d_\mu = -\nabla f(x^k) - A \lambda^k - G^T \mu^k \quad (3.39)$$

$$A d_x = -A x^k + b \quad (3.40)$$

$$-\chi_{\mathcal{A}}(i) g_i^T d_x + \chi_{\mathcal{I}}(i) e_i^T d_\mu = -\chi_{\mathcal{I}}(i) \mu_i^k - \chi_{\mathcal{A}}(i) (r_i - \langle g_i, x^k \rangle) \quad i = 1, \dots, p \quad (3.41)$$

Die Indexmengen \mathcal{A} und \mathcal{I} sind dabei genau wie die in der Aktive-Mengen-Strategie.

Bei dem halbglatten Newton-Verfahren werden in Schritt 4 die Variablen x^{k+1} , λ^{k+1} und μ^{k+1} gesetzt. Es gilt hiernach

$$d_x = x^{k+1} - x^k, \quad d_\lambda = \lambda^{k+1} - \lambda^k, \quad \text{und} \quad d_\mu = \mu^{k+1} - \mu^k. \quad (3.42)$$

Dann lautet Gleichung (3.39)

$$\begin{aligned} f''(x^k)(x^{k+1} - x^k) + A^T(\lambda^{k+1} - \lambda^k) + G^T(\mu^{k+1} - \mu^k) &= -\nabla f(x^k) - A^T \lambda^k - G^T \mu^k \\ f''(x^k)x^{k+1} + A^T \lambda^{k+1} + G^T \mu^{k+1} &= f''(x^k)x^k - \nabla f(x^k). \end{aligned} \quad (3.43)$$

Das ist genau die Gleichung (3.35) in der Aktive-Mengen-Strategie mit x^{k+1} , λ^{k+1} , μ^{k+1} anstatt x , λ , μ .

Für die Gleichung (3.40) gilt

$$A(x^{k+1} - x^k) = -A x^k + b \quad (3.44)$$

$$A x^{k+1} = b. \quad (3.45)$$

Sie entspricht also der Gleichung (3.36) in Verfahren 3.3.

Sei $i \in \mathcal{A}_k$. Die Gleichung (3.41) lässt sich somit schreiben als

$$-g_i^T d_x = -(r_i - \langle g_i, x^k \rangle) \quad (3.46)$$

$$-g_i^T (x^{k+1} - x^k) = -r_i + \langle g_i, x^k \rangle \quad (3.47)$$

$$r_i = \langle g_i, x^{k+1} \rangle. \quad (3.48)$$

Sie findet man auch in der Gleichung (3.37) in der Aktive-Mengen-Strategie.

Sei nun $i \in \mathcal{I}_k$. Dann sieht die Gleichung (3.41) wie folgt aus.

$$e_i^T d_\mu = -\mu_i^k \quad (3.49)$$

$$e_i^T (\mu^{k+1} - \mu^k) = -\mu_i^k \quad (3.50)$$

$$\mu_i^{k+1} - \mu_i^k = -\mu_i^k \quad (3.51)$$

$$\mu_i^{k+1} = 0. \quad (3.52)$$

Das ist die Gleichung (3.38) in der Aktive-Mengen-Strategie mit μ_i^{k+1} anstatt μ_i .

D. h., die beiden Verfahren berechnen eigentlich denselben nächsten Iterationspunkt x^{k+1} . Deswegen kann die Aktive-Mengen-Strategie als das halbglatte Newton-Verfahren interpretiert werden.

Kapitel 4

Vergleich

In diesem Kapitel werden numerische Ergebnisse der beiden Verfahren für verschiedene Testprobleme vorgestellt. Der Test wird auf Optimierungsprobleme mit linearen Restriktionen beschränkt, d. h., Optimierungsprobleme mit nichtlinearen Restriktionen werden wegen des hohen Aufwands nicht berücksichtigt.

4.1 Testprobleme

Die Testprobleme können in drei Gruppen eingeteilt werden. Die erste Gruppe besteht aus Problemen (PVB) mit nur oberen und unteren Schranken. Zur zweiten Gruppe gehören die Probleme (PLG) mit linearen Gleichungsnebenbedingungen. Die Probleme (PLU) mit linearen Ungleichungsnebenbedingungen kommen anschließend in der letzten Gruppe. In jeder Beschreibung der Testprobleme werden der zu verwendende Startpunkt x^0 und die Lösung x^* auch angegeben.

Testproblem 1

$$\begin{aligned} & \min_{x \in \mathbb{R}^2} (x_1 - 4)^2 + (x_2 - 7)^2 \\ & \text{bei} \quad -10 \leq x_i \leq 10, \quad i = 1, 2 \end{aligned} \tag{4.1}$$
$$x^0 = (8, 9)^T \quad \text{und} \quad x^* = (4, 7)^T.$$

Testproblem 2 (Rosenbrock-Funktion, vgl. Beispiel 1.4.1 in [Alt11, S. 14])

$$\begin{aligned} & \min_{x \in \mathbb{R}^2} 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \\ & \text{bei} \quad -10 \leq x_i \leq 10, \quad i = 1, 2 \end{aligned} \tag{4.2}$$
$$x^0 = (-1, 2)^T \quad \text{und} \quad x^* = (1, 1)^T.$$

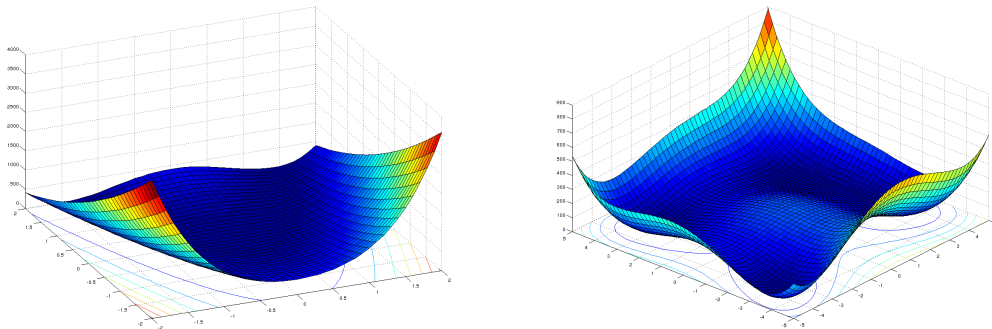


Abbildung 4.1: Rosenbrock- und Himmelblau-Funktion

Testproblem 3 (Himmelblau-Funktion, vgl. Beispiel 1.4.2 in [Alt11, S. 14 f.])

$$\begin{aligned} \min_{x \in \mathbb{R}^2} & (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2 \\ \text{bei} & -10 \leq x_i \leq 10, \quad i = 1, 2 \\ x^0 &= (5, 5)^T \quad \text{und} \quad x^* = (3, 2)^T. \end{aligned} \tag{4.3}$$

Testproblem 4 (Bazaraa-Shetty-Funktion, vgl. Beispiel 1.4.3 in [Alt11, S. 15 f.])

$$\begin{aligned} \min_{x \in \mathbb{R}^2} & (x_1 - 2)^4 + (x_1 - 2x_2)^2 \\ \text{bei} & -10 \leq x_i \leq 10, \quad i = 1, 2 \\ x^0 &= (5, 5)^T \quad \text{und} \quad x^* = (2, 1)^T. \end{aligned} \tag{4.4}$$

Testproblem 5 (Beale-Funktion, vgl. Aufgabe 2.2 in [Alt11, S. 39])

$$\begin{aligned} \min_{x \in \mathbb{R}^2} & (1.5 - x_1(1 - x_2))^2 + (2.25 - x_1(1 - x_2^2))^2 + (2.625 - x_1(1 - x_2^3))^2 \\ \text{bei} & -10 \leq x_i \leq 10, \quad i = 1, 2 \\ x^0 &= (5, 0)^T \quad \text{und} \quad x^* = (3, 0.5)^T. \end{aligned} \tag{4.5}$$

Testproblem 6 (Exponent-Funktion)

$$\begin{aligned} \min_{x \in \mathbb{R}^2} & e^{\|x\|^2} \\ \text{bei} & -10 \leq x_i \leq 10, \quad i = 1, 2 \\ x^0 &= (5, 4)^T \quad \text{und} \quad x^* = (0, 0)^T. \end{aligned} \tag{4.6}$$

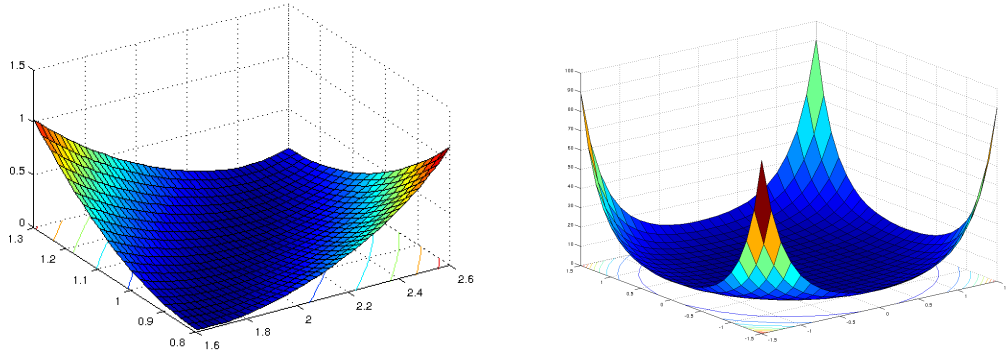


Abbildung 4.2: Bazaraa-Shetty- und Exponent-Funktion

Testproblem 7 (Vgl. Problem 38 in [HS81, S. 61])

$$\begin{aligned}
 \min_{x \in \mathbb{R}^4} & 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 90(x_4 - x_3^2)^2 + (1 - x_3)^2 \\
 & + 10.1((x_2 - 1)^2 + (x_4 - 1)^2) + 19.8(x_2 - 1)(x_4 - 1) \\
 \text{bei} & -10 \leq x_i \leq 10, \quad i = 1, \dots, 4 \\
 x^0 &= (3, -1, -3, -1)^T \quad \text{und} \quad x^* = (1, 1, 1, 1)^T.
 \end{aligned} \tag{4.7}$$

Testproblem 8 (Dixon-Funktion, vgl. Beispiel 1.4.5 in [Alt11, S. 16])

$$\begin{aligned}
 \min_{x \in \mathbb{R}^{10}} & (1 - x_1)^2 + \sum_{k=1}^9 (x_k^2 - x_{k+1})^2 + (1 - x_{10})^2 \\
 \text{bei} & -10 \leq x_i \leq 10, \quad i = 1, \dots, 10 \\
 x^0 &= (10, \dots, 10)^T \quad \text{und} \quad x^* = (1, \dots, 1)^T.
 \end{aligned} \tag{4.8}$$

Testproblem 9

$$\begin{aligned}
 \min_{x \in \mathbb{R}^3} & (x_1 - 4)^2 + (x_2 - 2)^2 + (x_3 - 7)^2 \\
 \text{bei} & 5 \leq x_i \leq 10, \quad i = 1, 2, 3 \\
 x^0 &= (7, 10, 9)^T \quad \text{und} \quad x^* = (5, 5, 7)^T.
 \end{aligned} \tag{4.9}$$

Testproblem 10 (Rosenbrock-Funktion)

$$\begin{aligned}
 \min_{x \in \mathbb{R}^2} & 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \\
 \text{bei} & -10 \leq x_1 \leq 10 \\
 & 1.5 \leq x_2 \leq 10
 \end{aligned} \tag{4.10}$$

$$x^0 = (2, 3)^T \quad \text{und} \quad x^* = (1.224, 1.5)^T.$$

Testproblem 11 (*Himmelblau-Funktion*)

$$\begin{aligned} \min_{x \in \mathbb{R}^2} & (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2 \\ \text{bei} & \quad -10 \leq x_1 \leq 10 \\ & \quad 5 \leq x_2 \leq 10 \end{aligned} \tag{4.11}$$

$$x^0 = (10, 10)^T \quad \text{und} \quad x^* = (-2.928, 5)^T.$$

Testproblem 12 (*Bazaraa-Shetty-Funktion*)

$$\begin{aligned} \min_{x \in \mathbb{R}^2} & (x_1 - 2)^4 + (x_1 - 2x_2)^2 \\ \text{bei} & \quad 4 \leq x_1 \leq 10 \\ & \quad -10 \leq x_2 \leq 10 \end{aligned} \tag{4.12}$$

$$x^0 = (7, 5)^T \quad \text{und} \quad x^* = (4, 2)^T.$$

Testproblem 13 (*Beale-Funktion*)

$$\begin{aligned} \min_{x \in \mathbb{R}^2} & (1.5 - x_1(1 - x_2))^2 + (2.25 - x_1(1 - x_2^2))^2 + (2.625 - x_1(1 - x_2^3))^2 \\ \text{bei} & \quad 3 \leq x_1 \leq 10 \\ & \quad -10 \leq x_2 \leq 10 \end{aligned} \tag{4.13}$$

$$x^0 = (5, 0)^T \quad \text{und} \quad x^* = (3, 0.5)^T.$$

Testproblem 14 (*Exponent-Funktion*)

$$\begin{aligned} \min_{x \in \mathbb{R}^3} & e^{\|x\|^2} \\ \text{bei} & \quad 0.5 \leq x_1 \leq 10 \\ & \quad 1 \leq x_2 \leq 10 \\ & \quad 1 \leq x_3 \leq 10 \end{aligned} \tag{4.14}$$

$$x^0 = (5, 2, 4)^T \quad \text{und} \quad x^* = (0.5, 1, 1)^T.$$

Testproblem 15 (Vgl. Problem 38 in [HS81, S. 61])

$$\begin{aligned} \min_{x \in \mathbb{R}^4} & 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 90(x_4 - x_3^2)^2 + (1 - x_3)^2 \\ & + 10.1((x_2 - 1)^2 + (x_4 - 1)^2) + 19.8(x_2 - 1)(x_4 - 1) \\ \text{bei} & \quad 2 \leq x_1 \leq 10 \\ & \quad -10 \leq x_i \leq 10, \quad i = 2, 3, 4 \end{aligned} \quad (4.15)$$

$$x^0 = (3, -1, -3, -1)^T \quad \text{und} \quad x^* = (2, 3.831, 0.03, -0.178)^T.$$

Testproblem 16 (Dixon-Funktion)

$$\begin{aligned} \min_{x \in \mathbb{R}^{10}} & (1 - x_1)^2 + \sum_{k=1}^9 (x_k^2 - x_{k+1})^2 + (1 - x_{10})^2 \\ \text{bei} & \quad 2 \leq x_i \leq 10, \quad i = 1, \dots, 10 \end{aligned} \quad (4.16)$$

$$x^0 = (10, \dots, 10)^T \quad \text{und} \quad x^* = (2, \dots, 2, 2.5)^T.$$

Testproblem 17 Lineare Regression (Vgl. Beispiel 1.3 auf Seite 6)

$$\begin{aligned} \min_{x \in \mathbb{R}^2} & \sum_{i=1}^{10} (x_1 \xi_i + x_2 - \eta_i)^2 \\ \text{bei} & \quad -10 \leq x_i \leq 10, \quad i = 1, 2 \\ & \quad x^0 = (-10, 10)^T. \end{aligned} \quad (4.17)$$

Die Messwerte $(\xi_i, \eta_i), i = 1, \dots, 10$, sind in Tabelle 4.1 zu finden.

$$x^* = (0.7, -4.3)^T \Rightarrow \tilde{\eta}(\xi) = 0.7 \xi - 4.3.$$

ξ_i	1	2	3	4	5	6	7	8	9	10
η_i	-0.5	-2	-3	-3	-2.5	-2	-1	1	3	5.5

Tabelle 4.1: Messwerte für Testprobleme 17 und 18

Testproblem 18 Nichtlineare Regression mit dem funktionalen Zusammenhang (1.9) (Vgl. Beispiel 1.4 auf Seite 6)

$$\begin{aligned} \min_{x \in \mathbb{R}^3} \quad & \sum_{i=1}^{10} (x_1(\xi_i - x_2)^2 + x_3 - \eta_i)^2 \\ \text{bei} \quad & -10 \leq x_i \leq 20, \quad i = 1, 2, 3 \\ & x^0 = (1, 3, -4)^T. \end{aligned} \quad (4.18)$$

Die Messwerte in Tabelle 4.1 sind zu benutzen.

$$x^* = (0.242, 4.056, -2.955)^T \Rightarrow \hat{\eta}(\xi) = 0.242 (\xi - 4.056)^2 - 2.955.$$

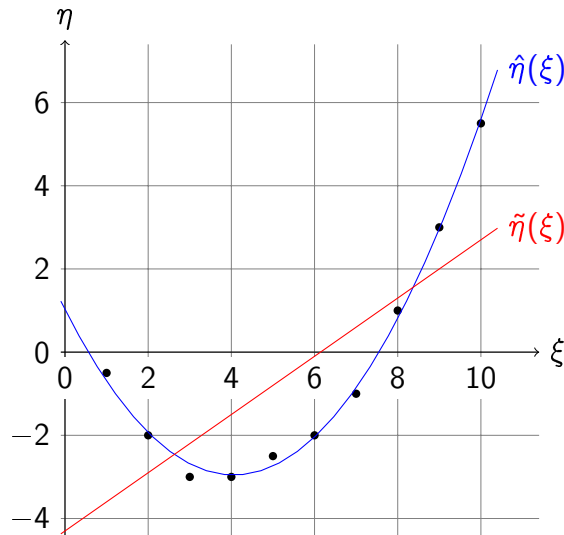


Abbildung 4.3: Ergebnisse der Testprobleme 17 und 18

Testproblem 19 Nichtlineare Regression mit dem funktionalen Zusammenhang (1.10) (Vgl. Beispiel 1.4 auf Seite 6)

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \quad & \sum_{i=1}^{10} (x_1 e^{\xi x_2} - \eta_i)^2 \\ \text{bei} \quad & -10 \leq x_i \leq 10, \quad i = 1, 2 \\ & x^0 = (0.2, 0.5)^T. \end{aligned} \quad (4.19)$$

Die Messwerte $(\xi_i, \eta_i), i = 1, \dots, 10$, sind in Tabelle 4.2 gegeben.

$$x^* = (0.632, 0.195)^T \Rightarrow \bar{\eta}(\xi) = 0.632 e^{0.195 \xi}.$$

ξ_i	1	2	3	4	5	6	7	8	9	10
η_i	1	1.1	1.2	1.35	1.55	1.75	2.5	3	3.7	4.5

Tabelle 4.2: Messwerte für Testproblem 19

Testproblem 20 (Vgl. Beispiel 16.2 in [NW06, S. 452])

$$\begin{aligned}
 \min_{x \in \mathbb{R}^3} \quad & 3x_1^2 + 2x_1x_2 + x_1x_3 + 2.5x_2^2 + 2x_2x_3 + 2x_3^2 - 8x_1 - 3x_2 - 3x_3 \\
 \text{bei} \quad & x_1 + x_3 = 3 \\
 & x_2 + x_3 = 0
 \end{aligned} \tag{4.20}$$

$$x^0 = (5, 2, -2)^T \quad \text{und} \quad x^* = (2, -1, 1)^T.$$

Testproblem 21 (Vgl. Beispiel 1.5 auf Seite 10)

$$\begin{aligned}
 \min_{x \in \mathbb{R}^5} \quad & \|x\|^2 \\
 \text{bei} \quad & x_1 + x_2 + x_3 + x_4 + x_5 = 1
 \end{aligned} \tag{4.21}$$

$$x^0 = (5, -1, -1, -1, -1)^T \quad \text{und} \quad x^* = (0.2, \dots, 0.2)^T.$$

Testproblem 22 (Vgl. Problem 28 in [HS81, S. 51])

$$\begin{aligned}
 \min_{x \in \mathbb{R}^3} \quad & (x_1 + x_2)^2 + (x_2 + x_3)^2 \\
 \text{bei} \quad & x_1 + 2x_2 + 3x_3 = 1
 \end{aligned} \tag{4.22}$$

$$x^0 = (-4, 1, 1)^T \quad \text{und} \quad x^* = (0.5, -0.5, 0.5)^T.$$

Testproblem 23 (Vgl. Problem 48 in [HS81, S. 71])

$$\begin{aligned}
 \min_{x \in \mathbb{R}^5} \quad & (x_1 - 1)^2 + (x_2 - x_3)^2 + (x_4 - x_5)^2 \\
 \text{bei} \quad & x_1 + x_2 + x_3 + x_4 + x_5 = 5 \\
 & x_3 - 2x_4 - 2x_5 = -3
 \end{aligned} \tag{4.23}$$

$$x^0 = (3, 5, -3, 2, -2)^T \quad \text{und} \quad x^* = (1, 1, 1, 1, 1)^T.$$

Testproblem 24 (Vgl. Problem 51 in [HS81, S. 74])

$$\begin{aligned} \min_{x \in \mathbb{R}^5} & (x_1 - x_2)^2 + (x_2 + x_3 - 2)^2 + (x_4 - 1)^2 + (x_5 - 1)^2 \\ \text{bei} & \quad x_1 + 3x_2 = 4 \\ & \quad x_3 + x_4 - 2x_5 = 0 \\ & \quad x_2 - x_5 = 0 \end{aligned} \quad (4.24)$$

$$x^0 = (2.5, 0.5, 2, -1, 0.5)^T \quad \text{und} \quad x^* = (1, 1, 1, 1, 1)^T.$$

Testproblem 25 (Vgl. Problem 52 in [HS81, S. 75])

$$\begin{aligned} \min_{x \in \mathbb{R}^5} & (4x_1 - x_2)^2 + (x_2 + x_3 - 2)^2 + (x_4 - 1)^2 + (x_5 - 1)^2 \\ \text{bei} & \quad x_1 + 3x_2 = 0 \\ & \quad x_3 + x_4 - 2x_5 = 0 \\ & \quad x_2 - x_5 = 0 \end{aligned} \quad (4.25)$$

$$x^0 = (-1.5, 0.5, 2, -1, 0.5)^T \quad \text{und} \quad x^* = (-0.095, 0.032, 0.516, -0.453, 0.032)^T.$$

Testproblem 26 (Vgl. Problem 53 in [HS81, S. 76])

$$\begin{aligned} \min_{x \in \mathbb{R}^5} & (x_1 - x_2)^2 + (x_2 + x_3 - 2)^2 + (x_4 - 1)^2 + (x_5 - 1)^2 \\ \text{bei} & \quad x_1 + 3x_2 = 0 \\ & \quad x_3 + x_4 - 2x_5 = 0 \\ & \quad x_2 - x_5 = 0 \\ & \quad -10 \leq x_i \leq 10, \quad i = 1, \dots, 5 \end{aligned} \quad (4.26)$$

$$x^0 = (-1.5, 0.5, 2, -1, 0.5)^T \quad \text{und} \quad x^* = (-0.767, 0.256, 0.628, -0.116, 0.256)^T.$$

Testproblem 27 (Vgl. Problem 49 in [HS81, S. 72])

$$\begin{aligned} \min_{x \in \mathbb{R}^5} & (x_1 - x_2)^2 + (x_3 - 1)^2 + (x_4 - 1)^2 + (x_5 - 1)^6 \\ \text{bei} & \quad x_1 + x_2 + x_3 + 4x_4 = 7 \\ & \quad x_3 + 5x_5 = 6 \end{aligned} \quad (4.27)$$

$$x^0 = (10, 7, 2, -3, 0.8)^T \quad \text{und} \quad x^* = (1, 1, 1, 1, 1)^T.$$

Testproblem 28 (Vgl. Problem 50 in [HS81, S. 73])

$$\begin{aligned}
 & \min_{x \in \mathbb{R}^5} (x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_4)^4 + (x_4 - x_5)^2 \\
 & \text{bei } x_1 + 2x_2 + 3x_3 = 6 \\
 & \quad x_2 + 2x_3 + 3x_4 = 6 \\
 & \quad x_3 + 2x_4 + 3x_5 = 6 \\
 & x^0 = (35, -31, 11, 5, -5)^T \quad \text{und} \quad x^* = (1, 1, 1, 1, 1)^T.
 \end{aligned} \tag{4.28}$$

Testproblem 29

$$\begin{aligned}
 & \min_{x \in \mathbb{R}^2} x_1^2 + x_2^2 - 4x_1 - 4x_2 \\
 & \text{bei } 2x_1 + x_2 \leq 2 \\
 & \quad x_1 - x_2 \leq 1 \\
 & \quad -x_1 - x_2 \leq 1 \\
 & \quad -2x_1 + x_2 \leq 2 \\
 & x^0 = (-1, 0)^T \quad \text{und} \quad x^* = (0.4, 1.2)^T.
 \end{aligned} \tag{4.29}$$

Testproblem 30 (Vgl. Beispiel 1.6 auf Seite 12)

$$\begin{aligned}
 & \min_{x \in \mathbb{R}^2} (x_1 - 1)^2 + (x_2 - 2.5)^2 \\
 & \text{bei } -x_1 + 2x_2 \leq 2 \\
 & \quad x_1 + 2x_2 \leq 6 \\
 & \quad x_1 - 2x_2 \leq 2 \\
 & \quad 0 \leq x_i, \quad i = 1, 2 \\
 & x^0 = (2, 0)^T \quad \text{und} \quad x^* = (1.4, 1.7)^T.
 \end{aligned} \tag{4.30}$$

Testproblem 31 (Vgl. Beispiel 1.7 auf Seite 12)

$$\begin{aligned}
 & \min_{x \in \mathbb{R}^4} (x_1 - x_3)^2 + (x_2 - x_4)^2 \\
 & \text{bei } -x_1 \leq 0 \\
 & \quad -x_2 \leq 0 \\
 & \quad x_1 + 2x_2 \leq 2 \\
 & \quad -x_4 \leq -2 \\
 & \quad -x_3 - x_4 \leq -3 \\
 & \quad x_3 + 2x_4 \leq 6 \\
 & x^0 = (0, 1, 2, 2)^T \quad \text{und} \quad x^* = (0.4, 0.8, 1, 2)^T.
 \end{aligned} \tag{4.31}$$

Testproblem 32 (Vgl. Problem 21 in [HS81, S. 44])

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \quad & 0.01x_1^2 + x_2^2 - 100 \\ \text{bei} \quad & -10x_1 + x_2 \leq -10 \\ & 2 \leq x_1 \leq 50 \\ & -50 \leq x_2 \leq 50 \end{aligned} \tag{4.32}$$

$$x^0 = (10, -10)^T \quad \text{und} \quad x^* = (2, 0)^T.$$

Testproblem 33 (Vgl. Problem 35 in [HS81, S. 58])

$$\begin{aligned} \min_{x \in \mathbb{R}^3} \quad & 2x_1^2 + 2x_1x_2 + 2x_1x_3 + 2x_2^2 + x_3^2 - 8x_1 - 6x_2 - 4x_3 + 9 \\ \text{bei} \quad & x_1 + x_2 + 2x_3 \leq 3 \\ & 0 \leq x_i, \quad i = 1, 2, 3 \end{aligned} \tag{4.33}$$

$$x^0 = (1.5, 0.5, 0.5)^T \quad \text{und} \quad x^* = (1.333, 0.778, 0.444)^T.$$

Testproblem 34 (Vgl. Problem 76 in [HS81, S. 96])

$$\begin{aligned} \min_{x \in \mathbb{R}^4} \quad & x_1^2 + 0.5x_2^2 + x_3^2 + 0.5x_4^2 - x_1x_3 + x_3x_4 - x_1 - 3x_2 + x_3 - x_4 \\ \text{bei} \quad & x_1 + 2x_2 + x_3 + x_4 \leq 5 \\ & 3x_1 + x_2 + 2x_3 - x_4 \leq 4 \\ & -x_2 - 4x_3 \leq -1.5 \\ & 0 \leq x_i, \quad i = 1, \dots, 4 \end{aligned} \tag{4.34}$$

$$x^0 = (0.25, 1.5, 0.25, 0.5)^T \quad \text{und} \quad x^* = (0.273, 2.091, 0, 0.545)^T.$$

Testproblem 35 (Steuerungsproblem, vgl. Beispiel 6.2.6 in [Alt11, S. 216 f.])

Es ist ein Produktionsplanungsproblem mit beschränktem Lager zu betrachten. In einem Zeitraum $[0, T]$ mit $T > 0$ soll eine Produktion so gesteuert werden, dass die Gesamtkosten minimal sind. Seien

- $z(t)$ die Anzahl der zur Zeit t im Lager vorhandenen Produkte,
- $r(t)$ die Nachfrage nach den Produkten zur Zeit t und
- $u(t)$ die Produktionsrate zur Zeit t .

Die Änderung der verfügbaren Lagermenge $z(t)$ sei durch die Lagerbilanzgleichung

$$\dot{z}(t) = u(t) - r(t) \quad \forall t \in [0, T] \quad (4.35)$$

mit dem Anfangslagerbestand $z(0) = a \geq 0$ gegeben. Die Produktions- und Lagerhaltungskosten zur Zeit t betragen $\frac{1}{2}\rho_u u(t)^2 + \rho_z z(t)$ mit Konstanten $\rho_u, \rho_z > 0$. Durch den Verkauf des Lagerrestbestandes $z(T)$ kommt ein Gewinn $\sigma z(T)$ mit $\sigma \geq 0$. Dabei sind $u(t)$ und $z(t)$ zu bestimmen, wobei die Lagerkapazität beschränkt ist, d. h. $z(t) \leq \zeta$ für alle $t \in [0, T]$ mit $\zeta > 0$. Damit ergibt sich das folgende zeitkontinuierliche Problem:

$$\begin{aligned} \min f(z, u) &:= \int_0^T (\frac{1}{2}\rho_u u(t)^2 + \rho_z z(t)) dt - \sigma z(T) \\ \text{bei} \quad \dot{z}(t) &= u(t) - r(t) \quad \forall t \in [0, T] \\ z(0) &= a \\ 0 \leq z(t) &\leq \zeta \quad \forall t \in [0, T] \\ 0 \leq u(t) &\quad \forall t \in [0, T]. \end{aligned} \quad (4.36)$$

Eine Diskretisierung soll durchgeführt werden, damit das Problem numerisch gelöst werden kann. Mit $N \in \mathbb{N}$ und $\tau := \frac{T}{N}$ wird das Intervall $[0, T]$ in N Teilintervalle $[t_0, t_1], \dots, [t_{N-1}, t_N]$ unterteilt, wobei $t_i = i\tau$ für alle $i = 0, \dots, N$. Die Produktionsrate $u(t)$ wird durch eine stückweise konstante Funktion $\hat{u}(t)$ mit

$$\hat{u}(t) := u_i \quad \forall t \in [t_i, t_{i+1}[, i = 0, \dots, N-1 \quad (4.37)$$

approximiert und die Lagermenge $z(t)$ durch eine lineare Interpolationsfunktion $\hat{z}(t)$ mit den Stützstellen $(t_0, a), (t_1, z_1), \dots, (t_N, z_N)$, d. h.

$$\hat{z}(t) := z_i + \frac{1}{\tau}(z_{i+1} - z_i)(t - t_i) \quad \forall t \in [t_i, t_{i+1}[, i = 0, \dots, N-1 \quad (4.38)$$

mit $z_0 := a$. Die diskrete Version des Problems (4.36) lässt sich dann wie folgt formulieren.

$$\begin{aligned} \min \quad & \frac{1}{2}\tau\rho_u \sum_{i=0}^{N-1} u_i^2 + \tau\rho_z \sum_{i=1}^N z_i - \sigma z_N \\ \text{bei} \quad & z_{i+1} - z_i - \tau(u_i - r_i) = 0 \quad \forall i = 0, \dots, N-1 \\ & 0 \leq z_i \leq \zeta \quad \forall i = 1, \dots, N \\ & 0 \leq u_i \quad \forall i = 0, \dots, N-1, \end{aligned} \quad (4.39)$$

wobei $r_i := r(t_i)$ für $i = 0, \dots, N-1$. Das Problem (4.39) ist ein Problem vom Typ (QLU) und $x := (u_0, z_1, u_1, z_2, \dots, u_{N-1}, z_N)^T$ sei als die Reihenfolge der Variablen gewählt. Konkrete Daten für den Test seien

$$N = 100, T = 12, \rho_u = 0.2, \rho_z = 0.1, \quad (4.40)$$

$$\sigma = 1.5, a = 5, \zeta = 6, r(t) = 10 + 5 \sin t. \quad (4.41)$$

Der Startvektor $x^0 \in \mathbb{R}^{2N}$ sei durch

$$x_{2i+1}^0 := r_i, \quad x_{2i+2}^0 := a \quad \forall i = 0, \dots, N-1 \quad (4.42)$$

definiert. Die Abbildung 4.4 zeigt das Ergebnis der berechneten Lösung.

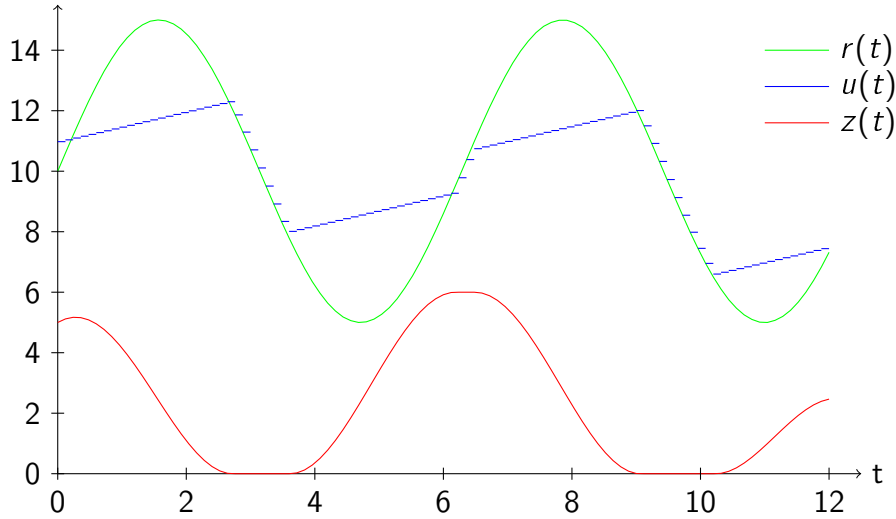


Abbildung 4.4: Ergebnis des Testproblems 35

4.2 Numerische Ergebnisse

Das halbglatte Newton-Verfahren und das SQP-Verfahren wurden in der Programmiersprache MATLAB implementiert. Bei allen im Abschnitt 4.1 vorgestellten Testproblemen waren die implementierten Programme in der Lage, die angegebene Lösung zu finden. Die entsprechenden Programme bezeichnen wir hier jeweils als SSN und SQP.

Die folgenden zwei Tabellen zeigen die Ergebnisse des Tests. Die beiden Programme wurden 100 Mal auf jedes Testproblem angewendet und der Durchschnitt der Laufzeiten ist als die ermittelte Laufzeit zu verstehen. Die Bezeichnung T steht für die Laufzeit des Programms zur Lösung des Problems in Millisekunden und $\#It$ bezeichnet die Anzahl der benötigten Iterationen. Als Abbruchschranke für die beiden Programme wurde $\varepsilon = 10^{-5}$ gewählt. Die Abbildung 4.5 stellt die Laufzeiten der beiden Programme graphisch dar.

Wie aus den Tabellen zu entnehmen ist, hat SSN eine bessere Leistung bezüglich der Laufzeit bei den Testproblemen mit weniger als 10 Variablen geliefert. Beim Testproblem 35 mit 200 Variablen war SQP schneller.

<i>Nr.</i>	$T_{SSN} (ms)$	$T_{SQP} (ms)$	$\#It_{SSN}$	$\#It_{SQP}$
1	1.38	1.94	2	1
2	3.87	6.94	6	5
3	4.80	8.88	7	6
4	9.12	36.33	15	29
5	19.82	24.85	9	8
6	35.54	64.52	48	47
7	171.86	195.18	27	26
8	148.26	210.62	36	31
9	2.56	6.17	3	1
10	5.78	13.16	9	6
11	6.22	21.10	9	10
12	3.98	5.30	6	3
13	22.27	21.36	10	5
14	50.24	91.33	49	46
15	70.10	85.54	11	10
16	41.64	97.56	10	8
17	15.24	18.42	3	2
18	141.55	151.97	14	13
19	75.10	87.43	21	20

Tabelle 4.3: Ergebnisse der Programme für Testprobleme 1 bis 19

Bei den meisten Testproblemen hat SQP eine geringere Anzahl von Iterationen als SSN benötigt. Theoretisch ist das auch zu erwarten, weil nach Satz 2.1 eine quadratische Konvergenz bei SQP-Verfahren im besten Fall vorkommen kann, bei halbglattem Newton-Verfahren ist es dagegen nach Satz 3.1 nur superlineare Konvergenz. Besonders bei Testproblemen mit quadratischer Zielfunktion hat SQP nur eine Iteration beansprucht.

Nur bei zwei Fällen hat SQP mehr Iterationen gebraucht, damit die Abbruchbedingung erfüllt wurde. Wenn man aber die Iterationen genauer betrachtet, hätte man das Programm eigentlich schon bei weniger Iterationen abbrechen können, weil die berechnete Lösung da schon erreicht wurde.

Bei den meisten Testproblemen hat SSN genauere Optimalwerte als SQP erzielt. In Tabellen 4.5, 4.6 und 4.7 werden die Funktionsauswertungen der Iterationen bei drei ausgewählten Testproblemen gezeigt.

Nr.	$T_{SSN} (ms)$	$T_{SQP} (ms)$	$\#It_{SSN}$	$\#It_{SQP}$
20	1.05	2.27	2	1
21	1.08	2.31	2	1
22	1.03	2.22	2	1
23	1.07	2.34	2	1
24	1.05	2.30	2	1
25	1.10	2.35	2	1
26	2.21	3.81	2	1
27	20.07	22.30	3	2
28	65.47	73.87	9	8
29	2.12	5.72	3	1
30	2.16	5.53	3	1
31	2.78	9.62	3	1
32	2.92	3.77	4	1
33	2.01	4.36	3	1
34	3.97	7.59	5	1
35	798.02	509.44	14	1

Tabelle 4.4: Ergebnisse der Programme für Testprobleme 20 bis 35

k	$f(x_{SSN}^k)$	$f(x_{SSN}^k)$
0	$2.545312500000000 \cdot 10^1$	$2.545312500000000 \cdot 10^1$
1	$5.919504078098388 \cdot 10^{-1}$	$4.965197557963974 \cdot 10^0$
2	$1.054707528786025 \cdot 10^0$	$1.579335910652743 \cdot 10^{-3}$
3	$8.858473238377809 \cdot 10^{-2}$	$1.529508105102504 \cdot 10^{-5}$
4	$5.393135687740480 \cdot 10^{-2}$	$4.533268471207106 \cdot 10^{-9}$
5	$3.482443869623315 \cdot 10^{-1}$	$2.955441591489031 \cdot 10^{-16}$
6	$1.111389669918921 \cdot 10^{-2}$	
7	$3.903668297871150 \cdot 10^{-5}$	
8	$4.550548206926666 \cdot 10^{-9}$	
9	$1.147312883953148 \cdot 10^{-15}$	
10	$1.529141090397863 \cdot 10^{-18}$	

Tabelle 4.5: Funktionsauwertungen der Iterationen bei Testproblem 13

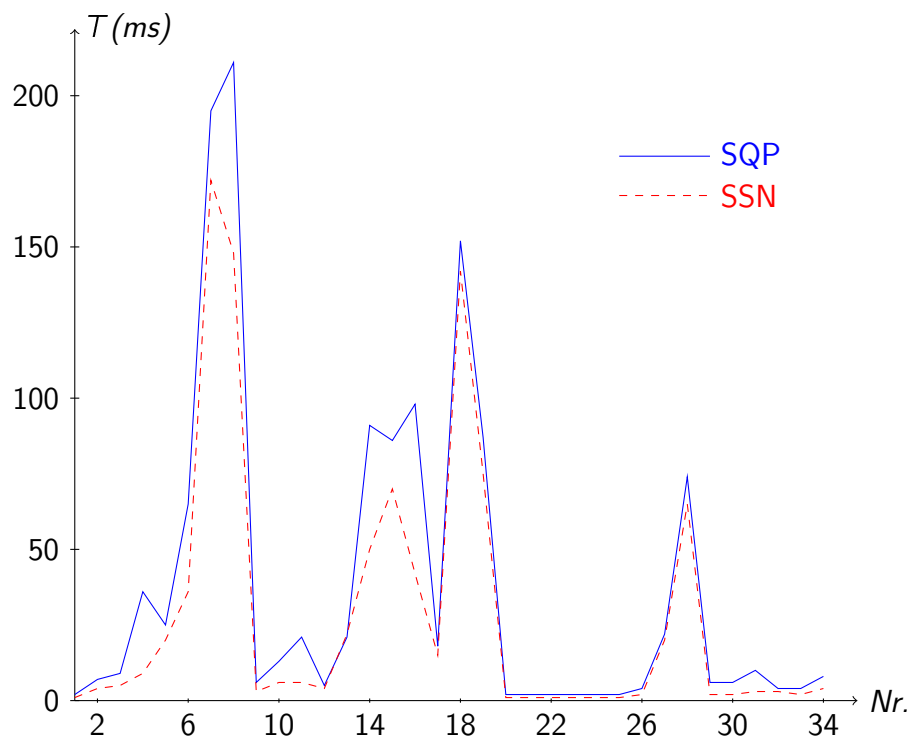


Abbildung 4.5: Laufzeiten der Programme für Testprobleme 1 bis 34

k	$f(x_{SSN}^k)$	$f(x_{SSN}^k)$
0	$7.516000000000000 \cdot 10^3$	$7.516000000000000 \cdot 10^3$
1	$2.767331665411093 \cdot 10^2$	$2.767331665411093 \cdot 10^2$
2	$5.773310883142562 \cdot 10^1$	$5.773256812579893 \cdot 10^1$
3	$1.254337618949399 \cdot 10^1$	$1.254324565470578 \cdot 10^1$
4	$2.779889090465643 \cdot 10^0$	$2.779857416624084 \cdot 10^0$
5	$5.287002889696653 \cdot 10^{-1}$	$5.286930595418454 \cdot 10^{-1}$
6	$4.147307881929466 \cdot 10^{-2}$	$4.147204949594635 \cdot 10^{-2}$
7	$7.875479554383774 \cdot 10^{-5}$	$7.874913302118516 \cdot 10^{-5}$
8	$6.419062610396878 \cdot 10^{-13}$	$6.417673709173576 \cdot 10^{-13}$
9	$1.232595164407831 \cdot 10^{-32}$	

Tabelle 4.6: Funktionsauwertungen der Iterationen bei Testproblem 28

k	$f(x_{SSN}^k)$	$f(x_{SSN}^k)$
0	$1.359016105161961 \cdot 10^2$	$1.359016105161961 \cdot 10^2$
1	$7.335302056475359 \cdot 10^1$	$1.197191744218137 \cdot 10^2$
2	$1.304845900695675 \cdot 10^2$	
3	$1.228854017478896 \cdot 10^2$	
4	$1.221620612725005 \cdot 10^2$	
5	$1.215130877159840 \cdot 10^2$	
6	$1.209513512807593 \cdot 10^2$	
7	$1.204858007269527 \cdot 10^2$	
8	$1.201621994947561 \cdot 10^2$	
9	$1.199331045245944 \cdot 10^2$	
10	$1.198027037119121 \cdot 10^2$	
11	$1.197394750363944 \cdot 10^2$	
12	$1.197199871731989 \cdot 10^2$	
13	$1.197191744218137 \cdot 10^2$	
14	$1.197191744218137 \cdot 10^2$	

Tabelle 4.7: Funktionsauwertungen der Iterationen bei Testproblem 35

Kapitel 5

Fazit und Ausblick

Es gibt mehrere Kriterien, um die Effizienz eines Verfahrens zu bewerten. Die Kriterien sind z. B. die Laufzeit des Programms und die Anzahl der benötigten Iterationen. Hier ziehen wir beim halbglatten Newton-Verfahren und beim SQP-Verfahren daher folgende Schlussfolgerungen:

- Falls die Laufzeit das wichtigste Kriterium sein soll, dann ist das halbglatte Newton-Verfahren das beste Verfahren für Probleme mit wenigen Variablen und das SQP-Verfahren das bessere Verfahren für Probleme mit vielen Variablen.
- Falls die Anzahl der Iterationen eine wichtigere Rolle spielt, dann ist das SQP-Verfahren das beste Verfahren.

Wenn die Implementierung des Verfahrens berücksichtigt werden soll, dann ist das halbglatte Newton-Verfahren die bessere Wahl, da die Implementierung von SQP-Verfahren aufwendig sein kann.

In dieser Arbeit wurden keine Testprobleme mit nichtlinearen Restriktionen berücksichtigt. Es wäre interessant, zu sehen, wie die Ergebnisse bei diesen Testproblemen aussehen. Ein anderer interessanter Aspekt ist die Eigenschaft der globalen Konvergenz beider Verfahren. Für jedes Testproblem wurde hier der Startpunkt in der Nähe der Lösung angegeben, damit die Eigenschaft der lokalen Konvergenz beider Verfahren verwendet werden konnte. Die beiden Verfahren können jedoch noch erweitert werden, damit ein beliebiger Startpunkt benutzt werden kann. Diese Varianten beider Verfahren mit globaler Konvergenz sollten noch untersucht und verglichen werden.

Eine Idee von Ulbrich ist, dass man beide Verfahren kombiniert [Ulbr11, S. 13]. Demnach sollte man versuchen, das Teilproblem in SQP-Verfahren mit halbglattem Newton-Verfahren anstatt mit dem Aktive-Mengen-Verfahren zu lösen. Es sollte untersucht werden, wie konkurrenzfähig diese Kombination ist.

Danksagung

An dieser Stelle möchte ich mich ganz herzlich bei Prof. Dr. Fredi Tröltzsch für die Anregung des Themas und die wissenschaftliche Betreuung bedanken.

Anschließend möchte ich meinen Eltern und meiner Schwester für die Unterstützung danken. Mein Dank gilt auch meinen Freunden, die mich während der Anfertigung dieser Arbeit unterstützt und motiviert haben. Besonders möchte ich Christian Cölln und Patrick Simon für die Hilfe beim Korrekturlesen danken.

Mein größter Dank gilt meinem Herrn und Erlöser Jesus Christus, der mir immer wieder Kraft geschenkt hat.

Literaturverzeichnis

- [AL07] Andreas Antoniou und Wu-Sheng Lu. *Practical Optimization: Algorithms and Engineering Applications*. Springer-Verlag New York Inc, 2007.
- [Alt11] Walter Alt. *Nichtlineare Optimierung: Eine Einführung in Theorie, Verfahren und Anwendungen*. Vieweg+Teubner, 2. Auflage, 2011.
- [HS81] Willi Hock und Klaus Schittkowski. *Test Examples for Nonlinear Programming Codes*. Volume 187 of Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, 1981.
- [IK08] Kazufumi Ito und Karl Kunisch. *Lagrange Multiplier Approach to Variational Problems and Applications*. Society for Industrial and Applied Mathematics, 2008.
- [NW06] Jorge Nocedal und Stephen J. Wright. *Numerical Optimization*. Springer-Verlag, 2. Auflage, 2006.
- [Trö11] Fredi Tröltzsch. *Nichtlineare Optimierung*. Vorlesungsskript, Technische Universität Berlin, 2011.
- [Ul11] Michael Ulbrich. *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*. Society for Industrial and Applied Mathematics and the Mathematical Optimization Society, 2011.
- [Wil63] R. B. Wilson. *A Simplicial Algorithm for Concave Programming*. PhD Thesis, Graduate School of Business Administration, Harvard University, Boston, 1963.