

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>ii</b>
<b>1 Einleitung</b>	<b>1</b>
<b>2 Grundlagen</b>	<b>2</b>
2.1 Allgemeine Optimierungsprobleme . . . . .	2
2.2 Optimierungsprobleme ohne Restriktionen . . . . .	5
2.3 Optimierungsprobleme mit linearen Restriktionen . . . . .	9
2.4 Optimierungsprobleme mit nichtlinearen Restriktionen . . . . .	14
<b>3 SQP-Verfahren</b>	<b>16</b>
3.1 Einführung . . . . .	16
3.2 Formulierung . . . . .	17
3.3 Aktive-Mengen-Verfahren . . . . .	18
3.4 Nullraum-Verfahren . . . . .	20
3.5 Lagrange-Newton-SQP-Verfahren . . . . .	21
<b>4 Das halbglatte Newton-Verfahren</b>	<b>22</b>
4.1 Grundlagen . . . . .	22
4.2 Formulierung . . . . .	24
4.3 Aktive-Menge-Strategie . . . . .	26
<b>5 Der Vergleich</b>	<b>29</b>
5.1 Testfunktionen . . . . .	29
5.2 Testprobleme . . . . .	32
<b>Literaturverzeichnis</b>	<b>37</b>

# Abbildungsverzeichnis

2.1	Die Funktion von Beispiel 2.1 . . . . .	4
2.2	Geometrische Darstellung des Beispiels 2.2 . . . . .	5
2.3	Geometrische Darstellung des Beispiels 2.3 . . . . .	10
2.4	Geometrische Darstellung des Beispiels 2.4 . . . . .	13
5.1	Exponentielle Funktion . . . . .	30
5.2	Rosenbrock-Funktion . . . . .	31
5.3	Himmelblau-Funktion . . . . .	32
5.4	Bazaraa-Shetty-Funktion . . . . .	33
5.5	$f(x) = 0.7x - 4.3$ und $g(x) = 0.242(x - 4.056)^2 - 2.955$ . . . . .	35
5.6	Restringiertes Problem . . . . .	36

# Kapitel 1

## Einleitung

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

Das SQP-Verfahren gehört zu den bekanntesten Verfahren der nichtlinearen Optimierung. In seiner Doktorarbeit im Jahr 1963 hat Wilson das erste SQP-Verfahren entwickelt. 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 im Kapitel 2 die wichtigsten Grundlagen der Optimierungstheorie eingeführt. Im Kapitel 3 wird das SQP-Verfahren betrachtet. Anschließend wird im Kapitel 4 das halbglatte Newton-Verfahren vorgestellt und untersucht. Im Kapitel 5 werden die Ergebnisse der beiden Verfahren für verschiedene Testprobleme aufgeführt.

# Kapitel 2

## Grundlagen

Wir werden erst mal in diesem Kapitel die grundlegenden Definitionen und die wichtigsten theoretischen Ergebnisse der Optimierungstheorie betrachten. Wir beschäftigen uns hier mit Optimierungsproblemen, die durch differenzierbare Funktionen charakterisiert werden. Wir werden verschiedene Optimalitätsbedingungen der Optimierungsprobleme sehen, die als Grundlage der numerischen Verfahren dienen werden. Dieses Kapitel orientiert sich an [Alt11, Trö11].

### 2.1 Allgemeine Optimierungsprobleme

**Definition 2.1** 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, dann bezeichnet man das Optimierungsproblem als unrestringiert. Es besitzt also keine Restriktionen. Ansonsten heißt es ein restringiertes Optimierungsproblem.

Man definiert in der Regel Optimierungsproblem als ein Minimierungsproblem, weil ein Maximierungsproblem  $\max g(x)$  zu dem Minimierungsproblem  $\min f(x) := -g(x)$  äquivalent ist.

**Definition 2.2** (Globale und lokale Lösung, vgl. Definition 1.1.4 in [Alt11, S. 2f])  
Ein Punkt  $x^* \in \mathcal{F}$  heißt globale Lösung des Problems (P) oder globales Minimum, wenn

$$f(x^*) \leq f(x) \quad \forall x \in \mathcal{F} \quad (2.1)$$

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

$$f(x^*) < f(x) \quad \forall x \in \mathcal{F} \setminus \{x^*\} \quad (2.2)$$

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

$$f(x^*) \leq f(x) \quad \forall x \in U(x^*) \cap \mathcal{F} \quad (2.3)$$

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

$$f(x^*) < f(x) \quad \forall x \in U(x^*) \cap \mathcal{F} \setminus \{x^*\} \quad (2.4)$$

gilt. Eine Umgebung  $U(x^*)$  von  $x^*$  ist eine offene Menge, die  $x^*$  beinhaltet.

Wegen dieser Definitionen kommt der Begriff *globale Optimierung*. 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 2.1** (vgl. Beispiel 1.1.2 in [Alt11, S. 1])

Ein Beispiel eines unrestringierten Optimierungsproblems ist

$$\min_{x \in \mathbb{R}} f(x) := (2x - 2)^2(3x + 3)^2 + 10x. \quad (2.5)$$

Dieses Problem hat eine strikte globale Lösung an der Stelle  $x^* = -1$  und eine strikte lokale Lösung an der Stelle  $x = 1$  (siehe Abbildung 2.1).

**Definition 2.3** (Nichtlineare Optimierungsprobleme) Die Aufgabenstellung bei der nichtlinearen Optimierung mit Gleichungs- und Ungleichungsrestriktionen kann man spezifischer wie folgt definieren:

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

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:  $\mathcal{F} = \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}$ .

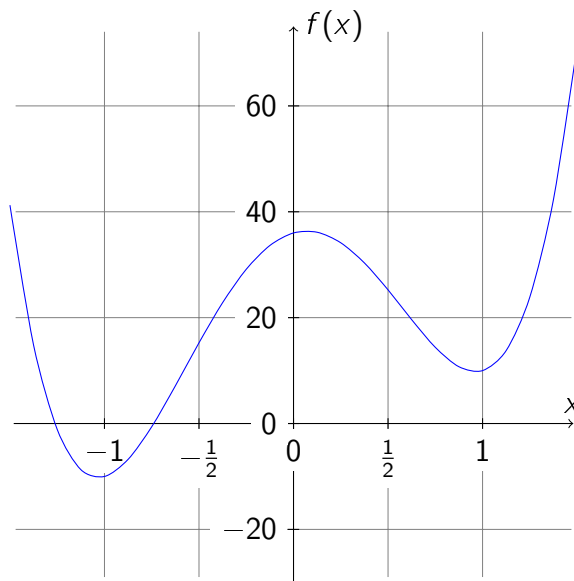


Abbildung 2.1: Die Funktion von Beispiel 2.1

Die Bezeichnung “Nb.” steht hier als Abkürzung für “unter der Nebenbedingung” oder “unter den Nebenbedingungen”. Die Operatoren  $\leq$  und  $=$  vergleichen hier die Vektoren miteinander elementenweise. Die Zahl 0 hier ist 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, so dass

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

Man bekommt dann das Problem

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

Man kann hierbei den Unterschied zwischen der linearen Optimierung und der nichtlinearen Optimierung gut erkennen. Bei der linearen Optimierung muss die Zielfunktion linear sein (d. h., die Zielfunktion muss in der Form  $f(x) = c^T x$ ,  $c \in \mathbb{R}^n$ , 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 2.2** (vgl. Problem 1.2 in [NW06, S. 3])

Ein Beispiel eines restringierten Optimierungsproblems ist

$$\begin{aligned} \min_{x \in \mathbb{R}^2} & (x_1 - 2)^2 + (x_2 - 1)^2 \\ \text{Nb. } & x_1^2 \leq x_2 \\ & x_1 + x_2 \leq 2. \end{aligned} \tag{2.6}$$

Die Abbildung 2.2 zeigt die zulässige Menge  $\mathcal{F}$ , die Höhenlinien der Zielfunktion und die optimale Lösung  $x^*$ .

Die Höhenlinie ist eine Menge von Punkten, wo die Zielfunktion einen konstanten Wert nimmt. In der Abbildung 2.2 sind die Höhenlinien in Form von Kreisen zu sehen und je größer der Kreis ist, desto größer ist der Wert der Zielfunktion.

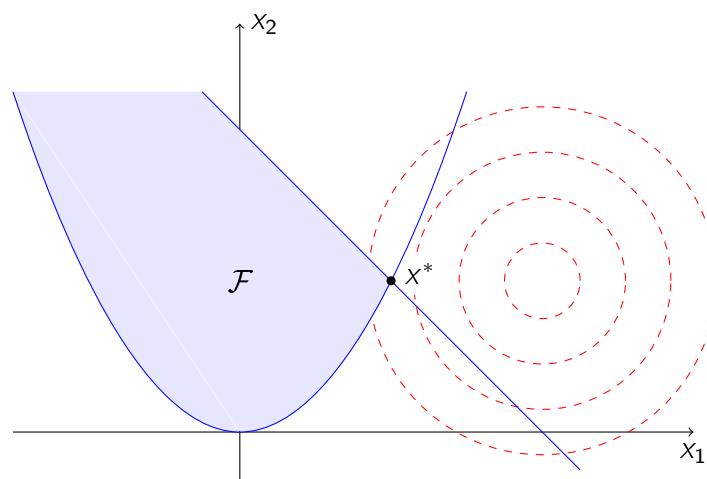


Abbildung 2.2: Geometrische Darstellung des Beispiels 2.2

## 2.2 Optimierungsprobleme ohne Restriktionen

Wir beginnen nun zuerst mit Optimierungsproblemen ohne Restriktionen und stellen die notwendige und hinreichende Optimalitätsbedingungen bereit. Wir werden auch zwei numerische Verfahren hierfür vorstellen.

**Definition 2.4** (Unrestringierte Optimierungsprobleme)

$$\min_{x \in \mathbb{R}^n} f(x) \quad (\text{PU})$$

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

**Satz 2.1** (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

$$\nabla f(x^*) = 0. \quad (2.7)$$

Die Bedingung (2.7) gilt aber nicht nur für lokales Minimum sondern auch für lokales Maximum von  $f$ .

**Definition 2.5** (Stationärer Punkt, vgl. Definition 3.1.4 in [Alt11, S. 42])

Die Funktion  $f$  sei in  $x^*$  differenzierbar. Der Punkt  $x^*$  heißt ein stationärer Punkt von  $f$ , wenn  $x^*$  die notwendige Bedingung (2.7) erfüllt.

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

**Satz 2.2** (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 (2.7) und

$$x^T f''(x^*) x \geq 0 \quad \forall x \in \mathbb{R}^n. \quad (2.8)$$

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

**Satz 2.3** (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 (2.7) sei erfüllt und  $f''(x^*)$  sei positiv definit, d. h.

$$x^T f''(x^*) x > 0 \quad \forall x \in \mathbb{R}^n. \quad (2.9)$$

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 2.6** (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

$$\nabla f(x)^T d < 0 \quad (2.10)$$

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 2.4** (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

$$f(x + \sigma d) < f(x) \quad \forall \sigma \in ]0, \hat{\sigma}[. \quad (2.11)$$

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 vorherige 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  $\sigma_k$  so, dass für  $x^{k+1} := x^k + \sigma_k d^k$

$$f(x^{k+1}) < f(x^k) \quad (2.12)$$

gilt.

## 2.2.1 Gradientenverfahren

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

**Verfahren 2.1** (Gradientenverfahren, vgl. Verfahren 4.2.39 in [Alt11, S. 98])

1. Wähle einen Startpunkt  $x^0$  und eine Abbruchschranke  $\varepsilon > 0$ . Setze  $k := 0$ .
2. Ist  $\|\nabla f(x^k)\| < \varepsilon \Rightarrow STOP$ .
3. Setze  $d^k := -\nabla f(x^k)$ .

4. Bestimme  $\sigma_k$  so, dass

$$f(x^k + \sigma_k d^k) < f(x^k + \sigma d^k) \quad \forall \sigma \geq 0. \quad (2.13)$$

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

Um die Schrittweite  $\sigma_k$  zu bestimmen, kann man das Schrittweitenverfahren von Armijo oder das Schrittweitenverfahren von Wolfe-Powell verwenden. Wir verweisen auf das Unterkapitel 4.2.7 über Schrittweitenverfahren in [Alt11, S. 90ff].

## 2.2.2 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 unsere unrestringierte Optimierungsprobleme können wir das Newton-Verfahren verwenden, um die Lösung der nichtlinearen Gleichung (2.7),  $\nabla f(x) = 0$ , zu finden. Wir müssen dabei voraussetzen, dass die Funktion  $f$  zweimal differenzierbar sei.

**Verfahren 2.2** (Newton-Verfahren, vgl. Verfahren 4.3.1 in [Alt11, S. 107])

1. Wähle einen Startpunkt  $x^0$  und eine Abbruchschranke  $\varepsilon > 0$ . Setze  $k := 0$ .
2. Ist  $\|\nabla f(x^k)\| < \varepsilon \Rightarrow \text{STOP}$ .
3. Berechne die Lösung  $d$  des linearen Gleichungssystems

$$f''(x^k)d = -\nabla f(x^k). \quad (2.14)$$

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 Iterationschritt ist eigentlich die Lösung des Minimierungsproblems, welches durch die quadratische Approximation von  $f$  in Punkt  $x^k$  definiert ist. In der Umgebung von  $x^k$  können wir die Funktion  $f$  wie folgt approximieren:

$$f(x) \approx f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2} (x - x^k)^T f''(x^k) (x - x^k). \quad (2.15)$$

Die Ableitung der rechten Seite ist

$$\nabla f(x^k) + f''(x^k)x - f''(x^k)x^k. \quad (2.16)$$

Setzen wir diese gleich null, dann bekommen wir

$$f''(x^k)(x - x^k) = -\nabla f(x^k) \quad (2.17)$$

$$x - x^k = -[f''(x^k)]^{-1} \nabla f(x^k) \quad (2.18)$$

$$x = x^k - \underbrace{[f''(x^k)]^{-1} \nabla f(x^k)}_d. \quad (2.19)$$

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

$$\min_{x \in \mathbb{R}^n} f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2} (x - x^k)^T f''(x^k) (x - x^k), \quad (2.20)$$

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

Wir werden später sehen, dass das SQP-Verfahren diese Idee auch gebrauchen wird.

Man kann auch noch eine Schrittweitensteuerung wie bei dem Gradientenverfahren durchführen. 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 man als das gedämpfte Newton-Verfahren bezeichnet. Wir verweisen auf das Unterkapitel 4.3.2 über das gedämpfte Newton-Verfahren in [Alt11, S. 111ff].

## 2.3 Optimierungsprobleme mit linearen Restriktionen

Nun kommen wir zu der Theorie der restringierten Optimierungsproblemen. Wir konzentrieren uns erst mal mit den linearen Restriktionen. D.h. die Nebenbedingungen sind durch lineare Gleichungen und Ungleichungen definiert.

### 2.3.1 Optimierungsprobleme mit linearen Gleichungsnebenbedingungen

**Definition 2.7** (Optimierungsprobleme mit linearen Gleichungsnebenbedingungen)

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

$A$  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 2.3** Ein einfaches Beispiel eines Optimierungsproblems PLG ist

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \quad & x_1^2 + x_2^2 \\ \text{Nb.} \quad & x_1 + x_2 = 1. \end{aligned} \tag{2.21}$$

Die zulässige Punkte dieses Problems liegen auf einer Gerade (Siehe Abbildung 2.3).

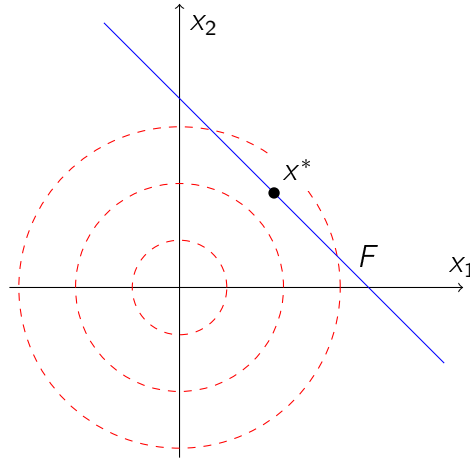


Abbildung 2.3: Geometrische Darstellung des Beispiels 2.3

**Satz 2.5** (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

$$\nabla f(x^*) + A^T \lambda = 0. \tag{2.22}$$

Hat  $A$  einen vollen Rang, dann ist  $\lambda$  eindeutig bestimmt.

Diese Bedingung heißt die Multiplikatoren-Regel von Lagrange. Man bezeichnet  $\lambda$  als Lagrange-Multiplikator.

**Satz 2.6** (Hinreichende Bedingung, vgl. Satz 5.3.8 in [Alt11, S. 177])

Sei  $f$  in  $x^*$  zweimal stetig differenzierbar. Die notwendige Bedingung in Satz 2.5 sei erfüllt. Es existiere eine Konstante  $\alpha > 0$  mit

$$d^T f''(x) d \geq \alpha \|d\|^2 \quad \forall d \in \text{Kern } A. \tag{2.23}$$

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

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

**Definition 2.8** (Nullraum-Matrix)

Eine  $(n \times l)$ -Matrix  $Z$  heißt Nullraum-Matrix von  $A$ , wenn für  $d \in \mathbb{R}^n$  gilt

$$d \in \text{Kern } A \iff d = Zz \text{ für ein } z \in \mathbb{R}^l. \quad (2.24)$$

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

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

$$\mathcal{F} = w + \text{Kern } A = w + \text{Im } Z = w + \{Zz \mid z \in \mathbb{R}^l\}. \quad (2.25)$$

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

$$\min_{z \in \mathbb{R}^l} F(z) := f(w + Zz). \quad (2.26)$$

Dieses Problem hat keine Nebenbedingung mehr, also unrestringiert. Man kann also Verfahren für unrestringierte Probleme anwenden.

### 2.3.2 Optimierungsprobleme mit linearen Ungleichungsnebenbedingungen

**Definition 2.9** (Optimierungsprobleme mit linearen Ungleichungsnebenbedingungen)

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

$A$  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.  $G$ , so dass

$$A = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix} \quad \text{und} \quad G = \begin{pmatrix} g_1^T \\ \vdots \\ g_p^T \end{pmatrix}. \quad (2.27)$$

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 können wir das Problem (PLU) wie folgt ausführlicher schreiben.

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

**Beispiel 2.4** Gegeben seien zwei Dreiecke  $R$  und  $S$  in der Abbildung 2.4. 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

$$\|r - s\|^2 = (x_1 - x_3)^2 + (x_2 - x_4)^2 = x^T H x \quad (2.28)$$

gegeben, wobei

$$H := \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}. \quad (2.29)$$

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

$$\begin{aligned} x_1, x_2 &\geq 0 \\ x_1 + 2x_2 &\leq 2 \\ x_4 &\geq 2 \\ x_3 + x_4 &\geq 3 \\ x_3 + 2x_4 &\leq 6 \end{aligned} \quad (2.30)$$

gegeben. D. h. das Problem kann als ein 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 2.7** (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

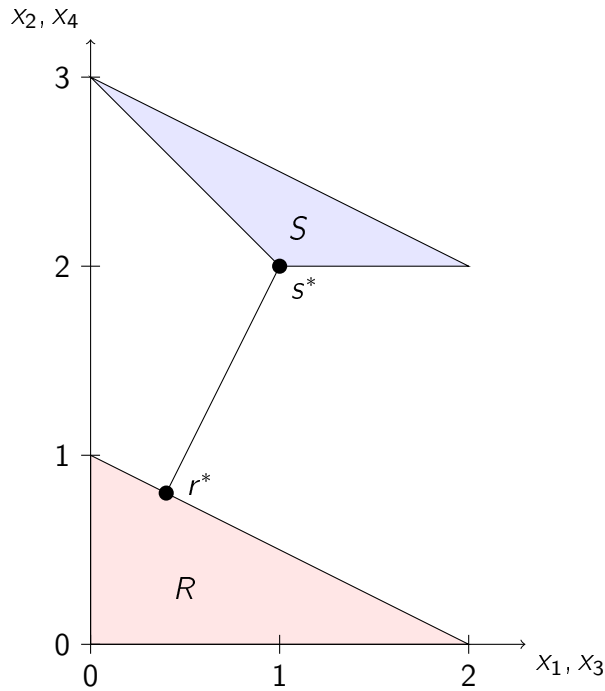


Abbildung 2.4: Geometrische Darstellung des Beispiels 2.4

Vektoren  $\lambda \in \mathbb{R}^m$  und  $\mu \in \mathbb{R}^p$  zu  $x^*$  mit

$$\nabla f(x^*) + A^T \lambda + G^T \mu = 0 \quad (2.31)$$

$$\mu_j (\langle g_j, x^* \rangle - r_j) = 0 \text{ für } j = 1, \dots, p \quad (2.32)$$

$$\mu \geq 0 \quad (2.33)$$

Die Vektoren  $\lambda$  und  $\mu$  heißen Lagrange-Multiplikatoren zu  $x^*$ .

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

$$J(x) := \{1 \leq j \leq p \mid \langle g_j, x \rangle = r_j\} \quad (2.34)$$

die Indexmenge der in  $x$  aktiven Ungleichungsrestriktionen.

**Satz 2.8** (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 2.7 seien erfüllt und es gelte mit  $\alpha > 0$

$$d^T f''(x^*) d \geq \alpha \|d\|^2 \quad \forall d \in \mathbb{R}^n : \begin{cases} Ad = 0, \\ \langle g_j, d \rangle \leq 0 & \text{für } j \in J(x^*) \text{ mit } \mu_j = 0, \\ \langle g_j, d \rangle = 0 & \text{für } j \in J(x^*) \text{ mit } \mu_j > 0. \end{cases} \quad (2.35)$$

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 2.10** (Optimierungsprobleme mit Variablenbeschränkungen)

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

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

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

## 2.4 Optimierungsprobleme mit nichtlinearen Restriktionen

Wir kommen nun zu unserem allgemeinen nichtlinearen Optimierungsproblem. Wir schreiben nochmal das Problem (PN).

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

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

**Definition 2.11** (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

$$\nabla h_j(x), \quad j = 1, \dots, m, \quad \text{linear unabhängig} \quad (2.36)$$

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

$$\nabla g_i(x)^T d < 0, \quad i \in \mathcal{I}(x) \quad \text{und} \quad \nabla h_j(x)^T d = 0, \quad j = 1, \dots, m \quad (2.37)$$

existiert.



Der folgende Satz gilt eigentlich mit der allgemeinen Regularitätsbedingung. Wir verwenden aber hier nur die Mangasarian-Fromowitz-Regularitätsbedingung.

**Satz 2.9** (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  $\mu \in \mathbb{R}^p$ , so dass*

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) = 0 \quad (2.38)$$

$$\mu_j \geq 0, \quad \mu_j g_j(x^*) = 0, \quad j = 1, \dots, p \quad (2.39)$$

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

# Kapitel 3

## 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 Verfahren für quadratische 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].

### 3.1 Einführung

Das SQP-Verfahren ist ein wichtiges Verfahren, um restringierte nichtlineare Optimierungsprobleme zu lösen. SQP ist eine Abkürzung für sequentielle quadratische Programmierung<sup>1</sup>. Die Hauptidee von SQP ist nämlich, iterativ Teilprobleme in Form quadratischer Optimierungsprobleme zu formulieren und zu lösen.

Wir haben im Unterkapitel 2.2.2 über das Newton-Verfahren gesehen, dass dieses Verfahren in jedem Iterationsschritt eigentlich ein unrestringiertes quadratisches Problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}(x - x^k)^T f''(x^k)(x - x^k) + \nabla f(x^k)^T (x - x^k) \quad (3.1)$$

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.

$$\min_{x \in \mathcal{F}} \frac{1}{2}(x - x^k)^T f''(x^k)(x - x^k) + \nabla f(x^k)^T (x - x^k). \quad (3.2)$$

---

<sup>1</sup>englisch: Sequential Quadratic Programming

## 3.2 Formulierung

Wir betrachten erstmal Optimierungsprobleme mit linearen Gleichungs- und Ungleichungsnebenbedingungen (PLU).

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

Das Teilproblem nach (3.2) sieht dann so aus:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} & \frac{1}{2}(x - x^k)^T f''(x^k)(x - x^k) + \nabla f(x^k)^T (x - x^k) \\ \text{Nb. } & Ax = b \\ & Gx \leq r \end{aligned} \quad (3.3)$$

Definieren wir  $d := x - x^k$ . Dann haben wir  $x = x^k + d$  und als Teilproblem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} d^T f''(x^k) d + \nabla f(x^k)^T d \quad (3.4)$$

$$\text{Nb. } Ax^k + Ad = b \quad (3.5)$$

$$Gx^k + Gd \leq r \quad (3.6)$$

Das ist ein quadratisches Optimierungsproblem mit linearen Restriktionen. Wir werden im nächsten Unterkapitel sehen, wie man es lösen kann.

Weil wir voraussetzen werden, dass  $x^k \in \mathcal{F}$  ist, d. h., es gilt  $Ax^k = b$ , können wir als Gleichungsnebenbedingung an der Stelle (3.5) die Gleichung  $Ad = 0$  schreiben. Nun sind wir bereit, das SQP-Verfahren zu formulieren.

**Verfahren 3.1** (SQP-Verfahren für (PLU), vgl. Verfahren 6.4.1 in [Alt11, S. 237])

1. Wähle einen zulässigen Startpunkt  $x^0$  und eine Abbruchschranke  $\varepsilon > 0$ . Setze  $k := 0$ .
2. Berechne die Lösung  $d$  des Problems

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} d^T f''(x^k) d + \nabla f(x^k)^T d \quad (3.7)$$

$$\text{Nb. } Ad = 0$$

$$Gx^k + Gd \leq r$$

Setze  $d^k := d$ .

3. Ist  $\|d^k\| < \varepsilon \Rightarrow \text{STOP}$ .

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

**Satz 3.1** (vgl. Satz 6.4.4 in [Alt11, S. 243])

Sei  $x^*$  lokale Lösung von (PLU). Sei  $f$  auf  $D$  zweimal stetig differenzierbar und es gäbe eine Konstante  $\alpha > 0$  mit

$$d^T f''(x) d \geq \alpha \|d\|^2 \quad \forall d \in \text{Kern } A. \quad (3.8)$$

Dann gibt es ein  $\delta > 0$ , so dass das SQP-Verfahren für jeden Startpunkt  $x^0 \in B(x^*, \delta)$  eine Folge  $\{x^k\}$  definiert, die superlinear gegen  $x^*$  konvergiert. Ist zusätzlich  $f''$  auf  $D$  Lipschitz-stetig, d. h. es gibt eine Konstante  $L > 0$  mit

$$\|f''(x) - f''(y)\| \leq L \|x - y\| \quad \forall x, y \in D, \quad (3.9)$$

dann konvergiert  $\{x^k\}$  für jeden Startpunkt  $x^0 \in B(x^*, \delta)$  quadratisch gegen  $x^*$ .

### 3.3 Aktive-Mengen-Verfahren

Wir haben gesehen, dass das SQP-Verfahren quadratische Optimierungsprobleme mit linearen Restriktionen als Teilprobleme hat. Wir werden jetzt aufklären, wie wir diese Probleme mit dem Aktive-Mengen-Verfahren lösen können.

**Definition 3.1** (Quadratische Optimierungsprobleme mit linearen Restriktionen)

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \langle Qx, x \rangle + \langle q, x \rangle \quad (\text{QLU})$$

$$\text{Nb. } \langle a_i, x \rangle = b_i \text{ für } i = 1, \dots, m$$

$$\langle g_j, x \rangle \leq r_j \text{ für } j = 1, \dots, p$$

Dabei sei  $Q$  eine symmetrische  $(n \times n)$ -Matrix und  $q \in \mathbb{R}^n$ .  $a_i \in \mathbb{R}^n$  und  $b_i \in \mathbb{R}$  für  $i = 1, \dots, m$ .  $g_j \in \mathbb{R}^n$  und  $r_j \in \mathbb{R}$  für  $j = 1, \dots, p$ .

Die Idee des Aktive-Mengen-Verfahrens ist, iterativ Probleme mit nur Gleichungsnebenbedingungen zu optimieren.

Sei  $x^k \in \mathbb{R}^n$  ein zulässiger Punkt des Problems (QLU). Angenommen, wir können eine Suchrichtung  $d \in \mathbb{R}^n$  finden, so dass der Punkt  $x^{k+1} := x^k + d$  ein besserer zulässiger Punkt ist. Diesen Vektor  $d$  können wir finden, indem wir das Problem

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} \langle Q(x^k + d), x^k + d \rangle + \langle q, x^k + d \rangle \quad (3.10)$$

$$\text{Nb. } \langle a_i, x^k + d \rangle = b_i \text{ für } i = 1, \dots, m$$

$$\langle g_j, x^k + d \rangle \leq r_j \text{ für } j = 1, \dots, p$$

lösen. 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.

Lassen wir die Konstanten in der Zielfunktion weg, dann bekommen wir als Zielfunktion

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

Weil  $x^k$  zulässig ist, d. h., es gilt  $\langle a_i, x^k \rangle = b_i$  für  $i = 1, \dots, m$ , können wir dann als Gleichungsnebenbedingungen  $\langle a_i, d \rangle = 0$  für  $i = 1, \dots, m$  schreiben.

Sei  $J(x^k)$  die Indexmenge der aktiven Ungleichungsrestriktionen zu  $x^k$ . Sei  $j \in J(x^k)$ , d. h., es gilt  $\langle g_j, x^k \rangle = r_j$ . Dann folgt aus  $\langle g_j, d \rangle = 0$  immer noch  $\langle g_j, x^k + d \rangle \leq r_j$ .

Seien  $J_k := J(x^k)$ ,  $p_k := |J_k|$  und  $j_1, \dots, j_{p_k} \in J_k$ . Wir definieren dann folgendes Teilproblem nur mit linearen Gleichungsnebenbedingungen:

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \frac{1}{2} \langle Qd, d \rangle + \langle Qx^k + q, d \rangle \\ \text{Nb.} \quad & \langle a_i, d \rangle = b_i \text{ für } i = 1, \dots, m \\ & \langle g_j, d \rangle = r_j \text{ für } j = j_1, \dots, j_{p_k} \in J_k \end{aligned} \tag{Q_k}$$

Dieses Problem werden wir mit dem Nullraum-Verfahren lösen, von dem wir auch den Lagrange-Multiplikator  $\tilde{\lambda}$  bekommen werden.  $\tilde{\lambda}$  können wir in  $\lambda \in \mathbb{R}^m$  und  $\mu \in \mathbb{R}^{p_k}$  zerlegen, so dass  $\tilde{\lambda} = \begin{pmatrix} \lambda \\ \mu \end{pmatrix}$ . Besonders nachher für die Abbruchbedingung des Verfahrens zu betrachten, ist der Vektor  $\mu$ .

Da wir nur  $j \in J_k$  betrachten, kann es sein, dass die Ungleichung  $\langle g_j, x^k + d \rangle \leq r_j$  für  $i \in I_k := \{1, \dots, p\} \setminus J_k$  verletzt wurde. Definiere Schrittweite durch

$$\tau_k := \begin{cases} \min \left\{ \frac{r_j - g_j^T x^k}{g_j^T d^k} \mid j \in I_k \right\}, & \text{falls } I_k \neq \emptyset \\ \infty, & \text{sonst.} \end{cases} \tag{3.11}$$

und

$$\sigma_k := \min\{1, \tau_k\}. \tag{3.12}$$

Diese Schrittweite brauchen wir, damit  $x^{k+1}$  ein zulässiger Punkt bleibt.

**Verfahren 3.2** (Aktive-Mengen-Verfahren für (QLU), vgl. Verfahren 6.2.1 in [Alt11, S. 213])

1. Wähle einen Startpunkt  $x^0$ . Setze  $k := 0$ .

2. Löse das Problem  $(Q_k) \Rightarrow d^k, \lambda^k, \mu^k$ .
3. Falls  $d^k = 0$ 
  - 3a. Falls  $\mu^k \geq 0 \Rightarrow \text{STOP}$ .
  - 3b. Falls  $\mu^k \not\geq 0$ 
    - i. Bestimme  $j \in J(x^k)$ , so dass  $\mu_j^k = \min\{\mu_i^k \mid i \in J(x^k)\}$ .
    - ii.  $J(x^k) := J(x^k) \setminus \{j\}$
    - iii. Streiche in  $G_k$  die  $j$ -te Zeile.
    - iv. Löse wieder das Problem  $(Q_k) \Rightarrow d^k \neq 0$
4. Berechne Schrittweite  $\sigma_k$ .
5. Setze  $x^{k+1} := x^k + \sigma_k d^k$  und  $k := k + 1$ .  $\Rightarrow$  Gehe zu Schritt 2.

### 3.4 Nullraum-Verfahren

**Definition 3.2** (Quadratische Optimierungsprobleme mit linearen Gleichungsnebenbedingungen)

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

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) werden wir auf ein unrestringiertes Verfahren reduzieren mit Hilfe einer Nullraum-Matrix von  $A$ . Daher kommt der Name *Nullraum-Verfahren*.

**Verfahren 3.3** (Nullraum-Verfahren für (QLG), vgl. Verfahren 6.1.1 in [Alt11, S. 208f])

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

$$HA^T = \begin{pmatrix} R \\ 0 \end{pmatrix}. \quad (3.13)$$

2. Berechne

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} := -Hq \quad \text{und} \quad B := HQH^T = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \quad (3.14)$$

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

3. Berechne den Vektor  $u_1$  als Lösung der Gleichung

$$R^T u_1 = b \quad (3.15)$$

und den Vektor  $u_2$  als Lösung der Gleichung

$$B_{22}u_2 = h_2 - B_{21}u_1 \quad (3.16)$$

4. Berechne  $x^*$

$$x^* := H^T \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \quad (3.17)$$

5. Der Multiplikator  $\lambda^*$  ist die Lösung der Gleichung

$$R\lambda^* = h_1 - B_{11}u_1 - B_{12}u_2 \quad (3.18)$$

## 3.5 Lagrange-Newton-SQP-Verfahren

Nun kommen wir zu den allgemeinen nichtlinearen Optimierungsproblemen.

**Definition 3.3** Wir definieren die Lagrange-Funktion  $\mathcal{L} : D \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$  durch

$$\mathcal{L}(x, \lambda, \mu) := f(x) + \lambda^T h(x) + \mu^T g(x). \quad (3.19)$$

**Verfahren 3.4** (Lagrange-Newton-SQP-Verfahren, vgl. Verfahren 8.1.4 in [Alt11, S. 294])

1. Wähle einen Startpunkt  $z^0 := (x^0, \lambda^0, \mu^0)$  und  $\varepsilon > 0$ . Setze  $k := 0$ .

2. Berechne die Suchrichtung  $d^k$  durch Lösung des Problems

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} d^T \mathcal{L}_{xx}(x^k, \lambda^k, \mu^k) d + \nabla f(x^k)^T d \quad (3.20)$$

$$\text{Nb. } h(x^k) + h'(x^k)d = 0 \quad (3.21)$$

$$g(x^k) + g'(x^k)d \leq 0 \quad (3.22)$$

3. Ist  $\|d^k\| < \varepsilon \Rightarrow \text{STOP}$ .

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

# Kapitel 4

## Das halbglatte Newton-Verfahren

### 4.1 Grundlagen

Wir betrachten erst mal die allgemeine Definition einer halbglatte<sup>1</sup> Funktion.

**Definition 4.1** Seien  $X, Z$  reelle Banachräume und sei  $D \subset X$  offen. Eine Funktion  $F : D \subset X \rightarrow Z$  heißt Newton-differenzierbar in Punkt  $x \in D$ , falls es eine Umgebung  $U(x) \subset D$  von  $x$  und eine Abbildung  $G : U(x) \rightarrow L(X, Z)$  gibt, so dass

$$\lim_{|h| \rightarrow 0} \frac{|F(x+h) - F(x) - G(x+h)h|_Z}{|h|_X} = 0 \quad (4.1)$$

Die Familie

$$\{G(u) \mid u \in U(x)\} \quad (4.2)$$

heißt  $N$ -Ableitung von  $F$  in  $x$ .

**Definition 4.2** Seien  $X, Z$  reelle Banachräume und sei  $D \subset X$  offen. Eine Funktion  $F : D \subset X \rightarrow Z$  heißt halbglatte in  $x \in D$ , falls  $F$  Newton-differenzierbar in  $x$  ist und

$$\lim_{t \rightarrow 0^+} G(x+th)h \quad (4.3)$$

einheitlich in  $|h| = 1$  existiert.

Nun kommt die Definition in endlichdimensionalen Räumen.

---

<sup>1</sup>englisch: semismooth



**Definition 4.3** Sei  $F : \mathbb{R}^m \rightarrow \mathbb{R}^n$  lokal Lipschitz-stetig. Sei  $D_F$  die Menge aller Punkte, wo  $F$  differenzierbar ist. Für  $x \in \mathbb{R}^m$  definieren wir

$$\partial_B F(x) := \left\{ J \mid J = \lim_{x_i \rightarrow x, x_i \in D_F} \nabla F(x_i) \right\} \quad (4.4)$$

und die allgemeine Ableitung in  $x$

$$\partial F(x) := \text{co } \partial_B F(x), \quad (4.5)$$

wobei  $\text{co}$  für die konvexe Hülle steht.

**Definition 4.4**  $F : \mathbb{R}^m \rightarrow \mathbb{R}^n$  heißt halbglatte in  $x \in \mathbb{R}^m$ , falls  $F$  lokal Lipschitz-stetig in  $x$  ist und

$$\lim_{V \in \partial F(x+th'), h' \rightarrow h, t \rightarrow 0^+} V h' \quad (4.6)$$

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

**Satz 4.1** Sei  $x^*$  eine Lösung des Problems  $F(x) = 0$  und  $F$  sei Newton-differenzierbar in  $x^*$  mit  $N$ -Ableitung  $G$ . Falls  $G$  nichtsingulär für alle  $x \in U(x^*)$  ist und  $\{\|G(x)^{-1}\| \mid x \in U(x^*)\}$  beschränkt ist, dann konvergiert die Newton-Iteration

$$x^{k+1} := x^k - G(x^k)^{-1} F(x^k) \quad (4.7)$$

superlinear gegen  $x^*$  unter der Bedingung, dass  $|x^0 - x^*|$  genügend klein ist.

**Verfahren 4.1** (Das halbglatte Newton-Verfahren)

1. Wähle einen Startpunkt  $x^0$  und eine Abbruchschranke  $\varepsilon > 0$ . Setze  $k := 0$ .
2. Ist  $\|\nabla f(x^k)\| < \varepsilon \Rightarrow \text{STOP}$ .
3. Wähle  $V_k \in \partial \nabla f(x^k)$
4. Berechne die Lösung  $d$  des linearen Gleichungssystems

$$V_k d = -\nabla f(x^k). \quad (4.8)$$

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 hat also das gleiche Ziel wie bei dem Newton-Verfahren, nämlich ein nichtlineares Gleichungssystem zu lösen. Der Unterschied liegt darin, dass hier die Funktion nicht unbedingt so glatt ist.

## 4.2 Formulierung

Wir werden jetzt das halbglatte Newton-Verfahren formulieren, um nichtlineare Optimierungsprobleme mit linearen Restriktionen zu lösen.

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

$$\begin{aligned} x_1, x_2 \geq 0 \\ x_1 x_2 = 0 \end{aligned} \Leftrightarrow \min(x_1, x_2) = 0 \quad (4.9)$$

Wenn wir also die Bedingungen

$$\mu \geq 0, \quad Gx \leq r \quad \text{und} \quad \mu^T(Gx - r) = 0 \quad (4.10)$$

haben, können wir sie als

$$\min(\mu, r - Gx) = 0 \quad (4.11)$$

schreiben. Hier arbeitet der Operator  $\min$  elementenweise.

Wir betrachten nun das Problem

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

Die Optimalitätsbedingungen für dieses Problem ist

$$\nabla f(x) + G^T \mu = 0 \quad (4.12)$$

$$\mu \geq 0, \quad Gx \leq r, \quad \mu^T(Gx - r) = 0 \quad (4.13)$$

Diese können wir nach unserer vorherigen Überlegung schreiben als

$$\nabla f(x) + G^T \mu = 0 \quad (4.14)$$

$$\min(\mu, r - Gx) = 0 \quad (4.15)$$

Wir definieren nun die Funktion

$$F(x, \mu) = \begin{pmatrix} \nabla f(x) + G^T \mu \\ \min(\mu, r - Gx) \end{pmatrix} \quad (4.16)$$

Wir müssen nur noch die Ableitung von  $F$  bestimmen. Dann können wir das Verfahren formulieren.

Wir definieren:

$$\mathcal{A} := \{ j \in \{1, \dots, p\} \mid r_j - g_j^T x < \mu_j \}, \quad (4.17)$$

$$\mathcal{I} := \{ j \in \{1, \dots, p\} \mid r_j - g_j^T x \geq \mu_j \} = \{1, \dots, p\} \setminus \mathcal{A} \quad (4.18)$$

$$\chi_M(m) := \begin{cases} 1 & \text{für } m \in M, \\ 0 & \text{sonst.} \end{cases} \quad (4.19)$$

D. h. es gilt

$$\min(\mu, r - Gx) = \begin{pmatrix} \chi_{\mathcal{I}}(1)\mu_1 + \chi_{\mathcal{A}}(1)(r_1 - g_1^T x) \\ \vdots \\ \chi_{\mathcal{I}}(p)\mu_p + \chi_{\mathcal{A}}(p)(r_p - g_p^T x) \end{pmatrix} \quad (4.20)$$

Die Ableitung ist dann in der Form:

$$F'(x, \mu) = \begin{pmatrix} f''(x) & G^T \\ -\chi_{\mathcal{A}}(1)g_1^T & \chi_{\mathcal{I}}(1)e_1^T \\ \vdots & \vdots \\ -\chi_{\mathcal{A}}(p)g_p^T & \chi_{\mathcal{I}}(p)e_p^T \end{pmatrix} \quad (4.21)$$

#### Verfahren 4.2 (Das halbglatte Newton-Verfahren)

1. Wähle  $x^0, \mu^0$  und eine Abbruchschranke  $\varepsilon > 0$ . Setze  $k := 0$ .

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

$$F'(x^k, \mu^k)d = -F(x^k, \mu^k). \quad (4.22)$$

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

3. Ist  $\|d_x^k\| < \varepsilon \Rightarrow \text{STOP}$ .

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

### 4.3 Aktive-Menge-Strategie

Wir betrachten in diesem Unterkapitel die Aktive-Menge-Strategie. Dieses Verfahren ist für uns interessant, weil sie als das halbglatte Newton-Verfahren interpretiert werden kann.

Wir betrachten wieder das Problem

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

Eine der Optimalitätsbedingungen dieses Problems ist:

$$\nabla f(x) + G^T \mu = 0. \quad (4.23)$$

Das ist ein nichtlineares Gleichungssystem, wenn  $f$  nichtlinear ist.

Wir versuchen nun, eine linearisierte Version zu bekommen.

Sei  $x^k$  ein zulässiger Punkt des Problems. Approximieren wir die Funktion  $f$  in der Umgebung von  $x^k$ :

$$f(x) \approx f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2} (x - x^k)^T f''(x^k) (x - x^k). \quad (4.24)$$

Der Gradient von  $f$  ist dann

$$\nabla f(x) \approx \nabla f(x^k) + f''(x^k)x - f''(x^k)x^k. \quad (4.25)$$

Ersetzen wir nun  $\nabla f(x)$  in der Optimalitätsbedingung mit seiner Approximation, dann bekommen wir die Gleichung

$$\nabla f(x^k) + f''(x^k)x - f''(x^k)x^k + G^T \mu = 0 \quad (4.26)$$

$$f''(x^k)x + G^T \mu = f''(x^k)x^k - \nabla f(x^k) \quad (4.27)$$

Das ist nun ein lineares Gleichungssystem. Dieses werden wir in unserer Aktive-Mengen-Strategie verwenden.

#### Verfahren 4.3 (Aktive-Menge-Strategie)

1. Wähle  $x^0$  und setze  $k := 0$ .
2. Bestimme

$$\mathcal{A}_k := \{i \in \{1, \dots, p\} \mid r_i - g_i^T x^k < \mu_i\}, \quad (4.28)$$

$$\mathcal{I}_k := \{i \in \{1, \dots, p\} \mid r_i - g_i^T x^k \geq \mu_i\} \quad (4.29)$$

### 3. Löse das Problem

$$f''(x^k)x + G^T \mu = f''(x^k)x^k - \nabla f(x^k) \quad (4.30)$$

$$g_i^T x = r_i \quad \text{für } i \in \mathcal{A}_k \quad (4.31)$$

$$\mu_i = 0 \quad \text{für } i \in \mathcal{I}_k \quad (4.32)$$

### 4. Falls

$$\nabla f(x) + \mu = 0 \quad (4.33)$$

$$\min(\mu, r - Gx) = 0 \quad (4.34)$$

$\Rightarrow \text{STOP}$

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

Wir werden nun das Verfahren genauer betrachten und zeigen, warum es als das halbglatte Newton-Verfahren interpretiert werden kann.

Sei  $k$  irgendeiner Iterationsschritt in den beiden Verfahren. Bei dem halbglatten Newton-Verfahren müssen wir dann folgendes Gleichungssystem lösen:

$$\begin{pmatrix} f''(x^k) & G^T \\ -\chi_{\mathcal{A}_k}(1)g_1^T & \chi_{\mathcal{I}_k}(1)e_1^T \\ \vdots & \vdots \\ -\chi_{\mathcal{A}_k}(p)g_p^T & \chi_{\mathcal{I}_k}(p)e_p^T \end{pmatrix} \begin{pmatrix} d_x \\ d_\mu \end{pmatrix} = - \begin{pmatrix} \nabla f(x^k) + G^T \mu^k \\ \chi_{\mathcal{I}_k}(1)\mu_1 + \chi_{\mathcal{A}_k}(1)(r_1 - g_1^T x^k) \\ \vdots \\ \chi_{\mathcal{I}_k}(p)\mu_p + \chi_{\mathcal{A}_k}(p)(r_p - g_p^T x^k) \end{pmatrix} \quad (4.35)$$

D. h.

$$f''(x^k)d_x + G^T d_\mu = -\nabla f(x^k) - G^T \mu^k \quad (4.36)$$

$$-\chi_{\mathcal{A}_k}(j)g_j^T d_x + \chi_{\mathcal{I}_k}(j)e_j^T d_\mu = -\chi_{\mathcal{I}_k}(j)\mu_j - \chi_{\mathcal{A}_k}(j)(r_j - g_j^T x^k) \quad \text{für } j = 1, \dots, p \quad (4.37)$$

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

Bei dem halbglatten Newton-Verfahren setzen wir dann  $x^{k+1} := x^k + d_x$  und  $\mu^{k+1} := \mu^k + d_\mu$ . Es gilt also  $d_x = x^{k+1} - x^k$  und  $d_\mu = \mu^{k+1} - \mu^k$ . D. h.

$$f''(x^k)(x^{k+1} - x^k) + G^T(\mu^{k+1} - \mu^k) = -\nabla f(x^k) - G^T \mu^k \quad (4.38)$$

$$\Rightarrow f''(x^k)x^{k+1} + G^T \mu^{k+1} = f''(x^k)x^k - \nabla f(x^k) \quad (4.39)$$

Das ist genau die Gleichung (4.30) in der Aktive-Mengen-Strategie.

Sei  $j \in \mathcal{A}_k$ . Die Gleichung (4.37) ist dann gleich

$$-g_j^T d_x = -(r_j - g_j^T x^k) \quad (4.40)$$

$$\Rightarrow -g_j^T (x^{k+1} - x^k) = -(r_j - g_j^T x^k) \quad (4.41)$$

$$\Rightarrow g_j^T x^{k+1} = r_j \quad (4.42)$$

Das ist genau die Gleichung (4.31) in der Aktive-Mengen-Strategie.

Sei nun  $j \in \mathcal{I}_k$ . Die Gleichung (4.37) ist dann gleich

$$e_j^T d_\mu = -\mu_j \quad (4.43)$$

$$\Rightarrow e_j^T (\mu^{k+1} - \mu^k) = -\mu_j \quad (4.44)$$

$$\Rightarrow \mu_j^{k+1} - \mu_j^k = -\mu_j \quad (4.45)$$

$$\Rightarrow \mu_j^{k+1} = 0 \quad (4.46)$$

Das ist genau die Gleichung (4.32) in der Aktive-Mengen-Strategie.

Somit sind die Iterationsschritte in den beiden Verfahren gleich. Und deswegen sind die beiden Verfahren äquivalent.

# Kapitel 5

## Der Vergleich

### 5.1 Testfunktionen

Wir betrachten hier einige Testfunktionen, die als Zielfunktion verschiedener restringierter Optimierungsaufgaben benutzt werden können.

**Testfunktion 1** *Quadratische Funktion in  $\mathbb{R}^n$*

$$f(x) := \|x - d\|^2 = \sum_{k=1}^n (x_k - d_k)^2$$

Die erste Testfunktion ist eine einfache mehrdimensionale Parabel mit Minimalstelle im Punkt  $d$  und Optimalwert 0.

**Testfunktion 2** *Exponentielle Funktion*

$$f(x) := e^{\|x\|^2} = e^{\sum_{k=1}^n x_k^2}$$

Die zweite Testfunktion ist eine exponentielle Funktion mit Minimalstelle im Ursprung und Optimalwert 1. Diese Funktion ist herausfordernd, weil ihre Funktionswerte sehr schnell groß werden können.

**Testfunktion 3** *Rosenbrock-Funktion (vgl. Beispiel 1.4.1 in [Alt11, S. 14])*

$$f(x) := 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

Die Rosenbrock-Funktion hat ein einziges Minimum an der Stelle  $x^* = (1, 1)^T$  mit  $f(x^*) = 0$ . Diese Funktion kann für manche Verfahren Schwierigkeiten anbieten, da ihr Minimum in einem schmalen „bananenförmig“ gekrümmten Tal liegt. Als Anfangspunkt wird normalerweise den Punkt  $x^0 = (-1, 1)^T$  genommen.

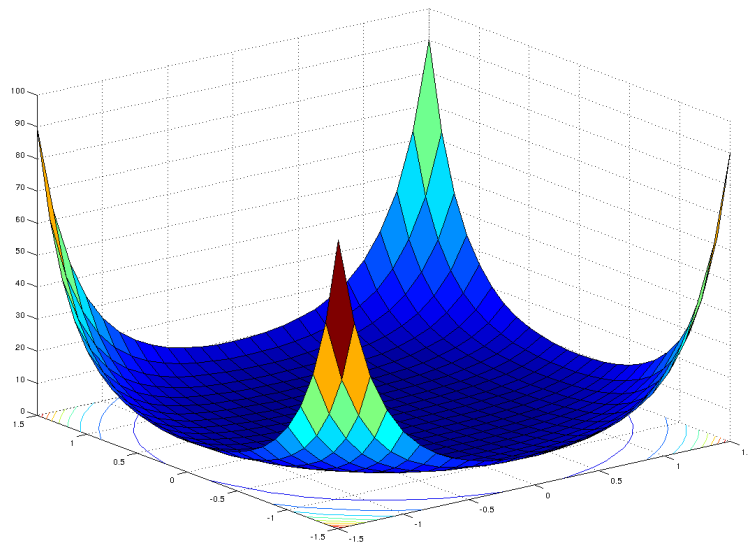


Abbildung 5.1: Exponentielle Funktion

**Testfunktion 4** *Himmelblau-Funktion* (vgl. Beispiel 1.4.2 in [Alt11, S. 14f])

$$f(x) := (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$$

Die Himmelblau-Funktion ist ein Polynom 4. Grades mit vier globale Minimalstellen. Ein Minimum ist z. B. an der Stelle  $x^* = (3, 2)^T$ . Der Optimalwert ist 0.

**Testfunktion 5** *Bazaraa-Shetty-Funktion* (vgl. Beispiel 1.4.3 in [Alt11, S. 15f])

$$f(x) := (x_1 - 2)^4 + (x_1 - 2x_2)^2$$

Die Bazaraa-Shetty-Funktion ist ein Polynom 4. Grades mit einem globalen Minimum in  $x^* = (2, 1)^T$  und Optimalwert 0.

**Testfunktion 6** (vgl. Beispiel 1.4.4 in [Alt11, S. 16])

$$f(x_1, \dots, x_5) := 2x_1^2 + 2x_2^2 + x_3^2 + x_4^2 + \frac{1}{2}x_5^2 - 4x_1 - 4x_2 - 2x_3 - 2x_4 - x_5 + 6\frac{1}{2}$$

Diese quadratische Funktion hat ein globales Minimum an der Stelle  $x^* = (1, 1, 1, 1, 1)^T$  mit  $f(x^*) = 0$ .



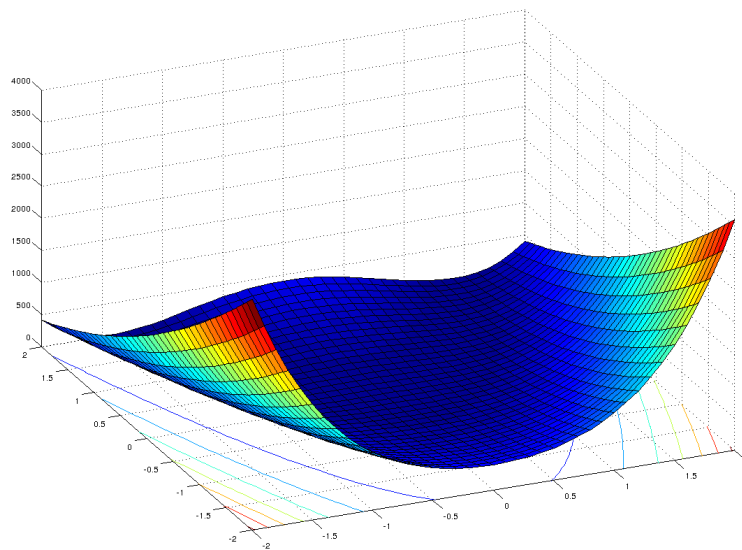


Abbildung 5.2: Rosenbrock-Funktion

**Testfunktion 7** *Dixon-Funktion* (vgl. Beispiel 1.4.5 in [Alt11, S. 16])

$$(1 - x_1)^2 + \sum_{k=1}^9 (x_k^2 - x_{k+1})^2 + (1 - x_{10})^2$$

Die Dixon-Funktion ist ein Polynom 4. Grades mit 10 Variablen. Das globale Minimum ist an der Stelle  $x^* = (1, \dots, 1)^T$  und hat den Optimalwert 0.

**Testfunktion 8** *Beale-Funktion* (vgl. Aufgabe 2.2 in [Alt11, S. 39])

$$f(x) := (1.5 - x_1(1 - x_2))^2 + (2.25 - x_1(1 - x_2^2))^2 + (2.625 - x_1(1 - x_2^3))^2$$

Das globale Minimum ist an der Stelle  $x^* = (3, 0.5)^T$ .

**Testfunktion 9** *Colville-Funktion*

$$f(x) := 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 90(x_4 - x_3^2)^2 + (1 - x_3)^2 \\ + 10.1((x_2 - 1)^2 + (x_4 - 1)^2) + 19.8(x_2 - 1)(x_4 - 1)$$

Die Colville-Funktion ist ein Polynom 4. Grades mit vier Variablen. Das globale Minimum ist in  $x^* = (1, 1, 1, 1)^T$  und Optimalwert 0.

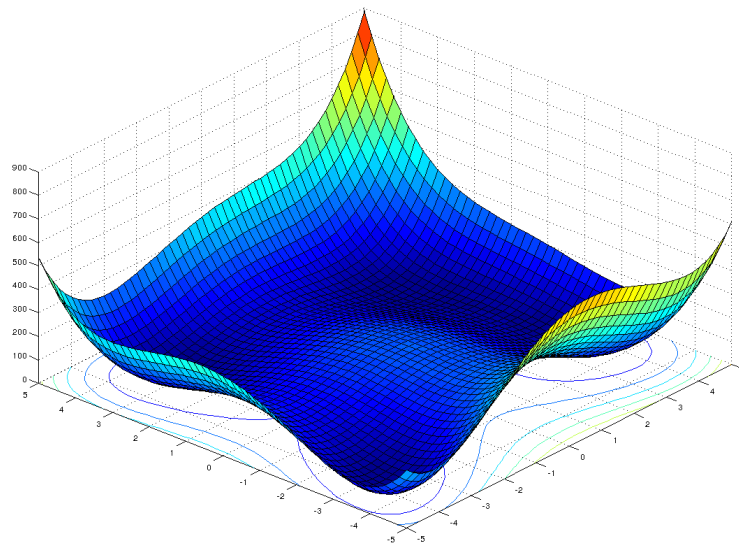


Abbildung 5.3: Himmelblau-Funktion

## 5.2 Testprobleme

Die folgende Tabelle zeigt die Ergebnisse der Optimierung der Testfunktionen mit Variablenbeschränkung, wobei die Minimalstellen immer noch in dem zulässigen Bereich bleiben.

Problem	SSN	SQP
	T (ms)	
Rosenbrock	16.72	25.62
Himmelblau	17.34	30.31
Bazaraa-Shetty	53.59	109.38
Colville	593.75	656.25

Tabelle 5.1: Vergleich

**Testproblem 1** (Lineare Regression, vgl. Beispiel 1.1.6 in [Alt11, S. 4f])  
 Gegeben seien  $m$ -Messwerte  $(\xi_i, \eta_i), i = 1, \dots, m$ . Gesucht ist eine Gerade

$$\eta(\xi) := x_1 \xi + x_2, \quad (5.1)$$

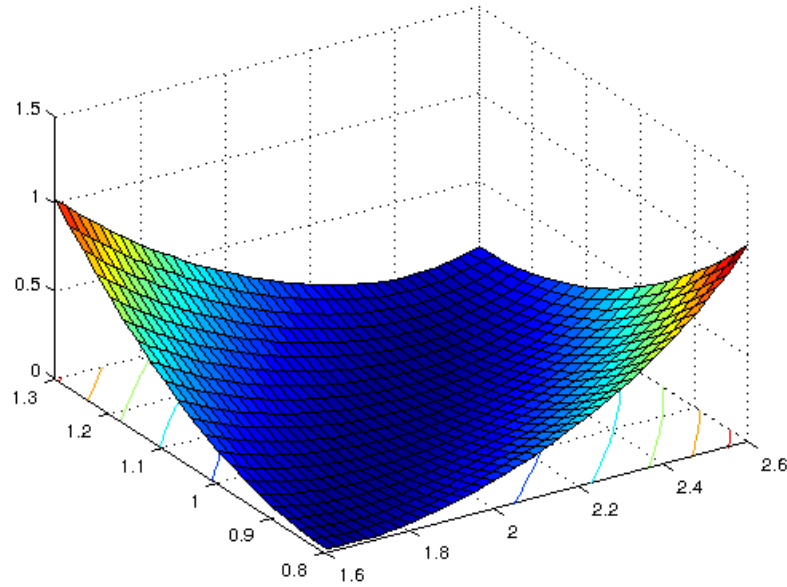


Abbildung 5.4: Bazaraa-Shetty-Funktion

die “optimal” zu den Messwerten passt. D. h. wir sollen den Parameter  $x = (x_1, x_2)^T \in \mathbb{R}^2$  so bestimmen, dass die Summe der Fehlerquadrate in den Messpunkten minimiert wird. Dazu definieren wir die Zielfunktion durch

$$f(x) := \sum_{i=1}^m (\eta(\xi_i) - \eta_i)^2 = \sum_{i=1}^m (x_1 \xi_i + x_2 - \eta_i)^2. \quad (5.2)$$

Es ist damit ein unrestringiertes Optimierungsproblem zu lösen.

**Testproblem 2** (Nichtlineare Regression, vgl. Abschnitt 2.3.2 in [Alt11, S. 30f])

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

$$\eta(\xi) := x_1 e^{\xi x_2} + x_3. \quad (5.3)$$

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

$\xi_i$	1	2	3	4	5	6	7	8	9	10
$\eta_i$	1	1.1	1.2	1.35	1.55	1.75	2.5	3	3.7	4.5

Tabelle 5.2: Gegebene Messwerte

$\xi_i$	1	2	3	4	5	6	7	8	9	10
$\eta_i$	$-\frac{1}{2}$	-2	-3	-3	$-2\frac{1}{2}$	-2	-1	1	3	$5\frac{1}{2}$

Tabelle 5.3: Gegebene Messwerte

**Testproblem 3** Löse folgendes, einfaches, 2-dimensionales Problem.

$$\min_{x \in \mathbb{R}^2} (x_1 - 1)^2 + (x_2 - 2.5)^2 \quad (5.4)$$

$$\text{Nb. } -x_1 + 2x_2 \leq 2 \quad (5.5)$$

$$x_1 + 2x_2 \leq 6 \quad (5.6)$$

$$x_1 - 2x_2 \leq 6 \quad (5.7)$$

$$0 \leq x_1, x_2 \quad (5.8)$$

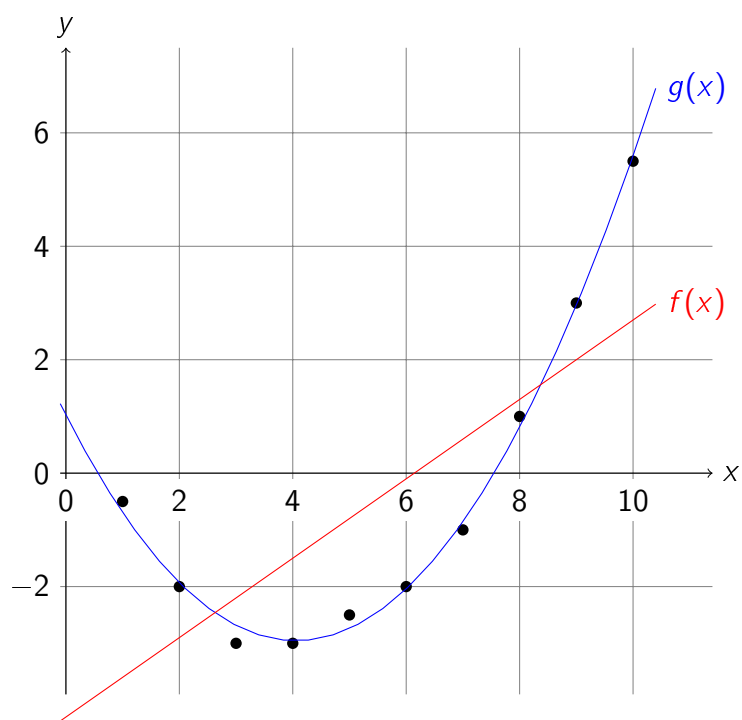


Abbildung 5.5:  $f(x) = 0.7x - 4.3$  und  $g(x) = 0.242(x - 4.056)^2 - 2.955$

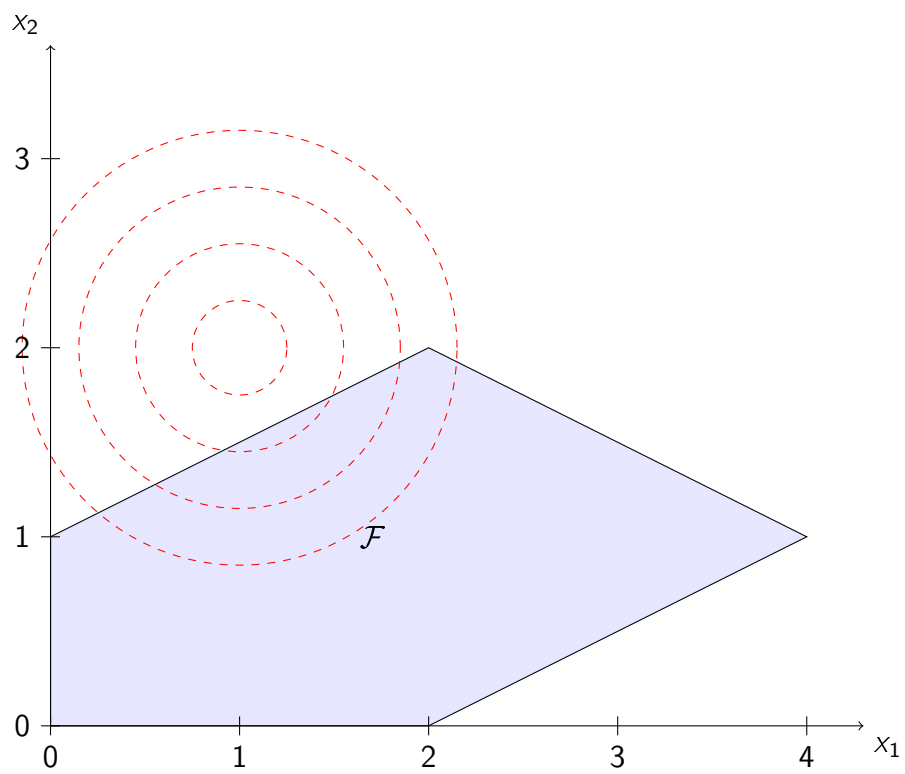


Abbildung 5.6: Restringiertes Problem

# Literaturverzeichnis

- [Alt11] Walter Alt. *Nichtlineare Optimierung: Eine Einführung in Theorie, Verfahren und Anwendungen*. Vieweg+Teubner, 2nd edition, 2011.
- [NW06] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 2nd edition, 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.