

Kapitel 1

Einführung

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

Das SQP-Verfahren gehört zu den bekanntesten Verfahren der nichtlinearen Optimierung. Es wurde schon seit den 60er Jahren entwickelt und wurde in vielen Optimierungsproblemen als Standardwerkzeug angewendet sowie weiterentwickelt. Das halbglatte Newton-Verfahren ist weniger bekannt als das SQP-Verfahren. Es basiert aber auf das bekannte Newton-Verfahren.

Bevor wir die beiden Verfahren näher betrachten, werden wir erst mal in diesem Kapitel die grundlegend Definitionen sowie die wichtigsten Ergebnisse betrachten.

1.1 Allgemeine Optimierungsprobleme

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

$$\min_{x \in \mathcal{F}} f(x) \tag{P}$$

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 definiert.

Ein einfaches Beispiel ist das Problem

$$\min_{x \in \mathbb{R}} (x - 1)^2.$$

Kapitel 1 Einführung

Falls $\mathcal{F} = D$ gilt, dann bezeichnet man das Optimierungsproblem als unrestringiert. Es besitzt also keine Nebenbedingungen. 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 1.2 (Globale und lokale Lösung) *Ein Punkt $x^* \in \mathcal{F}$ heißt globale Lösung des Problems (P) oder globales Minimum von f , 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 von f , 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 von f , wenn für eine Umgebung $U(x^*)$ 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 von f , wenn für eine Umgebung $U(x^*)$ von x^**

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

gilt. Eine Umgebung $U(x^)$ von x^* ist einfach 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.

Die Aufgabenstellung bei der nichtlinearen Optimierung kann man spezifischer wie folgt definieren:

Definition 1.3 (Nichtlineare Optimierungsprobleme)

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

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.

Kapitel 1 Einführung

D.h., die Menge \mathcal{F} sieht hier so aus $\mathcal{F} := \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}$.

Man kann dieses Problem ausführlicher schreiben, indem man die Funktionen g und h in mehreren skalaren Funktionen $g_1, \dots, g_p : \mathbb{R}^n \rightarrow \mathbb{R}$ und $h_1, \dots, h_m : \mathbb{R}^n \rightarrow \mathbb{R}$ zerlegen, 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

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

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.

1.2 Optimierungsprobleme ohne Restriktionen

Wir betrachten nun zuerst unrestringierte Optimierungsprobleme.

Definition 1.4 (Unrestringierte Optimierungsprobleme)

$$\min_{x \in \mathbb{R}^n} f(x). \tag{1.5}$$

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

Satz 1.1 (Notwendige Bedingung erster Ordnung) *Sei x^* eine lokale Lösung des Problems (1.5) und sei f einmal stetig differenzierbar in einer Umgebung von x^* , dann gilt*

$$\nabla f(x^*) = 0 \tag{1.6}$$

Kapitel 1 Einführung

Diese Bedingung gilt aber nicht nur für lokales Minimum sondern auch für lokales Maximum von f .

Definition 1.5 (Stationärer Punkt) f sei in x^* differenzierbar. x^* heißt ein stationärer Punkt von f , wenn x^* die notwendige Bedingung (1.6) 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.

Satz 1.2 (Notwendige Bedingung zweiter Ordnung) Sei x^* eine lokale Lösung des Problems (1.5) und sei f zweimal stetig differenzierbar in einer Umgebung von x^* , dann gilt (1.6) und

$$x^T f''(x^*)x \geq 0 \quad \forall x \in \mathbb{R}^n. \quad (1.7)$$

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

Durch diese notwendige Bedingung kann man zwischen einem lokalen Minimum und einem lokalen Maximum unterscheiden.

Satz 1.3 (Hinreichende Bedingung zweiter Ordnung) Sei f zweimal stetig differenzierbar in einer Umgebung von x^* . Die notwendige Bedingung (1.6) sei erfüllt und $f''(x^*)$ sei positiv definit, d.h.

$$x^T f''(x^*)x > 0 \quad \forall x \in \mathbb{R}^n. \quad (1.8)$$

Dann ist x^* eine strikte Lösung des Problems (1.5).

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 von der Abstiegsrichtung.

Definition 1.6 (Abstiegsrichtung) Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ differenzierbar in x . Ein Vektor $d \in \mathbb{R}^n \setminus \{0\}$ heißt Abstiegsrichtung von f in x , wenn

$$\nabla f(x)^T d < 0 \quad (1.9)$$

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 .

Kapitel 1 Einführung

Satz 1.4 Seien $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ differenzierbar in x und d eine Abstiegsrichtung von f in x . Dann gibt es ein $\hat{\sigma} > 0$ mit

$$f(x + \sigma d) < f(x) \quad \forall \sigma \in]0, \hat{\sigma}[. \quad (1.10)$$

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 σ_k , sodass für $x^{k+1} := x^k + \sigma_k d^k$

$$f(x^{k+1}) < f(x^k) \quad (1.11)$$

gilt.

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

Verfahren 1.1 (Gradientenverfahren)

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

$$f(x^k + \sigma_k d^k) < f(x^k + \sigma d^k) \quad \forall \sigma \geq 0. \quad (1.12)$$

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

Eine andere Möglichkeit ist, dass man das Newton-Verfahren verwendet, um die Lösung der Gleichung (1.6) zu finden. Man muss dabei voraussetzen, dass die Funktion f zweimal differenzierbar sei.

Verfahren 1.2 (Newton-Verfahren)

1. Wähle einen Startpunkt x^0 und ein Abbruchkriterium $\epsilon > 0$. Setze $k := 0$.
2. Ist $\|\nabla f(x^k)\| < \epsilon \Rightarrow STOP$.
3. Berechne die Lösung d des linearen Gleichungssystems

$$f''(x^k)d = -\nabla f(x^k). \quad (1.13)$$

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 können wir die Funktion f wie folgt approximieren:

$$f(x) \approx f(x^k) + \nabla f(x^k)^T(x - x^k) + \frac{1}{2}(x - x^k)^T f''(x^k)(x - x^k). \quad (1.14)$$

Die Ableitung der rechten Seite ist

$$f''(x^k)x + \nabla f(x^k) - f''(x^k)x^k. \quad (1.15)$$

Setzen wir diese gleich null, dann bekommen wir

$$f''(x^k)(x - x^k) = -\nabla f(x^k) \quad (1.16)$$

$$x - x^k = -[f''(x^k)]^{-1}\nabla f(x^k) \quad (1.17)$$

$$x = x^k - [f''(x^k)]^{-1}\nabla f(x^k). \quad (1.18)$$

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

$$\min_{x \in \mathbb{R}^n} f(x^k) + \nabla f(x^k)^T(x - x^k) + \frac{1}{2}(x - x^k)^T f''(x^k)(x - x^k). \quad (1.19)$$

Diese Idee wird auch bei dem SQP-Verfahren benutzt.

Man kann außerdem auch Schrittweitensteuerung durchführen, dann bekommt man ein Abstiegsverfahren, welches man als das gedämpfte Newton-Verfahren bezeichnet.

1.3 Optimierungsprobleme mit linearen Restriktionen

1.3.1 Optimierungsprobleme mit linearen Gleichungsnebenbedingungen

Definition 1.7 Optimierungsproblem mit linearen Gleichungsnebenbedingungen

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

Nb. $Ax = b$

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

Kapitel 1 Einführung

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

Satz 1.5 (Notwendige Bedingung erster Ordnung) *Sei x^* eine lokale Lösung des Problems (PLG) und f sein in x^* differenzierbar. Dann gibt es ein $\lambda \in \mathbb{R}^m$ mit*

$$\nabla f(x^*) + A^T \lambda = 0. \quad (1.20)$$

Hat A einen vollen Rang, dann ist λ eindeutig zu bestimmen.

Diese Bedingung heißt die Multiplikatoren Regel von Lagrange. Man bezeichnet λ als die Lagrange-Multiplikator.

Satz 1.6 (Hinreichende Bedingung zweiter Ordnung) *Sei f in x^* zweimal stetig differenzierbar. Die notwendige Bedingung erster Ordnung sei erfüllt. Es gäbe eine Konstante $\alpha > 0$ mit*

$$d^T f''(x) d \geq \alpha \|d\|^2 \quad \forall d \in \text{Kern } A. \quad (1.21)$$

Dann ist x^ eine strikte Lösung des linearen restringierten Problems.*

Definition 1.8 (Nullraum-Matrix) *Eine $(n \times l)$ -Matrix Z heißt Nullraum-Matrix von A , wenn für $d \in \mathbb{R}^n$ gilt*

$$d \in \text{Kern } A \quad \Leftrightarrow \quad d = Zz \text{ für ein } z \in \mathbb{R}^l. \quad (1.22)$$

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

Sei w eine Lösung von der Gleichung $Ax = b$. Man kann nun für \mathcal{F} so schreiben:

$$\mathcal{F} = w + \text{Kern } A = w + \text{Im } Z = w + \{Zz \mid z \in \mathbb{R}^l\}. \quad (1.23)$$

Das Problem (PLG) ist dann äquivalent zu

$$\min_{z \in \mathbb{R}^l} F(z) := f(w + Zz). \quad (1.24)$$

Dieses Problem hat keine Nebenbedingung mehr, also unrestringiert. Man kann also die Verfahren für unrestringierte Probleme anwenden. Wir werden aber nachher sehen, wie man dieses Problem effektiv lösen kann, wenn die Zielfunktion quadratisch ist.

1.3.2 Optimierungsprobleme mit linearen Ungleichungsnebenbedingungen

Definition 1.9 (Optimierungsproblem mit linearen Ungleichungsnebenbedingungen)

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

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 $g_j \in \mathbb{R}^n, j = 1, \dots, p$, die Vektoren in der Matrix G , so dass

$$G = \begin{pmatrix} (g_1)^T \\ \vdots \\ (g_p)^T \end{pmatrix}. \quad (1.25)$$

Satz 1.7 (Karush-Kuhn-Tucker-Satz) *Sei x^* lokale Lösung des Problems (PLN) und f sei in x^* differenzierbar. Dann existieren die Vektoren $\lambda \in \mathbb{R}^m$ und $\mu \in \mathbb{R}^p$ zu x^* mit*

$$\nabla f(x^*) + A^T \lambda + G^T \mu = 0 \quad (1.26)$$

$$\mu \geq 0, \quad \mu_j (\langle g^j, x^* \rangle - r_j) = 0, \quad j = 1, \dots, p. \quad (1.27)$$

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

Definition 1.10 *Sei $x \in \mathcal{F}$. Wir bezeichnen mit*

$$J(x) := \{1 \leq j \leq p \mid \langle g_j, x \rangle = r_j\} \quad (1.28)$$

die Indexmenge der in x aktiven Ungleichungsrestriktionen.

Satz 1.8 (Hinreichende Optimalitätsbedingungen) *Sei f in $x^* \in \mathcal{F}$ zweimal stetig differenzierbar. Die notwendige Bedingung von Satz 1.7 sei erfüllt und es gelte mit $\alpha > 0$*

$$d^T f''(x^*) d \geq \alpha \|d\|^2 \quad \forall d \in \mathbb{R}^n : \begin{cases} Ad = 0, \\ \langle g_j, d \rangle \leq 0 & \text{für } j \in J(x^*) \text{ mit } \mu_j = 0, \\ \langle g_j, d \rangle = 0 & \text{für } j \in J(x^*) \text{ mit } \mu_j > 0. \end{cases} \quad (1.29)$$

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

Kapitel 1 Einführung

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

Definition 1.11 (Optimierungsproblem mit Variablenbeschränkungen)

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

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

Mit $G := \begin{pmatrix} -I \\ I \end{pmatrix}$ und $r := \begin{pmatrix} -v \\ w \end{pmatrix}$ hat \mathcal{F} die Form $\mathcal{F} := \{x \in \mathbb{R}^n \mid Gx \leq r\}$.

1.4 Optimierungsprobleme mit nichtlinearen Restriktionen

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

$$\begin{aligned} \min f(x) \\ \text{Nb. } g(x) \leq 0 \\ h(x) = 0 \end{aligned}$$

Definition 1.12 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. x erfüllt die Mangasarian-Fromowitz-Regularitätsbedingung, wenn die Gradienten $\nabla h_j(x)$, $j = 1, \dots, m$, linear unabhängig sind und ein Vektor $d \in \mathbb{R}^n$ existiert mit

$$\nabla g_i(x)^T d < 0, \quad i \in \mathcal{I}(x) \quad \text{und} \quad \nabla h_j(x)^T d = 0, \quad j = 1, \dots, m. \tag{1.30}$$

Satz 1.9 Sei x^* eine lokale Lösung des Problems (PN) und x erfülle die Mangasarian-Fromowitz-Regularitätsbedingung. Dann existieren die Vektoren $\lambda \in \mathbb{R}^m$ und $\mu \in \mathbb{R}^p$, so dass

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) = 0 \tag{1.31}$$

$$h(x^*) = 0 \tag{1.32}$$

$$\mu \geq 0, \quad \mu^T g(x^*) = 0, \quad g(x^*) \leq 0 \tag{1.33}$$

Kapitel 2

SQP-Verfahren

2.1 Einführung

Das SQP-Verfahren ist ein wichtiges Verfahren, um restringierte nichtlineare Probleme zu lösen. SQP ist eine Abkürzung für sequentielle quadratische Programmierung (englisch: Sequential Quadratic Programming). Die Idee ist nämlich, dass Teilprobleme in Form quadratischer Optimierungsprobleme iterativ formuliert und gelöst werden.

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \frac{1}{2} d^T f''(x^k) d + \nabla f(x^k)^T d \\ \text{Nb. } & Ad = 0, \\ & Gx^k + Gd \leq r \end{aligned} \tag{2.1}$$

Verfahren 2.1 (SQP-Verfahren)

1. Berechne einen zulässigen Startpunkt x^0 und setze $k := 0$.
2. Berechne die Lösung d des Problems (2.1). Setze $d^k := d$.
3. Setze $x^{k+1} := x^k + d^k$ und $k := k + 1$. \Rightarrow Gehe zu Schritt 2

Definition 2.1

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2} \langle Qx, x \rangle + \langle q, x \rangle \\ \text{Nb. } & Ax = b \\ & Gx \leq r \end{aligned} \tag{2.2}$$