

Kapitel 1

Grundlagen

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 grundlegenden Definitionen sowie die wichtigsten Ergebnisse betrachten.

1.1 Allgemeine Optimierungsprobleme

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

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

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 definiert.

Ein einfaches Beispiel ist das Problem

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

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, 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 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, 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 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)

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

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

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

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

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)

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

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 (PU) und sei f einmal stetig differenzierbar in einer Umgebung von x^* , dann gilt

$$\nabla f(x^*) = 0. \tag{1.5}$$

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

Definition 1.5 (Stationärer Punkt) *Die Funktion f sei in x^* differenzierbar. Der Punkt x^* heißt ein stationärer Punkt von f , wenn x^* die notwendige Bedingung (1.5) 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 1.2 (Notwendige Bedingung zweiter Ordnung) *Sei x^* eine lokale Lösung des Problems (PU) und sei f zweimal stetig differenzierbar in einer Umgebung von x^* , dann gilt (1.5) und*

$$x^T f''(x^*)x \geq 0 \quad \forall x \in \mathbb{R}^n. \quad (1.6)$$

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

Satz 1.3 (Hinreichende Bedingung zweiter Ordnung) *Sei f zweimal stetig differenzierbar in einer Umgebung von x^* . Die notwendige Bedingung (1.5) sei erfüllt und $f''(x^*)$ sei positiv definit, d. h.*

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

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

$$\nabla f(x)^T d < 0 \quad (1.8)$$

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 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

$$f(x + \sigma d) < f(x) \quad \forall \sigma \in]0, \hat{\sigma}[. \quad (1.9)$$

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

$$f(x^{k+1}) < f(x^k) \quad (1.10)$$

gilt.

1.2.1 Gradientenverfahren

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 $\|\nabla f(x^k)\| < \epsilon \Rightarrow \text{STOP}$.
3. Setze $d^k := -\nabla f(x^k)$.
4. Bestimme σ_k so, dass

$$f(x^k + \sigma_k d^k) < f(x^k + \sigma d^k) \quad \forall \sigma \geq 0. \quad (1.11)$$

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

Um die Schrittweite σ_k zu bestimmen, kann man das Armijo-Verfahren oder das Powell-Verfahren verwenden.

1.2.2 Newton-Verfahren

Ein bekanntes Verfahren der numerischen Mathematik, um ein nichtlineares Gleichungssystem zu lösen, ist das Newton-Verfahren. Für unsere unrestringierte Optimierungsprobleme können wir das Newton-Verfahren verwenden, um die Lösung des nichtlinearen Gleichungssystems (1.5), $\nabla f(x) = 0$, zu finden. Wir müssen 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 $\|\nabla f(x^k)\| < \epsilon \Rightarrow \text{STOP}$.
3. Berechne die Lösung d des linearen Gleichungssystems

$$f''(x^k)d = -\nabla f(x^k). \quad (1.12)$$

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:

$$f(x) \approx f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2}(x - x^k)^T f''(x^k)(x - x^k). \quad (1.13)$$

Die Ableitung der rechten Seite ist

$$\nabla f(x^k) + f''(x^k)x - f''(x^k)x^k. \quad (1.14)$$

Setzen wir diese gleich null, dann bekommen wir

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

$$x - x^k = -[f''(x^k)]^{-1} \nabla f(x^k) \quad (1.16)$$

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

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

$$\min_{x \in \mathbb{R}^n} f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2}(x - x^k)^T f''(x^k)(x - x^k), \quad (1.18)$$

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.

1.3 Optimierungsprobleme mit linearen Restriktionen

1.3.1 Optimierungsprobleme mit linearen Gleichungsnebenbedingungen

Definition 1.7 (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: $\mathcal{F} = \{x \in \mathbb{R}^n \mid Ax = b\}$.

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

$$\nabla f(x^*) + A^T \lambda = 0. \quad (1.19)$$

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)

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

$$d^T f''(x) d \geq \alpha \|d\|^2 \quad \forall d \in \text{Kern } A. \quad (1.20)$$

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

Eine wichtige Grundlage zum Lösen des Problems (PLG) ist die Definition der Nullraum-Matrix.

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

$$d \in \text{Kern } A \Leftrightarrow d = Zz \text{ für ein } z \in \mathbb{R}^l. \quad (1.21)$$

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:

$$\mathcal{F} = w + \text{Kern } A = w + \text{Im } Z = w + \{Zz | z \in \mathbb{R}^l\}. \quad (1.22)$$

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

$$\min_{z \in \mathbb{R}^l} F(z) := f(w + Zz). \quad (1.23)$$

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

1.3.2 Optimierungsprobleme mit linearen Ungleichungsnebenbedingungen

Definition 1.9 (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: $\mathcal{F} = \{x \in \mathbb{R}^n | Ax = b, Gx \leq r\}$.

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

$$A = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix} \quad \text{und} \quad G = \begin{pmatrix} g_1^T \\ \vdots \\ g_p^T \end{pmatrix}. \quad (1.24)$$

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) ausführlicher schreiben.

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

Die Notwendige Bedingung für eine lokale Lösung des Problems (PLU) ergibt sich aus dem folgenden Satz.

Satz 1.7 (Karush-Kuhn-Tucker-Satz) *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*

$$\nabla f(x^*) + A^T \lambda + G^T \mu = 0 \quad (1.25)$$

$$\mu_j (\langle g_j, x^* \rangle - r_j) = 0 \quad \forall j = 1, \dots, p \quad (1.26)$$

$$\mu \geq 0 \quad (1.27)$$

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

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

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

die Indexmenge der in x aktiven Ungleichungsrestriktionen.

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

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

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

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

Definition 1.10 (Optimierungsprobleme mit Variablenbeschränkungen)

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

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

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

1.4 Optimierungsprobleme mit nichtlinearen Restriktionen

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

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

Definition 1.11 (Mangasarian-Fromowitz-Regularitätsbedingung)

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 $\nabla h_j(x)$, $j = 1, \dots, m$, linear unabhängig sind und ein Vektor $d \in \mathbb{R}^n$ existiert mit

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

Satz 1.9 (Notwendige Bedingung) 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

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

$$h(x^*) = 0 \quad (1.32)$$

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

Kapitel 2

SQP-Verfahren

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 Unterkapitel 1.2.2 über das Newton-Verfahren gesehen, dass dieses Verfahren in jedem Iterationschritt 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.

$$\min_{x \in \mathcal{F}} \frac{1}{2}(x - x^k)^T f''(x^k)(x - x^k) + \nabla f(x^k)^T (x - x^k). \quad (2.2)$$

2.2 Formulierung

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

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

¹englisch: Sequential Quadratic Programming

Das Teilproblem nach (2.2) sieht dann so aus:

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

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

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

$$\text{Nb. } Ax^k + Ad = b \quad (2.5)$$

$$Gx^k + Gd \leq r \quad (2.6)$$

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 (2.5) die Gleichung $Ad = 0$ schreiben. Nun sind wir bereit, das SQP-Verfahren zu formulieren.

Verfahren 2.1 (SQP-Verfahren)

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

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2} d^T f''(x^k) d + \nabla f(x^k)^T d \\ \text{Nb. } & Ad = 0 \\ & Gx^k + Gd \leq r \end{aligned} \quad (2.7)$$

Setze $d^k := d$.

3. Ist $\|d^k\| < \epsilon \Rightarrow \text{STOP}$.
4. Setze $x^{k+1} := x^k + d^k$ und $k := k + 1$. \Rightarrow Gehe zu Schritt 2.

2.3 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 2.1 (Quadratische Optimierungsprobleme mit linearen Restriktionen)

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2} \langle Qx, x \rangle + \langle q, x \rangle & (\text{QLU}) \\ \text{Nb.} \quad & \langle a_i, x \rangle = b_i \text{ für } i = 1, \dots, m \\ & \langle g_j, x \rangle \leq r_j \text{ für } j = 1, \dots, p \end{aligned}$$

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

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \frac{1}{2} \langle Q(x^k + d), x^k + d \rangle + \langle q, x^k + d \rangle & (2.8) \\ \text{Nb.} \quad & \langle a_i, x^k + d \rangle = b_i \text{ für } i = 1, \dots, m \\ & \langle g_j, x^k + d \rangle \leq r_j \text{ für } j = 1, \dots, p \end{aligned}$$

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 mit nur linearen Gleichungsnebenbedingungen:

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

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 μ .

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. Definieren Schrittweite durch

$$\tau_k := \begin{cases} \min \left\{ \frac{r_j - g_j^T x^k}{g_j^T d^k} \mid j \in I_k \right\}, & \text{falls } I_k \neq \emptyset \\ \infty, & \text{sonst.} \end{cases} \quad (2.9)$$

und

$$\sigma_k := \min\{1, \tau_k\}. \quad (2.10)$$

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

Verfahren 2.2 (Aktive-Mengen-Methode für (QLU))

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 σ_k .

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

2.4 Nullraum-Verfahren

Definition 2.2 (Quadratische Optimierungsprobleme mit linearen Gleichungsnebenbedingungen)

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

Nb. $Ax = b$

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 2.3 (Nullraum-Verfahren)

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

$$HA^T = \begin{pmatrix} R \\ 0 \end{pmatrix}. \quad (2.11)$$

2. Berechne

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} := -Hq \quad \text{und} \quad B := HQH^T = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \quad (2.12)$$

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

$$R^T u_1 = b \quad (2.13)$$

und den Vektor u_2 als Lösung der Gleichung

$$B_{22}u_2 = h_2 - B_{21}u_1 \quad (2.14)$$

4. Berechne x^*

$$x^* := H^T \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \quad (2.15)$$

5. Der Multiplikator λ^* ist die Lösung der Gleichung

$$R\lambda^* = h_1 - B_{11}u_1 - B_{12}u_2 \quad (2.16)$$

2.5 Lagrange-Newton-SQP-Verfahren

Nun kommen wir zu den allgemeinen nichtlinearen Optimierungsproblemen.

Verfahren 2.4 (Lagrange-Newton-Verfahren)

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

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

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} d^T L_{xx}(x^k, \lambda^k, \mu^k) d + \nabla f(x^k)^T d \quad (2.17)$$

$$\text{Nb. } h(x^k) + h'(x^k)d = 0 \quad (2.18)$$

$$g(x^k) + g'(x^k)d \leq 0 \quad (2.19)$$

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

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

Kapitel 3

Das halbglatte Newton-Verfahren

3.1 Grundlagen

Seien X, Z reelle Banachräume und sei $D \subset X$ offen.

Definition 3.1 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

$$\lim_{|h| \rightarrow 0} \frac{|F(x+h) - F(x) - G(x+h)h|_Z}{|h|_X} = 0 \quad (3.1)$$

Die Familie

$$\{G(u) \mid u \in U(x)\} \quad (3.2)$$

heißt *N-Ableitung* von F in x .

Definition 3.2 Eine Funktion $F : D \subset X \rightarrow Z$ heißt *halbglatte*¹ in $x \in D$, falls F Newton-differenzierbar in x ist und

$$\lim_{t \rightarrow 0^+} G(x + th)h \quad (3.3)$$

einheitlich in $|h| = 1$ existiert.

Definition 3.3 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

$$\partial_B F(x) := \left\{ J \mid J = \lim_{x_i \rightarrow x, x_i \in D_F} \nabla F(x_i) \right\} \quad (3.4)$$

und die allgemeine Ableitung in x

$$\partial F(x) := \text{co } \partial_B F(x), \quad (3.5)$$

wobei co für die konvexe Hülle steht.

¹englisch: semismooth

Definition 3.4 $F : \mathbb{R}^m \rightarrow \mathbb{R}^n$ heißt *halbglatt* in $x \in \mathbb{R}^m$, falls F lokal Lipschitz-stetig in x ist und

$$\lim_{V \in \partial F(x+th'), h' \rightarrow h, t \rightarrow 0^+} Vh' \quad (3.6)$$

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

Satz 3.1 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

$$x^{k+1} := x^k - G(x^k)^{-1}F(x^k) \quad (3.7)$$

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

Verfahren 3.1 (Das halbglatte Newton-Verfahren)

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

$$V_k d = -\nabla f(x^k). \quad (3.8)$$

Setze $d^k := d$.

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

Das Aktive-Menge-Verfahren kann als das halbglatte Newton-Verfahren interpretiert werden.

3.2 Active-Set-Strategie

Wir betrachten in diesem Kapitel das Problem

Definition 3.5

$$\min_{x \in \mathbb{R}^n} f(x) + \frac{\lambda}{2}|x|^2 \quad (3.9)$$

$$\text{Nb. } a \leq x \leq b \quad (3.10)$$

Sei x^k ein zulässiger Punkt des Problems. Wir approximieren die Funktion f in der Umgebung von x^k .

$$f(x) \approx f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2} (x - x^k)^T f''(x^k) (x - x^k). \quad (3.11)$$

Der Gradient von f ist

$$\nabla f(x) \approx \nabla f(x^k) + f''(x^k)x - f''(x^k)x^k. \quad (3.12)$$

Die notwendige Bedingung für x^{k+1} wäre

$$\nabla f(x^{k+1}) + \lambda x^{k+1} + \mu^{k+1} = 0 \quad (3.13)$$

Ersetzen wir $\nabla f(x^{k+1})$ mit seiner Approximation, bekommen wir die Gleichung

$$\nabla f(x^k) + f''(x^k)x^{k+1} - f''(x^k)x^k + \lambda x^{k+1} + \mu^{k+1} = 0 \quad (3.14)$$

$$f''(x^k)x^{k+1} + \lambda x^{k+1} + \mu^{k+1} = f''(x^k)x^k - \nabla f(x^k) \quad (3.15)$$

Verfahren 3.2 (Aktive-Menge-Strategie)

1. $k := 0, x^k := x^0$

2. *Bestimme*

$$\mathcal{A}_+^k := \{i \mid x_i^k > b_i\} \quad (3.16)$$

$$\mathcal{A}_-^k := \{i \mid x_i^k < a_i\} \quad (3.17)$$

$$\mathcal{I}^k := \{i \mid a_i \leq x_i^k \leq b_i\} \quad (3.18)$$

3. *Löse das Problem*

$$f''(x^k)x^{k+1} + \lambda x^{k+1} + \mu^{k+1} = f''(x^k)x^k - \nabla f(x^k) \quad (3.19)$$

$$x_i^{k+1} = b_i \quad \text{für } i \in \mathcal{A}_+^k \quad (3.20)$$

$$x_i^{k+1} = a_i \quad \text{für } i \in \mathcal{A}_-^k \quad (3.21)$$

$$\mu^{k+1} = 0 \quad \text{für } i \in \mathcal{I}^k \quad (3.22)$$

4. *Falls*

$$0 = \nabla f(x^{k+1}) + \lambda x^{k+1} + \mu^{k+1} \quad (3.23)$$

$$\mu^{k+1} = \max(0, \mu^{k+1} + \lambda(x^{k+1} - b)) + \min(0, \mu^{k+1} + \lambda(x^{k+1} - a)) \quad (3.24)$$

$\Rightarrow \text{STOP}$

5. $x^k := x^{k+1}, k := k + 1 \Rightarrow 2.$