NICHTLINEARE OPTIMIERUNG

Vorlesungsskript, Wintersemester 2025/26

Christian Clason

Stand vom 5. August 2026

Institut für Mathematik und wissenschaftliches Rechnen
Universität Graz

Inhaltsverzeichnis

ÜBERBLICK

Die mathematische Optimierung beschäftigt sich mit der Aufgabe, Minima bzw. Maxima von Funktionen zu bestimmen. Konkret seien eine (nicht notwendigerweise echte) Teilmenge $X \subset \mathbb{R}^n$ und eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ gegeben. Gesucht ist ein $\bar{x} \in X$ mit

$$f(\bar{x}) \leq f(x) \quad \text{für alle } x \in X,$$

geschrieben

$$f(\bar{x}) = \min_{x \in X} f(x).$$

Die Fragen, die wir uns dabei stellen müssen, sind:

1. Hat dieses Problem eine Lösung?
2. Gibt es eine intrinsische Charakterisierung von $\bar{x}$, d. h. ohne Vergleich mit allen anderen $x \in X$?
3. Wie kann dieses $\bar{x}$ (effizient) berechnet werden?

Aus der Vielzahl der möglichen Beispiele sollen nur kurz folgende erwähnt werden:

(i) In *Transport- und Produktionsproblemen* sollen Kosten für Transport minimiert bzw. Gewinn aus Produktion maximiert werden. Dabei beschreibt $x \in \mathbb{R}^n$ die Menge der zu transportierenden bzw. produzierenden verschiedenen Güter und $f(x)$ die dafür nötigen Kosten bzw. aus dem Verkauf erzielten Gewinne. Die Nebenbedingung $x \in X$ beschreibt dabei, dass ein Mindestbedarf gedeckt werden muss bzw. nur endlich viele Rohstoffe zur Produktion zur Verfügung stehen.

(ii) In *inversen Problemen* sucht man einen Parameter u (zum Beispiel Röntgenabsorption von Gewebe in der Computertomographie), hat aber nur eine (gestörte) Messung y^δ zur Verfügung. Ist ein Modell bekannt, das für gegebenen Parameter u die entsprechende Messung $y = Kx$ liefert, so kann man den unbekannten Parameter näherungsweise rekonstruieren, indem man das Problem

$$\min_{x \in X} \|Kx - y^\delta\|^2 + \alpha \|x\|^2$$

für geeignet gewählte Normen und $\alpha > 0$ löst. Die Menge X kann dabei bekannte Einschränkungen an den Parameter (z. B. Nichtnegativität) beschreiben.

(iii) In der *optimalen Steuerung* ist man zum Beispiel daran interessiert, ein Auto oder eine Raumsonde möglichst effizient von einem Punkt y_0 zu einem anderen Punkt y_1 zu steuern. Beschreibt $y(t) \in \mathbb{R}^3$ die Position zum Zeitpunkt $t \in [0, T]$, so gehorcht $y(t)$ der Differenzialgleichung

$$\begin{cases} y'(t) = f(t, y(t), u(t)), \\ y(0) = y_0, \end{cases} \tag{1}$$

wobei $u(t)$ die Rolle der Steuerung spielt. Will man dabei die Energie minimieren, führt das auf das Problem

$$\min_{(y,u)\in X} |y(T) - y_1|^2 + \alpha \int_0^T |u(t)|^2 \quad \text{mit} \quad X = \{(y, u) \mid (1) \text{ ist erfüllt}\}.$$

In dieser Vorlesung behandeln wir *nichtlineare Optimierungsprobleme*, in denen zunächst $f : \mathbb{R}^n \to \mathbb{R}$ differenzierbar ist und U durch (differenzierbare) Gleichungen und Ungleichungen beschrieben werden kann. In diesem Fall ist die Antwort auf die Frage nach der Existenz relativ einfach zu beantworten; uns werden daher vor allem die beiden restlichen Fragen beschäftigen. Dabei wird das Konzept der *Abstiegsrichtung* in beiden Fällen fundamental sein: Grob gesprochen befinden wir uns in einem Minimum, falls keine Abstiegsrichtung existiert; ansonsten wählen wir eine und folgen ihr einen Schritt weit. Die wesentliche Schwierigkeit ist dabei die (oft komplizierte) Rolle der Nebenbedingung, da ein Minimierer in der Regel auf ihrem Rand liegen wird.

Dieses Skriptum basiert vor allem auf den folgenden Werken:

[i] W. Alt (2011), *Nichtlineare Optimierung. Eine Einführung in Theorie, Verfahren und Anwendungen*, 2. Aufl., Vieweg+Teubner, Wiesbaden

[ii] C. Geiger & C. Kanzow (2002), *Theorie und Numerik restringierter Optimierungsaufgaben*, Springer, Berlin, DOI: 10.1007/978-3-642-56004-0

[iii] A. Ruszczyński (2006), *Nonlinear Optimization*, Princeton University Press, Princeton, NJ

[iv] M. Ulbrich & S. Ulbrich (2012), *Nichtlineare Optimierung*, Birkhäuser, Basel, DOI: 10.1007/978-3-0346-0654-7

[v] S. J. Wright & B. Recht (2022), *Optimization for Data Analysis*, Cambridge: Cambridge University Press, DOI: 10.1017/9781009004282

Teil I

GRUNDLAGEN DER OPTIMIERUNG

1 THEORETISCHE GRUNDLAGEN

In diesem Kapitel fassen wir die wesentlichen Begriffe und Resultate aus der Theorie der nichtlinearen Optimierung ohne Nebenbedingungen zusammen.

1.1 ELEMENTARE DEFINITIONEN

Wir beginnen mit einigen elementaren Definitionen. Sei im folgenden stets $X \subset \mathbb{R}^n$ eine (nicht notwendigerweise echte) Teilmenge und $f : X \to \mathbb{R}$. Gesucht ist ein $\bar{x} \in X$ mit

$$f(\bar{x}) \leq f(x) \quad \text{für alle } x \in X,$$

geschrieben

$$f(\bar{x}) = \min_{x \in X} f(x).$$

Man nennt X *zulässige Menge* und einen Punkt $x \in X$ *zulässigen Punkt*; die Forderung $\bar{x} \in X$ wird *Nebenbedingung* genannt. Ist $X = \mathbb{R}^n$, so spricht man auch von *unrestringierter Optimierung* (d. h. *ohne Nebenbedingungen*), ansonsten von *restringierter Optimierung* (d. h. *mit Nebenbedingungen*). Oft wird f als *Zielfunktion* bezeichnet. Der optimale Wert $f(\bar{x})$ wird als *Minimum* bezeichnet, $\bar{x}$ selber als *Minimierer*, geschrieben $\bar{x} = \arg\min_{x \in X} f(x)$. Analog spricht man von *Maximum* und *Maximierer*, wenn $f(\bar{x}) \geq f(x)$ für alle $x \in X$ ist. Da gilt

$$\max_{x \in X} f(x) = -\min_{x \in X} -f(x),$$

werden wir in der Regel ohne Beschränkung der Allgemeinheit Minimierer suchen, können aber, wenn es bequemer ist, auch das äquivalente Maximierungsproblem betrachten.

Wir unterscheiden weiter: Die Funktion f hat in $\bar{x} \in X$

(i) ein *globales Minimum*, falls gilt $\bar{x} \in X$ und

$$f(\bar{x}) \leq f(x) \qquad \text{für alle } x \in X,$$

(ii) ein *striktes globales Minimum*, falls gilt $\bar{x} \in X$ und

$$f(\bar{x}) < f(x) \qquad \text{für alle } x \in X \setminus \{\bar{x}\},$$

(iii) ein *lokales Minimum*, falls gilt $\bar{x} \in X$ und ein $\varepsilon > 0$ existiert mit

$$f(\bar{x}) \leq f(x) \qquad \text{für alle } x \in X \cap B_\varepsilon(\bar{x}),$$

(iv) ein *striktes lokales Minimum*, falls gilt $\bar{x} \in X$ und ein $\varepsilon > 0$ existiert mit

$$f(\bar{x}) < f(x) \qquad \text{für alle } x \in (X \cap B_\varepsilon(\bar{x})) \setminus \{\bar{x}\}.$$

Entsprechend spricht man von (strikten) lokalen oder globalen Minimierern. Offensichtlich ist jedes (strikte) globale Minimum auch ein (striktes) lokales Minimum, jedoch nicht umgekehrt. Dabei sind strikte globale Minima eindeutig, während strikte lokale Minima lediglich isoliert sein müssen.

Wir werden sehen, dass wir nur lokale Minima mit vertretbarem Aufwand finden können. Eine Ausnahme bilden konvexe Funktionen. Zur Erinnerung: Eine Menge $X \subset \mathbb{R}^n$ heißt *konvex*, wenn für alle $x, y, \in X$ und alle $\lambda \in (0,1)$ gilt $\lambda x + (1-\lambda)y \in X$. Anschaulich bedeutet dies, dass für je zwei Punkte in X auch ihre Verbindungsstrecke in X liegt.

Sei nun $X \subset \mathbb{R}^n$ konvex. Dann heißt eine Funktion $f : X \to \mathbb{R}$

(i) *konvex* (auf X), wenn für alle $x, y \in X$ und alle $\lambda \in [0,1]$ gilt

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

(ii) *strikt konvex* (auf X), wenn für alle $x, y \in X$ mit $x \neq y$ und alle $\lambda \in (0,1)$ gilt

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y).$$

(iii) *gleichmäßig konvex* (auf X) oder kurz *μ-konvex*, wenn es ein *Konvexitätsmodul* $\mu > 0$ gibt, so dass für alle $x, y \in X$ und alle $\lambda \in [0,1]$ gilt

$$f(\lambda x + (1-\lambda)y) + \mu\lambda(1-\lambda)\|x-y\|^2 \leq \lambda f(x) + (1-\lambda) f(y).$$

Anschaulich bedeutet dies, dass für eine konvexe Funktion kein Punkt einer Verbindungsstrecke von zwei Punkten auf dem Graphen der Funktion unterhalb des Graphen liegt; für strikt konvexe Funktionen darf die Strecke darüber hinaus nicht mit dem Graphen zusammenfallen. Offensichtlich ist jede gleichmäßig konvexe Funktion strikt konvex und jede strikt konvexe Funktion konvex.

Satz 1.1. *Sei $X \subset \mathbb{R}^n$ eine konvexe Menge und sei $f : X \to \mathbb{R}$ eine konvexe Funktion. Dann gilt:*

(i) Jedes lokale Minimum von f ist auch ein globales Minimum.

(ii) Ist f strikt konvex, so besitzt f höchstens ein lokales Minimum, das dann sogar ein striktes globales Minimum ist.

Für stetig differenzierbare Funktionen lässt sich Konvexität über den Gradienten charakterisieren.

Satz 1.2. *Sei $X \subset \mathbb{R}^n$ offen und konvex und sei $f : X \to \mathbb{R}$ stetig differenzierbar. Dann ist*

(i) f genau dann konvex, wenn für alle $x, y \in X$ gilt

$$\nabla f(y)^T (x - y) \leq f(x) - f(y),$$

(ii) f genau dann strikt konvex, wenn für alle $x, y \in X$ mit $x \neq y$ gilt

$$\nabla f(y)^T (x - y) < f(x) - f(y),$$

(iii) f genau dann gleichmäßig konvex, wenn ein $\mu > 0$ existiert so dass für alle $x, y \in X$ gilt

$$\nabla f(y)^T (x - y) + \mu \|x - y\|^2 \leq f(x) - f(y).$$

Ist f sogar zweimal stetig differenzierbar, lässt sich die Konvexität auch über die Hesse-Matrix charakterisieren.

Satz 1.3. *Sei $X \subset \mathbb{R}^n$ offen und konvex und sei $f : X \to \mathbb{R}$ zweimal stetig differenzierbar. Dann ist*

(i) f genau dann konvex, wenn $\nabla^2 f(x)$ für alle $x \in X$ positiv semidefinit *ist, d. h. wenn gilt*

$$d^T \nabla^2 f(x) d \geq 0 \qquad \text{für alle } x \in X, d \in \mathbb{R}^n,$$

(ii) f strikt konvex, wenn $\nabla^2 f(x)$ für alle $x \in X$ positiv definit ist, d. h. wenn gilt

$$d^T \nabla^2 f(x) d > 0 \qquad \text{für alle } x \in X, d \in \mathbb{R}^n \setminus \{0\},$$

(iii) f genau dann gleichmäßig konvex, wenn $\nabla^2 f(x)$ für alle $x \in X$ gleichmäßig positiv definit ist, d. h. wenn ein $\mu > 0$ existiert mit

$$d^T \nabla^2 f(x) d \geq \mu \|d\|^2 \qquad \text{für alle } x \in X, d \in \mathbb{R}^n,$$

1.2 EXISTENZ

Wir betrachten nun die Frage der Existenz von Minimierern, die wir mit Hilfe des folgenden recht allgemeinen Satzes beantworten.

Satz 1.4. *Sei $X \subset \mathbb{R}^n$ nichtleer und abgeschlossen und $f : X \to \mathbb{R}$ stetig. Gilt*

(i) X ist beschränkt oder

(ii) f ist koerziv *auf X, d. h. für jede Folge $\{x^k\}_{k\in\mathbb{N}} \subset X$ mit $\|x^k\| \to \infty$ gilt $f(x^k) \to \infty$,*

so besitzt f einen globalen Minimierer $\bar{x} \in X$.

Ist X konvex und f strikt konvex, so ist der Minimierer eindeutig.

Ab nun werden wir stillschweigend voraussetzen, dass ein Minimierer existiert, und uns auf die Frage nach der Charakterisierung und Berechnung konzentrieren.

1.3 OPTIMALITÄTSBEDINGUNGEN

Wir betrachten in Folge unrestringierte Optimierungsprobleme, d. h. für $X = \mathbb{R}^n$, und leiten zunächst notwendige und hinreichende Bedingungen dafür her, dass ein Punkt $\bar{x} \in \mathbb{R}^n$ ein Minimierer ist. Die fundamentale Einsicht ist dabei, dass wir uns in einem Minimum befinden, wenn der Funktionswert bei Bewegung in jede Richtung zunehmen würde, d. h. wenn die Steigung in jede Richtung positiv ist.

Satz 1.5. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ differenzierbar auf der offenen Menge $U \subset \mathbb{R}^n$ und sei $\bar{x} \in U$ ein lokaler Minimierer von f. Dann gilt*

$$\nabla f(\bar{x})^T d \geq 0 \qquad \text{für alle } d \in \mathbb{R}^n. \tag{1.1}$$

Gilt (1.1), so folgt durch Einsetzen von $d = -\nabla f(\bar{x}) \in \mathbb{R}^n$ sofort $\|\nabla f(\bar{x})\|^2 \leq 0$ und damit die folgende Optimalitätsbedingung.

Satz 1.6 (notwendige Optimalitätsbedingung 1. Ordnung). *Sei $f : \mathbb{R}^n \to \mathbb{R}$ differenzierbar auf der offenen Menge $U \subset \mathbb{R}^n$ und sei $\bar{x} \in U$ ein lokaler Minimierer von f. Dann gilt*

$$\nabla f(\bar{x}) = 0. \tag{1.2}$$

Ein Punkt $\bar{x}$, der (1.2) erfüllt, heißt *stationärer Punkt*. Man spricht von einer *Bedingung 1. Ordnung*, da sie nur erste Ableitungen verwendet; die Bedingung ist lediglich notwendig, da auch Maximierer oder Sattelpunkte stationäre Punkte sind. Um diese auszuschließen, muss man zweite Ableitungen zu Rate ziehen.

Satz 1.7 (notwendige Optimalitätsbedingung 2. Ordnung). *Sei $f : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar auf der offenen Menge $U \subset \mathbb{R}^n$ und sei $\bar{x} \in U$ ein lokaler Minimierer von f. Dann ist $\nabla^2 f(\bar{x})$ positiv semidefinit.*

Auch diese Bedingung ist nur notwendig, da sie auch in Sattelpunkten erfüllt sein kann (betrachte $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^3$). Um diese auszuschließen, müssen wir die Bedingung verschärfen.

Satz 1.8 (hinreichende Optimalitätsbedingung 2. Ordnung). *Sei $f : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar auf der offenen Menge $U \subset \mathbb{R}^n$ und sei $\bar{x} \in U$ mit*

(i) $\nabla f(\bar{x}) = 0$ und

(ii) $\nabla^2 f(\bar{x})$ gleichmäßig positiv definit.

Dann hat f in $\bar{x}$ ein striktes lokales Minimum.

Dabei ist (ii) (nur) im Endlichdimensionalen genau dann erfüllt, wenn $\nabla^2 f(\bar{x})$ positiv definit ist. Diese Bedingung ist umgekehrt nur hinreichend, aber nicht notwendig, wie das Beispiel $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^4$, zeigt.

Beachte, dass Ableitungen immer nur lokale Informationen liefern und alle diese Bedingungen daher nur *lokale* Minimierer charakterisieren; ähnliche Bedingungen sind für *globale* Minimierer in der Regel nicht möglich! Eine Ausnahme bilden (mal wieder) konvexe Funktionen.

Satz 1.9 (notwendige und hinreichende Bedingung für konvexe Funktionen). *Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex und auf der offenen Menge $U \subset \mathbb{R}^n$ differenzierbar. Dann hat f in $\bar{x} \in U$ ein globales Minimum genau dann, wenn $\nabla f(\bar{x}) = 0$ ist.*

Die Beweise beruhen jeweils auf der Taylor-Entwicklung in Gestalt der folgenden *Mittelwertsätze.*

Satz 1.10 (Mittelwertsatz I). *Seien $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und $x, y \in \mathbb{R}^n$ gegeben. Dann existiert ein $\xi := y + \vartheta(x - y)$ mit $\vartheta \in (0, 1)$ und*

$$f(x) = f(y) + \nabla f(\xi)^T (x - y).$$

Satz 1.11 (Mittelwertsatz II). *Seien $f : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar und $x, y \in \mathbb{R}^n$ gegeben. Dann existiert ein $\xi := y + \vartheta(x - y)$ mit $\vartheta \in (0, 1)$ und*

$$f(x) = f(y) + \nabla f(y)^T (x - y) + \frac{1}{2}(x - y)^T \nabla^2 f(\xi)(x - y).$$

Für vektorwertige Funktionen $F : \mathbb{R}^n \to \mathbb{R}^m$ gelten diese Mittelwertsätze nicht direkt (das Problem ist, für alle Komponenten ein einheitliches ϑ zu finden). Es gilt aber der folgende Mittelwertsatz in Integralform (den wir zumeist auf $F(x) = \nabla f(x)$ anwenden werden).

Satz 1.12 (Mittelwertsatz III). *Seien $F : \mathbb{R}^n \to \mathbb{R}^m$ stetig differenzierbar und $x, y \in \mathbb{R}^n$ gegeben. Dann gilt*

$$F(x) = F(y) + \int_0^1 \nabla F\left(y + \vartheta(x - y)\right)^T (x - y)\, d\vartheta.$$

Ist ∇f sogar Lipschitz-stetig mit Konstante $L > 0$ (man nennt f dann *Lipschitz-stetig differenzierbar* oder, wenn man die Konstante betonen möchte, kurz *L-glatt*), folgt daraus eine für Abstiegsverfahren wichtige Abschätzung.

Folgerung 1.13 (Abstiegslemma). *Sei $f : \mathbb{R}^n \to \mathbb{R}$ L-glatt und $x, y \in \mathbb{R}^n$ gegeben. Dann gilt*

$$f(x) \leq f(y) + \nabla f(y)^T (x - y) + \frac{L}{2}\|x - y\|^2.$$

2 NUMERISCHE VERFAHREN

In diesem Kapitel wiederholen wir die grundlegenden Verfahren der nichtrestringierten Optimierung und ihre wesentlichen Eigenschaften.

2.1 ABSTIEGSVERFAHREN

Satz 1.6 legt folgendes iterative Verfahren zur Bestimmung eines Minimierers nahe:

Algorithmus 2.1 : Allgemeines Abstiegsverfahren

1 Wähle einen *Startpunkt* $x^0 \in \mathbb{R}^n$, setze $k = 0$
2 **while** $\nabla f(x^k) \neq 0$ **do**
3 Wähle eine *Suchrichtung* $s^k \in \mathbb{R}^n$ mit $\nabla f(x^k)^T s^k < 0$
4 Wähle eine *Schrittweite* $\sigma_k > 0$ mit $f(x^k + \sigma_k s^k) < f(x^k)$
5 Setze $x^{k+1} = x^k + \sigma_k s^k$, $k \leftarrow k + 1$

Der Beweis von Satz 1.6 zeigt, dass wir stets eine Suchrichtung und eine Schrittweite mit den gewünschten Eigenschaften finden können, solange x^k kein stationärer Punkt ist. Das einzige, was noch schief gehen kann, ist, dass wir vor Erreichen eines stationären Punkts „verhungern“, d. h. dass entweder s^k oder σ_k zu schnell zu klein werden. Wir suchen also Bedingungen, die das verhindern (und die wir für konkrete Vorschriften zur Berechnung von s^k und σ_k nachprüfen können), für die also das Verfahren *konvergiert*. Dies wollen wir wie folgt verstehen: Ein Verfahren *konvergiert global*, wenn jeder Häufungspunkt $\bar{x}$ einer entsprechend erzeugten Folge $\{x^k\}_{k\in\mathbb{N}}$ für beliebigen Startpunkt $x^0 \in \mathbb{R}^n$ ein stationärer Punkt ist. Beachte, dass man nicht mehr von einem Verfahren, das nur erste Ableitungen verwendet, erwarten kann. Insbesondere bedeutet globale Konvergenz *nicht*, dass das Verfahren gegen einen *globalen* Minimierer konvergiert! Dagegen sprechen wir von *lokaler Konvergenz*, wenn dies nur für Startwerte $x^0 \in B_\varepsilon(\bar{x})$ für ein $\varepsilon > 0$ und einen stationären Punkt $\bar{x}$ gelten muss.

Wir beginnen mit Bedingungen an die Suchrichtungen s^k. Eine Folge $\{s^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ nennen wir Folge von *zulässigen Suchrichtungen*, wenn gilt:

(i) Alle s^k sind *Abstiegsrichtungen*, d. h. $\nabla f(x^k)^T s^k < 0$ für alle $k \in \mathbb{N}$;

(ii) Aus $\dfrac{\nabla f(x^k)^T s^k}{\|s^k\|} \to 0$ folgt $\nabla f(x^k) \to 0$.

Zum Beispiel erzeugt die Wahl $s^k := -\nabla f(x^k)$ wegen $-\nabla f(x^k)^T s^k = \|s^k\|^2 = \|\nabla f(x^k)\|^2$ stets zulässige Suchrichtungen.

Nun zu den Schrittweiten σ_k. Eine Folge $\{\sigma^k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{>0}$ nennen wir Folge von *zulässigen Schrittweiten* für $\{s^k\}_{k\in\mathbb{N}}$, wenn gilt:

(i) Alle σ_k führen zu einem Abstieg, d. h. $f(x^k + \sigma_k s^k) \le f(x^k)$ für alle $k \in \mathbb{N}$;

(ii) Aus $f(x^k + \sigma_k s^k) - f(x^k) \to 0$ folgt $\dfrac{\nabla f(x^k)^T s^k}{\|s^k\|} \to 0$.

Ein Beispiel ist die *Armijo-Regel*, die Schrittweiten $\sigma \in (0, 1]$ generiert, die die *Armijo-Bedingung* erfüllen: Für eine gegebene Richtung s und $\gamma \in (0, 1)$ soll gelten

$$f(x + \sigma s) - f(x) \le \gamma\sigma\nabla f(x)^T s. \tag{2.1}$$

Anschaulich bestimmt diese Regel die größte Schrittweite zwischen 0 und 1, die mindestens den gleichen Abstieg wie die linearisierte Funktion $\varphi(\sigma) = f(x) + \sigma\gamma\nabla f(x)^T s$ erreicht. Üblicherweise wird γ klein gewählt, z. B. $\gamma = 10^{-2}$. Die Armijo-Regel erzeugt für $s^k = -\nabla f(x^k)$ stets zulässige Schrittweiten; die Kombination nennt man *Gradientenverfahren.*

Für zulässige Suchrichtungen und Schrittweiten konvergiert Algorithmus 2.1 global.

Satz 2.1. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar. Dann bricht Algorithmus 2.1 entweder nach endlich vielen Schritten ab, oder er erzeugt Folgen $\{x^k\}_{k\in\mathbb{N}}$, $\{s^k\}_{k\in\mathbb{N}}$, und $\{\sigma_k\}_{k\in\mathbb{N}}$, die nicht endlich sind. Sind die Suchrichtungen $\{s^k\}_{k\in\mathbb{N}}$ und die Schrittweiten $\{\sigma_k\}_{k\in\mathbb{N}}$ zulässig, so ist jeder Häufungspunkt von $\{x^k\}_{k\in\mathbb{N}}$ ein stationärer Punkt von f.*

2.2 NEWTON-ARTIGE VERFAHREN

Newton-artige Verfahren verwenden statt dem negativen Gradienten die Lösung eines Gleichungssystems mit dieser rechten Seite:

Algorithmus 2.2 : Allgemeines Newton-artiges Verfahren

1 Wähle einen *Startpunkt* $x^0 \in \mathbb{R}^n$, setze $k = 0$
2 **while** $\|\nabla f(x^k)\| > 0$ **do**
3 Wähle eine invertierbare Matrix $H_k \in \mathbb{R}^{n\times n}$
4 Berechne s^k als Lösung von $H_k s^k = -\nabla f(x^k)$
5 Setze $x^{k+1} = x^k + s^k$, $k \leftarrow k + 1$

Die Wahl der Schrittweite σ_k steckt dabei als Skalierung in der Wahl der Matrix H_k. Motivation ist hier natürlich das Newton-Verfahren mit $H_k := \nabla^2 f(x^k)$.

Für die Durchführbarkeit des Newton-Verfahrens sind die folgenden Hilfsresultate wichtig.

Lemma 2.2 (Banach-Lemma). *Seien $A, B \in \mathbb{R}^{n\times n}$ mit $\|I - BA\| < 1$. Dann sind A und B invertierbar, und es gilt*

$$\|A^{-1}\| \leq \frac{\|B\|}{1 - \|I - BA\|}.$$

Eine analoge Abschätzung gilt für B^{-1}.

Lemma 2.3. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar und sei $\bar{x} \in \mathbb{R}^n$ mit $\nabla^2 f(\bar{x})$ invertierbar. Dann existieren Konstanten $\delta > 0$ und $c > 0$, so dass gilt*

$$\|\nabla^2 f(x)^{-1}\| \leq c \qquad \text{für alle } x \in B_\delta(\bar{x}).$$

Insbesondere ist $\nabla^2 f(x)$ invertierbar für alle $x \in B_\delta(\bar{x})$.

Lemma 2.4. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar und sei $\bar{x} \in \mathbb{R}^n$ mit $\nabla^2 f(\bar{x})$ positiv definit. Dann existieren Konstanten $\delta > 0$ und $\mu > 0$, so dass gilt*

$$d^T \nabla^2 f(x) d \geq \mu \|d\|^2 \qquad \text{für alle } x \in B_\delta(\bar{x}), d \in \mathbb{R}^n.$$

Newton-artige Verfahren sind im allgemeinen deutlich aufwändiger als Abstiegsverfahren, da in jedem Schritt ein lineares Gleichungssystem gelöst werden muss. Damit sich der Aufwand lohnt, sollten also deutlich weniger Iterationen notwendig sein, um einen vorgegebenen Abstand $\|x^k - \bar{x}\| \leq \varepsilon$ zu einem stationären Punkt $\bar{x}$ zu erreichen. Dies kann mit dem Begriff der *Konvergenzgeschwindigkeit* einer Folge mathematisch präzisiert werden.

Wir sagen, eine Folge $\{x^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ konvergiert gegen $\bar{x} \in \mathbb{R}^n$

(i) *linear*, falls ein $c \in (0, 1)$ existiert mit

$$\|x^{k+1} - \bar{x}\| \leq c\|x^k - \bar{x}\| \qquad \text{für alle } k \in \mathbb{N} \text{ hinreichend groß,}$$

(ii) *superlinear*, falls eine Nullfolge $\{\varepsilon_k\}_{k\in\mathbb{N}}$ existiert mit

$$\|x^{k+1} - \bar{x}\| \leq \varepsilon_k\|x^k - \bar{x}\| \qquad \text{für alle } k \in \mathbb{N} \text{ hinreichend groß,}$$

(iii) *quadratisch*, falls $x^k \to \bar{x}$ und ein $C > 0$ existiert mit

$$\|x^{k+1} - \bar{x}\| \leq C\|x^k - \bar{x}\|^2 \qquad \text{für alle } k \in \mathbb{N} \text{ hinreichend groß.}$$

(Diese Begriffe werden in der Literatur auch als *q-lineare* (-superlineare, -quadratische) Konvergenz – im Gegensatz zu der weniger häufig verwendeten *r-linearen* (-superlinearen, -quadratischen) Konvergenz – bezeichnet.) Die letzten beiden Bedingungen kann man auch mit Hilfe der *Landau-Symbole* formulieren: $\|x^{k+1} - \bar{x}\| = o(\|x^k - \bar{x}\|)$ für superlineare Konvergenz bzw. $\|x^{k+1} - \bar{x}\| = O(\|x^k - \bar{x}\|^2)$ für quadratische Konvergenz.

Für Newton-artige Verfahren kann man die superlineare Konvergenz durch eine entsprechende Approximationseigenschaft der H_k charakterisieren.

Satz 2.5 (Dennis–Moré-Bedingungen). *Sei $f : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar und sei $\{x^k\}_{k\in\mathbb{N}}$ eine durch Algorithmus 2.2 erzeugte Folge, die gegen ein $\bar{x} \in \mathbb{R}^n$ mit $\nabla^2 f(\bar{x})$ invertierbar und $x^k \neq \bar{x}$ für alle $k \in \mathbb{N}$ konvergiert. Dann sind äquivalent:*

(i) $\{x^k\}_{k\in\mathbb{N}}$ *konvergiert superlinear gegen* $\bar{x}$ *und* $\nabla f(\bar{x}) = 0$,

(ii) $\|(H_k - \nabla^2 f(x^k))(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\|)$,

(iii) $\|(H_k - \nabla^2 f(\bar{x}))(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\|)$.

Eine analoge Aussage gilt für die quadratische Konvergenz, wenn $x \mapsto \nabla^2 f(x)$ Lipschitz-stetig ist.

Für das Newton-Verfahren folgt daraus sofort die lokale(!) superlineare Konvergenz.

Satz 2.6. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar und sei $\bar{x} \in \mathbb{R}^n$ ein stationärer Punkt von f mit $\nabla^2 f(\bar{x})$ invertierbar. Dann existiert ein $\varepsilon > 0$, so dass Newton-Verfahren für alle Startwerte $x^0 \in B_\varepsilon(\bar{x})$ superlinear gegen $\bar{x}$ konvergiert. Ist $\nabla^2 f$ darüber hinaus lokal Lipschitz-stetig, so ist die Konvergenz sogar quadratisch.*

Für die globale Konvergenz kombiniert man das Verfahren üblicherweise mit einer Liniensuche.

Algorithmus 2.3 : Globalisiertes Newton-Verfahren

Input : $\rho > 0$, $p > 2$, $\gamma \in (0, 1/2)$, $x^0 \in \mathbb{R}^n$

Setze $k = 0$
while $\|\nabla f(x^k)\| > 0$ **do**
  Versuche, Newton-Schritt d^k mit $\nabla^2 f(x^k)d^k = -\nabla f(x^k)$ zu berechnen
  if $\nabla f(x^k)^T d^k \leq -\rho\|d^k\|^p$ **then**
    Setze $s^k = d^k$
  else
    Setze $s^k = -\nabla f(x^k)$
  Bestimme $\sigma_k > 0$ mit Armijo-Regel für $\gamma \in (0, 1/2)$
  Setze $x^{k+1} = x^k + \sigma_k s^k, \quad k \leftarrow k + 1$

Die Bedingung in Schritt 4 setzt dabei stillschweigend voraus, dass eine Lösung des Newton-Systems $\nabla^2 f(x^k)d^k = -\nabla f(x^k)$ gefunden wurde. Beachte auch die Einschränkung $\gamma < \frac{1}{2}$; dies ist wichtig, um superlineare Konvergenz zu erhalten, indem für den vollen Newton-Schritt die Schrittweite 1 akzeptiert wird.

Satz 2.7. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar, und sei $\bar{x} \in \mathbb{R}^n$ ein Häufungspunkt der durch Algorithmus 2.3 erzeugten Folge $\{x^k\}_{k\in\mathbb{N}}$ mit $\nabla^2 f(\bar{x})$ positiv definit. Dann ist $\bar{x}$ ein strikter lokaler Minimierer und $\{x^k\}_{k\in\mathbb{N}}$ konvergiert gegen $\bar{x}$ superlinear. Ist $\nabla^2 f$ darüber hinaus lokal Lipschitz-stetig, so ist die Konvergenz quadratisch.*

Anstelle der Hesse-Matrix kann man eine Finite-Differenzen-Approximation verwenden, die die *Quasi-Newton-Gleichung*

$$H_{k+1}(x^{k+1} - x^k) = \nabla f(x^{k+1}) - \nabla f(x^k)$$

erfüllt. Diese Gleichung ist unterbestimmt; man bemüht sich üblicherweise zusätzlich, den Rang von $H_{k+1} - H_k$ minimal zu halten. Mit der Notation

$$H := H_k, \qquad H_+ := H_{k+1}, \qquad s := x^{k+1} - x^k, \qquad y := \nabla f(x^{k+1}) - \nabla f(x^k),$$

sind die folgenden Verfahren verbreitet:

(i) *Broyden-Update*

$$H_+^B = H + \frac{(y - Hs)s^T}{s^T s},$$

(ii) *SR-1-Update*

$$H_+^{SR1} = H + \frac{(y - Hs)(y - Hs)^T}{(y - Hs)^T s},$$

welches für symmetrische H ebenfalls symmetrisch ist,

(iii) *BFGS-Update*

$$H_+^{BFGS} = H + \frac{yy^T}{s^T y} - \frac{Hss^T H}{s^T Hs},$$

welches für symmetrisch und positiv definite H und $y^T s > 0$ ebenfalls symmetrisch und positiv definit ist.

3 LINEARE OPTIMIERUNG

Als Einstieg in die Optimierung mit Nebenbedingungen betrachten wir *lineare* Zielfunktionen mit (affin-)*linearen* Nebenbedingungen.

3.1 KONVEXE MENGEN UND POLYEDER

Zur Erinnerung: eine Menge $C \subset \mathbb{R}^n$ heißt *konvex*, falls gilt

$$\lambda x + (1-\lambda)\, y \in C \qquad \text{für alle } x, y \in C \text{ und } \lambda \in [0,1].$$

Ein fundamentales Hilfsmittel ist das folgende prototypische restringierte konvexe Optimierungsproblem.

Lemma 3.1. *Sei $C \subset \mathbb{R}^n$ nichtleer, konvex, und abgeschlossen und $z \in \mathbb{R}^n$ beliebig. Dann hat das Problem*

$$\min_{x \in C} \|x - z\|_2$$

eine eindeutige Lösung $\bar{x} \in C$, genannt Projektion von z auf C.

Die folgende Charakterisierung der Projektion ist unser erstes Beispiel einer Optimalitätsbedingung für restringierte Optimierungsprobleme (vergleiche die Aussage von Satz 1.5).

Satz 3.2 (Projektionssatz). *Sei $C \subset \mathbb{R}^n$ nichtleer, konvex, und abgeschlossen und $z \in \mathbb{R}^n$ beliebig. Dann ist $\bar{x} \in C$ die Projektion von z auf C genau dann, wenn gilt*

$$(\bar{x} - z)^T (x - \bar{x}) \geq 0 \qquad \textit{für alle } x \in C.$$

Damit kann man den folgenden fundamentalen Satz über konvexe Mengen beweisen, der einen Spezialfall der *Trennungssätze von Hahn–Banach* darstellt. Man zeigt leicht durch einfache geometrische Beispiele, dass keine der Voraussetzungen fallengelassen werden kann.

Satz 3.3 (Trennungssatz). *Sei $C \subset \mathbb{R}^n$ nichtleer, abgeschlossen, und konvex und $x_0 \in \mathbb{R}^n \setminus C$. Dann existieren ein $a \in \mathbb{R}^n \setminus \{0\}$ und ein $\alpha \in \mathbb{R}$ mit*

$$a^T x_0 > \alpha \geq a^T x \qquad \text{für alle } x \in C.$$

Ein Spezialfall konvexer Mengen sind *Polyeder*, die durch endlich viele linear-affine Gleichungen und Ungleichungen (insbesondere der Form $x \geq 0$) beschrieben werden können. Das zentrale Resultat über Polyeder ist der folgende Alternativsatz, der als *Farkas-Lemma* bekannt ist.

Lemma 3.4 (Farkas). *Seien $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$. Dann sind äquivalent:*

(i) Es existiert ein $x \in \mathbb{R}^n$ mit $Ax = b$ und $x \geq 0$.

(ii) Für alle $d \in \mathbb{R}^m$ mit $A^T d \geq 0$ gilt $b^T d \geq 0$.

3.2 DUALITÄT

Wir betrachten nun für $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, und $c \in \mathbb{R}^n$ das lineare Optimierungsproblem

$$\text{(LP)} \qquad \begin{cases} \min\limits_{x \in \mathbb{R}^n} c^T x \\ \text{mit } Ax \leq b. \end{cases}$$

Aus Satz 1.4 folgt sofort, dass (LP) eine Lösung besitzt, falls die zulässige Menge $P(A, b) := \{x \in \mathbb{R}^n \mid Ax \leq b\}$ nichtleer und $c^T x$ dort nach unten beschränkt ist. Man rechnet leicht nach, dass jeder zulässige Punkt $y \in P^=(A^T, -c) := \left\{y \in \mathbb{R}^m \;\middle|\; A^T y = -c, y \geq 0\right\}$ des *dualen Problems*

$$\text{(LD)} \qquad \begin{cases} \max\limits_{y \in \mathbb{R}^m} -b^T y \\ \text{mit } A^T y = -c, \\ \qquad\quad y \geq 0, \end{cases}$$

eine untere Schranke liefert, und dass die beste Schranke genau die Lösungen von (LD) charakterisiert.

Satz 3.5 (schwache Dualität). *Für alle $x \in P(A, b)$ und $y \in P^=(A^T, -c)$ ist*

$$c^T x \geq -b^T y. \tag{3.1}$$

Gilt Gleichheit für ein $\bar{x} \in P(A, b)$ und ein $\bar{y} \in P^=(A^T, -c)$, so ist $\bar{x}$ Lösung von (LP) und $\bar{y}$ Lösung von (LD).

Der *Fundamentalsatz der linearen Optimierung* besagt, dass Gleichheit in (3.1) *immer* gilt, solange die beiden zulässigen Mengen nichtleer sind; der Beweis beruht auf dem Farkas-Lemma 3.4.

Satz 3.6 (starke Dualität). *Beide Probleme* (LP) *und* (LD) *haben eine Lösung* $\bar{x}$ *bzw.* $\bar{y}$ *genau dann, wenn die zulässigen Mengen* $P(A, b)$ *bzw.* $P^{=}(A^T, -c)$ *nichtleer sind. In diesem Fall gilt*

$$c^T \bar{x} = -b^T \bar{y}.$$

Aus der Dualität erhält man auch eine erste Optimalitätsbedingung für restringierte Optimierungsprobleme. (Beachte, dass der Gradient der Zielfunktion hier konstant gleich c ist und damit überall oder nirgends verschwindet.)

Satz 3.7 (schwache Komplementarität). *Es ist* $\bar{x}$ *Lösung von* (LP) *und* $\bar{y}$ *Lösung von* (LD) *genau dann, wenn gilt*

$$\begin{cases} A\bar{x} \le b, & \text{(primale Zulässigkeit)} \\ A^T \bar{y} = -c, \quad \bar{y} \ge 0, & \text{(duale Zulässigkeit)} \\ \bar{y}_i (b_i - A_i \bar{x}) = 0 \quad \text{für alle } i = 1, \dots, m. & \text{(Komplementarität)} \end{cases} \tag{3.2}$$

Die Komplementaritätsbedingung sagt, dass $\bar{y}_i = 0$ oder $b_i = A_i \bar{x}$ für jedes $1 \le i \le m$ gilt. Dies ist aber kein exklusives Oder – es ist also zugelassen, dass beide Gleichungen simultan gelten. Man kann jedoch zeigen, dass unter allen Lösungen auch ein Paar existiert, für das immer nur genau eine Gleichheit gilt.

Satz 3.8 (strikte Komplementarität). *Sind* $P(A, b)$ *und* $P^{=}(A^T, -c)$ *nichtleer, so existiert eine Lösung* $\bar{x}$ *von* (LP) *und eine Lösung* $\bar{y}$ *von* (LD) *mit*

$$\bar{y}_i = 0 \quad \text{genau dann, wenn} \quad A_i \bar{x} < b_i$$

für alle $i = 1, \dots, m$.

Teil II

OPTIMIERUNG MIT NEBENBEDINGUNGEN

4 OPTIMALITÄTSBEDINGUNGEN

Wir betrachten nun für $X \subsetneq \mathbb{R}^n$ und $f : X \to \mathbb{R}$ das *restringierte* Optimierungsproblem

$$\min_{x \in X} f(x) \tag{4.1}$$

und leiten dafür zunächst Optimalitätsbedingungen her. Wie schon für den Fall $n = 1$ bekannt, ist das Verschwinden des Gradienten *keine* notwendige Optimalitätsbedingung für einen lokalen Minimierer, wenn dieser auf dem Rand der zulässigen Menge liegt. Für (vernünftige) $X \subset \mathbb{R}$ ist dies durch einfaches Nachprüfen endlich vieler Randpunkte noch leicht zu handhaben. Ist $n > 1$ aber $X \subset \mathbb{R}^n$ ein Polyeder (d. h. durch endlich viele *lineare* Gleichungs- und Ungleichungsnebenbedingungen beschrieben), so hat der Rand von X ebenfalls eine spezielle Struktur (was in der linearen Optimierung weidlich ausgenutzt wird). Für allgemeine Teilmengen $X \subsetneq \mathbb{R}^n$ kann der Rand aber beliebig kompliziert sein, was die Schwierigkeit der restringierten Optimierung ausmacht. Hier und in Folge nehmen wir an, dass f auf ganz $\mathbb{R}^n$ definiert ist, um Ableitungen auf dem Rand von X betrachten zu können.

4.1 TANGENTIALKEGEL

Wir orientieren uns an Satz 1.5, der besagt, dass für unrestringierte Probleme in einem lokalen Minimierer $\bar{x}$ keine Richtungen Abstiegsrichtungen sein dürfen, d. h. $\nabla f(\bar{x})^T d \geq 0$ für alle $d \in \mathbb{R}^n$ gilt. Für ein restringiertes Problem spielen dagegen nur die Richtungen eine Rolle, die nicht sofort(!) aus X hinausführen; für solch eine Richtung $d \in \mathbb{R}^n$ muss also $x := \bar{x} + td \in X$ für alle $t > 0$ klein genug gelten. Dies motiviert die folgende Definition: Wir nennen $d \in \mathbb{R}^n$ *Tangentialrichtung* an X in x, falls Folgen $\{x^k\}_{k \in \mathbb{N}} \subset X$ und $\{t_k\}_{k \in \mathbb{N}} \subset (0, \infty)$ existieren mit

$$x^k \to x, \qquad t_k \to 0, \qquad \frac{x^k - x}{t_k} \to d.$$

Die Menge aller Tangentialrichtungen bezeichnen wir als *Tangentialkegel*

$$T_X(x) := \{d \in \mathbb{R}^n \mid d \text{ ist Tangentialrichtung an } X \text{ in } x\}.$$

Ist x ein innerer Punkt von X, so gilt $T_X(x) = \mathbb{R}^n$ (wähle $x^k = x + t_k d \in X$ für d beliebig und t_k klein genug). Weiter gilt stets $0 \in T_X(x)$ (wähle $x^k := x$) sowie für $d \in T_X(x)$ und $\alpha > 0$ auch $\tilde{d} := \alpha d \in T_X(x)$ (wähle $\tilde{t}_k := \alpha^{-1} t_k$), was die Bezeichnung *Kegel* rechtfertigt.

Dass wir bei der Konstruktion von Tangentialrichtungen Folgen von Punkten in X und nicht nur feste Punkte zulassen, liegt daran, dass der so definierte Tangentialkegel stets abgeschlossen ist. Dies wird später von Bedeutung sein.

Lemma 4.1. *Seien $X \subset \mathbb{R}^n$ nichtleer und $x \in X$. Dann ist $T_X(x)$ abgeschlossen.*

Beweis. Wir müssen zeigen, dass für $\{d^k\}_{k\in\mathbb{N}} \subset T_X(x)$ mit $d^k \to d$ auch $d \in T_X(x)$ gilt. Für jede Tangentialrichtung d^k existieren gemäß Definition Folgen $\{x^{k,l}\}_{l\in\mathbb{N}} \subset X$ und $\{t_{k,l}\}_{l\in\mathbb{N}} \subset (0,\infty)$ so, dass für alle $k \in \mathbb{N}$ ein $l(k) \in \mathbb{N}$ existiert mit

$$\|x^{k,l(k)} - x\| \le \frac{1}{k}, \qquad t_{k,l(k)} \le \frac{1}{k}, \qquad \left\| \frac{x^{k,l(k)} - x}{t_{k,l(k)}} - d^k \right\| \le \frac{1}{k}.$$

Für die entsprechenden Diagonalfolgen $\{x^{k,l(k)}\}_{k\in\mathbb{N}} \subset X$ und $\{t_{k,l(k)}\}_{k\in\mathbb{N}} \subset (0,\infty)$ gilt daher $x^{k,l(k)} \to x$, $t_{k,l(k)} \to 0$, sowie wegen $d^k \to d$

$$\left\| \frac{x^{k,l(k)} - x}{t_{k,l(k)}} - d \right\| \le \left\| \frac{x^{k,l(k)} - x}{t_{k,l(k)}} - d^k \right\| + \|d^k - d\| \to 0$$

für $k \to \infty$. Also ist auch d eine Tangentialrichtung. □

Analog zu Satz 1.5 erhalten wir eine abstrakte notwendige Optimalitätsbedingung erster Ordnung.

Satz 4.2. *Seien $X \subset \mathbb{R}^n$ eine nichtleere Menge und $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar. Hat f in $\bar{x} \in X$ ein lokales Minimum, so gilt*

$$\nabla f(\bar{x})^T d \ge 0 \qquad \text{für alle } d \in T_X(\bar{x}). \tag{4.2}$$

Beweis. Sei $d \in T_X(\bar{x})$ beliebig. Dann existieren nach Definition Folgen $\{x^k\}_{k\in\mathbb{N}} \subset X$ und $\{t_k\}_{k\in\mathbb{N}} \subset (0,\infty)$ mit $x^k \to \bar{x}$, $t_k \to 0$, und $\frac{x^k - \bar{x}}{t_k} \to d$. Weiter existiert nach Satz 1.10 ein $\xi^k = \bar{x} + \vartheta_k(x^k - \bar{x})$ mit $\vartheta_k \in (0,1)$ und

$$\nabla f(\xi^k)^T (x^k - \bar{x}) = f(x^k) - f(\bar{x}) \ge 0$$

für $k \in \mathbb{N}$ hinreichend groß (denn $\bar{x}$ ist lokaler Minimierer und $x^k \in X$ mit $x^k \to \bar{x}$). Da mit $x^k \to \bar{x}$ auch $\xi^k \to \bar{x}$ konvergiert, können wir durch $t_k > 0$ dividieren und erhalten wegen der Stetigkeit von ∇f durch Grenzübergang

$$\nabla f(\bar{x})^T d = \lim_{k\to\infty} \nabla f(\xi^k)^T \left(\frac{x^k - \bar{x}}{t_k} \right) \ge 0,$$

was zu beweisen war. □

Ist $K \subset \mathbb{R}^n$ ein nichtleerer Kegel, so bezeichnet man die Menge

$$K^o := \left\{ x \in \mathbb{R}^n \;\middle|\; x^T d \leq 0 \text{ für alle } d \in K \right\}$$

als *Polarkegel* von K. Die notwendige Optimalitätsbedingung (4.2) kann man damit kompakt schreiben als

$$-\nabla f(\bar{x}) \in T_X(\bar{x})^o. \tag{4.3}$$

Dies ist die Verallgemeinerung der notwendigen Bedingung aus Satz 1.6, welche wir für $X = \mathbb{R}^n$ wegen $(\mathbb{R}^n)^o = \{0\}$ als Spezialfall wiedergewinnen.

Analog zu Satz 1.9 gelten für konvexe Probleme stärkere Aussagen: Ist f konvex, so ist die notwendige Bedingung (4.2) auch hinreichend; ist auch X konvex, so hat sie eine besonders einfache Form.

Satz 4.3. *Seien $X \subset \mathbb{R}^n$ eine nichtleere konvexe Menge und $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und konvex. Dann hat f in $\bar{x} \in X$ ein globales Minimum genau dann wenn gilt*

$$\nabla f(\bar{x})^T (x - \bar{x}) \geq 0 \qquad \text{für alle } x \in X. \tag{4.4}$$

Beweis. Sei zuerst $\bar{x}$ ein globaler Minimierer von f auf X. Dann genügt nach Satz 4.2 zu zeigen, dass $d := x - \bar{x} \in T_X(\bar{x})$ für alle $x \in X$ gilt. Sei dafür $x \in X$ beliebig und wähle $t_k := \frac{1}{k} \in (0, 1]$ und

$$x^k := \bar{x} + t_k(x - \bar{x}) = t_k x + (1 - t_k)\bar{x} \in X$$

wegen der Konvexität von X. Wegen $t_k^{-1}(x^k - \bar{x}) = x - \bar{x} = d$ für alle $k \in \mathbb{N}$ folgt daraus nach Definition die Behauptung.

Umgekehrt folgt aus Satz 1.2 (i) und der Annahme für alle $x \in X$

$$f(x) - f(\bar{x}) \geq \nabla f(\bar{x})^T (x - \bar{x}) \geq 0$$

und damit die globale Optimalität von $\bar{x}$. □

Man nennt die Menge

$$N_X(x) := \left\{ d \in \mathbb{R}^n \;\middle|\; d^T (y - x) \leq 0 \text{ für alle } y \in X \right\}$$

auch *Normalenkegel* an X in x; die Optimalitätsbedingung (4.4) kann man damit kompakt schreiben als

$$-\nabla f(\bar{x}) \in N_X(\bar{x}). \tag{4.5}$$

(Tatsächlich gilt für konvexe Mengen stets $N_X(x) = T_X(x)^o$.) Ist f nicht konvex, so zeigt der Beweis, dass (4.2) immer noch eine notwendige Optimalitätsbedingung für lokale Minimierer ist.

4.2 REGULARITÄTSBEDINGUNGEN

Der Rest des Kapitels ist nun der Aufgabe gewidmet, konkrete Darstellungen sowohl des Tangentialkegels $T_X(x)$ als auch der notwendigen Optimalitätsbedingung (4.3) herzuleiten für (nicht notwendigerweise konvexe) Mengen der speziellen Form

$$X := \left\{ x \in \mathbb{R}^n \;\middle|\; g_i(x) \leq 0,\ 1 \leq i \leq m, \quad h_j(x) = 0,\ 1 \leq j \leq p \right\} \tag{4.6}$$

für $g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar.

4.2.1 UNGLEICHUNGSNEBENBEDINGUNGEN

Wir betrachten zuerst den Fall von reinen Ungleichungsnebenbedingungen, d. h.

$$X = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, \quad 1 \leq i \leq m\}$$

für $g_i : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar. Da wir Abstiegsrichtungen durch die Ableitung (und damit Linearisierung) der Zielfunktion charakterisieren können, ist es naheliegend zu versuchen, Tangentialrichtungen über die Ableitung der Nebenbedingungen zu charakterisieren. Weiterhin ist zu erwarten, dass bei der Charakterisierung eines lokalen Minimierers $\bar{x}$ nur die *aktiven* Nebenbedingungen mit $g_i(\bar{x}) = 0$ eine Rolle spielen.

Lemma 4.4. *Für $x \in X$ und $d \in T_X(x)$ gilt*

$$\nabla g_i(x)^T d \leq 0 \qquad \text{für alle } i \in \{1, \ldots, m\} \text{ mit } g_i(x) = 0.$$

Beweis. Für $d \in T_X(x)$ existieren nach Definition Folgen $\{x^k\}_{k\in\mathbb{N}} \subset X$ und $\{t_k\}_{k\in\mathbb{N}} \subset (0, \infty)$ mit $x^k \to x$, $t_k \to 0$, und $\frac{x^k - x}{t_k} \to d$. Sei nun $i \in \{1, \ldots, m\}$ mit $g_i(x) = 0$ beliebig. Dann folgt für alle $k \in \mathbb{N}$ wegen $x^k \in X$ und Satz 1.10

$$0 \geq g_i(x^k) - g_i(x) = \nabla g_i(\xi^k)^T (x^k - x)$$

für ein $\xi^k := x + \vartheta_k(x^k - x)$ mit $\vartheta_k \in (0, 1)$. Wieder folgt aus $x^k \to x$ auch $\xi^k \to x$. Division durch $t_k > 0$ und Grenzübergang $k \to \infty$ ergibt dann wegen der stetigen Differenzierbarkeit von g_i die Behauptung. □

Definieren wir die *Menge der aktiven Nebenbedingungen*

$$\mathcal{A}_X(x) := \{i \in \{1, \ldots, m\} \mid g_i(x) = 0\}$$

und den *Linearisierungskegel*

$$L_X(x) := \left\{ d \in \mathbb{R}^n \;\middle|\; \nabla g_i(x)^T d \leq 0 \text{ für alle } i \in \mathcal{A}_X(x) \right\},$$

so haben wir gerade gezeigt, dass $T_X(x) \subset L_X(x)$ gilt. Für Satz 4.2 ist das aber die falsche Richtung: Da $L_X(x)$ *größer* ist als $T_X(x)$, ist die Bedingung $\nabla f(x)^T d \geq 0$ für alle $d \in L_X(x)$ *stärker* als (4.2) und daher nicht unbedingt für jeden lokalen Minimierer erfüllt – es ist also keine notwendige Optimalitätsbedingung mehr (und eine hinreichende sowieso nicht). Beachte auch, dass der Tangentialkegel nur von der Menge X abhängt und damit unabhängig ist von ihrer Beschreibung durch konkrete Ungleichungen $g_i(x) \leq 0$, der Linearisierungskegel dagegen sehr wohl von der Wahl der g_i abhängt.

Leider gilt die umgekehrte Inklusion $L_X(x) \subset T_X(x)$ im Allgemeinen nicht.

Beispiel 4.5. Wir betrachten die Menge

$$X := \left\{ x \in \mathbb{R}^2 \;\middle|\; g_1(x) = x_2 - x_1^3 \leq 0,\ g_2(x) = -x_2 \leq 0 \right\}$$

sowie den zulässigen Punkt $x = (0,0)^T$, in dem beide Nebenbedingungen aktiv sind. Der zugehörige Linearisierungskegel ist

$$L_X(0) = \left\{ d \in \mathbb{R}^2 \;\middle|\; d_2 \leq 0,\ -d_2 \leq 0 \right\} = \left\{ d \in \mathbb{R}^2 \;\middle|\; d_2 = 0 \right\}.$$

Der Tangentialkegel ist nach Lemma 4.4 eine Teilmenge dieses Kegels. Allerdings gilt für alle $x \in X$ stets $x_1^3 \geq x_2 \geq 0$ und damit auch $x_1 \geq 0$. Also gilt nach Definition von Tangentialrichtungen in $x = 0$ auch $d_1 = \lim_{k\to\infty} [x^k]_1 / t_k \geq 0$ für entsprechende Folgen $\{x^k\}_{k\in\mathbb{N}}$ und $\{t_k\}_{k\in\mathbb{N}}$. Der Tangentialkegel ist daher die echte Teilmenge

$$T_X(0) = \left\{ d \in \mathbb{R}^2 \;\middle|\; d_1 \geq 0,\ d_2 = 0 \right\} \subsetneq L_X(0).$$

Das Problem ist hier, dass die Nebenbedingungen in $x = (0,0)^T$ lokal nicht von der Bedingung $x_2 = 0$ unterscheidbar sind (X ist dort zu "spitz"); um durch Linearisierung genau den Tangentialkegel zu bekommen, brauchen wir also mehr "Luft" in X.

Satz 4.6. *Sei $x \in X$. Existiert ein $v \in \mathbb{R}^n$ mit*

$$\nabla g_i(x)^T v < 0 \qquad \text{für alle } i \in \mathcal{A}_X(x), \tag{4.7}$$

dann ist $T_X(x) = L_X(x)$.

Beweis. Nach Lemma 4.4 ist nur die Inklusion $L_X(x) \subset T_X(x)$ zu zeigen. Sei dazu $d \in L_X(x)$ beliebig. Wir konstruieren nun geeignete Folgen $\{x^k\}_{k\in\mathbb{N}} \subset X$ und $\{t_k\}_{k\in\mathbb{N}} \subset (0,\infty)$. Dafür betrachten wir für festes $\alpha > 0$ und $t > 0$ beliebig den Vektor

$$x_t := x + t(d + \alpha v).$$

Wir zeigen zuerst, dass für t hinreichend klein $x_t \in X$, d. h. $g_i(x_t) \leq 0$ für alle $1 \leq i \leq m$, gilt. Dafür machen wir eine Fallunterscheidung:

(i) $i \in \mathcal{A}_X(x)$: Wegen $d \in L_X(x)$ gilt zunächst $\nabla g_i(x)^T d \leq 0$ und daher

$$\lim_{t \to 0^+} \frac{g_i(x_t) - g_i(x)}{t} = \nabla g_i(x)^T (d + \alpha v) = \nabla g_i(x)^T d + \alpha \nabla g_i(x)^T v < 0.$$

Da der Grenzwert strikt negativ ist, muss wegen $g_i(x) = 0$ für $i \in \mathcal{A}_X(x)$ auch gelten

$$g_i(x_t) = g_i(x_t) - g_i(x) < 0$$

für $t > 0$ klein genug.

(ii) $i \notin \mathcal{A}_X(x)$: Dann ist $g_i(x) < 0$, und wegen der Stetigkeit von g_i und $x_t \to x$ für $t \to 0$ gilt auch $g_i(x_t) < 0$ für $t > 0$ klein genug.

Für alle $1 \leq i \leq m$ existiert also ein $t_i > 0$ mit $g_i(x_t) \leq 0$ für alle $t < t_i$. Also existiert ein $s := \min \{t_i \mid 1 \leq i \leq m\} > 0$ mit $x_t \in X$ für alle $t < s$.

Wir wählen nun eine beliebige Nullfolge $\{t_k\}_{k \in \mathbb{N}} \subset (0, s)$ und setzen $x^k := x_{t_k} \to x$. Für alle $k \in \mathbb{N}$ groß genug ist dann $x^k \in X$; außerdem gilt

$$\frac{x^k - x}{t_k} = d + \alpha v \qquad \text{für alle } k \in \mathbb{N}.$$

Also ist nach Definition $d + \alpha v \in T_X(x)$ für alle $\alpha > 0$. Da nach Lemma 4.1 Tangentialkegel stets abgeschlossen sind, folgt durch Grenzübergang $\alpha \to 0$ auch $d \in T_X(x)$. □

Ein Punkt $x \in X$, für den $T_X(x) = L_X(x)$ gilt, heißt *regulär*; eine Bedingung wie (4.7), die die Regularität eines Punktes garantiert, nennt man *Regularitätsbedingung* oder (auch im Deutschen) *constraint qualification*. Die "triviale" Regularitätsbedingung $T_X(x) = L_X(x)$ wird auch als *Adabie constraint qualification* bezeichnet.

Eine stärkere, aber leichter überprüfbare, Bedingung ist die sogenannte *linear independence constraint qualification* (LICQ).

Folgerung 4.7. *Sei $x \in X$. Ist die Menge $\{\nabla g_i(x) \mid i \in \mathcal{A}_X(x)\} \subset \mathbb{R}^n$ linear unabhängig, dann ist x regulär.*

Beweis. Unter dieser Voraussetzung hat das lineare Gleichungssystem

$$\nabla g_i(x)^T d = \alpha_i, \qquad i \in \mathcal{A}_X(x),$$

vollen (Zeilen-)Rang und damit für beliebige rechte Seite eine (nicht unbedingt eindeutige) Lösung. Wählt man $\alpha_i < 0$ für alle $i \in \mathcal{A}_X(x)$, ist somit (4.7) erfüllt, und die Aussage folgt aus Satz 4.6. □

Wieder sind für konvexe Funktionen stärkere Aussagen möglich.

Folgerung 4.8. *Sei g_i konvex für alle $1 \le i \le m$. Existiert ein $\tilde{x} \in X$ mit*

$$g_i(\tilde{x}) < 0 \qquad \text{für alle } 1 \le i \le m, \tag{4.8}$$

dann ist jeder *Punkt $x \in X$ regulär.*

Beweis. Für $x = \tilde{x}$ ist wegen (4.8) die aktive Menge $\mathcal{A}_X(x)$ leer und damit (4.7) trivialerweise erfüllt. Sei daher $x \neq \tilde{x}$. Aus der Konvexität folgt mit Satz 1.2 (i) für alle $i \in \mathcal{A}_X(x)$

$$\nabla g_i(x)^T(\tilde{x} - x) \le g_i(\tilde{x}) - g_i(x) = g_i(\tilde{x}) < 0$$

und daraus (4.7) mit $v := \tilde{x} - x$. □

Die Bedingung (4.8) wird *Slater-Bedingung* genannt.

Für lineare Nebenbedingungen könnte man erwarten, dass automatisch $T_X(x) = L_X(x)$ gilt, und in der Tat stellt die Linearität an sich eine Reguläritsbedingung dar.

Satz 4.9. *Seien alle g_i affin-linear für alle $1 \le i \le m$, d. h. es existieren $a_i \in \mathbb{R}^n$ und $\alpha_i \in \mathbb{R}$ mit $g_i(x) = a_i^T x - \alpha_i$. Dann ist jeder Punkt $x \in X$ regulär.*

Beweis. Der Beweis ist ein stark vereinfachter Spezialfall von Satz 4.6. Für $x \in X$ beliebig und $d \in L_X(x)$ setze $x^k := x + t_k d \to x$ für $t_k \ge 0$ beliebig. Wieder machen wir die Fallunterscheidung

(i) $i \in \mathcal{A}_X(x)$: Dann gilt $a_i^T x = \alpha_i$ und zusammen mit $a_i^T d \le 0$ wegen $d \in L_X(x)$ folgt

$$a_i^T x^k = a_i^T x + t_k a_i^T d \le \alpha_i \qquad \text{für alle } k \in \mathbb{N}, t_k \ge 0.$$

(ii) $i \notin \mathcal{A}_X(x)$: Dann gilt $a_i^T x < \alpha_i$ und damit auch

$$a_i^T x^k = a_i^T x + t_k a_i^T d < \alpha_i \qquad \text{für alle } k \in \mathbb{N}$$

mit t_k hinreichend klein.

Wählen wir also $\{t_k\}_{k\in\mathbb{N}}$ als hinreichend kleine Nullfolge, dann ist $x^k \in X$ für alle $k \in \mathbb{N}$ sowie

$$\frac{x^k - x}{t_k} = d \qquad \text{für alle } k \in \mathbb{N},$$

d. h. $d \in T_X(x)$. □

Auch Kombinationen dieser Regularitätsbedingungen sind möglich – es reicht zum Beispiel, dass nur diejenigen g_i, die nicht affin-linear sind, die Slater-Bedingung erfüllen.

4.2.2 GLEICHUNGSNEBENBEDINGUNGEN

Wir betrachten nun den Fall von reinen Gleichungsnebenbedingungen, d. h.

$$X = \left\{ x \in \mathbb{R}^n \;\middle|\; h_j(x) = 0, \quad 1 \le j \le p \right\}$$

für $h_j : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar. Da Gleichungsnebenbedingungen stets aktiv sind, ist der entsprechende Linearisierungskegel für $x \in X$ definiert durch

$$L_X(x) = \left\{ d \in \mathbb{R}^n \;\middle|\; \nabla h_j(x)^T d = 0,\ 1 \le j \le p \right\}.$$

Völlig analog zu Lemma 4.4 (nur mit Gleichheit an Stelle der Ungleichung) beweist man nun die folgende Inklusion.

Lemma 4.10. *Für alle $x \in X$ gilt $T_X(x) \subset L_X(x)$.*

Für die umgekehrte Inklusion benötigt man wieder eine Regularitätsbedingung.

Satz 4.11. *Sei $x \in X$. Ist die Menge $\left\{ \nabla h_j(x) \;\middle|\; 1 \le j \le p \right\} \subset \mathbb{R}^n$ linear unabhängig, dann ist x regulär.*

Beweis. Der Beweis folgt im Prinzip dem von Satz 4.6; die Schwierigkeit besteht dabei darin, dass wir mit der zu konstruierenden Folge $\{x^k\}_{k\in\mathbb{N}} \subset X$ den (nichtlinearen) Gleichungen $h_j(x) = 0$ folgen müssen. Dazu addieren wir zu der üblichen "linearen" Konstruktion $x_t = x + td$ einen nichtlinearen (in t) Korrekturterm, den wir – unter der genannten Voraussetzung – mit Hilfe des Satzes über implizite Funktionen erhalten.

Sei dafür $d \in L_X(x)$ beliebig. Ziel ist zu zeigen, dass für $\varepsilon > 0$ klein genug eine Kurve $x : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ existiert mit $x(0) = x$, $x'(0) = d$ und $h_i(x(t)) = 0$ für alle $t \in (-\varepsilon, \varepsilon)$ und $1 \le i \le p$. Zu diesem Zweck definieren wir zunächst die Funktion

$$h : \mathbb{R}^n \to \mathbb{R}^p, \qquad x \mapsto (h_1(x), \dots, h_p(x))^T,$$

und mit ihrer Hilfe die Funktion

$$H : \mathbb{R}^{p+1} \to \mathbb{R}^p, \qquad (y, t) \mapsto h(x + td + \nabla h(x)\, y),$$

wobei $\nabla h(x)$ die Jacobi-Matrix von h bezeichnet (deren Spalten genau die $\nabla h_i(x)$ sind). Beachte, dass $d \in L_X(x) = \ker \nabla h(x)^T$ und $\nabla h(x)\, y \in \operatorname{ran} \nabla h(x) = (\ker \nabla h(x)^T)^\perp$ liegen; wir zerlegen also die Korrektur in zwei orthogonale Anteile.

Wir betrachten nun das nichtlineare Gleichungssystem $H(y, t) = 0$, das wegen $x \in X$ offensichtlich die Lösung $(\bar{y}, \bar{t}) = (0, 0)$ besitzt. Auf diese Gleichung wenden wir nun

den Satz über implizite Funktionen an, um eine Kurve $y(t)$ zu erhalten. Zunächst hat die Funktion $y \mapsto H(y,t)$ für $t > 0$ fest nach der Kettenregel die Jacobi-Matrix

$$H_y(y,t) = h'(x + td + \nabla h(x)y)\nabla h(x) = \nabla h(x + td + \nabla h(x)y)^T \nabla h(x)$$

wegen $\nabla h(x) = h'(x)^T$; analog hat die Funktion $t \mapsto H(y,t)$ für y fest die Ableitung

$$H_t(y,t) = \nabla h(x + td + \nabla h(x)y)^T d.$$

Speziell ist also

$$H_y(0,0) = \nabla h(x)^T \nabla h(x) \in \mathbb{R}^{p \times p}$$

invertierbar, da $\nabla h(x)$ nach Voraussetzung vollen Rang hat. Nach dem Satz über implizite Funktionen existiert daher ein $\varepsilon > 0$ und eine stetig differenzierbare Funktion $y : (-\varepsilon, \varepsilon) \to \mathbb{R}^p$ mit $y(0) = 0$ und $H(y(t),t) = 0$ für alle $t \in (-\varepsilon, \varepsilon)$ sowie

$$0 = H_y(y(t),t)y'(t) + H_t(y(t),t) \qquad \text{für alle } t \in (-\varepsilon, \varepsilon).$$

Aus der Invertierbarkeit von H_y in $(0,0)$ folgt daraus

$$y'(0) = -H_y(0,0)^{-1} H_t(0,0) = -H_y(0,0)^{-1} \nabla h(x)^T d = 0$$

wegen $\nabla h_j(x)^T d = 0$ für alle $1 \le j \le p$ nach Annahme an $d \in L_X(x)$.

Wir definieren nun die gesuchte Kurve durch

$$x(t) := x + td + \nabla h(x)y(t).$$

Dann gilt nach Konstruktion von $y(t)$

$$h(x(t)) = H(y(t),t) = 0 \qquad \text{für alle } t \in (-\varepsilon, \varepsilon)$$

sowie $x(0) = 0$ und $x'(0) = d + \nabla h(x)^T y'(0) = d$. Wählen wir also eine beliebige Nullfolge $\{t_k\}_{k \in \mathbb{N}} \subset (0, \varepsilon)$ und setzen $x^k := x(t_k)$, so gilt $x^k \in X$ für alle $k \in \mathbb{N}$, $x^k \to x(0) = x$ wegen $y(0) = 0$, sowie

$$\lim_{k \to \infty} \frac{x^k - x}{t_k} = \lim_{k \to \infty} \frac{x(t_k) - x(0)}{t_k} = x'(0) = d,$$

d. h. $d \in T_X(x)$. □

Wieder ist die Linearität an sich eine Regularitätsbedingung.

Satz 4.12. *Seien alle h_j affin-linear für alle $1 \le j \le p$, d. h. es existieren $b_j \in \mathbb{R}^n$ und $\beta_j \in \mathbb{R}$ mit $h_j(x) = b_j^T x - \beta_j$. Dann ist jeder Punkt $x \in X$ regulär.*

Beweis. Der Beweis ist völlig analog zu dem von Satz 4.9: Für $x \in X$ und $d \in L_X(x)$ wählen wir wieder $t_k := \frac{1}{k} \to 0$ und $x^k := x + t_k d \to x$. Dann folgt mit $b_j^T x = h_j(x) + \beta_j = \beta_j$ und $b_j^T d = \nabla h_j(x)^T d = 0$ sofort

$$b_j^T x^k = b_j^T x + t_k b_j^T d = \beta_j.$$

Also ist $x^k \in X$ für alle $k \in \mathbb{N}$ sowie $\frac{x^k - x}{t_k} = d$, d. h. $d \in T_X(x)$. □

4.2.3 GEMISCHTE NEBENBEDINGUNGEN

Wir kommen nun zum allgemeinen Fall,

$$X = \left\{ x \in \mathbb{R}^n \;\middle|\; g_i(x) \leq 0,\ 1 \leq i \leq m, \quad h_j(x) = 0,\ 1 \leq j \leq p \right\}$$

für $g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar, für den der Linearisierungskegel gegeben ist durch

$$L_X(x) = \left\{ d \in \mathbb{R}^n \;\middle|\; \nabla g_i(x)^T d \leq 0,\ i \in \mathcal{A}_X(x), \quad \nabla h_j(x)^T d = 0,\ 1 \leq j \leq p \right\}.$$

Die "einfache" Inklusion beweist man wieder völlig analog zu Lemma 4.4, indem man die Nebenbedingungen einzeln betrachtet.

Lemma 4.13. *Für alle $x \in X$ gilt $T_X(x) \subset L_X(x)$.*

Für die andere Richtung kombiniert man nun die obigen Ansätze, wobei man nur darauf achten muss, dass sich Gleichungs- und Ungleichungsrestriktionen nicht in die Quere kommen. Die entsprechende Regularitätsbedingung nennt man *Mangasarian–Fromowitz constraint qualification* (MFCQ).

Satz 4.14. *Sei $x \in X$. Gilt*

(i) die Menge $\left\{ \nabla h_j(x) \;\middle|\; 1 \leq j \leq p \right\} \subset \mathbb{R}^n$ ist linear unabhängig,

(ii) es gibt ein $v \in \mathbb{R}^n$ mit

$$\begin{aligned} \nabla g_i(x)^T v &< 0 \qquad \text{für alle } i \in \mathcal{A}_X(x), \\ \nabla h_j(x)^T v &= 0 \qquad \text{für alle } j \in \{1, \dots, p\}, \end{aligned}$$

dann ist x regulär.

Beweis. Sei $d \in L_X(x)$ beliebig. Dann gilt insbesondere $\nabla h_j(x)^T d = 0$ und damit nach Annahme (ii) auch $\nabla h_j(x)^T (d + \alpha v) = 0$ für alle $\alpha > 0$ und $1 \leq j \leq p$. Wie im Beweis von Satz 4.11 konstruiert man nun mit Hilfe von Annahme (i) für $d + \alpha v$ eine Kurve $x : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ mit $x(0) = x$, $x'(0) = d + \alpha v$, und $h_j(x(t)) = 0$ für alle $1 \leq j \leq p$ und $t \in (-\varepsilon, \varepsilon)$.

Weiter ist $\nabla g_i(x)^T d \leq 0$ und damit nach Annahme (ii) auch $\nabla g_i(x)^T (d + \alpha v) < 0$ für alle $i \in \mathcal{A}_X(x)$. Also gilt für $x(t)$ nach Konstruktion

$$\lim_{t \to 0^+} \frac{g_i(x(t)) - g_i(x(0))}{t} = \nabla g_i(x(0))^T x'(0) = \nabla g_i(x)^T (d + \alpha v) < 0.$$

Für $i \notin \mathcal{A}_X(x)$ folgt aus $g_i(x(0)) = g_i(x) < 0$ und der Stetigkeit von g_i und $x(t)$ auch $g_i(x(t)) < 0$ für alle t klein genug. Wie im Beweis von Satz 4.6 existiert daher ein $s \in (0, \varepsilon)$ mit $g_i(x(t)) < 0$ für alle $t \in (0, s)$ und $1 \leq i \leq m$.

Durch Wahl von $t_k \to 0$, $t_k \in (0, s)$, und $x^k := x(t_k) \to x$ mit $x^k \in X$ für alle $k \in \mathbb{N}$ folgt nun wie zuvor, dass $d + \alpha v \in T_X(x)$ für alle $\alpha > 0$ ist. Aus der Abgeschlossenheit von Tangentialkegeln folgt daher $d \in T_X(x)$. □

Weitere Regularitätsbedingungen erhält man ebenfalls durch geeignete Kombinationen, etwa die "volle" LICQ.

Folgerung 4.15. *Sei $x \in X$. Ist die Menge*

$$\left\{\nabla g_i(x), \nabla h_j(x) \;\middle|\; i \in \mathcal{A}_X(x), 1 \leq j \leq p\right\} \subset \mathbb{R}^n$$

linear unabhängig, dann ist x regulär.

Beweis. Unter dieser Voraussetzung hat das lineare Gleichungssystem

$$\begin{aligned} \nabla g_i(x)^T v &= \alpha_i, \qquad && i \in \mathcal{A}_X(x), \\ \nabla h_j(x)^T v &= \beta_j, && j \in \{1, \dots, p\}, \end{aligned}$$

vollen (Zeilen-)Rang und damit für beliebige $\alpha_i, \beta_j \in \mathbb{R}$ eine (nicht unbedingt eindeutige) Lösung. Wählt man $\alpha_i < 0$ und $\beta_j = 0$, ist somit die MFCQ erfüllt, und die Aussage folgt aus Satz 4.14. □

Ebenso kombiniert man die "globalen" Regularitätsbedingungen.

Folgerung 4.16. *Seien alle g_i konvex und alle h_j affin-linear, d. h. es existieren $b_j \in \mathbb{R}^n$ und $\beta_j \in \mathbb{R}$ mit $h_j(x) = b_j^T x - \beta_j$. Gilt:*

(i) die Menge $\left\{b_j \;\middle|\; 1 \leq j \leq p\right\}$ ist linear unabhängig,

(ii) es existiert ein $\tilde{x} \in X$ mit

$$g_i(\tilde{x}) < 0 \qquad \text{für alle } 1 \leq i \leq m,$$

dann ist jeder *Punkt $x \in X$ regulär.*

Beweis. Wir müssen lediglich noch Annahme (ii) von Satz 4.14 nachprüfen. Dafür betrachten wir wieder für $x \in X$ beliebig die Richtung $v := \tilde{x} - x$. Genau wie im Beweis von Folgerung 4.8 erhält man aus der Konvexität von g_i

$$\nabla g_i(x)^T v \leq g_i(\tilde{x}) - g_i(x) = g_i(\tilde{x}) < 0 \qquad \text{für alle } i \in \mathcal{A}_X(x).$$

Außerdem gilt für alle $x \in X$ wegen $\tilde{x} \in X$ und damit insbesondere $h_j(\tilde{x}) = 0$ auch

$$\nabla h_j(x)^T v = b_j^T \tilde{x} - b_j^T x = \beta_j - \beta_j = 0 \qquad \text{für alle } 1 \leq j \leq p.$$

Also ist die MFCQ erfüllt, und die Aussage folgt aus Satz 4.14. □

Durch Kombination der Beweise von Satz 4.9 und Satz 4.12 erhält man schließlich den folgenden Satz, der erklärt, warum in der linearen Optimierung keine Regularitätsbedingung notwendig war.

Satz 4.17. *Seien alle g_i und h_j affin-linear. Dann ist jeder Punkt $x \in X$ regulär.*

4.3 DIE KKT-BEDINGUNGEN

Ist ein lokaler Minimierer $\bar{x}$ regulär, so kann man aus der abstrakten Optimalitätsbedingung (4.2) eine explizite Charakterisierung von $\bar{x}$ gewinnen. Um dies zu motivieren, betrachten wir den Fall einer einzigen Gleichungsnebenbedingung, $X = \{x \in \mathbb{R}^n \mid h(x) = 0\}$ für $h : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar. Dann ist $L_X(x) = \left\{d \in \mathbb{R}^n \;\middle|\; \nabla h(x)^T d = 0\right\} = \ker \nabla h(x)^T$ ein linearer Unterraum (nämlich der *Tangentialraum* von X in x). Da für Unterräume der Polarkegel mit dem orthogonalen Komplement übereinstimmt, erhalten wir für einen regulären Minimierer $\bar{x}$ aus der äquivalenten Schreibweise (4.3) die Bedingung

$$-\nabla f(\bar{x}) \in T_X(\bar{x})^o = L_X(\bar{x})^o = (\ker \nabla h(x)^T)^\perp = \operatorname{ran} \nabla h(x),$$

denn für jede lineare Abbildung A gilt $(\ker A)^\perp = \operatorname{ran} A^T$. Nach Definition des Bildes existiert also ein $\bar{\lambda} \in \mathbb{R}$ mit

$$-\nabla f(\bar{x}) = \nabla h(\bar{x})\bar{\lambda},$$

womit wir (im Fall $n = 1$) die klassische Lagrange-Multiplikator-Regel für die Minimierung reeller Funktionen unter Gleichungsnebenbedingungen wiedergewonnen haben. (Die dafür "offensichtlich" hinreichende Bedingung $T_X(x)^o = L_X(x)^o$ wird auch *Guignard constraint qualification* genannt.)

Für den allgemeinen Fall inklusive Ungleichungsnebenbedingungen verwenden wir statt der Gleichheit $(\ker A)^\perp = \operatorname{ran} A^T$ das Farkas-Lemma 3.4.

Satz 4.18. *Seien $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und $X \subset \mathbb{R}^n$ von der Form (4.6). Sei $\bar{x} \in X$ ein lokaler Minimierer von f. Ist $\bar{x}$ regulär, so existieren $\bar{\mu} \in \mathbb{R}^m$ und $\bar{\lambda} \in \mathbb{R}^p$ mit*

$$\begin{cases} \nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{\mu}_i \nabla g_i(\bar{x}) + \sum_{j=1}^{p} \bar{\lambda}_j \nabla h_j(\bar{x}) = 0, \\ h_j(\bar{x}) = 0, \quad 1 \le j \le p, \\ \bar{\mu}_i \ge 0, \quad g_i(\bar{x}) \le 0, \quad \bar{\mu}_i g_i(\bar{x}) = 0, \quad 1 \le i \le m. \end{cases} \tag{4.9}$$

Beweis. Aus (4.2) und der Regularität von $\bar{x}$ folgt

$$\nabla f(\bar{x})^T d \ge 0 \qquad \text{für alle } d \in L_X(\bar{x}).$$

Schreiben wir $\nabla g_{\mathcal{A}}(\bar{x})$ für die Matrix, deren Spalten durch $\nabla g_i(\bar{x})$, $i \in \mathcal{A}_X(\bar{x})$, gegeben ist und $\nabla h(\bar{x})$ für die Matrix, deren Spalten aus $\nabla h_j(\bar{x})$, $1 \le j \le p$, bestehen, ist dies äquivalent zu

$$\nabla f(\bar{x})^T d \ge 0 \quad \text{für alle } d \in \mathbb{R}^n \text{ mit } \begin{pmatrix} -\nabla g_{\mathcal{A}}(\bar{x})^T & -\nabla h(\bar{x})^T & \nabla h(\bar{x})^T \end{pmatrix}^T d \ge 0.$$

Nach Lemma 3.4 existiert also ein $u := (w, y^+, y^-) \ge 0$ mit

$$-\nabla g_{\mathcal{A}}(\bar{x})^T w - \nabla h(\bar{x})^T y^+ + \nabla h(\bar{x})^T y^- = \nabla f(\bar{x}).$$

Setzen wir

$$\bar{\lambda}_j := y_j^+ - y_j^- \quad \text{für } 1 \le j \le p, \qquad \bar{\mu}_i := \begin{cases} w_i & i \in \mathcal{A}_X(\bar{x}), \\ 0 & i \notin \mathcal{A}_X(\bar{x}), \end{cases}$$

ist dies genau die erste Zeile von (4.9). Nach Definition gilt dann auch $\bar{\mu}_i g_i(\bar{x}) = 0$ für alle $1 \le i \le m$, woraus zusammen mit der Zulässigkeit von $\bar{x} \in X$ die restlichen Zeilen folgen. □

Analog zur Minimierung reeller Funktionen nennt man $\bar{\lambda}, \bar{\mu}$ *Lagrange-Multiplikatoren*; die Bedingungen (4.9) werden *Karush–Kuhn–Tucker-* oder kurz *KKT-Bedingungen* genannt. Die erste Zeile nennt man auch *Stationaritätsbedingung*, die letzte Zeile *Komplementaritätsbedingung*; man spricht von *strikter Komplementarität*, falls für alle $i \in \mathcal{A}_X(\bar{x})$ gilt $\bar{\mu}_i \ne 0$, d. h. es gilt $\bar{\mu}_i = 0$ *genau* dann, wenn $g_i(\bar{x}) < 0$.

Unter stärkeren Regularitätsbedingungen kann man mehr über die Lagrange-Multiplikatoren aussagen.

Folgerung 4.19. *Sei $\bar{x} \in X$ ein lokaler Minimierer, der die LICQ erfüllt. Dann sind die zugehörigen Lagrange-Multiplikatoren $\bar{\lambda} \in \mathbb{R}^p$, $\bar{\mu} \in \mathbb{R}^m$ eindeutig bestimmt.*

Beweis. Zunächst muss für alle $i \notin \mathcal{A}_X(\bar{x})$ wegen der Komplementaritätsbedingung $\bar{\mu}_i = 0$ gelten. Damit reduziert sich die erste Zeile von (4.9) auf

$$\sum_{i \in \mathcal{A}_X(\bar{x})} \bar{\mu}_i \nabla g_i(\bar{x}) + \sum_{j=1}^{p} \bar{\lambda}_j \nabla h_j(\bar{x}) = -\nabla f(\bar{x}).$$

Da nach Voraussetzung die Vektoren $\nabla g_i(\bar{x})$, $i \in \mathcal{A}_X(\bar{x})$ und $\nabla h_j(\bar{x})$, $1 \le j \le p$, linear unabhängig sind, hat dieses Gleichungssystem eine eindeutige Lösung $\bar{\mu}_i$, $i \in \mathcal{A}_X(\bar{x})$, $\bar{\lambda}_j$, $1 \le j \le p$. Also sind die Lagrange-Multiplikatoren eindeutig. □

Konvexe Probleme (d. h. f *und* X konvex) sind die KKT-Bedingungen sogar hinreichend. Dafür ist nicht mal eine Slater-Bedingung erforderlich.

Satz 4.20. *Seien f und alle g_i konvex und alle h_j affin-linear, d. h. es existieren $b_j \in \mathbb{R}^n$ und $\beta_j \in \mathbb{R}$ mit $h_j(x) = b_j^T x - \beta_j$. Erfüllt ein Tripel $(\bar{x}, \bar{\mu}, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ die KKT-Bedingungen (4.9), dann ist $\bar{x}$ ein globaler Minimierer von f in X.*

Beweis. Aus den KKT-Bedingungen folgt direkt die Zulässigkeit von $\bar{x}$. Sei nun $x \in X$ beliebig. Aus der Konvexität von g_i folgt dann mit Satz 1.2

$$\nabla g_i(\bar{x})^T (x - \bar{x}) \le g_i(x) - g_i(\bar{x}) \le 0 \qquad \text{für alle } i \in \mathcal{A}_X(\bar{x}),$$

und aus der affinen Linearität der h_j

$$\nabla h_j(\bar{x})^T (x - \bar{x}) = b_j^T x - b_j^T \bar{x} = \beta_j - \beta_j = 0 \quad \text{für } 1 \le j \le p.$$

Damit folgt aus der Konvexität von f mit Satz 1.2, der ersten Zeile von (4.9), und der Komplementarität

$$\begin{aligned} f(x) &\ge f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) \\ &= f(\bar{x}) - \sum_{i=1}^{m} \bar{\mu}_i \nabla g_i(\bar{x})^T (x - \bar{x}) - \sum_{j=1}^{p} \bar{\lambda}_j b_j^T (x - \bar{x}) \\ &= f(\bar{x}) - \sum_{i \in \mathcal{A}_X(\bar{x})} \bar{\mu}_i \nabla g_i(\bar{x})^T (x - \bar{x}) \\ &\ge f(\bar{x}). \end{aligned}$$

Also ist $\bar{x}$ ein globaler Minimierer von f in X. □

Sind schließlich auch f und alle g_i affin-linear, so entsprechen die Lagrange-Multiplikatoren genau den dualen Variablen in der linearen Optimierung, und die KKT-Bedingungen liefern eine weitere Herleitung von Satz 3.7.

4.4 BEDINGUNGEN 2. ORDNUNG

Zum Abschluss betrachten wir Optimalitätsbedingungen zweiter Ordnung, wobei wir uns zunächst auf die in der Praxis wesentlich relevanteren hinreichenden Bedingungen konzentrieren. Während im Falle der unrestringierten Optimierung die Krümmung der Zielfunktion (über die positive Definitheit der Hessematrix) ausschlaggebend ist, müssen wir hier gegebenfalls auch die Krümmung der Nebenbedingungen berücksichtigen. Wir betrachten dafür statt f die *Lagrange-Funktion*

$$L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}, \qquad (x, \mu, \lambda) \mapsto f(x) + \sum_{i=1}^{m} \mu_i g_i(x) + \sum_{j=1}^{p} \lambda_j h_j(x) \tag{4.10}$$

und halten für den späteren Gebrauch ein paar wesentliche Eigenschaften fest. Zunächst gilt für alle zulässigen $x \in X$, $\mu \geq 0$, und $\lambda \in \mathbb{R}^p$

$$L(x, \mu, \lambda) = f(x) + \sum_{i=1}^{m} \mu_i g_i(x) + \sum_{j=1}^{p} \lambda_j h_j(x) \leq f(x) \tag{4.11}$$

wegen $g_i(x) \leq 0$ und $h_j(x) = 0$. Erfüllen $(\bar{x}, \bar{\mu}, \bar{\lambda})$ die KKT-Bedingungen (4.9), so gilt wegen der Komplementarität sogar

$$L(\bar{x}, \bar{\mu}, \bar{\lambda}) = f(\bar{x}), \tag{4.12}$$

und die Stationaritätsbedingung in (4.9) ist äquivalent zu

$$\nabla_x L(\bar{x}, \bar{\mu}, \bar{\lambda}) = \nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{\mu}_i \nabla g_i(\bar{x}) + \sum_{j=1}^{p} \bar{\lambda}_j \nabla h_j(\bar{x}) = 0. \tag{4.13}$$

Die Krümmung der Lagrange-Funktion müssen wir jetzt nur entlang geeigneter, „kritischer“, Richtungen untersuchen. Die Grundidee ist anschaulich die folgende: Aktive Nebenbedingungen, für die strikte Komplementarität gilt, sind in gewisser Weise “stark aktiv”; wir erwarten daher, dass sie bei kleinen Änderungen immer noch aktiv bleiben. Wir betrachten also nur solche Richtungen, für die diese Nebenbedingungen aktiv bleiben – in anderen Worten, wir behandeln sie wie Gleichungsnebenbedingungen. Wir zerlegen also für einen KKT-Punkt $(\bar{x}, \bar{\mu}, \bar{\lambda})$ die aktive Menge $\mathcal{A}_X(\bar{x})$ in

$$\begin{aligned} \mathcal{A}_+(\bar{x}, \bar{\mu}) &:= \{ i \in \mathcal{A}_X(\bar{x}) \mid \bar{\mu}_i > 0 \}, \\ \mathcal{A}_0(\bar{x}, \bar{\mu}) &:= \{ i \in \mathcal{A}_X(\bar{x}) \mid \bar{\mu}_i = 0 \}, \end{aligned}$$

d. h. in die Menge der aktiven Nebenbedingungen, für die strikte Komplementarität gilt bzw. nicht gilt, und definieren den *kritischen Kegel*

$$K_X(\bar{x}, \bar{\mu}) := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \nabla g_i(\bar{x})^T d = 0, \; i \in \mathcal{A}_+(\bar{x}, \bar{\mu}), \\ \nabla g_i(\bar{x})^T d \leq 0, \; i \in \mathcal{A}_0(\bar{x}, \bar{\mu}), \\ \nabla h_j(\bar{x})^T d = 0, \; 1 \leq j \leq p \end{array} \right\} \subset L_X(\bar{x}).$$

Ist nun in einem Punkt die Krümmung der Lagrange-Funktion entlang kritischer Richtungen stets positiv, haben wir ein striktes Minimum.

Satz 4.21 (hinreichende Bedingungen 2. Ordnung). *Seien f, g_i, h_j zweimal stetig differenzierbar. Erfüllt $(\bar{x}, \bar{\mu}, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ die KKT-Bedingungen (4.9) und gilt*

$$d^T \nabla^2_{xx} L(\bar{x}, \bar{\mu}, \bar{\lambda}) d > 0 \qquad \text{für alle } d \in K_X(\bar{x}, \bar{\mu}) \setminus \{0\}, \tag{4.14}$$

dann ist $\bar{x}$ ein strikter lokaler Minimierer von f in X.

Beweis. Angenommen, die Voraussetzungen sind erfüllt, aber $\bar{x} \in X$ ist kein strikter lokaler Minimierer. Dann existiert eine Folge $\{x^k\}_{k\in\mathbb{N}} \subset X \setminus \{\bar{x}\}$ mit $x^k \to \bar{x}$ und $f(x^k) \leq f(\bar{x})$ für alle $k \in \mathbb{N}$. Wir definieren nun eine Folge $\{d^k\}_{k\in\mathbb{N}}$ durch

$$d^k := \frac{x^k - \bar{x}}{\|x^k - \bar{x}\|}.$$

Da wegen $\|d^k\| = 1$ diese Folge beschränkt ist, existiert eine konvergente Teilfolge $\{d^k\}_{k\in K}$, $K \subset \mathbb{N}$ unendlich, mit $d^k \to d$ für ein $d \in \mathbb{R}^n$ mit $\|d\| = 1$, d. h. $d \neq 0$. Mit $t_k := \|x^k - \bar{x}\| \to 0$ erfüllt d also die Definition einer Tangentialrichtung; aus Lemma 4.13 folgt daher $d \in T_X(\bar{x}) \subset L_X(\bar{x})$. Insbesondere gilt

$$\nabla g_i(\bar{x})^T d \leq 0 \qquad \text{für alle } i \in \mathcal{A}_X(\bar{x}), \tag{4.15}$$

$$\nabla h_j(\bar{x})^T d = 0 \qquad \text{für alle } 1 \leq j \leq p. \tag{4.16}$$

Wir zeigen nun, dass sogar $d \in K_X(\bar{x}, \bar{\mu})$ gilt. Angenommen, das ist nicht der Fall, d. h. es existiert ein $i_+ \in \mathcal{A}_+(\bar{x}, \bar{\mu})$ mit $\nabla g_{i_+}(\bar{x})^T d < 0$. Dann folgt aus Satz 1.11 und $f(x^k) \leq f(\bar{x})$

$$\nabla f(\xi^k)^T (x^k - \bar{x}) = f(x^k) - f(\bar{x}) \leq 0.$$

Wieder gilt $\xi^k \to \bar{x}$ wegen $x^k \to \bar{x}$, und Division durch $\|x^k - \bar{x}\| > 0$ und Grenzübergang $K \ni k \to \infty$ liefert

$$0 \geq \nabla f(\bar{x})^T d = -\sum_{i=1}^{m} \bar{\mu}_i \nabla g_i(\bar{x})^T d - \sum_{j=1}^{p} \bar{\lambda}_j \nabla h_j(\bar{x})^T d \geq -\bar{\mu}_{i_+} \nabla g_{i_+}(\bar{x})^T d > 0$$

wegen $\nabla_x L(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0$ und (4.15), (4.16) und damit ein Widerspruch.

Also ist $d \in K_X(\bar{x}, \bar{\mu}) \setminus \{0\}$. Wegen $x^k \in X$ folgt dann aus Satz 1.11 für $x \mapsto L(x, \bar{\mu}, \bar{\lambda})$ zusammen mit (4.11)–(4.13)

$$\begin{aligned} f(\bar{x}) \geq f(x^k) &\geq L(x^k, \bar{\mu}, \bar{\lambda}) \\ &= L(\bar{x}, \bar{\mu}, \bar{\lambda}) + \nabla_x L(\bar{x}, \bar{\mu}, \bar{\lambda})^T (x^k - \bar{x}) + \frac{1}{2}(x^k - \bar{x})^T \nabla^2_{xx} L(\xi^k, \bar{\mu}, \bar{\lambda})^T (x^k - \bar{x}) \\ &= f(\bar{x}) + \frac{1}{2}(x^k - \bar{x})^T \nabla^2_{xx} L(\xi^k, \bar{\mu}, \bar{\lambda})^T (x^k - \bar{x}). \end{aligned}$$

Division durch $\|x^k - \bar{x}\|^2 > 0$ und Grenzübergang $K \ni k \to \infty$ liefert dann

$$d^T \nabla^2_{xx} L(\bar{x}, \bar{\mu}, \bar{\lambda}) d \leq 0 \qquad \text{für } d \in K_X(\bar{x}, \bar{\mu}),$$

im Widerspruch zu (4.14). □

Notwendige Optimalitätsbedingungen bedürfen wieder einer Regularitätsbedingung.

Satz 4.22. *Seien f, g_i, h_j zweimal stetig differenzierbar. Sei $\bar{x} \in X$ ein lokaler Minimierer von f, der die LICQ erfüllt. Dann gilt für die zugehörigen eindeutigen Lagrange-Multiplikatoren $\bar{\mu} \in \mathbb{R}^m$ und $\bar{\lambda} \in \mathbb{R}^p$*

$$d^T \nabla^2_{xx} L(\bar{x}, \bar{\mu}, \bar{\lambda}) d \geq 0 \qquad \textit{für alle } d \in K_X(\bar{x}, \bar{\mu}). \tag{4.17}$$

Beweis. Aus der LICQ folgt insbesondere die Regularität von $\bar{x}$ und damit aus Satz 4.18 die Existenz von Lagrange-Multiplikatoren, die nach Folgerung 4.19 eindeutig sind.

Angenommen, es existiert ein $d \in K_X(\bar{x}, \bar{\mu}) \setminus \{0\}$ mit

$$d^T \nabla^2_{xx} L(\bar{x}, \bar{\mu}, \bar{\lambda}) d < 0.$$

Ähnlich wie in Satz 4.14 konstruieren wir nun für dieses d eine zulässige Kurve $x(t) \in X$ so, dass für $t > 0$ klein genug $x(t) \in X$ aber $f(x(t)) < f(\bar{x})$ gilt. Dafür zerlegen wir die „biaktive Menge“ $\mathcal{A}_0(\bar{x}, \bar{\mu})$ weiter in

$$\begin{aligned}
\mathcal{A}_0^-(\bar{x}, \bar{\mu}, d) &:= \left\{ i \in \mathcal{A}_0(\bar{x}, \bar{\mu}) \;\middle|\; \nabla g_i(\bar{x})^T d < 0 \right\}, \\
\mathcal{A}_0^0(\bar{x}, \bar{\mu}, d) &:= \left\{ i \in \mathcal{A}_0(\bar{x}, \bar{\mu}) \;\middle|\; \nabla g_i(\bar{x})^T d = 0 \right\}.
\end{aligned}$$

Nun sind nach der LICQ insbesondere die Vektoren

$$\nabla g_i(\bar{x}), \quad i \in \mathcal{A}_+(\bar{x}, \bar{\mu}) \cup \mathcal{A}_0^0(\bar{x}, \bar{\mu}, d), \qquad \nabla h_j(\bar{x}), \quad 1 \leq j \leq p,$$

linear unabhängig. Weiter ist nach Definition $\nabla g_i(\bar{x})^T d < 0$ für $i \in \mathcal{A}_0^-(\bar{x}, \bar{\mu}, d)$. Wir können daher wie in Satz 4.14 eine Kurve $x : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ konstruieren mit $x(0) = \bar{x}$, $x'(0) = d$, und

$$\begin{cases}
h_j(x(t)) = 0, & 1 \leq j \leq p, \\
g_i(x(t)) = 0, & i \in \mathcal{A}_0^0(\bar{x}, \bar{\mu}, d) \cup \mathcal{A}_+(\bar{x}, \bar{\mu}) \\
g_i(x(t)) < 0, & i \in \mathcal{A}_0^-(\bar{x}, \bar{\mu}, d), \\
g_i(x(t)) < 0, & i \notin \mathcal{A}_X(\bar{x}),
\end{cases} \tag{4.18}$$

für alle $t \in (-\varepsilon, \varepsilon)$. Insbesondere ist dann $x(t) \in X$.

Wir setzen nun $\varphi(t) := L(x(t), \bar{\mu}, \bar{\lambda})$. Da f, g_i, und h_j zweimal stetig differenzierbar sind, gilt dies nach dem Satz über implizite Funktionen auch für $x(t)$ und damit für φ. Mit Hilfe der Ketten- und Produktregel erhalten wir daher

$$\begin{aligned}\varphi'(t) &= x'(t)^T \nabla_x L(x(t), \bar{\mu}, \bar{\lambda}),\\ \varphi''(t) &= x''(t)^T \nabla_x L(x(t), \bar{\mu}, \bar{\lambda}) + x'(t) \nabla^2_{xx} L(x(t), \bar{\mu}, \bar{\lambda}) x'(t).\end{aligned}$$

Für $t = 0$ folgt daraus mit (4.11)–(4.13) und der Wahl von d

$$\begin{aligned}\varphi(0) &= L(\bar{x}, \bar{\mu}, \bar{\lambda}) = f(\bar{x}),\\ \varphi'(0) &= d^T \nabla_x L(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0,\\ \varphi''(0) &= d^T \nabla^2_{xx} L(x(t), \bar{\mu}, \bar{\lambda}) d < 0.\end{aligned}$$

Wegen der Stetigkeit muss daher auch $\varphi''(t) < 0$ für $t > 0$ klein genug gelten. Aus Satz 1.11 folgt daher die Existenz eines $\vartheta_t \in (0, t)$ mit

$$\varphi(t) = \varphi(0) + t\varphi'(0) + \frac{t^2}{2}\varphi''(\vartheta_t) < \varphi(0).$$

Zusammen mit (4.11) sowie (4.18) (beachte $\bar{\mu}_i = 0$ für $i \in \mathcal{A}_0(\bar{x}, \bar{\mu})$ nach Definition und für $i \notin \mathcal{A}_X(\bar{x})$ nach Komplementarität aus den ebenfalls notwendigen KKT-Bedingungen) erhalten wir also

$$f(\bar{x}) = \varphi(0) > \varphi(t) = L(x(t), \bar{\mu}, \bar{\lambda}) = f(x(t))$$

für alle $t > 0$ hinreichend klein, weshalb $\bar{x}$ kein lokaler Minimierer sein kann. □

5 LAGRANGE-DUALITÄT

Ähnlich wie für lineare Optimierungsprobleme in Kapitel 3 kann man auch für nichtlineare Optimierungsprobleme eine Dualitätstheorie herleiten (die freilich durch die deutlich schwächere Struktur im Allgemeinen weniger befriedigend bleibt). Ausgangspunkt ist wieder die Lagrange-Funktion

$$L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}, \qquad (x, \mu, \lambda) \mapsto f(x) + \sum_{i=1}^{m} \mu_i g_i(x) + \sum_{j=1}^{p} \lambda_j h_j(x),$$

für die nach (4.11) gilt

$$L(x, \mu, \lambda) \leq f(x) \qquad \text{für alle } x \in X, \mu \geq 0, \lambda \in \mathbb{R}^p. \tag{5.1}$$

Nehmen wir nun das Infimum über alle $x \in X$, folgt

$$\inf_{x \in \mathbb{R}^n} L(x, \mu, \lambda) \leq \inf_{x \in X} L(x, \mu, \lambda) \leq \inf_{x \in X} f(x) \qquad \text{für alle } \mu \geq 0, \lambda \in \mathbb{R}^p.$$

Die linke Seite stellt daher eine untere Schranke für f auf X dar, und die beste Schranke erhalten wir, indem wir das Supremum über alle $\mu \geq 0$ und $\lambda \in \mathbb{R}^p$ bilden.

Satz 5.1 (schwache Dualität). *Setze*

$$q(\mu, \lambda) := \inf_{x \in \mathbb{R}^n} L(x, \mu, \lambda).$$

Dann gilt

$$\sup_{\mu \geq 0, \lambda} q(\mu, \lambda) \leq \inf_{x \in X} f(x). \tag{5.2}$$

Man nennt auch hier wieder $\sup_{\mu \geq 0, \lambda} q(\mu, \lambda)$ das *(Lagrange-)duale Problem*. Der Vorteil hier ist die deutlich einfachere Struktur der Nebenbedingung $\mu \geq 0$. Beachte aber, dass das Infimum in der Definition von q nicht endlich sein muss, d. h. $q(\mu, \lambda) = -\infty$ ist möglich – für das Supremum ist das aber nur ein Problem, wenn $q \equiv -\infty$ ist, die Lagrange-Funktion also für *alle* $\mu \geq 0, \lambda$ bezüglich x nach unten unbeschränkt ist. Wir bezeichnen

$$\operatorname{dom} q := \left\{ (\mu, \lambda) \in \mathbb{R}^m \times \mathbb{R}^p \;\middle|\; \mu \geq 0,\ q(\mu, \lambda) > -\infty \right\}$$

als den *effektiven Definitionsbereich* von q.

Im Fall linearer Zielfunktion und Nebenbedingung erhalten wir daraus wieder die klassische Dualität.

Beispiel 5.2. Betrachte das lineare Optimierungsproblem

$$\min_{x\in\mathbb{R}^n} c^T x \qquad \text{mit } Ax \le b.$$

Da wir keine Gleichungsnebenbedingungen haben, vereinfacht sich die zugehörige Lagrange-Funktion dann zu

$$L(x,\mu) = c^T x + \mu^T(Ax - b) = (c + A^T\mu)^T x - \mu^T b.$$

Da $x \in \mathbb{R}^n$ beliebig sein kann, folgt daraus offensichtlich

$$q(\mu) = \inf_{x\in\mathbb{R}^n} L(x,\mu) = \begin{cases} -\mu^T b & \text{falls } A^T\mu = -c, \\ -\infty & \text{falls } A^T\mu \neq -c. \end{cases}$$

Das Supremum wird also sicher nur für $\mu \in \operatorname{dom} q = \left\{\mu \in \mathbb{R}^m \;\middle|\; \mu \ge 0, A^T\mu = c\right\}$ angenommen. Damit ist das Lagrange-duale Problem genau (LD),

$$\max_{\mu\in\mathbb{R}^m} -b^T\mu \qquad \text{mit } A^T\mu = -c, \mu \ge 0.$$

Im allgemeinen ist die explizite Berechnung der dualen Funktion wegen der fehlenden Nebenbedingung $x \in X$ zwar einfacher als die Lösung des ursprünglichen (primalen) Minimierungsproblem aber trotzdem nur in Spezialfällen möglich. Man kann aber zumindest elementare Eigenschaften zeigen.

Lemma 5.3. *Es gilt stets*

(i) $\operatorname{dom} q$ *ist konvex;*

(ii) $q : \operatorname{dom} q \to \mathbb{R}$ *ist konkav (d. h.* $-q$ *ist konvex).*

Beweis. Seien $(\mu,\lambda), (\tilde\mu,\tilde\lambda) \in \operatorname{dom} q$ und $t \in (0,1)$ beliebig. Da die Lagrange-Funktion für festes $x \in \mathbb{R}^n$ linear in μ und λ ist, gilt wegen $f(x) = tf(x) + (1-t)f(x)$ offensichtlich

$$L(x, t\mu + (1-t)\tilde\mu, t\lambda + (1-t)\tilde\lambda) = tL(x,\mu,\lambda) + (1-t)L(x,\tilde\mu,\tilde\lambda).$$

Nehmen wir auf beiden Seiten das Infimum über alle $x \in \mathbb{R}^n$ und schätzen das Infimum über die rechte Summe nach unten durch die Summe der Infima ab, erhalten wir

$$\begin{aligned} q(t\mu + (1-t)\tilde\mu, t\lambda + (1-t)\tilde\lambda) &= \inf_{x\in\mathbb{R}^n} L(x, t\mu + (1-t)\tilde\mu, t\lambda + (1-t)\tilde\lambda) \\ &\ge t \inf_{x\in\mathbb{R}^n} L(x,\mu,\lambda) + (1-t) \inf_{x\in\mathbb{R}^n} L(x,\tilde\mu,\tilde\lambda) \\ &= tq(\mu,\lambda) + (1-t)q(\tilde\mu,\tilde\lambda) > -\infty \end{aligned}$$

wegen $(\mu,\lambda), (\tilde\mu,\tilde\lambda) \in \operatorname{dom} q$. Außerdem gilt $t\mu + (1-t)\tilde\mu \ge 0$. Also ist $\operatorname{dom} q$ konvex. Durch Umstellen folgt aus der Ungleichung auch die Konvexität von $-q$. □

Dies bedeutet, dass das duale Problem äquivalent ist zu dem *konvexen* Minimierungsproblem

$$\min_{\mu\geq 0,\lambda} -q(\mu,\lambda),$$

d. h. jede Lösung ist stets eine globale Lösung. Daraus folgt auch, dass anders als in der linearen Optimierung die weitere Dualisierung des dualen Problems im Allgemeinen *nicht* das primale (nichtkonvexe) Problem wiederherstellen wird. (Man kann dies aber zur Gewinnung einer in gewissem Sinne „optimalen“ konvexen Näherung anwenden.)

Ist das primale Problem ebenfalls konvex, sind stärkere Dualitätsaussagen möglich.

Satz 5.4. *Seien f und alle g_i konvex und alle h_j affin-linear. Dann sind äquivalent:*

(i) $(\bar{x},\bar{\mu},\bar{\lambda})$ *erfüllen die KKT-Bedingungen (4.9);*

(ii) $(\bar{x},\bar{\mu},\bar{\lambda})$ *ist ein* Sattelpunkt *der Lagrange-Funktion, d. h.* $\bar{\mu}\geq 0$ *und*

$$\sup_{\mu\geq 0,\lambda} L(\bar{x},\mu,\lambda) \leq L(\bar{x},\bar{\mu},\bar{\lambda}) \leq \inf_{x\in\mathbb{R}^n} L(x,\bar{\mu},\bar{\lambda}). \tag{5.3}$$

Beweis. (i) $\Rightarrow$ (ii): Erfüllen $(\bar{x},\bar{\mu},\bar{\lambda})$ die KKT-Bedingungen, so ist insbesondere $\bar{x}\in X$ und $\bar{\mu}\geq 0$. Aus (4.12) und (4.11) folgt dann

$$L(\bar{x},\bar{\mu},\bar{\lambda}) = f(\bar{x}) \geq L(\bar{x},\mu,\lambda) \qquad \text{für alle } \mu\geq 0, \lambda\in\mathbb{R}^p.$$

Supremum über alle $\mu\geq 0,\lambda\in\mathbb{R}^p$ ergibt die erste Ungleichung in (5.3).

Nach (4.13) gilt weiterhin $\nabla_x L(\bar{x},\bar{\mu},\bar{\lambda}) = 0$. Da nach Annahme die Funktion $x\mapsto L(x,\bar{\mu},\bar{\lambda})$ konvex ist, hat diese Funktion also nach Satz 1.9 ein *globales* Minimum in $\bar{x}$, d. h. es gilt

$$L(\bar{x},\bar{\mu},\bar{\lambda}) \leq L(x,\bar{\mu},\bar{\lambda}) \qquad \text{für alle } x\in\mathbb{R}^n.$$

Infimum über alle $x\in\mathbb{R}^n$ ergibt die zweite Ungleichung in (5.3).

(ii) $\Rightarrow$ (i): Ist umgekehrt $(\bar{x},\bar{\mu},\bar{\lambda})$ ein Sattelpunkt der Lagrange-Funktion, so folgt analog zu oben aus der zweiten Ungleichung in (5.3) und Satz 1.9, dass die notwendigen Optimalitätsbedingungen $\nabla_x L(\bar{x},\bar{\mu},\bar{\lambda}) = 0$ erfüllt sind. Weiter folgt aus der ersten Ungleichung in (5.3) für alle $\mu\geq 0,\lambda\in\mathbb{R}^p$

$$\sum_{i=1}^m \mu_i g_i(\bar{x}) + \sum_{j=1}^p \lambda_j h_j(\bar{x}) \leq \sum_{i=1}^m \bar{\mu}_i g_i(\bar{x}) + \sum_{j=1}^p \bar{\lambda}_j h_j(\bar{x}).$$

Also ist die linke Seite beschränkt für alle $\mu\geq 0,\lambda\in\mathbb{R}^p$, was nur möglich ist für $g_i(\bar{x})\leq 0$ und $h_j(\bar{x}) = 0$. Wähle nun $\mu = 0$ und $\lambda = \bar{\lambda}$. Dann erhalten wir $\sum_{i=1}^m \bar{\mu}_i g_i(\bar{x}) \geq 0$, was wegen $\bar{\mu}_i g_i(\bar{x}) \leq 0$ nur möglich ist für $\bar{\mu}_i g_i(\bar{x}) = 0$ für alle $1\leq i\leq m$. Also sind die KKT-Bedingungen (4.9) erfüllt. □

Daraus folgt sofort die starke Dualität für konvexe Optimierungsprobleme.

Satz 5.5 (starke Dualität). *Seien f und alle g_i konvex und alle h_j affin-linear. Existiert eine Lösung $\bar{x} \in X$ des primalen Problems $\min_{x \in X} f(x)$ die eine Regularitätsbedingung erfüllt, so hat auch das duale Problem eine Lösung $\bar{\mu} \geq 0, \bar{\lambda} \in \mathbb{R}^p$, und es gilt*

$$\max_{\mu \geq 0, \lambda} q(\mu, \lambda) = \min_{x \in X} f(x).$$

Beweis. Nach Annahme und Satz 4.18 existieren $(\bar{x}, \bar{\mu}, \bar{\lambda})$, die die KKT-Bedingungen erfüllen. Nach Satz 5.4 ist dies auch ein Sattelpunkt der Lagrange-Funktion, und aus (5.3) zusammen mit (4.12) folgt daher

$$\inf_{x \in X} f(x) = f(\bar{x}) = L(\bar{x}, \bar{\mu}, \bar{\lambda}) \leq \inf_{x \in \mathbb{R}^n} L(x, \bar{\mu}, \bar{\lambda}) = q(\bar{\mu}, \bar{\lambda}) \leq \sup_{\mu \geq 0, \lambda} q(\mu, \lambda) \leq \inf_{x \in X} f(x),$$

wobei die letzte Ungleichung aus der schwachen Dualität (Satz 5.1) folgt. Also gelten alle Ungleichungen mit Gleichheit, und das Maximum des dualen Problems wird in den Lagrange-Multiplikatoren $\bar{\mu} \geq 0, \bar{\lambda}$ angenommen. □

6 GRADIENTENVERFAHREN

Wir kommen nun zu Verfahren zur numerischen Lösung von Optimierungsproblemen mit Nebenbedingungen. Die Schwierigkeit hier ist natürlich, die Zulässigkeit der Lösung zu garantieren, was mit klassischen Verfahren der unbeschränkten Optimierung wie dem Gradientenverfahren nicht garantiert ist. In Spezialfällen ist es jedoch möglich, dies durch eine geeignete Modifikation zu erreichen.

6.1 PROJIZIERTE GRADIENTENVERFAHREN

Ist die zulässige Menge C nichtleer, konvex, und abgeschlossen, dann existiert nach Satz 3.2 für jedes $x \in \mathbb{R}^n$ eine eindeutige Projektion $\operatorname{proj}_C(x) \in C$. Eine naheliegende Idee ist in dem Fall, nach jedem Gradientenschritt einfach wieder auf die zulässige Menge zu projizieren. Um zu zeigen, wann das funktioniert, formulieren wir zunächst die notwendige Optimalitätsbedingung um.

Satz 6.1. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und $C \subset \mathbb{R}^n$ nichtleer, konvex, und abgeschlossen. Ist $\bar{x} \in C$ ein lokaler Minimierer von f auf C, dann gilt für jedes $\gamma > 0$*

$$\bar{x} = \operatorname{proj}_C(\bar{x} - \gamma \nabla f(\bar{x})). \tag{6.1}$$

Beweis. Nach dem Projektionssatz (Satz 3.2) ist $\bar{x} = \operatorname{proj}_C(\bar{x} - \gamma \nabla f(\bar{x}))$ genau dann, wenn für alle $x \in C$ gilt

$$0 \leq \langle \bar{x} - (\bar{x} - \gamma \nabla f(\bar{x})), x - \bar{x} \rangle = \gamma \langle \nabla f(\bar{x}), x - \bar{x} \rangle.$$

Aber das ist wegen $\gamma > 0$ und Satz 4.3 eine notwendige Bedingung für den Minimierer $\bar{x} \in C$ (und sogar hinreichend, falls f konvex ist). □

Der gesuchte Minimierer kann also durch eine Fixpunktgleichung charakterisiert werden; die zugehörige Fixpunktiteration lautet

$$x^{k+1} = \operatorname{proj}_C(x^k - \gamma \nabla f(x^k)).$$

Lässt man eine variable Schrittweite γ_k zu, erhält man daraus das *projizierte Gradientenverfahren.*

Algorithmus 6.1 : Projiziertes Gradientenverfahren

Input : $x^0 \in C$, $\varepsilon > 0$

for $k = 0, \ldots$ **do**
  Wähle $\gamma_k > 0$
  Setze $x^{k+1} = \operatorname{proj}_C(x^k - \gamma_k \nabla f(x^k))$
  if $\|x^{k+1} - x^k\| < \varepsilon$ **then**
    return x^k

Das Abbruchkriterium ist natürlich motiviert durch Satz 6.1: ist $x^{k+1} = x^k$, so erfüllt x^k nach Konstruktion die Optimalitätsbedingung (6.1). Die folgende Eigenschaft der Projektion ist für das Funktionieren des Verfahrens wesentlich.

Lemma 6.2. *Sei $C \subset \mathbb{R}^n$ nichtleer, konvex, und abgeschlossen. Dann gilt*

$$\|\operatorname{proj}_C(x) - \operatorname{proj}_C(y)\|_2 \leq \|x - y\|_2 \qquad \text{für alle } x, y \in \mathbb{R}^n.$$

Beweis. Für $x, y \in \mathbb{R}^n$ können wir mit Hilfe der produktiven Null schreiben

$$x - y = \operatorname{proj}_C(x) - \operatorname{proj}_C(y) + (x - \operatorname{proj}_C(x)) + (\operatorname{proj}_C(y) - y).$$

Aus dem Satz von Pythagoras folgt dann

$$\begin{aligned}\|x - y\|_2^2 = \|\operatorname{proj}_C(x) - \operatorname{proj}_C(y)\|_2^2 + \|(x - \operatorname{proj}_C(x)) + (\operatorname{proj}_C(y) - y)\|_2^2 \\ + 2\langle x - \operatorname{proj}_C(x), \operatorname{proj}_C(x) - \operatorname{proj}_C(y)\rangle \\ + 2\langle \operatorname{proj}_C(y) - y, \operatorname{proj}_C(x) - \operatorname{proj}_C(y)\rangle.\end{aligned}$$

Aus Satz 3.2 folgt aber für $z = x$ und $\tilde{x} = \operatorname{proj}_C(y) \in C$ bzw. $z = y$ und $\tilde{x} = \operatorname{proj}_C(x) \in C$

$$\begin{aligned}\langle \operatorname{proj}_C(x) - x, \operatorname{proj}_C(y) - \operatorname{proj}_C(x)\rangle \geq 0, \\ \langle \operatorname{proj}_C(y) - y, \operatorname{proj}_C(x) - \operatorname{proj}_C(y)\rangle \geq 0,\end{aligned}$$

und damit

$$\|x - y\|_2^2 \geq \|\operatorname{proj}_X(x) - \operatorname{proj}_X(y)\|_2^2. \qquad \square$$

Die Projektion ist also Lipschitz-stetig mit Konstante 1 (man sagt dafür auch *nichtexpansiv.*

Die Durchführbarkeit des Verfahrens steht und fällt natürlich damit, die Projektion auf C effizient berechnen zu können. In einfachen Fällen kann man sogar geschlossene Lösungen angeben.

Beispiel 6.3. Die Projektion für $z \in X$ ist explizit gegeben für

(i) Quader

$$C = \{x \in \mathbb{R}^n \mid x_i \in [a_i, b_i],\ i = 1, \dots, n\}.$$

Wegen

$$\min_{x \in C} \|x - z\|_2^2 = \min_{x_i \in [a_i, b_i]} \sum_{i=1}^{n} (x_i - z_i)^2$$

kann das Minimum komponentenweise bestimmt werden durch Minimierung einer quadratischen Funktion auf dem Intervall $[a_i, b_i]$; eine einfache Rechnung zeigt

$$[\mathrm{proj}_C(z)]_i = \begin{cases} a_i & \text{falls } z_i < a_i, \\ z_i & \text{falls } z_i \in [a_i, b_i], \\ b_i & \text{falls } z_i > b_i. \end{cases}$$

(ii) Kugeln mit Radius $r > 0$:

$$C = \{x \in \mathbb{R}^n \mid \|x\|_2 \le r\}.$$

Ist $z \in C$, so gilt offensichtlich $\mathrm{proj}_C(z) = z$. Andererseits folgt aus der Cauchy–Schwarz-Ungleichung

$$\|x - z\|_2^2 = \|x\|_2^2 - 2\langle x, z\rangle + \|z\|_2^2 \ge \|x\|_2^2 - 2\|x\|_2\|z\|_2 + \|z\|_2^2$$

mit Gleichheit, falls $x = \lambda z$ für ein $\lambda \in \mathbb{R}$ gilt. Ist nun $z \notin C$, so gilt $\|z\|_2 > r$, weshalb im Minimum $r = \|\mathrm{proj}_C(z)\|_2 = \lambda\|z\|_2$ gelten muss (sonst folgt aus der Projektionsungleichung $x = z \notin C$, was unmöglich ist). Also ist

$$\mathrm{proj}_C(z) = \frac{rz}{\max\{r, \|z\|_2\}} = \begin{cases} z & \text{falls } \|z\|_2 \le r, \\ \frac{r}{\|z\|_2} z & \text{sonst.} \end{cases}$$

(Dies ist erstmal nur heuristisch, aber man prüft leicht nach, dass diese Wahl die Ungleichung im Projektionssatz erfüllt.)

Viele weitere Projektionsformeln findet man in [Cegielski 2012].

Es bleibt, die Liniensuche für γ_k durchzuführen, etwa mit der Armijo-Regel. Dies ist deutlich komplizierter als für das unrestringierte Gradientenverfahren, da die Projektion einen hinreichenden Abstieg des reinen Gradientschrittes reduzieren kann; siehe zum Beispiel [Geiger & Kanzow 2002, Kapitel 5.8]. Hier machen wir uns die Sache einfach und setzen Lipschitz-stetige Differenzierbarkeit (L-Glattheit) von f voraus, da wir dann konstante Schrittweiten verwenden können.

Satz 6.4. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ L-glatt und $C \subset \mathbb{R}^n$ nichtleer, konvex, und abgeschlossen. Sei*

$\{x^k\}_{k\in\mathbb{N}}$ erzeugt aus Algorithmus 6.1 mit

$$0 < \varepsilon \le \gamma_k \le \frac{1}{L} \qquad \text{für alle } k \in \mathbb{N}.$$

Existiert ein Minimierer $\bar{x} \in C$ von f über C, dann erfüllt jeder Häufungspunkt von $\{x^k\}_{k\in\mathbb{N}}$ die Optimalitätsbedingung (4.4) und es gilt

$$\min_{0\le m\le k} \|x^m - \operatorname{proj}_C(x^m - \gamma\nabla f(x^m))\| \le \sqrt{\frac{2(f(x^0) - f(\bar{x}))}{L(k+1)}}.$$

Beweis. Der Beweis ist völlig analog zum klassischen Gradientenverfahren. Da f L-glatt ist, können wir das Abstiegslemma (Folgerung 1.13) anwenden auf $x = x^{k+1}$ und $y = x^k$ und erhalten

$$f(x^{k+1}) \le f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k\rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2.$$

Weiter folgt nach Definition von x^{k+1} aus dem Projektionssatz für $x = x^k \in C$

$$\langle x^{k+1} - x^k + \gamma_k \nabla f(x^k), x^k - x^{k+1}\rangle \ge 0$$

und damit nach Umsortieren

$$\langle \nabla f(x^k), x^{k+1} - x^k\rangle \le -\frac{1}{\gamma_k}\|x^{k+1} - x^k\|^2.$$

Zusammen ergibt das

$$f(x^{k+1}) \le f(x^k) - \left(\frac{1}{\gamma_k} - \frac{L}{2}\right)\|x^{k+1} - x^k\|^2 \le -\frac{L}{2}\|x^{k+1} - x^k\|^2 \tag{6.2}$$

nach Annahme an γ_k. Also ist $\{f(x^k)\}_{k\in\mathbb{N}}$ monoton fallend. Weiter folgt mit Teleskopsumme

$$\sum_{m=0}^{k} \|x^{m+1} - x^m\|^2 \le \frac{2}{L}(f(x^0) - f(x^k)) \le \frac{2}{L}(f(x^0) - f(\bar{x})).$$

Abschätzen aller Summanden auf der linken Seite durch deren Minimum ergibt die behauptete Ungleichung. Außerdem ist die Summe beschränkt für $k \in \mathbb{N}$, woraus $\|x^{k+1} - x^k\|^2 \to 0$ für $k \to \infty$ folgt. Sei nun $\hat{x}$ ein Häufungspunkt von $\{x^k\}_{k\in\mathbb{N}}$, d. h. $x^k \to \hat{x} \in C$ (wegen $x^k \in C$ und C abgeschlossen) für eine Teilfolge $K \ni k \to \infty$. Da $\{\gamma_k\}_{k\in\mathbb{N}} \subset [\varepsilon, \frac{1}{L}]$ beschränkt ist, können wir – notfalls nach Übergang zu einer weiteren Teilfolge – annehmne, dass auch $\lim_{K\ni k\to\infty} \gamma_k = \hat{\gamma} \ge \varepsilon > 0$ konvergiert. Da ∇f nach Annahme und die Projektion nach Lemma 6.2 stetig sind, folgt daraus

$$\begin{aligned}\|\hat{x} - \operatorname{proj}_C(\hat{x} - \hat{\gamma}\nabla f(\hat{x}))\| &= \lim_{K\ni k\to\infty} \|x^k - \operatorname{proj}_C(x^k - \gamma_k \nabla f(x^k))\| \\ &= \lim_{K\ni k\to\infty} \|x^k - x^{k+1}\| = 0,\end{aligned}$$

was nach Satz 6.1 äquivalent ist zu (4.4). □

Wie für das Gradientenverfahren können wir unter zusätzlicher Konvexitätsannahme eine (bessere) Rate für den Funktionswert zeigen.

Satz 6.5. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex und L-glatt und $C \subset \mathbb{R}^n$ nichtleer, konvex, und abgeschlossen. Sei $\{x^k\}_{k\in\mathbb{N}}$ erzeugt aus Algorithmus 6.1 mit $\gamma_k = \frac{1}{L}$ für alle $k \in \mathbb{N}$. Existiert ein Minimierer $\bar{x} \in C$ von f über C, dann gilt*

$$f(x^k) - f(\bar{x}) \leq \frac{L\|x^0 - \bar{x}\|^2}{2k} \qquad \text{für } k = 1, \dots$$

Beweis. Wir kombinieren das Abstiegslemma mit Satz 1.2 (i) und erhalten die Abschätzung

$$\begin{aligned} f(x^{k+1}) &\leq f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2 \\ &\leq f(\bar{x}) + \langle \nabla f(x^k), x^k - \bar{x} \rangle + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2 \\ &= f(\bar{x}) + \langle \nabla f(x^k), x^{k+1} - \bar{x} \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2. \end{aligned}$$

Aus dem Projektionssatz folgt diesmal für $x = \bar{x} \in C$ durch Umsortieren

$$\langle \nabla f(x^k), x^{k+1} - \bar{x} \rangle \leq \frac{1}{\gamma} \langle x^k - x^{k+1}, x^{k+1} - \bar{x} \rangle.$$

Zusammen erhalten wir mit $L = \frac{1}{\gamma}$ und dem Satz von Pythagoras

$$f(x^{k+1}) \leq f(\bar{x}) + \frac{L}{2}\left(\|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2\right).$$

Wie wir bereits in Satz 6.4 gezeigt haben, ist $\{f(x^k)\}_{k\in\mathbb{N}}$ monoton fallend. Mit Teleskopsumme folgt daher

$$\|x^{k+1} - \bar{x}\|^2 - \|x^0 - \bar{x}\|^2 \leq \frac{2}{L}\sum_{m=0}^{k}(f(\bar{x}) - f(x^{m+1}) \leq \frac{2}{L}(k+1)(f(\bar{x}) - f(x^{k+1})$$

bzw.

$$f(x^{k+1}) - f(\bar{x}) \leq \frac{L}{2(k+1)}(\|x^0 - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2) \leq \frac{L}{2(k+1)}\|x^0 - \bar{x}\|^2. \qquad \square$$

6.2 BEDINGTES GRADIENTENVERFAHREN

Der Nachteil des projizierten Gradientenverfahrens ist, dass für manche Mengen die Projektion keine geschlossene Form hat oder nur aufwändig zu berechnen ist. In diesen Fällen kann es vorteilhaft sein, die Zulässigkeit in die Berechnung der Suchrichtung

einzubauen. Die Idee ist dabei, in jeder Iteration ein *linearisiertes* restringiertes Problem zu lösen, d. h.

$$\bar{x}^k \in \arg\min_{x \in C} f(x^k) + \langle \nabla f(x^k), x - x^k \rangle = \arg\min_{x \in C} \langle \nabla f(x^k), x \rangle$$

zu berechnen, und dann x^{k+1} als Konvexkombination von x^k und $\bar{x}^k$ zu bestimmen. Dies führt auf das *bedingte Gradientenverfahren* („*conditional gradient method*“) oder *Frank–Wolfe-Verfahren.*

Algorithmus 6.2 : Bedingtes Gradientenverfahren

Input : $x^0 \in C$

1 **for** $k = 0, \ldots$ **do**

2 Bestimme $\bar{x}^k \in \arg\min_{x \in C} \langle \nabla f(x^k), x \rangle$

3 Wähle $\gamma_k \in (0, 1]$

4 Setze $x^{k+1} = x^k + \gamma_k(\bar{x}^k - x^k)$

Die einzelnen Schritte verdienen Bemerkungen.

Zu Schritt 2: Hier müssen wir eine lineare Funktion über eine Menge minimieren; man spricht daher auch von einem *linearen Minimierungs-Orakel* (da wir nur an der Lösung interessiert sind, nicht wo sie herkommt). Damit dieses Teilproblem garantiert eine Lösung besitzt, muss die Menge nichtleer, abgeschlossen, und beschränkt sein, d. h. endlichen *Durchmesser*

$$D := \max_{x,y \in C} \|x - y\| < \infty$$

haben. Natürlich soll dieser Schritt auch einfacher zu lösen sein als das ursprüngliche Problem. Dies ist insbesondere dann der Fall, wenn C ein Polyeder ist, denn dann liegt ein klassisches lineares Optimierungsproblem vor, das mit effizienten Standard-Verfahren gelöst werden kann. Mehr noch: Da wir wissen, dass lineare Optimierungsprobleme stets eine Lösung besitzen, die in einer Ecke liegt, genügt es, diese endlich vielen Ecken durchzuprobieren und diejenige mit dem kleinsten Skalarprodukt zu nehmen. Hat der Polyeder eine geeignete Struktur, kann man so eine Lösung sogar direkt angeben.

Beispiel 6.6. Das lineare Minimierungs-Orakel für einen gegebenen Vektor $v \in \mathbb{R}^n$ hat eine explizite Lösung zum Beispiel in folgenden Fällen:

(i) Für Quader

$$C = \{x \in \mathbb{R}^n \mid x_i \in [a_i, b_i],\ i = 1, \ldots, n\}.$$

Wegen

$$\min_{x \in C} \langle v, x \rangle = \min_{x_i \in [a_i, b_i]} \sum_{i=1}^{n} v_i x_i$$

kann der Minimierer wieder komponentenweise berechnet werden, und eine einfache Fallunterscheidung zeigt

$$\left[\arg\min_{x\in C}\langle v, x\rangle\right]_i = \begin{cases} \{a_i\} & \text{falls } v_i < 0, \\ \{b_i\} & \text{falls } v_i > 0, \\ [a_i, b_i] & \text{falls } v_i = 0. \end{cases}$$

Um $\bar{x}^k$ zu bestimmen, müssen wir also lediglich das Vorzeichen der Komponenten von $\nabla f(x^k)$ betrachten.

(ii) Für die Kugel

$$C = \{x \in \mathbb{R}^n \mid \|x\|_2 \leq 1\}.$$

Wegen

$$\min_{x\in C}\langle v, x\rangle = -\max_{\|x\|_2\leq 1}\langle -v, x\rangle = -\|-v\|_2 = -\|v\|_2$$

wird das Minimum angenommen in

$$\arg\min_{x\in C}\langle v, x\rangle = \begin{cases} \left\{-\frac{v}{\|v\|_2}\right\} & \text{falls } v \neq 0, \\ C & \text{falls } v = 0. \end{cases}$$

Also können wir für $\bar{x}^k$ einfach den *normierten* negativen Gradienten wählen.

(iii) Für den Einheitssimplex

$$C = \left\{ x \in \mathbb{R}^n \;\middle|\; x \geq 0,\ \sum_{i=1}^n x_i = 1 \right\}.$$

Hier sind die Ecken genau die Einheitsvektoren e_i, $i = 1, \ldots, n$, im $\mathbb{R}^n$, und deshalb gilt

$$\min_{x\in C}\langle v, x\rangle = \min_{i=1,\ldots,n}\langle v, e_i\rangle = \min_{i=1,\ldots,n} v_i.$$

bzw.

$$\arg\min_{x\in C}\langle v, x\rangle = \left\{ e_j \;\middle|\; v_j = \min_{i=1,\ldots,n} v_i \right\}.$$

Wir müssen für die Berechnung von $\bar{x}^k$ also lediglich den kleinsten Eintrag von $\nabla f(x^k)$ suchen. (Dagegen existiert für die Projektion auf C keine geschlossene Formel.)

(iv) Eine ähnliche Formel gilt für $C = \{x \in \mathbb{R}^n \mid \|x\|_1 \leq 1\}$.

Wie das Beispiel zeigt, ist die Wahl von $\bar{x}^k$ in vielen Fällen nicht eindeutig. Auch wenn jede

zulässige Wahl zur Konvergenz führt, kann eine bestimmte Auswahl der Lösung in der Praxis besser funktionieren als eine andere.

Zu Schritt 3: Für die Schrittweitensuche existieren verschiedene Regeln; die wichtigsten sind

(i) *Liniensuche*: Wähle
$$\gamma_k \in \operatorname*{arg\,min}_{\gamma\in(0,1]} f(x^k + \gamma(\bar{x}^k - x^k)).$$
Diese Regel ist in Allgemeinen nur für "einfache" Funktionen umsetzbar.

(ii) Für L-glatte Funktionen ist der *short step*
$$\gamma_k = \min\left\{1, \frac{\langle \nabla f(x^k), x^k - \bar{x}^k\rangle}{L\|x^k - \bar{x}^k\|^2}\right\}$$
eine Näherung, die statt $f(x^{k+1})$ eine geeignete obere Schranke minimiert. Ist f konvex, so folgt aus Satz 1.2 (i) zusammen mit der Wahl von $\bar{x}^k$
$$\langle \nabla f(x^k), x^k - \bar{x}^k\rangle \geq \langle \nabla f(x^k), x^k - \bar{x}\rangle \geq f(x^k) - f(\bar{x}) > 0$$
und damit $\gamma_k \in (0,1]$, falls x^k nicht bereits ein Minimierer ist.

(iii) Die *agnostische Schrittweite*
$$\gamma_k = \frac{2}{k+2} \in (0,1]$$
ist selten optimal aber unabhängig von der konkreten Funktion und einfach umzusetzen.

Zu Schritt 4: Ist C konvex und $x^k \in C$, so gilt wegen $\gamma_k \in (0,1]$ und $\bar{x}^k \in C$ auch
$$x^{k+1} = x^k + \gamma_k(\bar{x}^k - x^k) = \gamma_k\bar{x}^k + (1-\gamma_k)x^k \in C.$$
Startet man mit $x^0 \in C$, so bleiben also alle Iterierten garantiert zulässig. Eine Variante, das *voll-korrigierte* („*fully corrected*") Frank–Wolfe-Verfahren, ersetzt diesen Schritt und die Liniensuche durch
$$x^{k+1} \in \operatorname*{arg\,min}_{x\in\operatorname{co}\{x^0,\dots,x^k,\bar{x}^k\}} f(x),$$
wobei $\operatorname{co} S$ die konvexe Hülle der Menge S bezeichnet. Dies ist nur in Spezialfällen praktikabel, und auch dann nur in Verbindung mit Regeln, die in jedem Schritt geeignete Iterierte x^j, $j < k$, aus der Berechnung der konvexen Hülle entfernen („*pruning*").

Wir untersuchen nun die Konvergenz des Frank–Wolfe-Verfahrens, wobei wir uns auf konvexe Lipschitz-stetig differenzierbare Funktionen beschränken, für die wir auch Konvergenzraten zeigen können.

Satz 6.7. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex und L-glatt mit Minimierer $\bar{x} \in C$ und $C \subset \mathbb{R}^n$ nichtleer, abgeschlossen, konvex, und beschränkt. Dann gilt für $\{x^k\}_{k\in\mathbb{N}}$ erzeugt durch Algorithmus 6.2 mit γ_k gewählt durch Liniensuche,* short step, *oder agnostisch*

$$f(x^k) - f(\bar{x}) \leq \frac{2LD^2}{k+2} \qquad \text{für alle } k \in \mathbb{N}.$$

Beweis. Zunächst folgt aus dem Abstiegslemma (Folgerung 1.13), der Wahl von $\bar{x}^k$, und Satz 1.2 (i) für alle $k \geq 0$ und $\gamma \in (0,1]$ die fundamentale Abschätzung

$$\begin{aligned} f(x^k + \gamma(\bar{x}^k - x^k)) - f(x^k) &\leq \gamma\langle \nabla f(x^k), \bar{x}^k - x^k\rangle + \frac{L\gamma^2}{2}\|\bar{x}^k - x^k\|^2 \\ &\leq \gamma\langle \nabla f(x^k), \bar{x} - x^k\rangle + \gamma^2\frac{L}{2}\|\bar{x}^k - x^k\|^2 \\ &\leq \gamma(f(\bar{x}) - f(x^k)) + \gamma^2\frac{L}{2}\|\bar{x}^k - x^k\|^2. \end{aligned}$$

Speziell für $\gamma = \gamma_k$ folgt daraus durch Abschätzen $\|\bar{x}^k - x^k\| \leq D$ wegen $\bar{x}^k, x^k \in C$, Subtraktion von $f(\bar{x})$ auf beiden Seiten, und Umstellen die Rekursion

$$f(x^{k+1}) - f(\bar{x}) \leq (1-\gamma_k)(f(x^k) - f(\bar{x})) + \gamma_k^2\frac{LD^2}{2}. \tag{6.3}$$

Wir zeigen nun für die agnostische Schrittweite $\gamma_k^a = \frac{2}{k+2}$ die gewünschte Abschätzung durch Induktion nach k. Für $k = 0$ folgt diese aus $\gamma_0^a = 1$ und damit

$$f(x^1) - f(\bar{x}) \leq \frac{LD^2}{2} \leq \frac{2LD^2}{0+2}.$$

Sei jetzt angenommen, dass die Abschätzung gilt für $k \in \mathbb{N}$. Dann folgt aus (6.3) zusammen mit der Induktionsannahme

$$\begin{aligned} f(x^{k+1}) - f(\bar{x}) \leq \left(1 - \frac{2}{k+2}\right)\frac{2LD^2}{k+2} + \frac{4}{(k+2)^2}\frac{LD^2}{2} &= 2LD^2\left(\frac{k}{(k+2)^2} + \frac{1}{(k+2)^2}\right) \\ &= 2LD^2\frac{k+1}{k^2+4k+4} \\ &\leq 2LD^2\frac{k+1}{k^2+4k+3} \\ &= 2LD^2\frac{k+1}{(k+1)(k+3)} \end{aligned}$$

und damit die Behauptung.

Die Abschätzung für die restlichen Schrittweiten erhalten wir daraus, indem wir zeigen, dass diese mindestens so großen Abstieg wie die agnostische Schrittweite erzielen:

- Für die Liniensuche γ_k^l folgt dies sofort aus

$$f(x^{k+1}) - f(\bar{x}) = \min_{\gamma\in(0,1]} f\left(x^k + \gamma(\bar{x}^k - x^k)\right) - f(\bar{x}) \le f\left(x^k + \gamma_k^a(\bar{x}^k - x^k)\right) - f(\bar{x}),$$

  so dass auch in diesem Fall (6.3) gilt mit $\gamma_k = \gamma_k^a$, woraus folgt

$$f(x^{k+1}) - f(\bar{x}) \le (1-\gamma_k^a)(f(x^k) - f(\bar{x})) + (\gamma_k^a)^2\frac{LD^2}{2} \le \frac{2LD^2}{k+3}.$$

  (Ein ähnliches Argument zeigt, dass das voll-korrigierte Frank–Wolfe-Verfahren mit der selben Rate konvergiert.)

- Für den *short step* γ_k^s verwenden wir stattdessen die Abschätzung

$$\begin{aligned} f(x^{k+1}) - f(x^k) &\le \gamma_k^s\langle\nabla f(x^k), \bar{x}^k - x^k\rangle + \frac{L(\gamma_k^s)^2}{2}\|\bar{x}^k - x^k\|^2 \\ &= \min_{\gamma\in(0,1]} \gamma\langle\nabla f(x^k), \bar{x}^k - x^k\rangle + \frac{L\gamma^2}{2}\|\bar{x}^k - x^k\|^2 \\ &\le \gamma_k^a\langle\nabla f(x^k), \bar{x}^k - x^k\rangle + \frac{L(\gamma_k^a)^2}{2}\|\bar{x}^k - x^k\|^2, \end{aligned}$$

  da eine einfache Rechnung zeigt, dass wegen der bereits gezeigten Nichtnegativität des Skalarprodukts γ_k^s genau die Lösung des quadratischen Minimierungsproblems in der zweiten Zeile ist. Also gilt (6.3) mit $\gamma_k = \gamma_k^a$, woraus wieder folgt

$$f(x^{k+1}) - f(\bar{x}) \le (1-\gamma_k^a)(f(x^k) - f(\bar{x})) + (\gamma_k^a)^2\frac{LD^2}{2} \le \frac{2LD^2}{k+3}.$$

In allen Fällen erhalten also die behauptete obere Schranke. □

Es existieren zahlreiche weitere Varianten des Frank–Wolfe-Verfahrens, die sich in den Annahmen an f und C unterscheiden und verschiedene zusätzliche Schritte beinhalten; siehe z. B. [Braun u. a. 2025].

6.3 REDUZIERTE VERFAHREN

Wir betrachten nun Optimierungsprobleme mit Gleichheitsnebenbedingungen speziell der Form

$$\min_{y\in\mathbb{R}^m, u\in\mathbb{R}^n} f(y,u) \quad \text{mit} \quad e(y,u) = 0 \in \mathbb{R}^m. \tag{6.4}$$

Setzt man $x = (y,u) \in \mathbb{R}^{m+n}$, $f(x) = f(y,u)$, und $h(x) = e(y,u)$, so fällt dies genau in die Klasse der Probleme, die wir bislang behandelt haben. In vielen Fällen – insbesondere,

wenn die Probleme wie die Beispiele aus der Einleitung aus inversen Problemen oder der optimalen Steuerung stammen – weiß man aber, dass für festes $u \in \mathbb{R}^n$ genau ein $y = y(u) \in \mathbb{R}^m$ existiert, so dass $e(y, u) = 0$ gilt (man nennt dann y *abhängige Variable* und u *Entscheidungsvariable*). Es gibt also für y „nichts zu optimieren“, und die Betrachtung von $x = (y, u)$ wäre verschwenderisch, besonders falls $m \gg n$ ist.

Die Alternative ist, das unbeschränkte *reduzierte Problem*

$$\min_{u \in \mathbb{R}^n} j(u) := f(y(u), u)$$

zu betrachten. (Man nennt $j(u)$ entsprechend das *reduzierte Funktional*.) Dies ist nun ein *unbeschränktes* Minimierungsproblem, so dass wir alle bekannten Verfahren aus Kapitel 2 anwenden können. Die Frage ist nun, wie man – möglichst effizient – den dafür benötigten *reduzierten Gradienten* $\nabla j(u)$ berechnen kann.

Dazu nehmen wir zuerst an, dass die Abbildung $u \mapsto y(u)$ (Fréchet-)differenzierbar ist mit Ableitung $y'(u) : \mathbb{R}^n \to \mathbb{R}^m$; in diesem Fall folgt aus der Kettenregel

$$j'(u)h = f_y(y(u), u)\, y'(u)h + f_u(y(u), u)h \qquad \text{für alle } h \in \mathbb{R}^n$$

mit den partiellen Ableitungen $f_y(y, u) : \mathbb{R}^m \to \mathbb{R}$ und $f_u(y, u) : \mathbb{R}^n \to \mathbb{R}$ von f, deren Existenz wir annehmen. Der Gradient ist dann gegeben durch

$$\nabla j(u) = j'(u)^T = y'(u)^T \nabla_y f(y(u), u) + \nabla_u f(y(u), u) \in \mathbb{R}^n \tag{6.5}$$

mit $\nabla_y f(y, u) = f_y(y, u)^T \in \mathbb{R}^m$ und analog für $\nabla_u f(y, u) \in \mathbb{R}^n$.

Es bleibt die Frage, wie wir die Ableitung $y'(u)$ erhalten. Ist e stetig differenzierbar und die partielle Ableitung $e_y(y, u)$ in einem Punkt (y, u) mit $e(y, u) = 0$ (d. h. in $(y(u), u)$) invertierbar, so garantiert der Satz über implizite Funktionen, dass $u \mapsto y(u)$ in u stetig differenzierbar ist und es gilt

$$e_y(y(u), u)\, y'(u)h + e_u(y(u), u)h = 0 \qquad \text{für alle } h \in \mathbb{R}^n.$$

Aus der Invertierbarkeit folgt dann insbesondere, dass wir für alle $h \in \mathbb{R}^n$ nach

$$y'(u)h = -e_y(y(u), u)^{-1} e_u(y(u), u)h \tag{6.6}$$

auflösen können. (Man nennt $y'(u)h \in \mathbb{R}^m$ auch *Sensitivität* von y in u in Richtung h.) Aus der Linearen Algebra ist nun bekannt, dass $y'(u)$ die Matrixdarstellung

$$y'(u) = y'(u)\,\mathrm{Id} = [y'(u)e_1 \mid \ldots \mid y'(u)e_n]$$

für die Einheitsvektoren $e_i \in \mathbb{R}^n$ hat; wir können die Jacobi-Matrix $y'(u) \in \mathbb{R}^{m \times n}$ also spaltenweise durch Lösen des linearen Gleichungssystems in (6.6) für $e_1, \ldots, e_n$ berechnen, das Ergebnis transponieren, und dann in (6.5) einsetzen.

Beispiel 6.8. Wir betrachten das Modellproblem (6.4) mit

$$f(y,u) = \frac{1}{2}\|y-z\|_2^2 + \frac{\alpha}{2}\|u\|_2^2$$

für $z \in \mathbb{R}^m$ und $\alpha \geq 0$ beliebig sowie

$$e(y,u) = Ay + y^3 + Bu - c \in \mathbb{R}^m$$

für $A \in \mathbb{R}^{m\times m}$ positiv definit, $B \in \mathbb{R}^{n\times m}$ und $c \in \mathbb{R}^m$ beliebig, wobei y^3 komponentenweise zu verstehen ist. Man kann zeigen, dass die Gleichung $Ay + y^3 = c - Bu$ für alle u, c eine eindeutige Lösung $y = y(u)$ besitzt. Man rechnet nun leicht nach, dass die partiellen Ableitungen gegeben sind durch

$$e_y(y,u)\tilde{y} = A\tilde{y} + 3y^2\tilde{y}, \qquad e_u(y,u)\tilde{u} = B\tilde{u}$$

für alle $\tilde{y} \in \mathbb{R}^m$ und $\tilde{u} \in \mathbb{R}^n$, wobei $3y^2\tilde{y}$ wieder als komponentenweise Multiplikation zu verstehen ist. Da A positiv definit ist und der lineare Operator $\tilde{y} \mapsto 3y^2\tilde{y}$ durch eine Diagonalmatrix mit den Einträgen $3y^2 \geq 0$ auf der Diagonalen repräsentiert werden kann, ist $e_y(y,u) = A + \operatorname{diag}(3y^2) \in \mathbb{R}^{m\times m}$ auch positiv definit und damit für alle $y \in \mathbb{R}^m$ invertierbar. Die Sensitivitäten $y_i' := [y'(u)]_i = y'(u)e_i$, $i = 1,\dots,m$, können wir daher berechnen durch Lösung des linearen Gleichungssystems

$$Ay_i' + 3y(u)^2 y_i' = -Be_i = -b_i.$$

Für große n ist dies aber sehr aufwändig – und auch gar nicht notwendig, denn (6.5) benötigt nur den *Vektor* $y'(u)^T\nabla_y f(y(u),u) \in \mathbb{R}^n$. Dieser kann tatsächlich direkt berechnet werden, ohne $y'(u)$ explizit aufzustellen: Nehmen wir in (6.6) das Skalarprodukt mit $\nabla_y f(y(u),u)$, so folgt für alle $h \in \mathbb{R}^n$

$$\begin{aligned}\langle y'(u)^T\nabla_y f(y(u),u), h\rangle &= \langle \nabla_y f(y(u),u), y'(u)h\rangle \\ &= \langle \nabla f(y(u),u), -e_y(y(u),u)^{-1}e_u(y(u),u)h\rangle \\ &= \langle -e_u(y(u),u)^T[e_y(y(u),u)^{-1}]^T\nabla_y f(y(u),u), h\rangle\end{aligned}$$

und damit (wegen $(A^{-1})^T = (A^T)^{-1}$)

$$y'(u)^T\nabla_y f(y(u),u) = -e_u(y(u),u)^T[e_y(y(u),u)^T]^{-1}\nabla_y f(y(u),u).$$

Wir müssen daher nur *ein* lineares Gleichungssystem

$$e_y(y(u),u)^T p = -\nabla_y f(y(u),u) \tag{6.7}$$

lösen (welches wegen der angenommenen Invertierbarkeit von $e_y(y(u),u)$ möglich ist) um mit deren Lösung $p(u)$

$$y'(u)^T\nabla_y f(y(u),u) = e_u(y(u),u)^T p(u)$$

zu erhalten. (Man nennt (6.7) die *adjungierte Gleichung*; die Lösung p entsprechend die *Adjungierte* (Lösung)).

Beispiel 6.9. Wir knüpfen an an Beispiel 6.8. Es gilt

$$e_y(y,u) = A + \operatorname{diag}(3y^2)$$

und damit

$$e_y(y,u)^T = A^T + \operatorname{diag}(3y^2).$$

Weiter rechnet man leicht nach, dass gilt

$$\nabla_y f(y,u) = y - z, \qquad \nabla_u f(y,u) = \alpha u.$$

Die adjungierte Gleichung (6.7) hat hier also die Form

$$A^T p + 3y(u)^2 p = -(y(u) - z)$$

für die Lösung $y(u) \in \mathbb{R}^m$ von $Ay + y^3 = c - Bu$, wobei das Produkt wieder komponentenweise zu verstehen ist. Diese Gleichung hat ebenfalls eine eindeutige Lösung $p(u) \in \mathbb{R}^m$. Weiter ist

$$e_u(y,u) = B.$$

Der reduzierte Gradient ist dann gegeben durch

$$\nabla j(u) = e_u(y(u),u)^T p + \nabla_u f(y(u),u) = B^T p(u) + \alpha u.$$

Eine alternative und oft übersichtlichere Möglichkeit, den reduzierten Gradienten zur berechnen, ist das *Lagrange-Kalkül*. (Wichtig: Dies ist nur eine formale Methode, mit deren Hilfe man eine adjungierte Lösung berechnen kann, *wenn* man bereits gezeigt hat, dass sie existiert!) Wir betrachten dafür wieder die (nicht reduzierte) Lagrange-Funktion

$$L : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m, \qquad (y,u,p) \mapsto f(y,u) + \langle p, e(y,u)\rangle,$$

für die mit $e(y(u),u) = 0$ gilt

$$L(y(u),u,p) = f(y(u),u) = j(u) \qquad \text{für alle } p \in \mathbb{R}^m.$$

Daraus folgt durch (formales) Differenzieren für alle $h \in \mathbb{R}^n$

$$\langle \nabla j(u), h\rangle = \langle \nabla_y L(y(u),u,p), y'(u)h\rangle + \langle \nabla_u L(y(u),u,p), h\rangle \tag{6.8}$$

für beliebige $p \in \mathbb{R}^m$. Wir wählen nun $p = p(u)$ speziell so, dass

$$\nabla_y L(y(u),u,p(u)) = 0$$

ist. (Dies ist der formale Part, da die Existenz so eines $p(u)$ erstmal nicht garantiert ist!) Wegen

$$\begin{aligned}\langle \nabla_y L(y,u,p), \tilde{y}\rangle &= \langle \nabla_y f(y,u), \tilde{y}\rangle + \langle p, e_y(y,u)\tilde{y}\rangle \\ &= \langle \nabla_y f(y,u) + e_y(y,u)^T p, \tilde{y}\rangle\end{aligned}$$

für alle $\tilde{y} \in \mathbb{R}^m$ und damit

$$0 = \nabla_y L(y(u),u,p) = \nabla_y f(y(u),u) + e_y(y(u),u)^T p$$

ist dies genau die adjungierte Gleichung. Einsetzen der Lösung $p(u)$ in (6.8) ergibt, weil $h \in \mathbb{R}^n$ beliebig war, dann wieder

$$\nabla j(u) = \nabla_u L(y(u),u,p(u) = \nabla_u f(y(u),u) + e_u(y(u),u)^T p(u).$$

Beispiel 6.10. Wir wenden das Lagrange-Kalkül an auf Beispiel 6.8. Die Lagrange-Funktion ist hier

$$L(y,u,p) = \frac{1}{2}\|y-z\|_2^2 + \frac{\alpha}{2}\|u\|_2^2 + \langle p, Ay + y^3 + Bu - c\rangle.$$

Wir setzen für gegebenes u nun einfach schrittweise die partiellen Ableitungen auf Null und lösen auf:

1. Schritt: Berechne $y(u)$ als Lösung nach y von

$$0 = \nabla_p L(y,u,p) = Ay + y^3 + Bu - c.$$

2. Schritt: Berechne $p(u)$ als Lösung von

$$0 = \nabla_y L(y(u),u,p) = y(u) - z + A^T p + 3y(u)^2 p;$$

diese Gleichung erhält man durch formales Durchdifferenzieren in

$$0 = \langle p, e_y(y,u)\tilde{y}\rangle = \langle p, A\tilde{y} + 3y^2\tilde{y}\rangle = \langle A^T p + 3y^2 p, \tilde{y}\rangle$$

für alle $\tilde{y} \in \mathbb{R}^m$.

3. Schritt: Setze

$$\nabla j(u) = \nabla_u L(y(u),u,p(u)) = \alpha u + B^T p(u),$$

wobei wir den zweiten Summanden analog erhalten haben durch Differenzieren in

$$\langle p, e_u(y,u)h\rangle = \langle p, Bh\rangle = \langle B^T p, h\rangle$$

für alle $h \in \mathbb{R}^n$.

Analog kann man mit Hilfe der (vollen) Hesse-Matrix der Lagrange-Funktion und geeigneter Rückwärtssubstitution die (Matrix-Vektor-Multiplikation mit) der reduzierten Hesse-Matrix $\nabla^2 j(u)$ berechnen; dies kann zum Beispiel verwendet werden, um den Newton-Schritt mit Hilfe des CG-Verfahrens zu berechnen („Krylov–Newton-Verfahren“).

7 STRAFVERFAHREN

Wir kommen nun zu Verfahren zur numerischen Lösung von Optimierungsproblemen mit allgemeinen Nebenbedingungen. Ein klassischer Ansatz besteht darin, das Problem durch eine Folge von unrestringierten Problemen zu approximieren, indem man die Zielfunktion so modifiziert, dass das Verlassen des zulässigen Bereichs X zunehmend "teuer" wird. Bei *Strafverfahren* (auch *Penalty-Verfahren*) wird dabei eine *Straffunktion* $\pi : \mathbb{R}^n \to \mathbb{R}$ mit $\pi(x) = 0$ für $x \in X$ und $\pi(x) > 0$ für $x \notin X$ addiert, die mit einem *Penalty-Parameter* $\alpha > 0$ gewichtet wird. Man betrachtet also das Problem

$$\min_{x \in \mathbb{R}^n} f(x) + \alpha\pi(x).$$

Je größer α gewählt ist, desto mehr Wert wird auf die (näherungsweise) Erfüllung der Nebenbedingung $x \in X$ gelegt. Man hofft daher, dass für $\alpha \to \infty$ die Folge $\{x_\alpha\}_{\alpha>0}$ der entsprechenden Minimierer gegen einen Minimierer $\bar{x} \in X$ von f konvergiert.

7.1 QUADRATISCHE STRAFVERFAHREN

Wir betrachten wieder den konkreten Fall

$$X = \left\{ x \in \mathbb{R}^n \;\middle|\; g_i(x) \le 0,\ 1 \le i \le m, \quad h_j(x) = 0,\ 1 \le j \le p \right\}$$

für $g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar, und bestrafen die Nebenbedingungen einzeln. Im *quadratischen Strafverfahren* wählt man dazu quadratische Funktionen, und zwar

(i) für die Gleichungsnebenbedingung $h_j(x) = 0$ die Funktion $x \mapsto \frac{1}{2}|h_j(x)|^2$;

(ii) für die Ungleichungsnebenbedingung $g_i(x) \le 0$ die Funktion $x \mapsto \frac{1}{2}|(g_i)^+|^2$, wobei $(t)^+ := \max\{0, t\}$ bezeichnet.

Man minimiert nun anstelle von f für $\alpha > 0$ die Funktion

$$\begin{aligned} P_\alpha(x) &:= f(x) + \frac{\alpha}{2} \sum_{i=1}^{m} |(g_i(x))^+|^2 + \frac{\alpha}{2} \sum_{j=1}^{p} |h_j(x)|^2 \\ &= f(x) + \frac{\alpha}{2} \|(g(x))^+\|^2 + \frac{\alpha}{2} \|h(x)\|^2, \end{aligned} \tag{7.1}$$

wobei wir wieder die Nebenbedingungen zu vektorwertigen Funktionen $g : \mathbb{R}^n \to \mathbb{R}^m$ und $h : \mathbb{R}^n \to \mathbb{R}^p$ zusammengesetzt haben, und $(v)^+$ für $v \in \mathbb{R}^m$ komponentenweise zu verstehen ist. Wir setzen in Folge kurz

$$\pi(x) := \frac{1}{2}\left(\|(g(x))^+\|^2 + \|h(x)\|^2\right).$$

Offensichtlich gilt $P_\alpha(x) = f(x)$ für alle $x \in X$ und $\alpha > 0$.

Wir fragen uns zuerst, wann das penalisierte Problem

$$\min_{x \in \mathbb{R}^n} P_\alpha(x) \tag{7.2}$$

eine Lösung besitzt. Dafür müssen wir natürlich annehmen, dass f auf ganz $\mathbb{R}^n$ wohldefiniert ist. Außerdem brauchen wir eine Annahme an die Darstellung der Menge X.

Satz 7.1. *Seien $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$ und $h_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \dots, p$ stetig. Gilt entweder*

(i) f koerziv oder

(ii) π koerziv und f nach unten beschränkt,

dann existiert für alle $\alpha > 0$ eine Lösung $x_\alpha \in \mathbb{R}^n$ von (7.2).

Beweis. Gilt (i) oder (ii), so ist $P_\alpha = f + \alpha\pi$ die Summe einer nach unten beschränkten und einer koerziven Funktion und damit koerziv. Da mit g und h (und $t \mapsto (t)^+$) auch $\pi(x)$ stetig ist, folgt die Existenz aus Satz 1.4. □

Wir nehmen in Folge an, dass die Bedingungen von Satz 7.1 erfüllt sind. Ein Strafverfahren hat dann die Form von

Algorithmus 7.1 : Quadratisches Strafverfahren

Input : $\alpha_0 > 0$, $x^0 \in \mathbb{R}^n$

for $k = 0, \dots$ **do**
  Berechne x^{k+1} als globalen Minimierer von (7.2) mit $\alpha = \alpha_k$ (und Startwert x^k)
  if $x^{k+1} \in X$ **then**
    return x^{k+1}
  else
    Wähle $\alpha_{k+1} > \alpha_k$

Um die Konvergenz dieses Verfahrens zu untersuchen, zeigen wir zuerst nützliche Eigenschaften der so erzeugten Folge $\{x^k\}_{k \in \mathbb{N}}$.

Lemma 7.2. *Sei $X \subset \mathbb{R}^n$ nichtleer. Angenommen, Algorithmus 7.1 erzeugt für eine streng monoton wachsende, unbeschränkte Folge $\{\alpha_k\}_{k\in\mathbb{N}} \subset (0,\infty)$ eine unendliche Folge $\{x^k\}_{k\in\mathbb{N}}$. Dann gilt*

(i) $\{P_{\alpha_k}(x^k)\}_{k\in\mathbb{N}}$ *ist monoton wachsend;*

(ii) $\{\pi(x^k)\}_{k\in\mathbb{N}}$ *ist monoton fallend;*

(iii) $\{f(x^k)\}_{k\in\mathbb{N}}$ *ist monoton wachsend;*

(iv) $\{(g_i(x^k))^+\}_{k\in\mathbb{N}}$, $1 \le i \le m$, *und* $\{h_j(x^k)\}_{k\in\mathbb{N}}$, $1 \le j \le p$, *sind Nullfolgen.*

Beweis. Zu (i): Aus der globalen Optimalität von x^k und $\alpha_k < \alpha_{k+1}$ folgt

$$\begin{aligned} P_{\alpha_k}(x^k) \le P_{\alpha_k}(x^{k+1}) &= f(x^{k+1}) + \alpha_k \pi(x^{k+1}) \\ &\le f(x^{k+1}) + \alpha_{k+1}\pi(x^{k+1}) = P_{\alpha_{k+1}}(x^{k+1}). \end{aligned}$$

Zu (ii): Aus $P_{\alpha_k}(x^k) \le P_{\alpha_k}(x^{k+1})$ und $P_{\alpha_{k+1}}(x^{k+1}) \le P_{\alpha_{k+1}}(x^k)$ folgt durch Addition

$$\alpha_k \pi(x^k) + \alpha_{k+1}\pi(x^{k+1}) \le \alpha_k \pi(x^{k+1}) + \alpha_{k+1}\pi(x^k),$$

was durch Umformen

$$(\alpha_k - \alpha_{k+1})\left(\pi(x^k) - \pi(x^{k+1})\right) \le 0$$

ergibt. Aus $\alpha_k < \alpha_{k+1}$ folgt daher $\pi(x^k) \ge \pi(x^{k+1})$.

Zu (iii): Aus der Optimalität von x^k und (ii) folgt sofort

$$f(x^k) + \alpha_k \pi(x^k) \le f(x^{k+1}) + \alpha_k \pi(x^{k+1}) \le f(x^{k+1}) + \alpha_k \pi(x^k)$$

und damit $f(x^k) \le f(x^{k+1})$.

Zu (iv): Da X nichtleer ist, existiert ein $\hat{x} \in X$. Aus der Optimalität von x^k und (iii) folgt dann

$$f(\hat{x}) = f(\hat{x}) + \alpha_k \pi(\hat{x}) \ge f(x^k) + \alpha_k \pi(x^k) \ge f(x^0) + \alpha_k \pi(x^k).$$

Wegen $\alpha_k \to \infty$ folgt daraus nun

$$\pi(x^k) \le \frac{1}{\alpha_k}\left(f(\hat{x}) - f(x^0)\right) \to 0.$$

Nach Definition von π müssen daher auch $(g_i(x^k))^+ \to 0$ und $h_j(x^k) \to 0$ gehen. □

Damit können wir nun die Konvergenz zeigen.

Satz 7.3. *Seien $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ stetig und $X \subset \mathbb{R}^n$ nichtleer. Dann bricht Algorithmus 7.1 entweder nach endlich vielen Schritten in einem globalen Minimierer ab oder erzeugt für eine streng monoton wachsende, unbeschränkte Folge $\{\alpha_k\}_{k\in\mathbb{N}} \subset (0,\infty)$ eine unendliche Folge $\{x^k\}_{k\in\mathbb{N}}$, für die jeder Häufungspunkt ein globaler Minimierer ist.*

Beweis. Nach Satz 7.1 kann das Strafverfahren nur im Fall $x^k \in X$ abbrechen. In diesem Fall gilt aber

$$f(x^k) = f(x^k) + \alpha_k \pi(x^k) \leq f(x) + \alpha_k \pi(x) = f(x)$$

für alle $x \in X$, d. h. x^k ist globaler Minimierer von f in X

Andernfalls sei $\bar{x} \in \mathbb{R}^n$ ein Häufungspunkt von $\{x^k\}_{k\in\mathbb{N}}$ und $\{x^k\}_{k\in K}$ eine gegen $\bar{x}$ konvergente Teilfolge. Da mit g und h (und $t \mapsto (t)^+$) auch $\pi(x)$ stetig ist, folgt aus Lemma 7.2 (iv) (angewendet auf die Teilfolge, für die ebenfalls $\lim_{K\ni k\to\infty} \alpha_k = \infty$ gilt)

$$\pi(\bar{x}) = \lim_{K\ni k\to\infty} \pi(x^k) = 0.$$

Nach Definition von π impliziert dies $\bar{x} \in X$. Für alle $x \in X$ und $k \in K$ gilt wegen $\pi \geq 0$ daher

$$f(x^k) \leq f(x^k) + \alpha_k \pi(x^k) \leq f(x) + \alpha_k \pi(x) = f(x).$$

Durch Grenzübergang $k \to \infty$ auf beiden Seiten folgt daraus $f(\bar{x}) \leq f(x)$ für alle $x \in X$, d. h. $\bar{x}$ ist globaler Minimierer von f in X. □

Um die in Schritt 3 benötigte Lösung von (7.2) zu berechnen, möchten wir Verfahren aus Kapitel 2 einsetzen, für die P_α zumindest stetig differenzierbar sein muss. Problematisch ist dabei nur die Penalisierung der Ungleichungsnebenbedingungen. Durch Fallunterscheidung $t > 0$, $t < 0$ und $t = 0$ in der Definition der reellen Ableitung vergewissert man sich aber schnell, dass gilt

$$\frac{d}{dt}\left(\tfrac{1}{2}|(t)^+|^2\right) = (t)^+.$$

Aus der Summen- und Kettenregel folgt daher

$$\nabla P_\alpha(x) = \nabla f(x) + \alpha \sum_{i=1}^{m} (g_i(x))^+ \nabla g_i(x) + \alpha \sum_{j=1}^{p} h_j(x) \nabla h_j(x). \tag{7.3}$$

Vergleicht man dies mit (4.10), so erhält man

$$\nabla P_\alpha(x) = \nabla_x L(x, \mu_\alpha, \lambda_\alpha) \qquad \text{für} \quad \mu_\alpha := \alpha(g(x))^+, \quad \lambda_\alpha := \alpha h(x).$$

Tatsächlich kann man (unter einer Regularitätsbedingung) zeigen, dass die zu $\{x^k\}_{k\in\mathbb{N}}$ gehörenden Folgen $\{\mu^k\}_{k\in\mathbb{N}}$, $\{\lambda^k\}_{k\in\mathbb{N}}$ gegen die entsprechenden Lagrange-Multiplikatoren des restringierten Problems konvergieren.

Satz 7.4. *Seien $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ und $h : \mathbb{R}^n \to \mathbb{R}^p$ stetig differenzierbar. Konvergiert die durch Algorithmus 7.1 erzeugte Folge $\{x^k\}_{k\in\mathbb{N}}$ gegen einen Punkt $\bar{x} \in X$, der die LICQ erfüllt, so konvergieren auch*

$$\mu^k := \alpha_k (g(x^k))^+ \to \bar{\mu}, \qquad \lambda^k := \alpha_k h(x^k) \to \bar{\lambda},$$

und $(\bar{x}, \bar{\mu}, \bar{\lambda})$ erfüllt die KKT-Bedingungen (4.9).

Beweis. Wir zeigen zuerst die Konvergenz von $\{\mu^k\}_{k\in\mathbb{N}}$ und $\{\lambda^k\}_{k\in\mathbb{N}}$. Wir unterscheiden wieder aktive und inaktive Nebenbedingungen: Für $i \notin \mathcal{A}_X(\bar{x})$ gilt $g_i(\bar{x}) < 0$. Aus der Stetigkeit von g_i folgt dann $g_i(x^k) < 0$ und damit auch $(g_i(x^k))^+ = 0$ für $k \in \mathbb{N}$ groß genug. Also gilt

$$\mu_i^k = \alpha_k (g_i(x^k))^+ \to 0 =: \bar{\mu}_i \qquad \text{für alle } i \notin \mathcal{A}_X(\bar{x}). \tag{7.4}$$

Für die restlichen Komponenten verwenden wir die LICQ. Wir bezeichnen mit A_k diejenige Matrix, die aus den Spalten $\nabla g_i(x^k)$, $i \in \mathcal{A}_X(\bar{x})$, und $\nabla h_j(x^k)$, $1 \le j \le p$, gebildet wird. Da nach Voraussetzung g und h stetig differenzierbar sind, konvergiert die Folge dieser Matrizen gegen die Matrix $\bar{A}$, die entsprechend aus den Spalten $\nabla g_i(\bar{x})$ und $\nabla h_j(\bar{x})$ gebildet wird. Weiterhin hat $\bar{A}$ aufgrund der LICQ vollen Spaltenrang, und damit ist $\bar{A}^T\bar{A}$ invertierbar. Nach dem Banach-Lemma 2.2 ist damit auch $A_k^T A_k$ für $k \in \mathbb{N}$ hinreichend groß invertierbar, und es gilt $(A_k^T A_k)^{-1} \to (\bar{A}^T\bar{A})^{-1}$.

Nun ist x^k ein unrestringierter Minimierer von P_{α_k}, erfüllt also die notwendige Optimalitätsbedingung $\nabla P_{\alpha_k}(x^k) = 0$. Für $k \in \mathbb{N}$ mit $g_i(x^k) < 0$ für alle $i \notin \mathcal{A}_X(\bar{x})$ folgt daraus (vergleiche (7.3))

$$0 = A_k^T \nabla P_{\alpha_k}(x^k) = A_k^T \nabla f(x^k) + A_k^T A_k \begin{pmatrix} \mu_{\mathcal{A}}^k \\ \lambda^k \end{pmatrix},$$

wobei $\mu_{\mathcal{A}}^k$ den Vektor bezeichnet, der aus den Komponenten μ_i^k, $i \in \mathcal{A}_X(\bar{x})$, besteht. Aufgrund der Stetigkeit von ∇f und der Konvergenz $A_k \to \bar{A}$ erhalten wir damit

$$\begin{pmatrix} \mu_{\mathcal{A}}^k \\ \lambda^k \end{pmatrix} = -(A_k^T A_k)^{-1} A_k^T \nabla f(x^k) \to -(\bar{A}^T\bar{A})^{-1}\bar{A}\nabla f(\bar{x}) =: \begin{pmatrix} \bar{\mu}_{\mathcal{A}} \\ \bar{\lambda} \end{pmatrix}.$$

Damit ist die Konvergenz der Folgen $\{\mu^k\}_{k\in\mathbb{N}}$ und $\{\lambda^k\}_{k\in\mathbb{N}}$ gezeigt.

Für die KKT-Bedingungen (4.9) folgt aus der Optimalität der x^k, der Definition der μ^k und λ^k sowie der Stetigkeit von ∇f, ∇g und ∇h

$$\nabla_x L(\bar{x}, \bar{\mu}, \bar{\lambda}) = \lim_{k\to\infty} \nabla_x L(x^k, \mu^k, \lambda^k) = \lim_{k\to\infty} \nabla P_{\alpha_k}(x^k) = 0$$

und damit die Stationaritätsbedingung. Weiter gilt nach Annahme $\bar{x} \in X$ und damit insbesondere $h_j(\bar{x}) = 0$ für $1 \le j \le p$, d. h. die Zulässigkeitsbedingung. Ebenso gilt $g_i(\bar{x}) \le 0$ sowie nach Definition

$$\bar{\mu}_i = \lim_{k\to\infty} \alpha_k (g_i(x^k))^+ \ge 0$$

für alle $1 \leq i \leq m$. Schließlich folgt aus (7.4) auch $\bar{\mu}_i g_i(\bar{x}) = 0$ für alle $1 \leq i \leq m$ und damit die Komplementaritätsbedingung. □

Beachte, dass wir hier (neben der Zulässigkeit des Häufungspunkts) nur die Stationarität von x^k bezüglich P_{α_k} verwendet haben, nicht die globale Optimalität. Dieses Resultat ist daher auch auf den (praktisch relevanteren) Fall anwendbar, dass wir in Algorithmus 7.1 nur stationäre Punkte berechnen können.

Ein Nachteil der quadratischen Penalisierung ist, dass in der Regel $x_\alpha \notin X$ gilt. Aus (7.3) folgt nämlich $\nabla P_\alpha(x) = \nabla f(x)$ für alle $x \in X$. Wäre also $x_\alpha \in X$ ein Minimierer von P_α, so würde aus der notwendigen Optimalitätsbedingung folgen

$$0 = \nabla P_\alpha(x_\alpha) = \nabla f(x_\alpha),$$

was nur möglich ist, wenn x_α stationärer Punkt von f im Inneren von X ist. Man muß also tatsächlich $\alpha \to \infty$ streben lassen; erschwerend kommt hinzu, dass in der Regel die Minimierung von P_α mit wachsendem α zunehmend schwieriger wird (z. B. durch wachsende Konditionszahl der Newton-Systeme).

7.2 EXAKTE STRAFVERFAHREN

Dieser Nachteil kann durch eine unterschiedliche Wahl der Straffunktion vermieden werden. Eine Straffunktion $\pi : \mathbb{R}^n \to \mathbb{R}$ heißt *exakt* in einem Minimierer $\bar{x} \in X$, wenn ein endliches $\bar{\alpha} > 0$ existiert, so dass $\bar{x}$ auch ein unrestringierter Minimierer von $f + \alpha\pi$ für alle $\alpha > \bar{\alpha}$ ist.

Eine Variante besteht darin, anstelle der Quadrate den Absolutbetrag der Nebenbedingungen zu penalisieren. Man betrachtet also die Straffunktion

$$\pi_1(x) := \sum_{i=1}^{m} (g_i(x))^+ + \sum_{j=1}^{p} |h_j(x)| = \|(g(x))^+\|_1 + \|h(x)\|_1.$$

Für konvexe Probleme kann man zeigen, dass π_1 in der Tat exakt ist.

Satz 7.5. *Seien $f : \mathbb{R}^n \to \mathbb{R}$ und $g_i : \mathbb{R}^n \to \mathbb{R}$, $1 \leq i \leq m$, stetig differenzierbar und konvex, und sei $h_j : \mathbb{R}^n \to \mathbb{R}$, $1 \leq j \leq p$, affin-linear. Es sei weiter die Slater-Bedingung (4.8) erfüllt. Dann ist π_1 exakt in jedem Minimierer $\bar{x} \in X$.*

Beweis. Unter den Annahmen an die Nebenbedingung sind alle Punkte $x \in X$ regulär; jeder Minimierer $\bar{x} \in X$ von f erfüllt also die KKT-Bedingungen (4.9). Insbesondere existieren Lagrange-Multiplikatoren $\bar{\mu} \in \mathbb{R}^m$ und $\bar{\lambda} \in \mathbb{R}^p$. Wir wählen nun

$$\bar{\alpha} := \max\{\bar{\mu}_1, \ldots, \bar{\mu}_m, |\bar{\lambda}_1|, \ldots, |\bar{\lambda}_p|\}. \tag{7.5}$$

Dann gilt für alle $\alpha \geq \bar{\alpha}$ und $x \in \mathbb{R}^n$ nach Satz 5.4 wegen $\bar{\mu}_i \geq 0$

$$\begin{aligned} f(\bar{x}) + \alpha\pi_1(\bar{x}) = f(\bar{x}) &= L(\bar{x}, \bar{\mu}, \bar{\lambda}) \leq L(x, \bar{\mu}, \bar{\lambda}) \\ &\leq f(x) + \sum_{i=1}^{m} \bar{\mu}_i (g_i(x))^+ + \sum_{j=1}^{p} |\bar{\lambda}_j||h_j(x)| \\ &\leq f(x) + \alpha \sum_{i=1}^{m} (g_i(x))^+ + \alpha \sum_{j=1}^{p} |h_j(x)| \\ &= f(x) + \alpha\pi_1(x). \end{aligned}$$

□

Allerdings gibt es nichts geschenkt; da der Betrag (bzw. die Funktion $t \mapsto (t)^+$) nicht differenzierbar ist, ist auch $f + \alpha\pi_1$ nicht differenzierbar. Die für das Strafverfahren benötigten Minimierer können also nicht mit den Methoden aus Kapitel 2 berechnet werden. Tatsächlich sind exakte Straffunktionen dieser Form notwendig nicht differenzierbar (außer eben in Minimierern, die im Inneren von X liegen).

Satz 7.6. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und sei $\bar{x} \in X$ ein Minimierer mit $\nabla f(\bar{x}) \neq 0$. Ist $\pi : \mathbb{R}^n \to \mathbb{R}$ eine differenzierbare Straffunktion, so ist sie nicht exakt in $\bar{x}$.*

Beweis. Angenommen, $\pi : \mathbb{R}^n \to \mathbb{R}$ wäre eine differenzierbare, exakte Straffunktion. Dann existiert ein $\bar{\alpha} > 0$, so dass $\bar{x}$ ein unrestringierter Minimierer von $f + \alpha\pi$ für alle $\alpha \geq \bar{\alpha}$ ist. Da $f + \alpha\pi$ nach Annahme differenzierbar ist, folgt aus der notwendigen Optimalitätsbedingung

$$\nabla f(\bar{x}) + \alpha \nabla \pi(\bar{x}) = 0. \tag{7.6}$$

Für beliebige $\alpha_1, \alpha_2 > \bar{\alpha}$ mit $\alpha_1 \neq \alpha_2$ gilt daher

$$\nabla f(\bar{x}) + \alpha_1 \nabla \pi(\bar{x}) = 0 = \nabla f(\bar{x}) + \alpha_2 \nabla \pi(\bar{x}),$$

was durch Umformen

$$(\alpha_1 - \alpha_2) \nabla \pi(\bar{x}) = 0$$

ergibt. Daraus folgt $\nabla \pi(\bar{x}) = 0$ und damit wegen (7.6) auch $\nabla f(\bar{x}) = 0$, im Widerspruch zur Annahme. □

Dieser Nachteil wird in dem im nächsten Abschnitt vorgestellten Verfahren vermieden.

7.3 MULTIPLIKATOR-STRAFVERFAHREN

Die Idee ist, ein quadratisches Strafverfahren auf die Lagrange-Funktion anstatt auf die Zielfunktion anzuwenden. Wir betrachten zunächst den Fall reiner Gleichungsnebenbedingungen und wenden diesen dann durch eine geeignete Umformulierung auch auf Ungleichungsnebenbedingungen an.

7.3.1 GLEICHUNGSNEBENBEDINGUNGEN

Wir betrachten das Optimierungsproblem

$$\min_{x\in\mathbb{R}^n} f(x) \qquad \text{mit } h(x) = 0 \tag{7.7}$$

für $f : \mathbb{R}^n \to \mathbb{R}$ und $h : \mathbb{R}^n \to \mathbb{R}^p$ stetig differenzierbar und definieren für $\alpha > 0$ die *erweiterte Lagrange-Funktion* (englisch: *augmented Lagrangian*)

$$L_\alpha : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}, \qquad L_\alpha(x,\lambda) = f(x) + \lambda^T h(x) + \frac{\alpha}{2}\|h(x)\|^2.$$

Wählen wir konkret einen Lagrange-Multiplikator $\bar\lambda$ für einen lokalen Minimierer $\bar x$ von (7.7), so ist dies trotz der stetigen Differenzierbarkeit eine exakte Penalisierung.

Satz 7.7. *Seien f, h zweimal stetig differenzierbar. Erfüllt $(\bar x, \bar\lambda) \in \mathbb{R}^n \times \mathbb{R}^p$ für (7.7) die hinreichende Optimalitätsbedingungen aus Satz 4.21, so existiert ein $\bar\alpha > 0$ so, dass $\bar x$ für alle $\alpha \geq \bar\alpha$ ein strikter lokaler Minimierer von $x \mapsto L_\alpha(x, \bar\lambda)$ ist.*

Beweis. Wir zeigen, dass für $\alpha > 0$ groß genug die hinreichenden Bedingungen 2. Ordnung für das restringierte Problem (7.7) diejenigen für das unrestringierte Problem $\min_{x\in\mathbb{R}^n} L_\alpha(x, \bar\lambda)$ implizieren.

Nach Annahme erfüllt $(\bar x, \bar\lambda)$ insbesondere die KKT-Bedingungen. Daraus folgt sofort für alle $\alpha > 0$

$$\nabla_x L_\alpha(\bar x, \bar\lambda) = \nabla_x L(\bar x, \bar\lambda) + \alpha h(\bar x)^T \nabla h(\bar x) = 0$$

wegen $h(\bar x) = 0$.

Weiter gilt

$$\begin{aligned}\nabla^2_{xx} L_\alpha(\bar x, \bar\lambda) &= \nabla^2_{xx} L(\bar x, \bar\lambda) + \alpha\left(h(\bar x)\nabla^2 h(\bar x) + \nabla h(\bar x)\nabla h(\bar x)^T\right)\\ &= \nabla^2_{xx} L(\bar x, \bar\lambda) + \alpha \nabla h(\bar x)\nabla h(\bar x)^T.\end{aligned}$$

Nach Annahme gilt außerdem

$$d^T \nabla^2_{xx} L(\bar x, \bar\lambda) d > 0 \qquad \text{für alle } d \neq 0 \text{ mit } \nabla h(\bar x)^T d = 0. \tag{7.8}$$

Wir zeigen nun durch Widerspruch, dass daraus für $\alpha > 0$ groß genug folgt

$$d^T\left(\nabla^2_{xx} L(\bar x, \bar\lambda) + \alpha \nabla h(\bar x)\nabla h(\bar x)^T\right) d > 0 \quad \text{für alle } d \neq 0. \tag{7.9}$$

Angenommen, dies gilt nicht. Dann können wir Folgen $\{d^k\}_{k\in\mathbb{N}}$ mit $\|d^k\| = 1$ und $\{\alpha_k\}_{k\in\mathbb{N}}$ mit $\alpha_k \to \infty$ finden so dass gilt

$$(d^k)^T \nabla^2_{xx} L(\bar x, \bar\lambda) d^k + \alpha_k \|\nabla h(\bar x)^T d^k\|^2 \leq 0 \qquad \text{für alle } k \in \mathbb{N}.$$

Da die Folge $\{d^k\}_{k\in\mathbb{N}}$ beschränkt ist, können wir eine Teilfolge extrahieren (die wir in der Notation nicht unterscheiden) mit $d^k \to d \neq 0$ und $\alpha_k \to \infty$. Grenzübergang ergibt dann

$$d^T \nabla^2_{xx} L(\bar{x}, \bar{\lambda}) d \leq d^T \nabla^2_{xx} L(\bar{x}, \bar{\lambda}) d + \limsup_{k\to\infty} \alpha_k \|\nabla h(\bar{x})^T d^k\|^2 \leq 0.$$

Der mittlere Term kann daher nur beschränkt bleiben, falls $\nabla h(\bar{x})^T d = \lim_{k\to\infty} \nabla h(\bar{x})^T d^k = 0$ gilt. Durch Grenzübergang erhalten wir also

$$d^T \nabla^2_{xx} L(\bar{x}, \bar{\lambda}) d \leq 0 \qquad \text{für } d \neq 0 \text{ mit } \nabla h(\bar{x})^T d = 0,$$

im Widerspruch zu (7.8).

Die Behauptung folgt nun aus Satz 1.8. □

Nun ist das praktisch noch unbefriedigend, weil weder der exakte Strafparameter $\bar{\alpha}$ noch – viel wichtiger – der Lagrange-Multiplikator $\bar{\lambda}$ bekannt sind. Dann können wir aber immer noch hoffen, analog zu Abschnitt 7.1 die Konvergenz zeigen zu können.

Satz 7.8. *Seien f, h stetig, $\{\lambda^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p$ beschränkt, und $\{\alpha_k\}_{k\in\mathbb{N}} \subset (0,\infty)$ mit $\alpha_k \to \infty$. Sei $x^k \in \mathbb{R}^n$ ein globaler Minimierer von $x \mapsto L_{\alpha_k}(x, \lambda^k)$ für alle $k \in \mathbb{N}$. Dann ist jeder Häufungspunkt von $\{x^k\}_{k\in\mathbb{N}}$ ein globaler Minimierer von (7.1).*

Beweis. Aus der Optimalität von x^k und (4.11) folgt für $k \in \mathbb{N}$ beliebig

$$f(x^k) + (\lambda^k)^T h(x^k) + \frac{\alpha_k}{2}\|h(x^k)\|^2 = L_{\alpha_k}(x^k, \lambda^k) \leq \inf_{x\in X} L_{\alpha_k}(x, \lambda^k) = \inf_{x\in X} f(x). \tag{7.10}$$

Sei nun $\bar{x} \in \mathbb{R}^n$ ein Häufungspunkt. Wegen der Beschränktheit von $\{\lambda^k\}_{k\in\mathbb{N}}$ können wir durch Übergang zu einer Teilfolge annehmen, dass $x^k \to \bar{x}$ und $\lambda^k \to \bar{\lambda}$ für ein $\bar{\lambda} \in \mathbb{R}^p$ gilt. Wieder folgt aus (7.10), dass gilt

$$f(\bar{x}) + \bar{\lambda}^T h(\bar{x}) + \limsup_{k\to\infty} \frac{\alpha_k}{2}\|h(x^k)\|^2 \leq \inf_{x\in X} f(x)$$

was wegen $\alpha_k \to \infty$ nur möglich ist für $h(\bar{x}) = \lim_{k\to\infty} h(x^k) = 0$. Also ist $\bar{x}$ zulässig für (7.1), woraus $f(\bar{x}) \geq \inf_{x\in X} f(x)$ folgt. Daraus erhalten wir

$$\begin{aligned}\inf_{x\in X} f(x) &\leq \inf_{x\in X} f(x) + \limsup_{k\to\infty} \frac{\alpha_k}{2}\|h(x^k)\|^2 \\ &\leq f(\bar{x}) + \bar{\lambda}^T h(\bar{x}) + \limsup_{k\to\infty} \frac{\alpha_k}{2}\|h(x^k)\|^2 \leq \inf_{x\in X} f(x),\end{aligned}$$

und damit zuerst $\limsup_{k\to\infty} \frac{\alpha_k}{2}\|h(x^k)\|^2 = 0$ und daher $f(\bar{x}) = \inf_{x\in X} f(x)$. □

Analog zu Satz 7.4 haben wir auch ein Konvergenzresultat für stationäre Punkte.

Satz 7.9. *Seien f, h stetig differenzierbar, $\{\lambda^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p$ beschränkt, und $\{\alpha_k\}_{k\in\mathbb{N}} \in (0,\infty)$ mit $\alpha_k \to \infty$. Sei $x^k \in \mathbb{R}^n$ ein stationärer Punkt von $x \mapsto L_{\alpha_k}(x, \lambda^k)$ für alle $k \in \mathbb{N}$. Ist $\bar{x}$ ein Häufungspunkt von $\{x^k\}_{k\in\mathbb{N}}$, der die LICQ erfüllt, so konvergiert für die entsprechende Teilfolge*

$$\lambda^k + \alpha_k h(x^k) \to \bar{\lambda},$$

und $(\bar{x}, \bar{\lambda})$ erfüllt die KKT-Bedingungen für (7.1).

Beweis. Da x^k stationär ist für $x \mapsto L_{\alpha_k}(x, \lambda^k)$, gilt

$$0 = \nabla_x L_{\alpha_k}(x^k, \lambda^k) = \nabla f(x^k) + (\lambda^k + \alpha_k h(x^k))^T \nabla h(x^k).$$

Da $\nabla h(\bar{x})$ nach Annahme vollen Rang hat und stetig ist, ist $\nabla h(x^k)\nabla h(x^k)^T$ für $k \in \mathbb{N}$ groß genug invertierbar, und es folgt

$$\lambda^k + \alpha_k h(x^k) = \left(\nabla h(x^k)\nabla h(x^k)^T\right)^{-1} \nabla h(x^k) \left(-\nabla f(x^k)^T\right).$$

Durch Grenzübergang $k \to \infty$ folgt daraus

$$\lambda^k + \alpha_k h(x^k) \to \bar{\lambda} := \left(\nabla h(\bar{x})\nabla h(\bar{x})^T\right)^{-1} \nabla h(\bar{x}) \left(-\nabla f(\bar{x})^T\right). \tag{7.11}$$

Daraus folgt wegen der Beschränktheit von $\{\lambda^k\}_{k\in\mathbb{N}}$ auch die von $\{\alpha_k h(x^k)\}_{k\in\mathbb{N}}$, was wegen $\alpha_k \to \infty$ impliziert $h(\bar{x}) = \lim_{k\to\infty} h(x^k) = 0$. Schließlich folgt aus (7.11) auch

$$\nabla h(\bar{x}) \left(\nabla h(\bar{x})^T \bar{\lambda} + \nabla f(\bar{x})^T\right) = 0.$$

Aufgrund der LICQ ist das nur möglich, wenn der Term in Klammern verschwindet, was (nach Transponieren) zusammen mit der Zulässigkeit die KKT-Bedingungen ergibt. □

Diese Resultate motivieren (aber beweisen natürlich nicht dessen Konvergenz) das folgende *Multiplikator-Strafverfahren* (englisch: *augmented Lagrangian method*).

Algorithmus 7.2 : Multiplikator-Strafverfahren für Gleichungen

Input : $\alpha_0 > 0$, $\lambda^0 \in \mathbb{R}^p$

1 **for** $k = 0, \ldots$ **do**
2 Berechne x^{k+1} als stationären Punkt von $L_{\alpha_k}(x, \lambda^k)$
3 **if** $x^{k+1} \in X$ **then**
4 **return** x^{k+1}
5 **else**
6 Setze $\lambda^{k+1} = \lambda^k + \alpha_k h(x^{k+1})$
7 Wähle $\alpha_{k+1} \geq \alpha_k$

Für die Wahl von α_{k+1} kann man dabei adaptiv vorgehen; z. B.

$$\alpha_{k+1} = \begin{cases} q\alpha_k & \text{falls } \|h(x^{k+1})\| \geq \rho\|h(x^k)\|, \\ \alpha_k & \text{sonst,} \end{cases}$$

für passend gewählte $q > 1$ und $\rho \in (0, 1)$.

Für den Beweis der Konvergenz muss man dabei insbesondere garantieren, dass die so erzeugte Folge $\{\lambda^k\}_{k\in\mathbb{N}}$ beschränkt bleibt, was unter einer hinreichenden Optimalitätsbedingung 2. Ordnung zumindest lokal und für $\alpha_k \geq \bar{\alpha}$ hinreichend groß möglich ist; siehe [Bertsekas 1982, Proposition 2.7].

7.3.2 UNGLEICHUNGSNEBENBEDINGUNGEN

Wir betrachten nun den Fall reiner Ungleichungsnebenbedingungen,

$$\min_{x\in\mathbb{R}^n} f(x) \qquad \text{mit } g(x) \leq 0$$

für $f : \mathbb{R}^n \to \mathbb{R}$ und $g : \mathbb{R}^n \to \mathbb{R}^m$ stetig differenzierbar. Um darauf den Ansatz aus dem letzten Abschnitt anwenden zu können, müssen wir die Ungleichung in eine äquivalente Gleichung umschreiben. Die Idee dafür ist, eine neue Schlupfvariable $z \geq 0$ einzuführen, so dass $g(x) \leq 0$ gilt genau dann, wenn $g(x) + z = 0$ ist. Um die Vorzeichenbedingung loszuwerden, schreiben wir außerdem $z_i = s_i^2$ für $s_i \in \mathbb{R}$. Die zugehörige erweiterte Lagrangefunktion ist dann

$$L_\alpha^= : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}, \qquad L_\alpha^=(x, s, \mu) := f(x) + \mu^T(g(x) + s^2) + \frac{\alpha}{2}\|g(x) + s^2\|^2.$$

(Wir verwenden hier bereits die Notation für Lagrange-Multiplikatoren zu Ungleichungsnebenbedingungen, auch wenn es sich hier eigentlich noch um Gleichungsnebenbedingungen handelt.)

Auf dieses Problem wenden wir nun Algorithmus 7.2 an. Dabei müssen wir in Schritt 2 den Minimierer bezüglich (x, s) bestimmen, d. h. das Problem

$$\min_{x\in\mathbb{R}^n, s\in\mathbb{R}^m} f(x) + \sum_{i=1}^m \left(\mu_i(g_i(x) + s_i^2) + \frac{\alpha}{2}(g_i(x) + s_i^2)^2\right)$$

lösen. Der Clou ist nun, dass wir das wieder auf ein Problem nur in x reduzieren können, da wir für festes x und μ das optimale $s(x, \mu)$ explizit ausrechnen können. Dafür verwenden wir, dass wir jeden Summanden unabhängig für s_i minimieren können. Außerdem kommt nur s_i^2 vor, was wir wieder durch $z_i := s_i^2 \geq 0$ ersetzen können. Wir müssen also lediglich für festes $x \in \mathbb{R}^n$ und $\mu_i \in \mathbb{R}$ das skalare Problem

$$\min_{z\geq 0} \mu_i(g_i(x) + z) + \frac{\alpha}{2}(g_i(x) + z)^2$$

lösen. Ohne die Nebenbedingung ist dies offensichtlich ein konvexes quadratisches Problem in z, so dass nach Satz 1.9 der unrestringierte Minimierer die notwendige Optimalitätsbedingung

$$\hat{z}_i = -\left(\frac{\mu_i}{\alpha} + g_i(x)\right)$$

erfüllt. Um das Einsetzen in die erweiterte Lagrange-Funktion zu erleichtern, machen wir gleich eine spezielle Fallunterscheidung.

(i) $\mu_i + \alpha g_i(x) \leq 0$: Dann ist $\hat{z}_i \geq 0$, und damit ist $\bar{z}_i = \hat{z}_i$ der gewünschte Minimierer. Also ist wegen $g_i(x) + \bar{z}_i = -\frac{\mu_i}{\alpha}$

$$\mu_i(g_i(x) + \bar{z}_i) + \frac{\alpha}{2}(g_i(x) + \bar{z}_i)^2 = -\frac{\mu_i^2}{\alpha} + \frac{\alpha}{2}\frac{\mu_i^2}{\alpha^2} = -\frac{1}{2\alpha}\mu_i^2.$$

(ii) $\mu_i + \alpha g_i(x) > 0$: Dann ist $\hat{z}_i < 0$, und daher wird das Minimum in $\bar{z} = 0$ angenommen. Also ist mit quadratischer Ergänzung

$$\mu_i g_i(x) + \frac{\alpha}{2} g_i(x)^2 = \frac{1}{2\alpha}\left((\mu_i + \alpha g_i(x))^2 - \mu_i^2\right).$$

In beiden Fällen können wir dies schreiben als

$$\mu_i(g_i(x) + \bar{z}_i) + \frac{\alpha}{2}(g_i(x) + \bar{z}_i)^2 = \frac{1}{2\alpha}\left(\max\{0, \mu_i + \alpha g_i(x)\}^2 - \mu_i^2\right).$$

Dies führt auf die reduzierte erweiterte Lagrange-Funktion

$$L_\alpha(x, \mu) := L_\alpha^=(x, s(x, \mu), \mu) = f(x) + \frac{1}{2\alpha}\sum_{i=1}^{m}\left(\max\{0, \mu_i + \alpha g_i(x)\}^2 - \mu_i^2\right).$$

(Für $\mu = 0$ entspricht dies genau der quadratischen Straffunktion.) Der letzte Term $-\mu_i^2$ hängt dabei nicht von x ab und kann daher bei der entsprechenden Minimierung vernachlässigt werden.

Betrachten wir nun wie im Beweis von Satz 7.9 die notwendige Optimalitätsbedingung

$$0 = \nabla_x L_\alpha(x, \mu) = \nabla f(x) + \sum_{i=1}^{m} \max\{0, \mu_i + \alpha g_i(x)\}\nabla g_i(x),$$

wobei das Maximum wieder komponentenweise zu verstehen ist, erhalten wir daraus die Aktualisierung

$$\mu^{k+1} := \max\{0, \mu^k + \alpha_k g(x^k)\} \geq 0.$$

Für die Aktualisierung von α_k muss man nun neben der Verletzung der Ungleichungsnebenbedingung auch die der für die KKT-Bedingungen notwendigen Komplementarität der

Lagrange-Multiplikatoren betrachten. Dies macht man üblicherweise nicht direkt, sondern über eine sogenannte *Komplementaritätsfunktion*: es gilt

$$0 = \min\{-g_i(x), \mu_i\} \qquad \text{genau dann wenn} \qquad \mu_i \geq 0, g_i(x) \leq 0, \mu_i g_i(x) = 0.$$

(Denn wenn das Minimum gleich Null ist, gilt entweder $0 = -g_i(x) \leq \mu_i$ oder $0 = \mu_i \leq -g_i(x)$ und damit in beiden Fällen die Komplementarität; die umgekehrte Richtung ist klar.)[1]

Also vergrößert man α_k, wenn gilt

$$\| \min\{-g(x^{k+1}), \mu^{k+1}\}\| \geq \rho \| \min\{-g(x^k), \mu^k\}\|.$$

Die Kombination mit Gleichungsnebenbedingungen ist nun trivial, und wir erhalten das volle Multiplikator-Strafverfahren.

Algorithmus 7.3 : Multiplikator-Strafverfahren

Input : $\alpha_0 > 0$, $\lambda^0 \in \mathbb{R}^p$, $\mu^0 \geq 0 \in \mathbb{R}^m$, $q > 1$, $\rho \in (0,1)$

1 **for** $k = 0, \ldots$ **do**

2 Berechne x^{k+1} als stationären Punkt von

$$f(x) + (\lambda^k)^T h(x) + \frac{\alpha_k}{2}\|h(x)\|^2 + \frac{1}{2\alpha_k}\|(\mu^k + \alpha_k g(x))^+\|^2$$

if $h(x^{k+1}) = 0$ *und* $\min\{-g(x^{k+1}), \mu^k\} = 0$ **then**

3 **return** x^{k+1}

4 **else**

5 Setze $\lambda^{k+1} = \lambda^k + \alpha_k h(x^{k+1})$

6 Setze $\mu^{k+1} = \max\{0, \mu^k + \alpha_k g(x^{k+1})\}$

7 Setze

$$\alpha_{k+1} = \begin{cases} q\alpha_k & \text{falls } \|h(x^{k+1})\| \geq \rho\|h(x^k)\| \\ & \text{oder } \| \min\{-g(x^{k+1}), \mu^{k+1}\}\| \geq \rho\| \min\{-g(x^k), \mu^k\}\|, \\ \alpha_k & \text{sonst,} \end{cases}$$

Ähnlich wie zuvor kann man nun zeigen, dass unter der Annahme der Regularität von $\bar{x}$ und der Beschränktheit der Folgen $\{\lambda^k\}_{k\in\mathbb{N}}$ und $\{\mu^k\}_{k\in\mathbb{N}}$ (z. B. unter einer hinreichenden Optimalitätsbedingung 2. Ordnung) die (Teil-)Folge (x^k, μ^k, λ^k) gegen einen KKT-Punkt $(\bar{x}, \bar{\mu}, \bar{\lambda})$ konvergiert.

[1] Eine alternative Komplementaritätsfunktion ist $\mu_i - \max\{0, \mu_i + \alpha g_i(x)\}$ für $\alpha > 0$ beliebig, was eine weitere Motivation für die Wahl von μ^{k+1} ist.

8 BARRIERE- UND INNERE-PUNKTE-VERFAHREN

Strafverfahren sind ungeeignet, wenn die Zielfunktion f für $x \notin X$ gar nicht definiert ist. In *Barriereverfahren* wird deshalb der Minimierer $\bar{x} \in X$ durch eine Folge von *inneren* Punkten angenähert. (Man spricht daher auch von *Innere-Punkte-Verfahren*.) Da für Gleichungsnebenbedingungen der zulässige Bereich keine inneren Punkte besitzt, betrachten wir hier nur reine Ungleichungsnebenbedingungen

$$X = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0,\ 1 \leq i \leq m\}$$

für $g_i : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar. (Zusätzliche Gleichungsnebenbedingungen werden entweder durch eine Straffunktion oder den in Abschnitt 6.3 vorgestellten Ansatz behandelt.) Anstelle einer Straffunktion, die erst außerhalb von X zu wirken beginnt, verwendet man nun eine *Barrierefunktion*, die bereits bei Annäherung an den Rand von X gegen unendlich strebt; verbreitet ist dabei die logarithmische Barrierefunktion $x \mapsto -\ln(-g_i(x))$. Anstelle von f minimiert man daher für $\alpha > 0$ die Funktion

$$B_\alpha(x) := f(x) + \alpha \sum_{i=1}^{m} [-\ln(-g_i(x))] =: f(x) + \alpha\beta(x).$$

Ist diese Funktion nicht definiert wegen $g_i(x) \geq 0$ für ein $1 \leq i \leq m$, so setzen wir $B_\alpha(x) := \infty$. Damit also eine Lösung von

$$\min_{x \in \mathbb{R}^n} B_\alpha(x) \tag{8.1}$$

existiert, muss es einen *strikt* zulässigen Punkt mit $g_i(x) < 0$ für alle $1 \leq i \leq m$ geben – dies ist genau die Slater-Bedingung (4.8).

Satz 8.1. *Sei $f : X \to \mathbb{R}$ stetig, $X \subset \mathbb{R}^n$ beschränkt, nichtleer und abgeschlossen. Gilt die Slater-Bedingung (4.8), dann existiert für alle $\alpha > 0$ eine Lösung $x_\alpha \in X$ von (8.1).*

Beweis. Die Slater-Bedingung garantiert die Existenz eines strikt zulässigen Punktes $\tilde{x} \in X$, für den

$$M := B_\alpha(\tilde{x}) < \infty$$

gilt. Wir betrachten nun die Menge

$$X_M := \{x \in X \mid B_\alpha(x) \leq M\}.$$

Offensichtlich muss ein Minimierer von B_α (falls existent) in X_M liegen, d. h. auch Lösung sein von

$$\min_{x \in X_M} B_\alpha(x). \tag{8.2}$$

Es genügt also zu zeigen, dass dieses Problem eine Lösung hat.

Wir zeigen zuerst, dass X_M kompakt ist. Nach Annahme ist mit X auch $X_M \subset X$ beschränkt. Für die Abgeschlossenheit sei $\{x^k\}_{k\in\mathbb{N}} \subset X_M \subset X$ eine konvergente Folge mit Grenzwert $x \in X$ wegen der angenommenen Abgeschlossenheit von X. Wegen der Stetigkeit von f, g_i, $1 \le i \le m$, sowie $t \mapsto \ln(t)$ ist B_α stetig auf X_M (beachte, dass $g_i(x) < 0$ für alle $x \in X_M$ gelten muss). Durch Grenzübergang folgt daher auch $B_\alpha(x) \le M$, d. h. $x \in X_M$. Also ist X_M kompakt, und aus der Stetigkeit von B_α folgt mit Satz 1.4 die Existenz einer Lösung von (8.2). □

Man kann nun hoffen, dass für $\alpha \to 0$ die Folge $\{x_\alpha\}_{\alpha>0}$ der entsprechenden Minimierer von (8.1) gegen einen Minimierer $\bar{x} \in X$ von f konvergiert. Entsprechend hat das Barriereverfahren nun die Form von

Algorithmus 8.1 : Barriereverfahren

Input : $\alpha_0 > 0$, $x^0 \in X$ strikt zulässig

1 **for** $k = 0, \ldots$ **do**

2 Berechne x^{k+1} als globalen Minimierer von (8.1) mit $\alpha = \alpha_k$ (und Startwert x^k)

3 Wähle $\alpha_{k+1} < \alpha_k$

Wieder wird man in Schritt 3 Verfahren aus Kapitel 2 anwenden, wobei man diesmal (etwa bei der Schrittweitensuche) aufpassen muss, dass nur strikt zulässige Iterierte erzeugt werden.

Auch hier kann man Monotonieeigenschaften der so erzeugten Folge zeigen.

Lemma 8.2. *Angenommen, Algorithmus 8.1 erzeugt für eine streng monoton fallende Nullfolge $\{\alpha_k\}_{k\in\mathbb{N}} \subset (0,\infty)$ eine unendliche Folge $\{x^k\}_{k\in\mathbb{N}} \subset X$. Dann gilt*

(i) $\{\beta(x^k)\}_{k\in\mathbb{N}}$ *ist monoton wachsend;*

(ii) $\{f(x^k)\}_{k\in\mathbb{N}}$ *ist monoton fallend.*

Beweis. Zu (i): Wir verwenden wieder die globale Optimalität zusammen mit der strikten Zulässigkeit aller x^k. Aus $B_{\alpha_k}(x^k) \le B_{\alpha_k}(x^{k+1})$ und $B_{\alpha_{k+1}}(x^{k+1}) \le B_{\alpha_{k+1}}(x^k)$ folgt durch Addition und Umformen

$$(\alpha_k - \alpha_{k+1})\left(\beta(x^k) - \beta(x^{k+1})\right) \le 0.$$

Aus $\alpha_k > \alpha_{k+1}$ folgt nun $\beta(x^k) \le \beta(x^{k+1})$.

Zu (ii): Aus der Optimalität folgt zusammen mit (i)

$$\begin{aligned} 0 &\leq B_{\alpha_{k+1}}(x^k) - B_{\alpha_{k+1}}(x^{k+1}) = f(x^k) - f(x^{k+1}) + \alpha_{k+1}\left(\beta(x^k) - \beta(x^{k+1})\right) \\ &\leq f(x^k) - f(x^{k+1}). \end{aligned}$$

□

Um die Konvergenz von Algorithmus 8.1 zeigen zu können, bedarf es einer Art Regularitätsbedingung. Dafür definieren wir das *strikte Innere*

$$X^- := \{x \in \mathbb{R}^n \mid g_i(x) < 0,\ 1 \leq i \leq m\} \subset X$$

der zulässigen Menge. Beachte, dass dies im Allgemeinen nur eine Teilmenge des topologischen Inneren sein muss! (Das strikte Innere hängt ja, im Gegensatz zum topologischen Inneren, von der konkreten Beschreibung von X durch die g_i ab.)

Beispiel 8.3. Für $g : \mathbb{R} \to \mathbb{R}$, $g(x) = \max\{0, x\}^2$, stetig differenzierbar ist $X = (-\infty, 0]$ und damit $\operatorname{int} X = (-\infty, 0)$, aber $X^- = \emptyset$.

Satz 8.4. *Seien $f : X \to \mathbb{R}$ und $g_i : \mathbb{R}^n \to \mathbb{R}$, $1 \leq i \leq m$, stetig. Es gelte $X^- \neq \emptyset$ sowie $\overline{X^-} = X$. Erzeugt Algorithmus 8.1 für eine streng monoton fallende Nullfolge $\{\alpha_k\}_{k\in\mathbb{N}} \subset (0, \infty)$ eine unendliche Folge $\{x^k\}_{k\in\mathbb{N}} \subset X$, so ist jeder Häufungspunkt ein globaler Minimierer von f in X.*

Beweis. Sei $\{x^k\}_{k\in K} \subset X^- \subset X$ eine konvergente Teilfolge mit Grenzwert $\bar{x} \in X$ wegen der Abgeschlossenheit von X. Angenommen, $\bar{x}$ wäre kein globaler Minimierer von f in X. Dann existiert ein $x \in X$ mit $f(x) < f(\bar{x})$. Wegen $\overline{X^-} = X$ und der Stetigkeit von f existiert daher ein $\hat{x} \in X^-$ nahe genug an x, so dass ebenfalls $f(\hat{x}) < f(\bar{x})$ gilt.

Aus Lemma 8.2 (i), der Optimalität von x^k, und $x^0, \hat{x} \in X^-$ folgt nun

$$f(x^k) + \alpha_k\beta(x^0) \leq f(x^k) + \alpha_k\beta(x^k) \leq f(\hat{x}) + \alpha_k\beta(\hat{x}) < \infty$$

für alle $k \geq 0$. Stetigkeit von f und $\alpha_k \to 0$ ergibt nun

$$f(\bar{x}) = \lim_{K\ni k\to\infty} f(x^k) \leq f(\hat{x}) + \lim_{K\ni k\to\infty} \alpha_k\left(\beta(\hat{x}) - \beta(x^0)\right) = f(\hat{x}),$$

im Widerspruch zu $f(\hat{x}) < f(\bar{x})$. Also ist $\bar{x} \in X$ ein globaler Minimierer. □

Für konvexe Ungleichungsnebenbedingungen ist die Annahme $\overline{X^-} = X$ immer erfüllt.

Für stetig differenzierbare g_i folgt aus $\frac{d}{dt}\ln(t) = \frac{1}{t}$ mit der Summen- und Kettenregel

$$\nabla B_\alpha(x) = \nabla f(x) - \alpha \sum_{i=1}^{m} \frac{\nabla g_i(x)}{g_i(x)} \qquad \text{für alle } x \in X^-. \tag{8.3}$$

Vergleicht man dies wieder mit (4.10), so erhält man

$$\nabla B_\alpha(x) = \nabla_x L(x, \mu_\alpha) \qquad \text{für} \quad \mu_\alpha := \frac{\alpha}{-g(x)} > 0,$$

und in der Tat kann man analog zu Satz 7.4 unter der LICQ zeigen, dass für $x^k \to \bar{x}$ die entsprechenden μ^k gegen den Lagrange-Multiplikator $\bar{\mu}$ konvergieren. Dies wird im *primal-dualen Innere-Punkte-Verfahren* ausgenutzt. Dafür schreibt man die notwendige Optimalitätsbedingung $\nabla B_\alpha(x_\alpha) = 0$ um in

$$\begin{cases} \nabla_x L(x_\alpha, \mu_\alpha) = 0, \\ -(\mu_\alpha)_i g_i(x_\alpha) = \alpha, \quad 1 \le i \le m, \end{cases}$$

(vergleiche die KKT-Bedingungen (4.9) – die Komplementaritätsbedingungen $\bar{\mu}_i g_i(\bar{x}) = 0$ werden hier "von innen" angenähert). Auf dieses System wird nun ein Newton-Verfahren angewendet, wobei nach jedem Newton-Schritt der Penalty-Parameter α geeignet reduziert wird; eine Schrittweitenregel sorgt dabei dafür, dass die neuen Iterierten sich nicht zu weit von (x_α, μ_α) entfernen. Für konvexe Optimierungsprobleme haben sich diese Verfahren als sehr leistungsfähig erwiesen; siehe z. B. [Boyd & Vandenberghe 2004, Kapitel 11.7]. Wir betrachten hier nur den Spezialfall von linearen Optimierungsproblemen, wo sich Innere-Punkte-Verfahren als moderne Alternative zu Simplex-Verfahren durchgesetzt haben.

INNERE-PUNKTE-VERFAHREN FÜR LINEARE OPTIMIERUNGSPROBLEME

Wir betrachten also wieder das Problem (LP), diesmal in der alternativen Form

$$\text{(LP)} \qquad \begin{cases} \min\limits_{x \in \mathbb{R}^n} c^T x \\ \text{mit } Ax = b, \\ \qquad x \ge 0. \end{cases}$$

Wir nehmen in Folge stets an, dass dieses Problem eine zulässige Lösung $\bar{x} \in P^=(A, b)$ besitzt. Da $\bar{x}$ nach Satz 4.17 regulär ist, existieren nach Satz 4.18 Lagrange-Multiplikatoren $s \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}^m$ mit

$$\begin{cases} A^T \lambda + s = c, \\ \qquad A\bar{x} = b, \\ \qquad \bar{x}_i s_i = 0, \\ \qquad \bar{x}, s \ge 0. \end{cases}$$

Wir ersetzen nun diese KKT-Bedingungen durch die KKT-Bedingungen des entsprechenden Barriereproblems

$$(8.4) \qquad \begin{cases} \min\limits_{x \in \mathbb{R}^n} c^T x - \alpha \sum\limits_{i=1}^{n} \ln(x_i) \\ \text{mit } Ax = b, \end{cases}$$

was ebenfalls ein (sogar strikt) konvexes Optimierungsproblem ist. Wir haben hier nur affin-lineare Gleichungsnebenbedingungen, also existiert nach Satz 4.12 für jede Lösung $x_\alpha \in \mathbb{R}^n$ mit notwendigerweise $x_\alpha > 0$ ein Lagrange-Multiplikator λ_α so, dass die KKT-Bedingungen

$$\begin{cases} A^T\lambda_\alpha + \dfrac{\alpha}{x_\alpha} = c, \\ Ax_\alpha = b, \end{cases}$$

erfüllt sind. Wir schreiben nun $s_\alpha := \frac{\alpha}{x_\alpha} > 0$ und erhalten die *zentralen-Pfad-Bedingungen*

$$\begin{cases} A^T\lambda_\alpha + s_\alpha = c, \\ Ax_\alpha = b, \\ [x_\alpha]_i[s_\alpha]_i = \alpha, \\ x_\alpha, s_\alpha > 0. \end{cases} \tag{8.5}$$

Wir müssen noch zeigen, dass überhaupt Lösungen von (8.5) existieren, was wir analog zu Satz 8.1 machen. (Beachte, dass $P^=(A,b)$ nicht als beschränkt vorausgesetzt war.) Dazu definieren wir die *primal-dual zulässige Menge*

$$Z := \left\{(x,\lambda,s) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \;\middle|\; Ax = b, A^T\lambda + s = c, x, s \geq 0\right\}$$

sowie die *strikt zulässige Menge*

$$Z^+ := \left\{(x,\lambda,s) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \;\middle|\; Ax = b, A^T\lambda + s = c, x, s > 0\right\}.$$

Satz 8.5. *Sei Z^+ nichtleer. Dann existiert für alle $\alpha > 0$ eine Lösung $(x_\alpha, \lambda_\alpha, s_\alpha)$ von (8.5). Dabei sind (x_α, s_α) eindeutig bestimmt. Hat A vollen Rang, ist auch λ_α eindeutig.*

Beweis. Da (8.4) strikt konvex ist, genügt es zu zeigen, dass ein Minimierer $x_\alpha > 0$ existiert. Dazu verwenden wir wieder einen strikt zulässigen Punkt $(\hat{x}, \hat{\lambda}, \hat{s}) \in Z^+$, für den insbesondere gilt

$$M := B_\alpha(\hat{x}) := c^T\hat{x} - \alpha\sum_{i=1}^{n}\ln(\hat{x}_i) < \infty,$$

und betrachten die Menge

$$X_M := \{x \in \mathbb{R}^n \mid Ax = b, B_\alpha(x) \leq M\}.$$

Wie im Beweis von Satz 8.1 folgt die Abgeschlossenheit von X_M; wir müssen also nur noch die Beschränktheit zeigen. Zunächst gilt für alle $x \in X_M$ und für $(\hat{x}, \hat{\lambda}, \hat{s}) \in Z^+$ nach

Definition

$$\begin{aligned} M &\geq c^T x - \alpha \sum_{i=1}^{n} \ln(x_i) \\ &= c^T x - (Ax - b)^T \hat{\lambda} - \alpha \sum_{i=1}^{n} \ln(x_i) \\ &= c^T x - (c - \hat{s})^T x + b^T \hat{\lambda} - \alpha \sum_{i=1}^{n} \ln(x_i) \\ &= x^T \hat{s} + b^T \hat{\lambda} - \alpha \sum_{i=1}^{n} \ln(x_i), \end{aligned}$$

d. h.

$$\sum_{i=1}^{n} (\hat{s}_i x_i - \alpha \ln(x_i)) \leq M - b^T \hat{\lambda} < \infty.$$

Nun gilt wegen $\hat{s}_i > 0$ sowohl $\lim_{t\to\infty}(s_i t - \alpha \ln(t)) = \infty$ als auch $\lim_{t\to 0}(s_i t - \alpha \ln(t)) = \infty$ (d. h. die Summe kann nicht beschränkt bleiben, wenn ein Summand divergiert), so dass die Annahme, X_M wäre unbeschränkt, zu einem Widerspruch führt. Also ist X_M kompakt, und $\min_{x\in X_M} B_\alpha(x)$ hat eine eindeutige Lösung $x_\alpha > 0$, die auch die eindeutige Lösung von (8.4) ist. Damit ist auch $s_\alpha := \frac{\alpha}{x_\alpha}$ eindeutig. Dagegen ist der Lagrange-Multiplikator λ_α aus (8.5) nur eindeutig, falls A vollen Rang hat; in diesem Fall ist er die eindeutige Lösung von $A^T\lambda = c - s_\alpha$. □

Hat also A vollen Rang (was in der linearen Optimierung keine große Einschränkung ist), ist der *zentrale Pfad* $\alpha \mapsto (x_\alpha, \lambda_\alpha, s_\alpha)$ wohldefiniert.

Die Idee ist nun, die Gleichungen in den zentralen-Pfad-Bedingungen (8.5) durch ein Newton-Verfahren zu lösen, wobei die Ungleichungen durch eine Liniensuche garantiert werden. Dafür ist etwas Notation hilfreich: wir definieren für Vektoren $u, v \in \mathbb{R}^n$ das komponentenweise *Hadamard-Produkt*

$$u \odot v := (u_1 v_1, \ldots, u_n v_n)^T \in \mathbb{R}^n,$$

die durch $v \in \mathbb{R}^n$ erzeugte Diagonalmatrix

$$D_v := \operatorname{diag}(v_1, v_2, \ldots, v_n) \in \mathbb{R}^{n\times n},$$

sowie $\mathbb{1} := (1, \ldots, 1)^T \in \mathbb{R}^n$. Dann ist die Lösung $(x_\alpha, \lambda_\alpha, s_\alpha)$ von (8.5) auch eine Nullstelle von

$$F_\alpha(x, \lambda, s) := \begin{pmatrix} A^T\lambda + s - c \\ Ax - b \\ x \odot s - \alpha\mathbb{1} \end{pmatrix},$$

und ein Schritt im Newton-Verfahren basiert auf der Lösung von

$$F'_\alpha(x^k, \lambda^k, s^k)(\Delta x, \Delta\lambda, \Delta s) = -F_\alpha(x^k, \lambda^k, s^k) \quad \text{für} \quad F'_\alpha(x, \lambda, s) = \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ D_s & 0 & D_x \end{pmatrix}.$$

Dazu müssen wir zuerst garantieren, dass $F'_\alpha(x^k, \lambda^k, s^k)$ stets invertierbar ist.

Satz 8.6. *Seien $(x, \lambda, s) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ mit $x, s > 0$. Hat A vollen Rang, dann ist $F'_\alpha(x, \lambda, s)$ regulär für alle $\alpha > 0$.*

Beweis. Es genügt zu zeigen, dass $F'_\alpha(x, \lambda, s)$ injektiv ist. Sei also $w = (w_1, w_2, w_3)$ mit $F'_\alpha(x, \lambda, s)w = 0$ gegeben, d. h.

$$\begin{aligned} A^T w_2 + w_3 &= 0, \\ Aw_1 &= 0, \\ D_s w_1 + D_x w_3 &= 0. \end{aligned}$$

Wegen $x > 0$ ist D_x invertierbar mit $D_x^{-1} = \operatorname{diag}(x_1^{-1}, \dots x_n^{-1})$; wir können also die dritte Gleichung nach w_3 auflösen und in die erste Gleichung einsetzen. Multiplizieren mit $-w_1^T$ ergibt dann zusammen mit der zweiten Gleichung

$$0 = -w_1^T A^T w_2 + w_1^T D_x^{-1} D_s w_1 = -(Aw_1)^T w_2 + w_1^T D_x^{-1} D_s w_1 = w_1^T D_x^{-1} D_s w_1.$$

Da $D_x^{-1} D_s = \operatorname{diag}(\frac{s_1}{x_1}, \dots, \frac{s_n}{x_n})$ wegen $\frac{s_i}{x_i} > 0$ positiv definit ist, folgt daraus $w_1 = 0$ und damit auch $w_3 = -D_x^{-1} D_s w_1 = 0$. Da A vollen Rang hat, folgt aus der ersten Gleichung schließlich $w_2 = 0$. □

Insbesondere ist $F'_\alpha(x^k, \lambda^k, s^k)$ regulär für alle $(x^k, \lambda^k, s^k) \in Z^+$, die nach Definition $A^T\lambda^k + s^k = c$ und $Ax^k = b$ erfüllen. Der Newton-Schritt vereinfacht sich dann zu

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ D_{s^k} & 0 & D_{x^k} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \\ \Delta s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -x^k \odot s^k + \alpha \mathbb{1} \end{pmatrix}.$$

Setzen wir für beliebiges $t_k > 0$

$$x^{k+1} = x^k + t_k \Delta x, \qquad \lambda^{k+1} = \lambda^k + t_k \Delta\lambda, \qquad s^{k+1} = s^k + t_k \Delta s,$$

so folgt aus der ersten Zeile des Newton-Schritts

$$\begin{aligned} A^T\lambda^{k+1} + s^{k+1} - c &= A^T(\lambda^k + t_k\Delta\lambda) + (s^k + t_k\Delta s) - c \\ &= (A^T\lambda^k + s^k - c) + t_k(A^T\Delta\lambda^k + \Delta s) \\ &= 0 \end{aligned}$$

und analog aus der zweiten Zeile

$$Ax^{k+1} - b = Ax^k - b + t_k(A\Delta x) = 0.$$

Wählen wir $t_k > 0$ klein genug, dass $x^{k+1}, s^{k+1} > 0$ ist, gilt daher auch $(x^{k+1}, \lambda^{k+1}, s^{k+1}) \in Z^+$.

Beachte, dass nur die rechte Seite explizit von α abhängt; anstelle die Konvergenz des Newton-Verfahrens abzuwarten, können wir im nächsten Schritt auch gleich α reduzieren. Statt mit konstantem α verwenden wir also eine Nullfolge $\{\alpha_k\}_{k\in\mathbb{N}}$, so dass die Iterierten (x^k, λ^k, s^k) näherungsweise dem zentralen Pfad $\alpha \mapsto (x_\alpha, \lambda_\alpha, s_\alpha)$ für $\alpha \to 0$ folgen; man spricht daher von einem *Pfadverfolgungsverfahren.* Analog zu inexakten Newton-Verfahren bietet sich hier die Wahl

$$\alpha_k = \sigma_k \mu_k \qquad \text{für } \mu_k := \frac{1}{n}(x^k)^T s^k, \qquad \sigma_k \in (0,1),$$

an, denn $x^k, s^k \geq 0$ erfüllen die Komplementaritätsbedingungen genau dann, wenn $(x^k)^T s^k = 0$ gilt. (Der *Zentrierungsparameter* σ_k kann verwendet werden, um das Verfahren nahtlos zwischen der Verfolgung des zentralen Pfads ($\sigma_k = 1$) und dem Newtonverfahren für die exakten Komplementaritätsbedingungen ($\sigma_k = 0$) zu steuern.)

Ähnlich wählt man auch die Schrittweite t_k als das maximale $t \in (0,1]$ so, dass für $\gamma \in (0,1)$ gilt

$$(x^{k+1}, \lambda^{k+1}, s^{k+1}) \in Z^\gamma := \left\{ (x,\lambda,s) \in Z^+ \;\middle|\; x_i s_i \geq \frac{\gamma}{n} x^T s, 1 \leq i \leq n \right\}.$$

Zusammen erhalten wir das folgende zulässige Pfadverfolgungsverfahren.

Algorithmus 8.2 : zulässiges Innere-Punkte-Verfahren

Input : $\gamma \in (0,1)$, $0 < \sigma_{\min} \leq \sigma_{\max} < 1$, $\varepsilon > 0$, $(x^0, \lambda^0, s^0) \in Z^\gamma$

1 **for** $k = 0, \ldots$ **do**

2 **if** $\mu_k := \frac{1}{n}(x^k)^T s^k \leq \varepsilon$ **then return** (x^k, λ^k, s^k)

3 Wähle $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$

4 Löse

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ D_{s^k} & 0 & D_{x^k} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -x^k \odot s^k + \sigma_k \mu_k \mathbb{1} \end{pmatrix} \tag{8.6}$$

5 Bestimme t_k als maximales $t \in (0,1]$ mit

$$(x^k + t\Delta x, \lambda^k + t\Delta\lambda, s^k + t\Delta s) \in Z^\gamma$$

6 Setze

$$x^{k+1} = x^k + t_k \Delta x, \qquad \lambda^{k+1} = \lambda^k + t_k \Delta\lambda, \qquad s^{k+1} = s^k + t_k \Delta s$$

Dabei kann für $(x^0, \lambda^0, s^0) \in Z^+$ stets $\gamma := \frac{n \min_i x_i^0 s_i^0}{(x^0)^T s^0} \in (0,1)$ gewählt werden, so dass auch $(x^0, \lambda^0, s^0) \in Z^\gamma$ gilt.

Wir zeigen nun die Konvergenz von Algorithmus 8.2, wofür wir wegen der nach Konstruktion stets zulässigen Iterierten nur garantieren müssen, dass $(x^k)^T s^k \to 0$ bzw. äquivalent $\mu^k \to 0$ konvergiert. (Dies rechtfertigt auch die Abbruchbedingung in Schritt 2.) Dafür verwenden wir eine Reihe technischer Lemmas sowie die Notation

$$(x^k(t), \lambda^k(t), s^k(t)) := (x^k + t\Delta x, \lambda^k + t\Delta\lambda, s^k + t\Delta s),$$
$$\mu_k(t) := \frac{1}{n} x^k(t)^T s^k(t),$$

für $t > 0$.

Lemma 8.7. *Sei $(\Delta x, \Delta\lambda, \Delta s)$ eine Lösung von (8.6). Dann gilt*

(i) $\Delta x^T \Delta s = 0$;

(ii) $\mu_k(t) = (1 - t(1 - \sigma_k))\mu_k$.

Beweis. Zu (i): Dies folgt wie in Satz 8.6 durch Multiplikation der ersten Gleichung von (8.6) mit Δx^T und Verwenden der zweiten Gleichung.

Zu (ii): Die dritte Gleichung von (8.6) lautet komponentenweise für $1 \le i \le n$

$$s_i^k \Delta x_i + x_i^k \Delta s_i = -x_i^k s_i^k + \sigma_k \mu_k,$$

woraus wir durch Summation über alle i erhalten

$$(s^k)^T \Delta x + (x^k)^T \Delta s = -(x^k)^T s^k + n\sigma_k\mu_k = -(1 - \sigma_k)(x^k)^T s^k$$

nach Definition von μ_k. Zusammen mit (i) folgt dann

$$x^k(t)^T s^k(t) = (x^k)^T s^k + t\left((x^k)^T \Delta s + \Delta x^T s^k\right) + t^2 \Delta x^T \Delta s \quad = (1 - t(1 - \sigma_k))(x^k)^T s^k,$$

und Division durch n ergibt die Behauptung. □

Das nächste Lemma ist eine Abschätzung für das Hadamard-Produkt.

Lemma 8.8. *Für $u, v \in \mathbb{R}^n$ mit $u^T v \ge 0$ gilt*

$$\|u \odot v\|_2 \le \frac{1}{2}\|u + v\|_2^2.$$

Beweis. Zunächst gilt für alle $a, b \in \mathbb{R}$

$$\frac{1}{4}(a + b)^2 = \frac{1}{4}(a - b)^2 + ab \ge ab.$$

Aus der Voraussetzung an u, v folgt weiter

$$0 \leq u^T v = \sum_{i=1}^{n} u_i v_i = \sum_{u_i v_i \geq 0} u_i v_i - \sum_{u_i v_i < 0} |u_i v_i|$$

und damit wegen $\|x\|_2 \leq \|x\|_1$ für alle $x \in \mathbb{R}^n$

$$\begin{aligned}\|u \odot v\|_2 \leq \|u \odot v\|_1 = \sum_{i=1}^{n} |u_i v_i| \leq 2 \sum_{u_i v_i \geq 0} u_i v_i \leq \frac{1}{2} \sum_{u_i v_i \geq 0} (u_i + v_i)^2 \leq \frac{1}{2} \sum_{i=1}^{n} (u_i + v_i)^2 \\ = \frac{1}{2} \|u + v\|_2^2.\end{aligned}$$

□

Damit können wir die folgende Abschätzung beweisen.

Lemma 8.9. *Sei $(x^k, \lambda^k, s^k) \in Z^\gamma$ und $(\Delta x, \Delta\lambda, \Delta s)$ Lösung von (8.6). Dann gilt*

$$\|\Delta x \odot \Delta s\|_2 \leq \frac{1}{2} \left(1 + \frac{1}{\gamma}\right) n \mu_k.$$

Beweis. Multiplizieren der letzten Gleichung in (8.6) mit

$$(D_{x^k} D_{s^k})^{-1/2} := \operatorname{diag}((x_1^k s_1^k)^{-1/2}, \dots (x_n^k s_n^k)^{-1/2})$$

ergibt für $A_k := D_{x^k}^{1/2} D_{s^k}^{-1/2}$ (da Diagonalmatrizen kommutieren)

$$A_k^{-1} \Delta x + A_k \Delta s = (D_{x^k} D_{s^k})^{-1/2} (-x^k \odot s^k + \sigma_k \mu_k \mathbb{1}).$$

Wegen $(A_k^{-1} \Delta x)^T (A_k \Delta s) = \Delta x^T \Delta s = 0$ nach Lemma 8.7 (i) folgt aus Lemma 8.8 dann

$$\begin{aligned}\|\Delta x \odot \Delta s\|_2 &= \|(A_k^{-1} \Delta x) \odot (A_k \Delta s)\|_2 \leq \frac{1}{2} \|A_k^{-1} \Delta x + A_k \Delta s\|_2^2 \\ &= \frac{1}{2} \|(D_{x^k} D_{s^k})^{-1/2} (-x^k \odot s^k + \sigma_k \mu_k \mathbb{1})\|_2^2 \\ &= \frac{1}{2} \sum_{i=1}^{n} \left(-\sqrt{x_i^k s_i^k} + \sigma_k \mu_k \frac{1}{\sqrt{x_i^k s_i^k}} \right)^2 = \frac{1}{2} \sum_{i=1}^{n} \left(x_i^k s_i^k - 2\sigma_k \mu_k + \frac{\sigma_k^2 \mu_k^2}{x_i^k s_i^k} \right) \\ &\leq \frac{1}{2} \left((x^k)^T s^k - 2n\sigma_k \mu_k + \sigma_k^2 \mu_k^2 \frac{n}{\gamma \mu_k} \right) \\ &\leq \frac{1}{2} \left(1 + \frac{1}{\gamma}\right) n \mu_k,\end{aligned}$$

wobei wir $x_i^k s_i^k \geq \gamma \mu_k$ für $(x^k, \lambda^k, s^k) \in Z^\gamma$, die Definition von μ_k, und $\sigma_k \in (0, 1)$ verwendet haben. □

Das nächste Lemma charakterisiert die maximal zulässige Schrittweite und ist der wesentliche Schritt im Konvergenzbeweis für Algorithmus 8.2.

Lemma 8.10. *Sei* $(x^k, \lambda^k, s^k) \in Z^\gamma$ *und* $(\Delta x, \Delta\lambda, \Delta s)$ *Lösung von* (8.6). *Dann gilt*

$$(x^k(t), \lambda^k(t), s^k(t)) \in Z^\gamma \qquad \textit{für} \quad 0 \le t \le t_k := 2\gamma\frac{\sigma_k}{n}\frac{1-\gamma}{1+\gamma}.$$

Beweis. Wir müssen insbesondere zeigen, dass $[x^k(t)]_i[s^k(t)]_i \ge \gamma\mu_k(t)$ für $t \le t_k$ gilt. Zunächst folgt aus komponentenweiser Betrachtung der letzten Gleichung von (8.6), dass gilt

$$s_i^k\Delta x_i + x_i^k\Delta s_i = -x_i^k s_i^k + \sigma_k\mu_k, \qquad 1 \le i \le n.$$

Weiter folgt aus Lemma 8.9

$$|\Delta x_i \Delta s_i| \le \|\Delta x \odot \Delta s\|_2 \le \frac{1}{2}\left(1+\frac{1}{\gamma}\right)n\mu_k.$$

Zusammen ergibt das

$$\begin{aligned} x_i^k(t)s_i^k(t) &= (x_i^k + t\Delta x_i)(s_i^k + t\Delta s_i) \\ &= x_i^k s_i^k + t(x_i^k\Delta s_i + s_i^k\Delta x_i) + t^2\Delta x_i\Delta s_i \\ &\ge (1-t)x_i^k s_i^k + t\sigma_k\mu_k - t^2|\Delta x_i\Delta s_i| \\ &\ge (1-t)\gamma\mu_k + t\sigma_k\mu_k - \frac{t^2}{2}\left(1+\frac{1}{\gamma}\right)n\mu_k, \end{aligned}$$

wobei wir im letzten Schritt wieder $x_i^k s_i^k \ge \gamma\mu_k$ für $(x^k, \lambda^k, s^k) \in Z^\gamma$ verwendet haben. Es genügt nach Lemma 8.7 (ii) also zu garantieren, dass gilt

$$(1-t)\gamma\mu_k + t\sigma_k\mu_k - \frac{t^2}{2}\left(1+\frac{1}{\gamma}\right)n\mu_k \ge \gamma\mu_k(t) = \gamma(1-t(1-\sigma_k))\mu_k.$$

Vereinfachen und Auflösen nach $t > 0$ ergibt dann genau $t \le t_k$.

Es bleibt noch zu zeigen, dass $(x^k(t), \lambda^k(t), s^k(t)) \in Z^+$ für alle $t \le t_k$ gilt. Da die Gleichungsnebenbedingungen für alle $t \ge 0$ erhalten bleiben, ist nur $x^k(t), s^k(t) > 0$ von Bedeutung. Aber das folgt aus

$$[x^k(t)]_i[s^k(t)]_i \ge \gamma(1-t(1-\sigma_k))\mu_k > 0$$

wegen $\mu_k = (x^k)^T(s^k)/n > 0$ für $x^k, s^k > 0$, $\gamma, \sigma_k \in (0,1)$, und $t \le t_k < 1$. □

Für die so gewählte maximale Schrittweite können wir das zentrale Konvergenzresultat für Algorithmus 8.2 beweisen.

Satz 8.11. *Sei $\{x^k, \lambda^k, s^k\}_{k\in\mathbb{N}}$ durch Algorithmus 8.2 mit t_k aus Lemma 8.10 erzeugt. Dann existiert ein $\delta > 0$ mit*

$$\mu_{k+1} \le \left(1 - \frac{\delta}{n}\right)\mu_k \qquad \text{für alle } k \in \mathbb{N}.$$

Beweis. Aus Lemma 8.7 (ii) folgt für $t = t_k$

$$\mu_{k+1} = \mu_k(t_k) = \left(1 - 2\gamma\frac{1-\gamma}{1+\gamma}\frac{\sigma_k}{n}(1-\sigma_k)\right)\mu_k.$$

Da die konkave quadratische Funktion $\sigma \mapsto \sigma(1-\sigma)$ ihr Minimum auf dem Rand des kompakten Intervalls $[\sigma_{\min}, \sigma_{\max}]$ annimmt, gilt

$$\sigma_k(1-\sigma_k) \ge \sigma^* := \min\{\sigma_{\min}(1-\sigma_{\min}), \sigma_{\max}(1-\sigma_{\max})\} > 0$$

wegen $0 < \sigma_{\min}, \sigma_{\max} < 1$. Damit erhalten wir die gewünschte Abschätzung für $\delta := 2\gamma\sigma^*\frac{1-\gamma}{1+\gamma} > 0$. □

Daraus folgt sofort, dass das Innere-Punkte-Verfahren polynomielle Komplexität (in n) hat, was ja für das Simplex-Verfahren bekannterweise nicht gilt.

Folgerung 8.12. *Sei $\{x^k, \lambda^k, s^k\}_{k\in\mathbb{N}}$ durch Algorithmus 8.2 mit t_k aus Lemma 8.10 erzeugt. Sei $\varepsilon > 0$ und $\kappa > 0$ mit $\mu^0 \le \varepsilon^{-\kappa}$. Dann gilt*

$$\mu_k \le \varepsilon \qquad \text{für alle } k \ge \left(\frac{1+\kappa}{\delta}\right) n|\log\varepsilon|.$$

Beweis. Nach Satz 8.11 und Annahme gilt

$$\mu_k \le \left(1 - \frac{\delta}{n}\right)^k \mu_0 \le \left(1 - \frac{\delta}{n}\right)^k \varepsilon^{-\kappa}$$

und daher (Monotonie des Logarithmus)

$$\log\mu_k \le k\log\left(1 - \frac{\delta}{n}\right) + \kappa\log\tfrac{1}{\varepsilon} \le k\left(-\frac{\delta}{n}\right) + \kappa\log\tfrac{1}{\varepsilon} \tag{8.7}$$

wegen $\log(1+t) \le t$ für alle $t > -1$ (z. B. aus der Bernoullischen Ungleichung).

Also ist hinreichend für $\mu_k \le \varepsilon$, dass die rechte Seite von (8.7) nicht größer als $\log\varepsilon = -\log\frac{1}{\varepsilon}$ ist; Auflösen nach k ergibt dann für $\varepsilon < 1$

$$k \ge (1+\kappa)\frac{n}{\delta}\log\tfrac{1}{\varepsilon} = \left(\frac{1+\kappa}{\delta}\right) n|\log\varepsilon|$$

und damit die Behauptung. □

Der Nachteil bei diesem *zulässigen Innere-Punkte-Verfahren* ist die Notwendigkeit, einen strikt zulässigen Startwert $(x^0, \lambda^0, s^0) \in Z^+$ finden zu müssen, der insbesondere die Gleichungsnebenbedingungen $Ax^0 = b$ und $A^T\lambda^0 + s^0 = c$ erfüllt. Zumindest letztere Bedingungen können aufgegeben werden; anstelle von (8.6) muss dann in jeder Iteration der volle Newton-Schritt

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ D_{s^k} & 0 & D_{x^k} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \\ \Delta s \end{pmatrix} = \begin{pmatrix} c - A^T\lambda^k - s^k \\ b - Ax^k \\ -x^k \odot s^k + \sigma_k \mu_k \mathbb{1} \end{pmatrix}$$

gelöst werden; die Schrittweitensuche ist dann so durchzuführen, dass gilt

$$(x^{k+1}, \lambda^{k+1}, s^{k+1}) \in Z^{\gamma,\beta} := \left\{ (x, \lambda, s) \;\middle|\; x, s > 0,\ \|(b - Ax, c - s - A^T\lambda)\| \le r^0 \beta x^T s,\ x \odot s \ge \frac{\gamma}{n} x^T s \right\}$$

für $\gamma \in (0, 1)$, $\beta \ge 1$, und $r^0 := \frac{\|(b - Ax^0, c - s^0 - A^T\lambda^0)\|}{(x^0)^T(s^0)}$. Dies verkompliziert natürlich die Schrittweitensuche sowie den Konvergenzbeweis deutlich; siehe [Geiger & Kanzow 2002, Kapitel 4.2.2].

9 SQP-VERFAHREN

Eine der leistungsfähigsten und flexibelsten Klassen von Verfahren für Optimierungsprobleme mit Nebenbedingungen sind die sogenannten *sequential quadratic programming-Verfahren*, kurz *SQP-Verfahren*. Dabei handelt es sich um die Erweiterung von (Quasi-) Newton-Verfahren für unrestringierte Probleme auf Gleichungs- und Ungleichungsnebenbedingungen. Wir führen diese zuerst für den Fall reiner Gleichungsnebenbedingungen ein, und erweitern dies dann auf den Fall von gemischten Nebenbedingungen.

9.1 LAGRANGE–NEWTON-VERFAHREN FÜR GLEICHUNGSNEBENBEDINGUNGEN

Wir betrachten das Problem

$$\min_{x\in\mathbb{R}^n} f(x) \qquad \text{mit} \quad h(x) = 0$$

für zweimal stetig differenzierbare Funktionen $f : \mathbb{R}^n \to \mathbb{R}$ und $h : \mathbb{R}^n \to \mathbb{R}^p$. Gilt eine Regularitätsbedingung, so erfüllt ein lokaler Minimierer $\bar{x} \in \mathbb{R}^n$ zusammen mit einem Lagrange-Multiplikator $\bar{\lambda} \in \mathbb{R}^p$ die KKT-Bedingungen

$$\begin{cases} \nabla f(\bar{x}) + \bar{\lambda}^T \nabla h(\bar{x}) = 0, \\ h(\bar{x}) = 0. \end{cases}$$

Dies sind $n + p$ nichtlineare Gleichungen für die $n + p$ unbekannten Komponenten von $(\bar{x}, \bar{\lambda})$, die wir mit Hilfe der Lagrange-Funktion

$$L : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}, \qquad (x, \lambda) \mapsto f(x) + \lambda^T h(x),$$

schreiben können als

$$\nabla L(\bar{x}, \bar{\lambda}) = \begin{pmatrix} \nabla_x L(\bar{x}, \bar{\lambda}) \\ \nabla_\lambda L(\bar{x}, \bar{\lambda}) \end{pmatrix} = 0.$$

Auf diese Gleichung wenden wir nun das Newton-Verfahren an: Für gegebene (x^k, λ^k) berechnen wir $s^k \in \mathbb{R}^{n+p}$ als Lösung von

$$\nabla^2 L(x^k, \lambda^k) s = -\nabla L(x^k, \lambda^k),$$

und setzen (nach Zerlegung von $s^k \in \mathbb{R}^{n+p}$ in $s_x^k \in \mathbb{R}^n$ und $s_\lambda^k \in \mathbb{R}^p$)

$$x^{k+1} := x^k + s_x^k, \qquad \lambda^{k+1} := \lambda^k + s_\lambda^k.$$

Für die lokale superlineare Konvergenz müssen wir zunächst nachweisen, dass die Hesse-Matrix $\nabla^2 L$ in einem KKT-Punkt $(\bar{x}, \bar{\lambda})$ invertierbar ist. Da die Lagrange-Funktion linear ist in λ, hat diese stets die Form

$$\nabla^2 L(x,\lambda) = \begin{pmatrix} \nabla_{xx}^2 L(x,\lambda) & \nabla_{x\lambda}^2 L(x,\lambda) \\ \nabla_{\lambda x}^2 L(x,\lambda) & \nabla_{\lambda\lambda}^2 L(x,\lambda) \end{pmatrix} = \begin{pmatrix} \nabla_{xx}^2 L(x,\lambda) & \nabla h(x) \\ \nabla h(x)^T & 0 \end{pmatrix}.$$

Diese Struktur können wir ausnutzen, um hinreichende Bedingungen für die Invertierbarkeit anzugeben.

Lemma 9.1. *Seien $f : \mathbb{R}^n \to \mathbb{R}$ und $h : \mathbb{R}^n \to \mathbb{R}^p$ zweimal stetig differenzierbar. Gilt in $(x,\lambda) \in \mathbb{R}^n \times \mathbb{R}^p$:*

(i) $\nabla h(x) \in \mathbb{R}^{n\times p}$ *hat vollen Spaltenrang p;*

(ii) $d^T \nabla_{xx}^2 L(x,\lambda) d > 0 \quad$ *für alle* $d \in \mathbb{R}^n \setminus \{0\}$ *mit* $\nabla h(x)^T d = 0$,

so ist $\nabla^2 L(x,\lambda) \in \mathbb{R}^{(n+p)\times(n+p)}$ *invertierbar.*

Beweis. Da $\nabla^2 L(x,\lambda)$ quadratisch ist, genügt es zu zeigen, dass $\nabla^2 L(x,\lambda)$ injektiv ist. Seien daher (s_x, s_λ) mit

$$\begin{pmatrix} \nabla_{xx}^2 L(x,\lambda) & \nabla h(x) \\ \nabla h(x)^T & 0 \end{pmatrix} \begin{pmatrix} s_x \\ s_\lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Aus der zweiten Zeile folgt sofort $\nabla h(x)^T s_x = 0$. Multiplizieren der ersten Zeile von links mit s_x^T ergibt dann

$$0 = s_x^T \nabla_{xx}^2 L(x,\lambda) s_x + s_x^T \nabla h(x) s_\lambda = s_x^T \nabla_{xx}^2 L(x,\lambda) s_x.$$

Wegen Annahme (ii) und $\nabla h(x)^T s_x = 0$ ist dies aber nur möglich, falls $s_x = 0$ gilt. Damit reduziert sich die erste Zeile zu $\nabla h(x) s_\lambda = 0$, was aber nach Annahme (i) nur für $s_\lambda = 0$ gilt. Also ist $\nabla^2 L(x,\lambda)$ injektiv und deshalb invertierbar. □

Für einen KKT-Punkt $(\bar{x}, \bar{\lambda})$ entspricht Annahme (i) genau der LICQ, während Annahme (ii) die hinreichende Bedingung zweiter Ordnung ist. Analog zu Lemma 2.3 mit L an Stelle von f folgt daraus, dass $\nabla^2 L(x,\lambda)$ für (x,λ) in einer hinreichend kleinen Umgebung eines solchen KKT-Punktes ebenfalls invertierbar ist. Ebenso folgt daraus wie in Satz 2.6 die lokal superlineare Konvergenz des Lagrange–Newton-Verfahrens.

Algorithmus 9.1 : Lagrange–Newton-Verfahren

Input : $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^p$

for $k = 0, \ldots$ **do**
  if $\nabla_x L(x^k, \lambda^k) = 0$ *und* $h(x^k) = 0$ **then**
    return (x^k, λ^k)
  Berechne (s_x^k, s_λ^k) als Lösung von
  $$\begin{pmatrix} \nabla^2_{xx} L(x^k, \lambda^k) & \nabla h(x^k) \\ \nabla h(x^k)^T & 0 \end{pmatrix} \begin{pmatrix} s_x \\ s_\lambda \end{pmatrix} = - \begin{pmatrix} \nabla_x L(x^k, \lambda^k) \\ h(x^k) \end{pmatrix}$$
  Setze $x^{k+1} := x^k + s_x^k, \quad \lambda^{k+1} := \lambda^k + s_\lambda^k$

Satz 9.2. *Seien $f : \mathbb{R}^n \to \mathbb{R}$ und $h : \mathbb{R}^n \to \mathbb{R}^p$ zweimal stetig differenzierbar, und sei $(\bar{x}, \bar{\lambda})$ ein KKT-Punkt, in dem die LICQ und die hinreichende Optimalitätsbedingung zweiter Ordnung gilt. Dann existiert ein $\delta > 0$ so, dass für alle Startwerte (x^0, λ^0) mit*

$$\|x^0 - \bar{x}\| + \|\lambda^0 - \bar{\lambda}\| \leq \delta$$

das Lagrange–Newton-Verfahren Folgen $\{x^k\}_{k\in\mathbb{N}}$ und $\{\lambda^k\}_{k\in\mathbb{N}}$ erzeugt, die superlinear gegen $\bar{x}$ bzw. $\bar{\lambda}$ konvergieren. Ist $\nabla^2 L$ lokal Lipschitz-stetig, so ist die Konvergenz sogar quadratisch.

Wie im unrestringierten Fall hat dieses Verfahren den Nachteil der lokalen Konvergenz (die darüberhinaus auch einen Maximierer als Grenzwert haben kann), den man mit einer Globalisierung in den Griff bekommen möchte. Die Frage ist auch, wie man dieses Verfahren auf Ungleichungen anwenden kann. Dafür ist eine eine alternative Sichtweise auf das Verfahren nützlich.

9.2 SQP-VERFAHREN FÜR GEMISCHTE NEBENBEDINGUNGEN

Wir erinnern uns aus der Herleitung der Trust-Region-Verfahren für unrestringierte Verfahren, dass der Newton-Schritt äquivalent als Minimierer einer geeigneten quadratischen Funktion charakterisiert werden kann. Analog kann man für Gleichungsnebenbedingungen die Lösung (s_x^k, s_λ^k) des Lagrange–Newton-Schritts

$$\begin{pmatrix} \nabla^2_{xx} L(x^k, \lambda^k) & \nabla h(x^k) \\ \nabla h(x^k)^T & 0 \end{pmatrix} \begin{pmatrix} s_x \\ s_\lambda \end{pmatrix} = \begin{pmatrix} -\nabla_x L(x^k, \lambda^k) \\ -h(x^k) \end{pmatrix}. \tag{9.1}$$

über ein quadratisches Minimierungsproblem mit *linearen* Beschränkungen charakterisieren. Wir betrachten für $(x^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^p$ das Problem

$$\begin{cases} \min\limits_{s\in\mathbb{R}^n} \nabla f(x^k)^T s + \dfrac{1}{2} s^T \nabla^2_{xx} L(x^k, \lambda^k) s \\ \text{mit } \; h(x^k) + \nabla h(x^k)^T s = 0. \end{cases} \tag{9.2}$$

Da die Linearität der Nebenbedingungen eine Regularitätsbedingung darstellt (Satz 4.12), existiert für jeden Minimierer $\bar{s}$ ein Lagrange-Multiplikator $\bar{\lambda}_{\text{lin}} \in \mathbb{R}^p$, so dass die KKT-Bedingungen

$$\begin{cases} \nabla f(x^k) + \nabla^2_{xx} L(x^k, \lambda^k)\bar{s} + \nabla h(x^k)\bar{\lambda}_{\text{lin}} = 0, \\ h(x^k) + \nabla h(x^k)^T \bar{s} = 0, \end{cases}$$

erfüllt sind. Setzen wir $s^k_x := \bar{s}$ und $s^k_\lambda := \bar{\lambda}_{\text{lin}} - \lambda^k$ und bringen alle Terme, die kein s^k enthalten, auf die rechte Seite, so erhalten wir genau (9.1). Umgekehrt ist für einen Lagrange–Newton-Schritt (s^k_x, s^k_λ) das Paar $(s^k_x, \lambda^k + s^k_\lambda)$ ein KKT-Punkt von (9.2). Anstelle eines Updates berechnet man hier also direkt die neue Näherung für den Lagrange-Multiplikator.

Für zusätzliche Ungleichungsnebenbedingungen

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{mit} \quad g(x) \leq 0, \quad h(x) = 0, \tag{9.3}$$

betrachtet man analog für $(x^k, \mu^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ das Problem

$$\begin{cases} \min\limits_{s \in \mathbb{R}^n} \nabla f(x^k)^T s + \dfrac{1}{2} s^T \nabla^2_{xx} L(x^k, \mu^k, \lambda^k) s \\ \text{mit} \;\; g(x^k) + \nabla g(x^k)^T s \leq 0, \\ \qquad h(x^k) + \nabla h(x^k)^T s = 0, \end{cases}$$

berechnet einen KKT-Punkt $(\bar{s}, \bar{\mu}_{\text{lin}}, \bar{\lambda}_{\text{lin}})$, und setzt dann

$$x^{k+1} = x^k + \bar{s}, \qquad \mu^{k+1} = \bar{\mu}_{\text{lin}}, \qquad \lambda^{k+1} = \bar{\lambda}_{\text{lin}}.$$

Existieren mehrere KKT-Punkte, wählen wir denjenigen, der am nächsten an (x^k, μ^k, λ^k) liegt; dies ist in der Regel nicht praktisch realisierbar (es sei denn, der KKT-Punkt ist eindeutig), erleichtert aber den Konvergenzbeweis wesentlich.

Algorithmus 9.2 : SQP-Verfahren

Input : $x^0 \in \mathbb{R}^n$, $\mu^0 \in \mathbb{R}^m$, $\lambda^0 \in \mathbb{R}^p$

1 **for** $k = 0, \ldots$ **do**

2 **if** (x^k, μ^k, λ^k) *ist KKT-Punkt von* (9.3) **then**

3 **return** (x^k, μ^k, λ^k)

4 Berechne $(x^{k+1}, \mu^{k+1}, \lambda^{k+1})$ als KKT-Punkt von

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n} \nabla f(x^k)^T (x - x^k) + \dfrac{1}{2}(x - x^k)^T \nabla^2_{xx} L(x^k, \mu^k, \lambda^k)(x - x^k) \\ \text{mit} \;\; g(x^k) + \nabla g(x^k)^T (x - x^k) \leq 0, \\ \qquad h(x^k) + \nabla h(x^k)^T (x - x^k) = 0, \end{cases} \tag{9.4}$$

  mit minimaler Norm $\|(x^{k+1}, \mu^{k+1}, \lambda^{k+1}) - (x^k, \mu^k, \lambda^k)\|$

Der Beweis der lokalen Konvergenz ist deutlich komplizierter als im Fall reiner Gleichungsbeschränkungen. Die Idee ist folgende: Wir gehen ähnlich wie im Multiplikator-Strafverfahren vor und definieren

$$H(x,\mu,\lambda) := \begin{pmatrix} \nabla_x L(x,\mu,\lambda) \\ h(x) \\ \min\{-g(x),\mu\} \end{pmatrix}, \tag{9.5}$$

wobei die Komplementaritätsfunktion wieder komponentenweise zu verstehen ist. Dann ist $(\bar{x},\bar{\mu},\bar{\lambda})$ ein KKT-Punkt von (9.3) genau dann, wenn $H(\bar{x},\bar{\mu},\bar{\lambda}) = 0$ gilt. Wir wenden nun auf H ein Newton-Verfahren an. Dafür müssen wir zeigen, dass H in (einer Umgebung von) $(\bar{x},\bar{\mu},\bar{\lambda})$ differenzierbar ist. Dies geht, wenn *strikte* Komplementarität (d. h. $\bar{\mu}_i = 0$ genau dann, wenn $g_i(\bar{x}) < 0$) gilt; in diesem Fall ist das Minimum in der letzten Gleichung eindeutig, und wegen der Komplementarität sind die KKT-Bedingungen für (9.3) äquivalent zum *Gleichungs*system

$$\begin{cases} \nabla_x L(\bar{x},\bar{\mu},\bar{\lambda}) = 0, & \\ h_j(\bar{x}) = 0, & 1 \le j \le p, \\ -g_i(\bar{x}) = 0, & i \in \mathcal{A}_X(\bar{x}), \\ \bar{\mu}_i = 0, & i \notin \mathcal{A}_X(\bar{x}), \end{cases}$$

denn die inaktiven Ungleichungsbedingungen sind für die Optimierung nicht relevant; die (unter einer LICQ) eindeutigen Lagrange-Multiplikatoren $\mu_i > 0$, $i \in \mathcal{A}_X(\bar{x})$, sind bereits durch die erste Gleichung festgelegt. Die eigentlich von x abhängende aktive Menge kann schließlich wegen der Stetigkeit der g_i „eingefroren“ werden.

Satz 9.3. *Sei $(\bar{x},\bar{\mu},\bar{\lambda})$ ein KKT-Punkt von (9.3), für den gilt*

(i) die LICQ-Bedingung ($\nabla g_i(\bar{x})$, $i \in \mathcal{A}_X(\bar{x})$, $\nabla h_j(\bar{x})$, $j = 1,\dots,p$ sind linear unabhängig);

(ii) strikte Komplementarität ($g_i(\bar{x}) + \bar{\mu}_i \neq 0$ für alle $1 \le i \le m$);

(iii) die hinreichende Bedingung 2. Ordnung ($d^T \nabla^2_{xx} L(\bar{x},\bar{\mu},\bar{\lambda}) d > 0$ für alle $d \in K_X(\bar{x},\bar{\mu}) \setminus \{0\}$).

Dann konvergiert für alle (x^0,μ^0,λ^0) hinreichend nahe an $(\bar{x},\bar{\mu},\bar{\lambda})$ die durch Algorithmus 9.2 erzeugte Folge superlinear.

Beweis. Nach der Grundannahme sind sowohl $\nabla_x L$ und h stetig differenzierbar; wegen der strikten Komplementarität existiert ein $\varepsilon_1 > 0$ so dass für alle

$$(x,\mu,\lambda) \in U_{\varepsilon_1}(\bar{x},\bar{\mu},\bar{\lambda}) := \left\{ (x,\mu,\lambda) \;\middle|\; \|(x,\mu,\lambda) - (\bar{x},\bar{\mu},\bar{\lambda})\| \le \varepsilon_1 \right\}$$

auch gilt

$$-g_i(x) \begin{cases} < \mu_i & \text{für } i \in \mathcal{A}_X(\bar{x}), \\ > \mu_i & \text{für } i \notin \mathcal{A}_X(\bar{x}), \end{cases} \tag{9.6}$$

und damit $\min\{-g_i(x), \mu_i\} = -g_i(x)$ genau dann, wenn $\min\{-g_i(\bar{x}), \bar{\mu}_i\} = -g_i(\bar{x})$ ist. Also ist H definiert in (9.5) stetig differenzierbar in $U_{\varepsilon_1}(\bar{x}, \bar{\mu}, \bar{\lambda})$. Außerdem gilt

$$H'(x,\mu,\lambda) = \begin{pmatrix} \nabla^2_{xx} L(x,\mu,\lambda) & \nabla g(x) & \nabla h(x) \\ \nabla h(x)^T & 0 & 0 \\ -D_{\mathcal{A}_X(\bar{x})} \nabla g(x)^T & I - D_{\mathcal{A}_X(\bar{x})} & 0 \end{pmatrix},$$

wobei $D_{\mathcal{A}_X(\bar{x})}$ eine Diagonalmatrix ist mit Diagonaleinträgen $[D_{\mathcal{A}_X(\bar{x})}]_{ii} = 1$ für $i \in \mathcal{A}_X(\bar{x})$ und 0 sonst. Analog zu Lemma 9.1 zeigt man mit Hilfe der Annahmen, dass $H'(\bar{x}, \bar{\mu}, \bar{\lambda})$ injektiv und daher invertierbar ist (denn wegen der strikten Komplementarität gilt $\mathcal{A}_+(\bar{x}, \bar{\mu}) = \mathcal{A}_X(\bar{x})$, und damit implizieren die zweite und dritte Gleichung $s_x \in K_X(\bar{x}, \bar{\mu})$); aus der Stetigkeit der Ableitung folgt dann auch, dass ein $\varepsilon_2 > 0$ existiert, so dass $H'(x, \mu, \lambda)$ ebenfalls invertierbar ist für alle $(x, \mu, \lambda) \in U_{\varepsilon_2}(\bar{x}, \bar{\mu}, \bar{\lambda})$. Daraus folgt mit den üblichen Argumenten die lokale superlineare Konvergenz des Newton-Verfahrens für H gegen $(\bar{x}, \bar{\mu}, \bar{\lambda})$.

Wir zeigen nun, dass die durch dieses Newton-Verfahren erzeugte Folge $\{(x^k, \mu^k, \lambda^k)\}_{k\in\mathbb{N}}$ mit der Folge aus Algorithmus 9.2 übereinstimmt, falls dessen Startvektor (x^0, μ^0, λ^0) hinreichend nahe an $(\bar{x}, \bar{\mu}, \bar{\lambda})$ liegt. Dafür verwenden wir, dass wegen (9.6) ein $\varepsilon_3 < \varepsilon_1$ existiert so, dass (x^k, μ^k) ebenfalls (9.6) erfüllt sowie

$$-g_i(x^k) - \nabla g_i(x^k)^T (x - x^k) \begin{cases} < \mu_i & \text{für } i \in \mathcal{A}_X(\bar{x}), \\ > \mu_i & \text{für } i \notin \mathcal{A}_X(\bar{x}), \end{cases} \tag{9.7}$$

für alle $(x^k, \mu^k, \lambda^k), (x, \mu, \lambda) \in U_{3\varepsilon_3}(\bar{x}, \bar{\mu}, \bar{\lambda})$ (beachte den größeren Radius $3\varepsilon_3$).

Sei nun $\varepsilon := \min\{\varepsilon_2, \varepsilon_3\}$ und $(x^k, \mu^k, \lambda^k) \in U_\varepsilon(\bar{x}, \bar{\mu}, \bar{\lambda})$. Ein Newton-Schritt für H ist dann die Lösung $(x^{k+1}, \mu^{k+1}, \lambda^{k+1}) \in U_\varepsilon(\bar{x}, \bar{\mu}, \bar{\lambda})$ von

$$\begin{aligned} \nabla^2_{xx} L(x^k, \mu^k, \lambda^k)(x^{k+1} - x^k) + \nabla g(x^k)(\mu^{k+1} - \mu^k) & \\ + \nabla h(x^k)(\lambda^{k+1} - \lambda^k) &= -\nabla_x L(x^k, \mu^k, \lambda^k), \\ \nabla h(x^k)^T (x^{k+1} - x^k) &= -h(x^k), \\ \nabla g_i(x^k)^T (x^{k+1} - x^k) &= -g_i(x^k), && i \in \mathcal{A}_X(\bar{x}), \\ (\mu_i^{k+1} - \mu_i^k) &= -\mu_i^k, && i \notin \mathcal{A}_X(\bar{x}), \end{aligned}$$

wobei wir (9.6) für die rechte Seite der letzten beiden Gleichungen verwendet haben. Nach Definition der Lagrange-Funktion vereinfacht sich nun die erste Zeile zu

$$\nabla f(x^k) + \nabla^2_{xx} L(x^k, \mu^k, \lambda^k)(x^{k+1} - x^k) + \nabla g(x^k)\mu^{k+1} + \nabla h(x^k)\lambda^{k+1} = 0,$$

und die zweite Zeile ist natürlich genau

$$h(x^k) + \nabla h(x^k)^T (x^{k+1} - x^k) = 0.$$

Die dritte Zeile garantiert analog

$$g_i(x^k) + \nabla g_i(x^k)^T (x^{k+1} - x^k) = 0 \qquad \text{für } i \in \mathcal{A}_X(\bar{x}).$$

Wegen $(x^{k+1}, \mu^{k+1}, \lambda^{k+1}) \in U_\varepsilon(\bar{x}, \bar{\mu}, \bar{\lambda})$ gilt dann (9.7) und damit insbesondere

$$\begin{aligned} \mu_i^{k+1} > -g_i(x^k) - \nabla g_i(x^k)^T(x^{k+1} - x^k) = 0 \qquad & \text{für } i \in \mathcal{A}_X(\bar{x}), \\ -g_i(x^k) - \nabla g_i(x^k)^T(x^{k+1} - x^k) > \mu_i^{k+1} = 0 \qquad & \text{für } i \notin \mathcal{A}_X(\bar{x}), \end{aligned}$$

wobei letztere Gleichung aus der vierten Zeile folgt. Dies sind genau die KKT-Bedingungen für (9.4). Insbesondere existiert damit ein KKT-Punkt von (9.4), so dass Algorithmus 9.2 wohldefiniert ist.

Es bleibt zu zeigen, dass $(x^{k+1}, \mu^{k+1}, \lambda^{k+1})$ der einzige KKT-Punkt in $U_{3\varepsilon_3}(\bar{x}, \bar{\mu}, \bar{\lambda})$ ist und damit insbesondere den Abstand zu $(x^k, \mu^k, \lambda^k) \in U_\varepsilon(\bar{x}, \bar{\mu}, \bar{\lambda})$ minimiert. Sei dazu $(\tilde{x}, \tilde{\mu}, \tilde{\lambda}) \in U_{3\varepsilon_3}(\bar{x}, \bar{\mu}, \bar{\lambda})$ ein KKT-Punkt von (9.4), d. h. erfüllt

$$\begin{aligned} \nabla f(x^k) + \nabla^2_{xx} L(x^k, \mu^k, \lambda^k)(\tilde{x} - x^k) + \nabla g(x^k)\tilde{\mu} + \nabla h(x^k)\tilde{\lambda} &= 0, \\ h(x^k) + \nabla h(x^k)^T(\tilde{x} - x^k) &= 0, \\ \min\{-g(x^k) - \nabla g(x^k)^T(\tilde{x} - x^k), \tilde{\mu}\} &= 0, \end{aligned}$$

wobei wir wieder die Komplementaritätsfunktion verwendet haben. Nach Annahme erfüllen $\tilde{x}$ und $\tilde{\mu}$ ebenfalls (9.7), so dass die letzte Zeile sich auch zerlegen lässt in

$$\begin{aligned} g_i(x^k) + \nabla g_i(x^k)^T(\tilde{x} - x^k) = 0 \qquad & \text{für } i \in \mathcal{A}_X(\bar{x}), \\ \tilde{\mu}_i = 0 \qquad & \text{für } i \notin \mathcal{A}_X(\bar{x}). \end{aligned}$$

Also ist $(\tilde{x}, \tilde{\mu}, \tilde{\lambda})$ eine Lösung der Newton-Gleichung

$$H'(x^k, \mu^k, \lambda^k) \begin{pmatrix} \tilde{x} - x^k \\ \tilde{\mu} - \mu^k \\ \tilde{\lambda} - \lambda^k \end{pmatrix} = -H(x^k, \mu^k, \lambda^k).$$

Da $H'(x^k, \mu^k, \lambda^k)$ invertierbar ist, ist die Lösung eindeutig und damit gilt in der Tat $(\tilde{x}, \tilde{\mu}, \tilde{\lambda}) = (x^{k+1}, \mu^{k+1}, \lambda^{k+1})$. Wegen

$$(x^k, \mu^k, \lambda^k), (x^{k+1}, \mu^{k+1}, \lambda^{k+1}) \in U_\varepsilon(\bar{x}, \bar{\mu}, \bar{\lambda}) \subset U_{\varepsilon_3}(\bar{x}, \bar{\mu}, \bar{\lambda})$$

ist $\|(x^{k+1}, \mu^{k+1}, \lambda^{k+1}) - (x^k, \mu^k, \lambda^k)\| \leq 2\varepsilon_3$, während für jeden weiteren KKT-Punkt gilt $(\tilde{x}, \tilde{\mu}, \tilde{\lambda}) \notin U_{3\varepsilon_3}(\bar{x}, \bar{\mu}, \bar{\lambda})$ und daher $\|(\tilde{x}, \tilde{\mu}, \tilde{\lambda}) - (x^k, \mu^k, \lambda^k)\| > 2\varepsilon_3$. Also ist $(x^{k+1}, \mu^{k+1}, \lambda^{k+1})$ in der Tat der nächste KKT-Punkt zu (x^k, μ^k, λ^k). □

ZUR GLOBALISIERUNG DES SQP-VERFAHRENS

Analog zum unrestringierten Newton-Verfahren kann man das SQP-Verfahren durch eine Schrittweitensuche globalisieren. Dabei ist es nicht ausreichend, nur die Zielfunktion f

zu betrachten; es müssen auch die Nebenbedingungen berücksichtigt werden. Dies kann durch Verwendung der exakten Straffunktion

$$f(x) + \alpha\pi_1(x)$$

für α hinreichend groß geschehen; dabei wird in der Praxis $\alpha = \alpha_k$ während des Verfahrens angepasst (etwa mit Hilfe von (7.5) für μ^k, λ^k anstelle von $\bar{\mu}, \bar{\lambda}$). Obwohl π_1 nicht differenzierbar ist, kann man eine Armijo-Bedingung mit Hilfe von Richtungsableitungen formulieren; siehe [Geiger & Kanzow 2002, Kapitel 5.5.4].

Allerdings reicht dann bei nichtkonvexen Problemen die hinreichende Bedingung 2. Ordnung nicht aus, um die positive Definitheit der Hesse-Matrizen in allen Iterierten und damit die Konvergenz zu garantieren; stattdessen wird man in (9.4) anstelle der exakten Hesse-Matrix $\nabla^2_{xx} L(x^k, \mu^k, \lambda^k)$ eine Quasi-Newton-Näherung H_k verwendet werden. Bei der Wahl des Updates sind wegen der gewählten Schrittweitensuche allerdings einige Schwierigkeiten zu beachten, um die positive Definitheit zu gewährleisten; siehe [Geiger & Kanzow 2002, Kapitel 5.5.5].

Um trotzdem lokal superlineare Konvergenz zu erhalten, muss irgendwann stets die Schrittweite $\sigma_k = 1$ akzeptiert werden. Für Probleme mit Gleichungsnebenbedingung kann es aber sein, dass das nie der Fall ist. Dies ist als *Maratos-Effekt* bekannt, und kann durch einen zusätzlichen Korrektur-Schritt behandelt werden, der zusätzlich $\nabla^2 h(x^k)(x^{k+1} - x^k)$ und $\nabla^2 g(x^k)(x^{k+1} - x^k)$ verwendet; siehe [Geiger & Kanzow 2002, Kapitel 5.5.6].

Schließlich kann es für nichtkonvexe Probleme vorkommen, dass die zulässige Menge für die quadratischen Teilprobleme leer ist. In diesem Fall muss man das SQP-Verfahren modifizieren, indem die Nebenbedingungen relaxiert werden: man fordert nur, dass gilt

$$\begin{aligned} g(x^k) + \nabla g(x^k)^T (x - x^k) &\leq \varepsilon, \\ h(x^k) + \nabla h(x^k)^T (x - x^k) &= \eta, \end{aligned}$$

wobei ε, η als neue Variablen mit einer exakten Straffunktion in die Zielfunktion des Teilproblems aufgenommen werden. Für positiv definite Matrizen H_k kann man dann (mit einigem Aufwand) globale Konvergenz zeigen; siehe [Geiger & Kanzow 2002, Kapitel 5.5.7-8].

Eine Alternative stellen Trust-Region-SQP-Verfahren dar, die in letzter Zeit vermehrt untersucht werden.

9.3 AKTIVE-MENGEN-STRATEGIE FÜR QUADRATISCHE PROBLEME

Für die Lösung der SQP-Teilprobleme (9.4) kann man analog zum Konvergenzbeweis die KKT-Bedingung als Gleichungssystem schreiben. Da die linearisierte aktive Menge

$$\mathcal{A}^k_{\text{lin}}(s^k) := \left\{ i \;\middle|\; g_i(x^k) + \nabla g_i(x^k)^T s^k = 0 \right\}$$

nicht bekannt ist, geht man iterativ vor: Man wählt eine Startschätzung $\mathcal{A}^0 \subset \{1,\dots,m\}$, löst das entsprechende Gleichungssystem mit $\mathcal{A}^0$ anstelle von $\mathcal{A}^k_{\text{lin}}(\bar{s})$, wählt (falls man nicht bereits einen KKT-Punkt zu (9.4) gefunden hat) eine geeignete neue Näherung $\mathcal{A}^1$, und wiederholt die Prozedur.

Wir betrachten dafür ein allgemeines quadratisches Optimierungsproblem

$$\text{(QP)} \qquad \begin{cases} \min\limits_{s\in\mathbb{R}^n} c^T s + \dfrac{1}{2} s^T H s \\ \text{mit} \quad a_i^T s + \alpha_i \le 0, \quad i = 1,\dots,m, \\ \qquad\; b_j^T s + \beta_j = 0, \quad j = 1,\dots,p, \end{cases}$$

für $H \in \mathbb{R}^{n\times n}$, $a_i, b_j, c \in \mathbb{R}^n$, $\alpha_i, \beta_j \in \mathbb{R}$. Wir nehmen an, dass ein zulässiges $s^0 \in \mathbb{R}^n$ existiert und H symmetrisch und positiv definit ist, so dass (QP) eine (sogar eindeutige) Lösung $\bar{s} \in \mathbb{R}^n$ besitzt. Wegen der Linearität der Nebenbedingungen ist dieser Punkt regulär, erfüllt also zusammen mit Lagrange-Multiplikatoren $\bar{\mu}, \bar{\lambda}$ die KKT-Bedingungen

$$\begin{aligned} c + H\bar{s} + \sum_{i=1}^m \bar{\mu}_i a_i + \sum_{j=1}^p \bar{\lambda}_j b_j &= 0, \\ a_i^T \bar{s} + \alpha_i &\le 0, \quad i = 1,\dots,m, \\ b_j^T \bar{s} + \beta_j &= 0, \quad j = 1,\dots,p, \\ \bar{\mu}_i \ge 0, \quad \bar{\mu}_i (a_i^T \bar{s} + \alpha_i) &= 0, \quad i = 1,\dots,m. \end{aligned}$$

Sei nun ein zulässiges(!) $s^k \in \mathbb{R}^n$ gegeben und definiere die *aktive* bzw. *inaktive Menge*

$$\mathcal{A}^k := \left\{ i \;\middle|\; a_i^T s^k + \alpha_i = 0 \right\}, \qquad \mathcal{I}^k := \{1,\dots,m\} \setminus \mathcal{A}^k.$$

Wir suchen dann eine Lösung des reduzierten Problems

$$\text{(QP}_k\text{)} \qquad \begin{cases} \min\limits_{s\in\mathbb{R}^n} c^T s + \dfrac{1}{2} s^T H s \\ \text{mit} \quad a_i^T s + \alpha_i = 0, \quad i \in \mathcal{A}^k, \\ \qquad\; b_j^T s + \beta_j = 0, \quad j = 1,\dots,p. \end{cases}$$

Auch dieses Problem hat eine Lösung s^{k+1}, da s^k nach Definition von $\mathcal{A}^k$ zulässig ist. Wieder sind alle (diesmal nur Gleichungs-)Nebenbedingungen affin-linear, so dass ein Lagrange-Multiplikator $(\mu^{k+1}, \lambda^{k+1}) \in \mathbb{R}^{|\mathcal{A}^k|+p}$ existiert, der die reduzierten KKT-Bedingungen erfüllt:

$$\begin{aligned} c + Hs^{k+1} + \sum_{i\in\mathcal{A}^k} \mu_i^{k+1} a_i + \sum_{j=1}^p \lambda_j^{k+1} b_j &= 0, \\ a_i^T s^{k+1} + \alpha_i &= 0, \quad i \in \mathcal{A}^k, \\ b_j^T s^{k+1} + \beta_j &= 0, \quad j = 1,\dots,p. \end{aligned}$$

Schreiben wir kurz A_k für die Matrix, deren Zeilen genau die a_i^T für $i \in \mathcal{A}^k$ sind, α^k für den Vektor mit Einträgen α_i, $i \in \mathcal{A}^k$, und B für die Matrix mit Zeilen b_j^T sowie β für den Vektor mit Einträgen β_j, $j = 1, \dots, p$, so haben die KKT-Bedingungen die Form

$$\text{(KKT}_k\text{)} \qquad \begin{pmatrix} H & A_k^T & B^T \\ A_k & 0 & 0 \\ B & 0 & 0 \end{pmatrix} \begin{pmatrix} s \\ \mu \\ \lambda \end{pmatrix} = \begin{pmatrix} -c \\ -\alpha_k \\ -\beta \end{pmatrix}.$$

Angenommen, wir haben nun eine Lösung $(s^{k+1}, \mu^{k+1}, \lambda^{k+1})$ von (KKT$_k$). Gilt dann $s^{k+1} = s^k$ und $\mu^{k+1} \geq 0$, so sind wir fertig: Es ist dann offensichtlich $\bar{s} := s^{k+1} = s^k$ zulässig für (QP) (nach Annahme), und $\bar{\mu}_i := \mu_i^{k+1}$ für $i \in \mathcal{A}^k = \mathcal{A}^{k+1}$ und $\bar{\mu}_i := 0$ sonst, erfüllt zusammen mit $\bar{\lambda} := \lambda^{k+1}$ die KKT-Bedingungen für (QP). Ansonsten gibt es ein $i \in \mathcal{A}^k$ mit $\mu_i^{k+1} < 0$, so dass die entsprechende Nebenbedingung in s^k doch nicht "wirklich" eine Ungleichung ist, die zufällig mit Gleichheit gilt. Wir entfernen also diese Bedingung aus der aktiven Menge und berechnen eine neue Lösung. (Gibt es mehrere Indizes mit $\mu_i^{k+1} < 0$, wählt man üblicherweise denjenigen mit der größten Verletzung der Bedingung $\mu \geq 0$.)

Gilt $s^{k+1} \neq s^k$, so ist der neue Punkt entweder weiterhin zulässig für (QP) oder nicht. Im ersten Fall merken wir ihn uns als neue Näherung, ändern aber die aktive Menge nicht. (Dieser Schritt macht keinen Fortschritt und dient nur der Buchhaltung.) Im zweiten Fall ist eine der Ungleichungen in der *inaktiven* Menge $\mathcal{I}^k$ verletzt; anstatt den vollen Schritt zu akzeptieren, machen wir nun (analog zum Simplex-Verfahren) einen Teilschritt $\hat{s}^{k+1} = s^k + t_k(s^{k+1} - s^k) =: s^k + t_k d^k$ mit $t_k \in (0, 1)$ so gewählt, dass alle Ungleichungen erfüllt sind, d. h. dass gilt

$$a_i^T \left(s^k + t_k d^k \right) + \alpha_i \leq 0, \quad \text{für alle } i = 1, \dots, m.$$

Dabei können wir verwenden, dass wegen der Zulässigkeit von s^k gilt $a_i^T s^k + \alpha_i \leq 0$. Ist nun $a_i^T d^k \leq 0$, so bleibt die Ungleichung für alle $t_k > 0$ erfüllt; wir müssen also nur solche $i \in \mathcal{I}^k$ mit $a_i^T d_k > 0$ betrachten. Auflösen nach t_k führt auf die maximale Wahl

$$t_k := \min \left\{ -\frac{a_i^T s^k + \alpha_i}{a_i^T d^k} \;\middle|\; i \in \mathcal{I}^k,\ a_i^T d^k > 0 \right\}.$$

Da nur endlich viele Indizes in der inaktiven Menge sein können, existiert dieses Minimum immer. Dann gilt nach Konstruktion Gleichheit für den Index i, in dem das Minimum angenommen wird, und daher werden wir i in die aktive Menge aufnehmen und eine neue Lösung berechnen.

Dieses Vorgehen führt auf die *Aktive-Mengen-Strategie*.

Algorithmus 9.3 : Aktive-Mengen-Strategie

Input : $s^0 \in \mathbb{R}^n$ zulässig für (QP)

Setze $\mathcal{A}^0 := \left\{ i \;\middle|\; a_i^T s^0 + \alpha_i = 0 \right\}$
for $k = 0, \ldots$ **do**
  Berechne $(\tilde{s}^{k+1}, \mu^{k+1}, \lambda^{k+1})$ als Lösung von (KKT_k)
  if $\tilde{s}^{k+1} = s^k$ **then**
    if $\mu^{k+1} \geq 0$ **then**
      return $(\tilde{s}^{k+1}, \mu^{k+1}, \lambda^{k+1})$
    else
      Wähle $r = \arg\min \left\{ \mu_i^{k+1} \;\middle|\; i \in \mathcal{A}^k,\ \mu_i^{k+1} < 0 \right\}$
      Setze $s^{k+1} = s^k$ und $\mathcal{A}^{k+1} = \mathcal{A}^k \setminus \{r\}$
  else
    if $\tilde{s}^{k+1}$ *zulässig für* (QP) **then**
      Setze $s^{k+1} = \tilde{s}^{k+1}$ und $\mathcal{A}^{k+1} = \mathcal{A}^k$
    else
      Setze $d^k := \tilde{s}^{k+1} - s^k$
      Wähle $r := \arg\min \left\{ -\frac{a_i^T s^k + \alpha_i}{a_i^T d^k} \;\middle|\; i \in \mathcal{I}^k,\ a_i^T d^k > 0 \right\}$
      Wähle $t_k := -\frac{a_r^T s^k + \alpha_r}{a_r^T d^k}$
      Setze $s^{k+1} = s^k + t_k d^k$ und $\mathcal{A}^{k+1} = \mathcal{A}^k \cup \{r\}$

Für den Algorithmus ist wichtig, dass der KKT-Punkt von (QP_k) eindeutig und daher als Lösung des linearen Gleichungssystems (KKT_k) berechnet werden kann. Dies kann man unter der folgenden Annahme garantieren.

Lemma 9.4. *Sei H symmetrisch und positiv definit. Sind die a_i, $i \in \mathcal{A}^0$, und b_j, $j = 1, \ldots, p$, linear unabhängig, dann hat für alle $k \in \mathbb{N} \cup \{0\}$ das Gleichungssystem (KKT_k) eine eindeutige Lösung.*

Beweis. Wir nehmen zunächst an, dass für $k \in \mathbb{N} \cup \{0\}$ die Vektoren a_i, $i \in \mathcal{A}^k$, und b_j, $j = 1, \ldots, p$, linear unabhängig sind. Es genügt wieder einmal zu zeigen, dass die Matrix in (KKT_k) injektiv ist. Sei dafür (s, μ, λ) Lösung des homogenen Systems

$$\begin{pmatrix} H & A_k^T & B^T \\ A_k & 0 & 0 \\ B & 0 & 0 \end{pmatrix} \begin{pmatrix} s \\ \mu \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Multiplizieren der ersten Gleichung mit s^T ergibt dann

$$0 = s^T H s + s^T A_k^T \mu + s^T B^T \lambda = s^T H s + (A_k s)^T \mu + (Bs)^T \lambda = s^T H s$$

unter Verwendung der zweiten und dritten Gleichung. Da H positiv definit ist, folgt $s = 0$. Damit reduziert sich die erste Gleichung auf

$$0 = A_k^T \mu + B^T \lambda = \sum_{i \in \mathcal{A}^k} \mu_i a_i + \sum_{j=1}^{p} \lambda_j b_j$$

und damit $\mu_i, \lambda_j = 0$ wegen der linearen Unabhängigkeit.

Bleibt zu zeigen, dass dann auch a_i, $i \in \mathcal{A}^{k+1}$ und b_j, $j = 1, \dots, p$, linear unabhängig sind. Dafür ist nur der Fall $\mathcal{A}^{k+1} \supsetneq \mathcal{A}^k$ interessant, d. h. $\mathcal{A}^{k+1} = \mathcal{A}^k \cup \{r\}$ mit $a_r^T d^k > 0$. Angenommen, a_r wäre linear abhängig von a_i, $i \in \mathcal{A}^k$, und b_j, $j = 1, \dots, p$. Dann existieren γ_i, δ_j so, dass gilt

$$0 < a_r^T d^k = \sum_{i \in \mathcal{A}^k} \gamma_i a_i^T (\tilde{s}^{k+1} - s^k) + \sum_{j=1}^{p} \delta_j b_j^T (\tilde{s}^{k+1} - s^k) = 0$$

da sowohl s^k (nach Definition von $\mathcal{A}^k$ bzw. der Zulässigkeit) als auch $\tilde{s}^{k+1}$ (wegen der zweiten bzw. dritten Gleichung von (KKT_k)) die Nebenbedingungen in (QP_k) erfüllen. Aber dies ist ein Widerspruch. □

Wir können nun eine Konvergenzaussage für Algorithmus 9.3 zeigen. Ähnlich wie im Simplex-Verfahren kann auch hier der Fall eines *Zykelns* nicht ausgeschlossen werden. Außer in diesem (praktisch seltenen) Fall terminiert das Verfahren aber nach endlich vielen Schritten. Beachte dabei, dass im letzten Fall ($\tilde{s}^{k+1} \neq s^k$ unzulässig für (QP)) die neue Iterierte s^{k+1} *kein* KKT-Punkt von (QP_k) ist.

Satz 9.5. *Sei H symmetrisch und positiv definit, $s^0 \in \mathbb{R}^n$ zulässig für (QP), und a_i, $i \in \mathcal{A}^0$, b_j, $j = 0, \dots, p$, linear unabhängig. Dann gilt für die Iterierten aus Algorithmus 9.3:*

(i) s^k ist zulässig sowohl für (QP) als auch für (QP_k) für alle $k \in \mathbb{N}$;

(ii) ist $s^{k+1} \neq s^k$ für ein $k \in \mathbb{N}$, so ist für $q(s) := c^T s + \frac{1}{2} s^T H s$

$$q(s^{k+1}) < q(s^k);$$

(iii) Bricht die Iteration nicht nach endlich vielen Schritten in einem KKT-Punkt von (QP) ab, dann existiert ein $K \in \mathbb{N}$ mit $s^k = s^K$ für alle $k \geq K$.

Beweis. Zu (i): Wir verwenden Induktion nach k. Nach Annahme ist s^0 zulässig für (QP). Weiter ist $\mathcal{A}^0$ definiert als die Menge der Ungleichungen, die in s^0 mit Gleichheit erfüllt sind; also ist s^0 auch zulässig für (QP_0). Sei nun $k \in \mathbb{N} \cup \{0\}$ und s^k zulässig für (QP) und (QP_k). Nach Konstruktion ist $\tilde{s}^{k+1}$ zulässig für (QP_k); wegen der Linearität der Nebenbedingungen ist dann auch die Konvexkombination $s^{k+1} = s^k + t_k(\tilde{s}^{k+1} - s^k) = t_k \tilde{s}^{k+1} + (1 - t_k) s^k$ für $t_k \in [0, 1]$ zulässig für (QP_k). Ist nun $\mathcal{A}^{k+1} \subseteq \mathcal{A}^k$, so ist s^{k+1} natürlich auch zulässig für

(QP_{k+1}). Andernfalls ist $\mathcal{A}^{k+1} = \mathcal{A}^k \cup \{r\}$ mit $a_r^T s^{k+1} + \alpha_r = 0$ und damit s^{k+1} ebenfalls zulässig für (QP_{k+1}). Schließlich ist entweder $s^{k+1} = s^k$ (nach Induktionsannahme), $s^{k+1} = \tilde{s}^{k+1}$ (nach Fallunterscheidung), oder $s^{k+1} = s^k + t_k(\tilde{s}^{k+1} - s^k)$ (nach Konstruktion) zulässig für (QP).

Zu (ii): Nach Voraussetzung an H ist (QP_k) strikt konvex; daher ist der KKT-Punkt $\tilde{s}^{k+1}$ der einzige globale Minimierer. Nach (i) ist nun s^k zulässig für (QP_k), und damit muss im Fall $\tilde{s}^{k+1} \neq s^k$ gelten $q(\tilde{s}^{k+1}) < q(s^k)$. Dann ist entweder $s^{k+1} = \tilde{s}^{k+1}$ und damit die Aussage erfüllt oder $s^{k+1} = t_k \tilde{s}^{k+1} + (1 - t_k)s^k$ für $t_k \in (0, 1)$ wie in Algorithmus 16 gewählt. Aus der strikten Konvexität von q folgt in diesem Fall

$$q(s^{k+1}) < t_k q(\tilde{s}^{k+1}) + (1 - t_k)q(s^k) < t_k q(s^k) + (1 - t_k)q(s^k) = q(s^k).$$

Zu (iii): Wir zeigen zuerst, dass für alle $k \in \mathbb{N}$ ein $\ell \geq k$ existiert, für das s^ℓ der eindeutige globale Minimierer von (QP_ℓ) ist, d. h. $s^\ell = \tilde{s}^\ell$ wird ohne Schrittweitensuche akzeptiert. Dazu machen wir eine Fallunterscheidung für $k \in \mathbb{N}$ beliebig:

- $s^{k+1} = \tilde{s}^{k+1} = s^k$: Dann ist $s^k = \tilde{s}^{k+1}$ die eindeutige Lösung der KKT-Bedingungen für (QP_k) und damit wegen der Konvexität auch der globale Minimierer für $\ell = k$.
- $s^{k+1} = \tilde{s}^{k+1} \neq s^k$: Wegen der Wahl von $\mathcal{A}^{k+1} = \mathcal{A}^k$ ist damit s^{k+1} auch die eindeutige Lösung der KKT-Bedingungen für (QP_{k+1}) und damit der globale Minimierer für $\ell = k + 1$.
- $s^{k+1} \neq \tilde{s}^{k+1} \neq s^k$: In diesem Fall ist $\mathcal{A}^{k+1} = \mathcal{A}^k \cup \{r\}$ eine strikte Obermenge. Da aber $\mathcal{A}^k \subset \{1, \ldots, m\}$ beschränkt bleibt, kann dieser Fall nur endlich oft hintereinander auftreten; es muss also spätestens für $\ell = k + m$ einer der anderen beiden Fälle eintreten.

Wenn Algorithmus 9.3 nicht nach endlich vielen Schritten abbricht (was unter den Voraussetzungen nach Lemma 9.4 nur in einem KKT-Punkt von (QP) der Fall ist), muss also eine unendliche Folge $\{k_n\}_{n \in \mathbb{N}}$ existieren so, dass s^{k_n} die eindeutige Lösung von (QP_{k_n}) ist für alle $n \in \mathbb{N}$. Da $\{1, \ldots, m\}$ und damit auch die Potenzmenge endlich ist, muss eine ebenfalls unendliche Teilfolge existieren (die wir von der Notation nicht unterscheiden) mit $\mathcal{A}^{k_n} = \mathcal{A}^{k_1}$ für alle $n \in \mathbb{N}$. Das bedeutet aber, dass die entsprechenden s^{k_n} (eindeutige) Lösung des gleichen Optimierungsproblems sind, woraus $s^{k_n} = s^{k_1}$ folgt. Angenommen, es existiert nun ein $k \in \mathbb{N}$ mit $k_n \leq k < k + 1 \leq k_{n+1}$ so, dass $s^{k+1} \neq s^k$ gilt. Dann folgt aus (ii)

$$q(s^{k_n}) = q(s^{k_{n+1}}) \leq q(s^{k+1}) < q(s^k) \leq q(s^{k_n}),$$

und damit ein Widerspruch. Also muss $s^k = s^{k_1}$ für alle $k \geq k_1 =: K$ gelten. □

10 SEMIGLATTE NEWTON-VERFAHREN

Die aktive-Mengen-Strategie kann man als Spezialfall eines "nichtglatten Newton-Verfahrens" interpretieren. In diesem Kapitel behandeln wir eine Verallgemeinerung, für die man ohne Annahme der zweimal stetigen Differenzierbarkeit (lokal) superlineare Konvergenz zeigen kann.

Als Motivation betrachten wir zuerst die allgemeine Form eines Newton-artigen Verfahrens. Sei $F : \mathbb{R}^n \to \mathbb{R}^n$ und gesucht sei ein $\bar{x} \in \mathbb{R}^n$ mit $F(\bar{x}) = 0$. Ein Newton-artiges Verfahren hat dann die folgende Form:

1. wähle $M_k := M(x^k) \in \mathbb{R}^{n\times n}$ invertierbar;
2. löse $M_k s^k = -F(x^k)$;
3. setze $x^{k+1} = x^k + s^k$.

Wir können uns nun fragen, unter welchen Bedingungen diese Iteration konvergiert, und insbesondere, wann die Konvergenz superlinear ist, d. h. wann gilt

$$\lim_{k\to\infty} \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} = 0. \tag{10.1}$$

Dafür setzen wir $e^k := x^k - \bar{x}$ und verwenden den Newton-Schritt sowie $F(\bar{x}) = 0$ um zu schreiben

$$\begin{aligned}
\|x^{k+1} - \bar{x}\| &= \|x^k - M(x^k)^{-1}F(x^k) - \bar{x}\| \\
&= \|M(x^k)^{-1}\left[F(x^k) - F(\bar{x}) - M(x^k)(x^k - \bar{x})\right]\| \\
&= \|M(\bar{x} + e^k)^{-1}\left[F(\bar{x} + e^k) - F(\bar{x}) - M(\bar{x} + e^k)e^k\right]\| \\
&\le \|M(\bar{x} + e^k)^{-1}\|\|F(\bar{x} + e^k) - F(\bar{x}) - M(\bar{x} + e^k)e^k\|.
\end{aligned}$$

Also gilt (10.1), falls erfüllt sind

(i) eine *Regularitätsbedingung*: es existiert ein $C > 0$ mit

$$\|M(\bar{x} + e^k)^{-1}\| \le C \qquad \text{für alle } k \in \mathbb{N},$$

(ii) eine *Approximationsbedingung*:

$$\lim_{k\to\infty} \frac{\|F(\bar{x}+e^k) - F(\bar{x}) - M(\bar{x}+e^k)e^k\|}{\|e^k\|} = 0.$$

Dies motiviert die folgende Definition: Wir nennen $F : \mathbb{R}^n \to \mathbb{R}^m$ *Newton-differenzierbar* in $x \in \mathbb{R}^n$ mit *Newton-Ableitung* $D_N F : \mathbb{R}^n \to \mathbb{R}^{n\times m}$, falls gilt

$$\lim_{\|h\|\to 0} \frac{\|F(x+h) - F(x) - D_N F(x+h)h\|}{\|h\|} = 0.$$

Beachte die Unterschiede zur Fréchet-Ableitung: Zum einen wird die Newton-Ableitung in $x + h$ anstelle von x ausgewertet. Wichtiger ist aber, dass kein expliziter Zusammenhang von $D_N F$ mit F in jedem Punkt gefordert wurde (im Gegensatz zur Fréchet-Ableitung, die eine Gateaux-Ableitung sein muss); eine Funktion ist also nur Newton-differenzierbar (oder nicht) in Bezug auf eine konkrete Wahl von $D_N F$. Insbesondere sind Newton-Ableitungen nicht eindeutig!

Ist F Newton-differenzierbar mit Newton-Ableitung $D_N F$, so erhält man das *semiglatte Newton-Verfahren*

$$x^{k+1} = x^k - D_N F(x^k)^{-1} F(x^k).$$

Direkt aus der Konstruktion folgt dann die lokal superlineare Konvergenz.

Satz 10.1. *Sei $F : \mathbb{R}^n \to \mathbb{R}^n$ und $\bar{x} \in \mathbb{R}^n$ mit $F(\bar{x}) = 0$. Ist F Newton-differenzierbar in $\bar{x}$ mit Newton-Ableitung $D_N F$, und existieren $\delta > 0$ und $C > 0$ mit $\|D_N F(x)^{-1}\| \le C$ für alle $x \in B_\delta(\bar{x})$, so konvergiert für alle x^0 hinreichend nahe an $\bar{x}$ das semiglatte Newton-Verfahren superlinear gegen $\bar{x}$.*

Beweis. Der Beweis ist völlig analog zum Konvergenzbeweis für das klassische Newton-Verfahren. Für $x^0 \in B_\delta(\bar{x})$ gilt wie schon gezeigt

$$\|e^1\| \le C\|F(\bar{x}+e^0) - F(\bar{x}) - D_N F(\bar{x}+e^0)e^0\|. \tag{10.2}$$

Sei nun $\varepsilon \in (0,1)$ beliebig. Aufgrund der Newton-Differenzierbarkeit existiert dann ein $\rho > 0$ mit

$$\|F(\bar{x}+h) - F(\bar{x}) - D_N F(\bar{x}+h)h\| \le \frac{\varepsilon}{C}\|h\| \qquad \text{für alle } \|h\| \le \rho.$$

Wählen wir also x^0 so, dass $\|\bar{x} - x^0\| \le \min\{\delta, \rho\}$ gilt, so folgt aus (10.2) die Abschätzung $\|\bar{x} - x^1\| \le \varepsilon\|\bar{x} - x^0\|$ und damit durch Induktion $\|\bar{x} - x^k\| \le \varepsilon^k\|\bar{x} - x^0\| \to 0$. Da ε beliebig war, kann man für jeden Schritt ein neues $\varepsilon_k \to 0$ wählen, und daher ist die Konvergenz superlinear. □

Der Rest des Kapitels ist nun der Konstruktion von Newton-Ableitungen gewidmet (wobei nicht verschwiegen werden soll, dass in der Praxis der Nachweis der Regularitätsbedingung die deutlich aufwändigere Aufgabe ist; insbesondere kann die *gleichmäßige* Invertierbarkeit aufgrund der nicht geforderten Stetigkeit nicht aus der Invertierbarkeit in $\bar{x}$ gefolgert werden!)

Wir beginnen mit dem offensichtlichen Zusammenhang mit der klassischen Differenzierbarkeit.

Satz 10.2. *Ist $F : \mathbb{R}^n \to \mathbb{R}^m$ stetig differenzierbar in $x \in \mathbb{R}^n$, so ist F Newton-differenzierbar in x mit Newton-Ableitung $D_N F = F'$.*

Beweis. Für beliebige $h \in \mathbb{R}^n$ gilt

$$\begin{aligned}\|F(x+h) - F(x) - F'(x+h)h\| \le \|&F(x+h) - F(x) - F'(x)h\| \\ &+ \|F'(x) - F'(x+h)\|\|h\|,\end{aligned}$$

wobei beide Summanden $o(\|h\|)$ sind: der erste nach Definition der (Fréchet-)Ableitung und der zweite wegen der Stetigkeit von F'. □

Rechenregeln beweist man nun analog zu denen für klassische Ableitungen. Die Summenregel ist offensichtlich; wir beweisen beispielhaft die Kettenregel.

Satz 10.3. *Sei $F : \mathbb{R}^n \to \mathbb{R}^p$ Newton-differenzierbar in $x \in \mathbb{R}^n$ mit Newton-Ableitung $D_N F$ und $G : \mathbb{R}^p \to \mathbb{R}^m$ Newton-differenzierbar in $y := F(x) \in \mathbb{R}^p$ mit Newton-Ableitung $D_N G$. Sind $D_N F$ und $D_N G$ gleichmäßig beschränkt in einer Umgebung von x bzw. y, so ist $G \circ F$ Newton-differenzierbar in x mit Newton-Ableitung*

$$D_N(G \circ F)(x) = D_N G(F(x)) D_N F(x), \qquad x \in \mathbb{R}^n.$$

Beweis. Wir gehen wie im Beweis der klassischen Kettelregel vor. Für $h \in \mathbb{R}^n$ und $g := F(x+h) - F(x)$ ist

$$(G \circ F)(x+h) - (G \circ F)(x) = G(y+g) - G(y).$$

Aus der Newton-Differenzierbarkeit von G folgt nun

$$\|(G \circ F)(x+h) - (G \circ F)(x) - D_N G(y+g)g\| = r_1(\|g\|)$$

mit $r_1(t)/t \to 0$ für $t \to 0$. Aus der Newton-Differenzierbarkeit von F folgt weiter

$$\|g - D_N F(x+h)h\| = r_2(\|h\|)$$

mit $r_2(t)/t \to 0$ für $t \to 0$. Insbesondere ist

$$\|g\| \le \|D_N F(x+h)\|\|h\| + r_2(\|h\|).$$

Wegen der gleichmäßigen Beschränktheit von $D_N F$ gilt also $\|g\| \to 0$ für $\|h\| \to 0$. Daher ist

$$\begin{aligned}\|(G \circ F)(x+h) - (G \circ F)(x) - D_N G(F(x+h)) D_N F(x+h)h\| \\ &\leq \|G(y+g) - G(g) - D_N G(y+g)g\| \\ &\quad + \|D_N G(y+g)\,[g - D_N F(x+h)h]\,\| \\ &\leq r_1(\|g\|) + \|D_N G(y+g)\| r_2(\|h\|),\end{aligned}$$

woraus wegen der gleichmäßigen Beschränktheit von $D_N G$ die gewünschte Aussage folgt. □

Schließlich folgt direkt aus der Definition der Produktnorm und der Newton-Differenzierbarkeit, dass für vektorwertige Funktionen die Newton-Ableitung komponentenweise berechnet werden kann.

Satz 10.4. *Seien $F_i : \mathbb{R}^n \to \mathbb{R}$ Newton-differenzierbar mit Newton-Ableitung $D_N F_i$ für $1 \leq i \leq m$. Dann ist*

$$F : \mathbb{R}^n \to \mathbb{R}^m, \qquad x \mapsto (F_1(x), \ldots, F_m(x))^T$$

Newton-differenzierbar mit Newton-Ableitung

$$D_N F(x) = (D_N F_1(x), \ldots, D_N F_m(x))^T, \qquad x \in \mathbb{R}^n.$$

Da die Definition keine konstruktive Vorschrift für die Newton-Ableitung enthält, bleibt die Frage, wie man dafür Kandidaten erhält, für die die Approximationsbedingung nachgeprüft werden kann. Wir betrachten hier nur eine spezielle Klasse von Funktionen, die in der nichtlinearen Optimierung häufig auftreten.

Eine Funktion $F : \mathbb{R}^n \to \mathbb{R}$ heißt *stückweise (stetig) differenzierbar* oder *PC^1-Funktion*, falls F stetig ist und für alle $x \in \mathbb{R}^n$ eine Umgebung $U_x \subset \mathbb{R}^n$ um x sowie eine endliche Menge $\{F_i : U_x \to \mathbb{R}\}_{i \in I_x}$ von stetig differenzierbaren Funktionen existiert mit

$$F(y) \in \{F_i(y)\}_{i \in I_x} \qquad \text{für alle } y \in U_x.$$

Wir nennen F dann eine *stetige Auswahl* der F_i in U_x. Die Menge

$$I_a(x) := \{i \in I_x \mid F(x) = F_i(x)\}$$

heißt *aktive Index-Menge*. Aufgrund der Stetigkeit der F_i gilt $F(y) \neq F_j(y)$ für alle $j \notin I_a(x)$ und y hinreichend nahe an x. Indizes, die nur auf einer Nullmenge aktiv sind, müssen wir in Folge nicht berücksichtigen. Wir definieren daher die *essentiell aktive Index-Menge*

$$I_e(x) := \{i \in I_x \mid x \in \operatorname{cl}(\operatorname{int}\{y \in U_x \mid F(y) = F_i(y)\})\} \subset I_a(x).$$

Beispiel 10.5. Betrachte die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \qquad t \mapsto \max\{0, t, t/2\},$$

d. h. $f_1(t) = 0$, $f_2(t) = t$, und $f_3(t) = t/2$. Dann ist $I_a(0) = \{1, 2, 3\}$ aber $I_e(0) = \{1, 2\}$, denn f_3 ist aktiv nur in $t = 0$ und daher $\operatorname{int}\{t \in \mathbb{R} \mid f(t) = f_3(t)\} = \emptyset = \operatorname{cl}\emptyset$.

Satz 10.6. *Sei $F : \mathbb{R}^n \to \mathbb{R}$ stückweise stetig differenzierbar. Dann ist F Newton-differenzierbar für alle $x \in \mathbb{R}^n$, und jedes*

$$D_N F(x) \in \operatorname{co}\left\{F_i'(x) \;\middle|\; i \in I_e(x)\right\}$$

ist eine Newton-Ableitung.

Beweis. Sei $x \in \mathbb{R}^n$ beliebig und $h \in \mathbb{R}^n$ mit $x + h \in U_x$. Nach Definition kann jedes Element in der konvexen Hülle geschrieben werden als

$$D_N F(x + h) = \sum_{i \in I_e(x+h)} t_i F_i'(x + h) \qquad \text{für} \sum_{i \in I_e(x+h)} t_i = 1, t_i \geq 0.$$

Da F stetig ist, gilt für alle $h \in \mathbb{R}^n$ mit $\|h\|$ hinreichend klein $I_e(x + h) \subset I_a(x + h) \subset I_a(x)$, wobei die zweite Inklusion aus der Tatsache folgt, dass wegen der Stetigkeit $F(x) \neq F_i(x)$ auch $F(x + h) \neq F_i(x + h)$ impliziert. Also ist sowohl $F(x + h) = F_i(x + h)$ als auch $F(x) = F_i(x)$ für alle $i \in I_e(x + h)$. Damit folgt aus Satz 10.2

$$\begin{aligned} |F(x + h) - F(x) - D_N F(x + h)h| &\leq \sum_{i \in I_e(x+h)} t_i |F_i(x + h) - F_i(x) - F_i' s(x + h)h| \\ &= o(\|h\|), \end{aligned}$$

da alle F_i nach Annahme stetig differenzierbar sind. □

Beispiel 10.7. Wir betrachten $f : \mathbb{R} \to \mathbb{R}$, $f(t) = \max\{0, t\}$, die offensichtlich stückweise stetig differenzierbar ist. Eine Newton-Ableitung ist daher jede Wahl von

$$D_N f(t) \in \begin{cases} \{0\} & \text{falls } t < 0, \\ \{1\} & \text{falls } t > 0, \\ [0, 1] & \text{falls } t = 0. \end{cases}$$

Nach Satz 10.4 erhalten wir daraus für das komponentenweise Maximum $F : \mathbb{R}^n \to \mathbb{R}^n$, $F(x) = \max\{0, x\}$, eine Newton-Ableitung komponentenweise durch

$$[D_N F(x)]_i = \begin{cases} 0 & \text{falls } x_i \leq 0, \\ 1 & \text{falls } x_i > 0, \end{cases}$$

(wobei wir den Fall $x_i = 0$ genausogut auch gleich 1 hätten setzen können). Diese Abbildung ist auch (sogar global) gleichmäßig beschränkt durch 1.

Zum Abschluss skizzieren wir noch die Anwendung auf restringierte Optimierungsprobleme. Dazu betrachten wir wieder das quadratische Teilproblem aus Abschnitt 9.3 (der Einfachheit ohne Gleichungsnebenbedingungen)

$$\text{(QP)} \qquad \begin{cases} \min\limits_{s\in\mathbb{R}^n} c^T s + \dfrac{1}{2} s^T H s \\ \text{mit } \ a_i^T s + \alpha_i \le 0, \quad i = 1, \dots, m, \end{cases}$$

mit KKT-Bedingungen

$$\begin{cases} Hs + c + \sum\limits_{i=1}^{m} \mu_i a_i = 0, \\ \mu_i \ge 0, a_i^T s + \alpha_i \le 0, \mu_i (a_i^T s + \alpha_i) = 0. \end{cases}$$

Die Komplementaritätsbedingungen formulieren wir wieder mit einer Komplementaritätsfunktion; diesmal verwenden wir

$$\mu - \max\{0, \mu + (As + \alpha)\} = 0$$

für A die Matrix mit Spalten a_i und α den Vektor mit Einträgen α_i. Man zeigt leicht durch Fallunterscheidung, dass diese Gleichheit äquivalent zu den Komplementaritätsbedingungen ist. Wir können also die KKT-Bedingungen schreiben als

$$M(s, \mu) := \begin{pmatrix} Hs + c + A^T \mu \\ \mu - \max\{0, \mu + (As + \alpha)\} \end{pmatrix} = 0.$$

Die erste Zeile ist affin-linear und damit stetig differenzierbar; die zweite Zeile ist stückweise stetig differenzierbar, so dass wir aus Beispiel 10.7 zusammen mit Satz 10.3 die Newton-Ableitung

$$D_N M(s, \mu) = \begin{pmatrix} H & A^T \\ -DA & I - D \end{pmatrix}$$

erhalten, wobei

$$D = D(\mu, s) = \operatorname{diag}(d_1, \dots, d_m), \qquad d_i = \begin{cases} 0 & \mu_i + (a_i^T s + \alpha_i) \le 0, \\ 1 & \mu_i + (a_i^T s + \alpha_i) > 0. \end{cases}$$

Ein semiglatter Newtonschritt ist dann gegeben durch Lösung von

$$D_N M(s^k, \mu^k) \begin{pmatrix} \Delta s \\ \Delta \mu \end{pmatrix} = -M(s^k, \mu^k)$$

und Setzen von $s^{k+1} = s^k + \Delta s$ und $\mu^{k+1} = \mu^k + \Delta\mu$. (Analog zur aktive-Mengen-Strategie kann man die stückweise Linearität von M und die Fallunterscheidung verwenden, um den Newton-Schritt in eine Gleichung für (s^{k+1}, μ^{k+1}) umzuschreiben.) Der wesentliche Unterschied zur aktive-Mengen-Strategie ist daher die Fallunterscheidung für die aktive Menge, die außerdem in jedem Schritt komplett neu berechnet wird.

Man zeigt nun relativ leicht mit dem üblichen Ansatz, dass $D_N M(s, \mu)$ für beliebige (s, μ) injektiv und damit invertierbar ist; die *gleichmäßig beschränkte* Invertierbarkeit, die Satz 10.1 benötigt, ist jedoch aufwändiger. Unter deutlich stärkeren Voraussetzungen kann man einen ähnlichen Ansatz auch direkt auf die KKT-Bedingungen (9.5) für das ursprüngliche nichtlineare Optimierungsproblem anwenden.

Literatur

W. Alt (2011), *Nichtlineare Optimierung. Eine Einführung in Theorie, Verfahren und Anwendungen*, 2. Aufl., Vieweg+Teubner, Wiesbaden.

D. P. Bertsekas (1982), *Constrained Optimization and Lagrange Multiplier Methods*, Computer Science and Applied Mathematics, Academic Press, Inc., New York-London.

S. Boyd & L. Vandenberghe (2004), *Convex Optimization*, Cambridge University Press, Cambridge, DOI: 10.1017/cbo9780511804441.

G. Braun, A. Carderera, C. W. Combettes, H. Hassani, A. Karbasi, A. Mokhtari & S. Pokutta (2025), *Conditional Gradient Methods*, Society for Industrial & Applied Mathematics, Philadelphia, DOI: 10.1137/1.9781611978568, arXiv: 2211.14103.

A. Cegielski (2012), *Iterative Methods for Fixed Point Problems in Hilbert Spaces*, Bd. 2057, Lecture Notes in Mathematics, Springer, Heidelberg, DOI: 10.1007/978-3-642-30901-4.

C. Geiger & C. Kanzow (2002), *Theorie und Numerik restringierter Optimierungsaufgaben*, Springer, Berlin, DOI: 10.1007/978-3-642-56004-0.

A. Ruszczyński (2006), *Nonlinear Optimization*, Princeton University Press, Princeton, NJ.

M. Ulbrich & S. Ulbrich (2012), *Nichtlineare Optimierung*, Birkhäuser, Basel, DOI: 10.1007/978-3-0346-0654-7.

S. J. Wright & B. Recht (2022), *Optimization for Data Analysis*, Cambridge: Cambridge University Press, DOI: 10.1017/9781009004282.