NICHTLINEARE OPTIMIERUNG

Vorlesungsskript, Wintersemester 2022/23

Christian Clason

Stand vom 2. Februar 2023

Institut für Mathematik und wissenschaftliches Rechnen
Universität Graz

INHALTSVERZEICHNIS

III KONVEXE OPTIMIERUNG

ÜBERBLICK

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

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

geschrieben

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

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 $x \in U$?
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 $x \in U$ 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 $y = Ku$ liefert, so kann man den unbekannten Parameter näherungsweise rekonstruieren, indem man das Problem

$$\min_{u \in U} \|Ku - y^\delta\|^2 + \alpha \|u\|^2$$

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

(iii) In der *optimalen Steuerung* ist man zum Beispiel daran interessiert, ein Auto oder eine Raumsonde möglichst effizient von einem Punkt x_0 zu einem anderen Punkt x_1 zu steuern. Beschreibt $x(t) \in \mathbb{R}^3$ die Position zum Zeitpunkt $t \in [0, T]$, so gehorcht $x(t)$ der Differenzialgleichung

$$\begin{cases} x'(t) = f(t, x(t), u(t)), \\ x(0) = x_0, \end{cases} \tag{1}$$

wobei $u(t)$ die Rolle der Steuerung spielt. Will man dabei den Treibstoffverbrauch (der proportional zu $|u(t)|$ ist) minimieren, führt das auf das Problem

$$\min_{(x,u)\in U} \int_0^T |u(t)| \quad \text{mit} \quad U = \{(x, u) : (1) \text{ ist erfüllt und } x(T) = x_1\}.$$

In dieser Vorlesung behandeln wir *nichtlineare Optimierungsprobleme*, in denen $f : \mathbb{R}^n \to \mathbb{R}$ 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] M. Ulbrich & S. Ulbrich (2012), *Nichtlineare Optimierung*, Birkhäuser, Basel, DOI: 10.1007/978-3-0346-0654-7

[iv] C. Clason & T. Valkonen (2020), Introduction to Nonsmooth Analysis and Optimization, ARXIV: 2020.00216

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 $\bar{x} \in X$ gilt 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 $\bar{x} \in X$ gilt 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), 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 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 jeder Richtung zunehmen würde, d. h. wenn die Steigung in jeder 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 $-d \in \mathbb{R}^n$ sofort $\nabla f(\bar{x})^T d = 0$ für alle $d \in \mathbb{R}^n$, was nur für $\nabla f(\bar{x}) = 0$ möglich ist. Wir erhalten also 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 $\xi := y + \theta(x - y)$ mit $\theta \in (0, 1)$ 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 $\xi := y + \theta(x - y)$ mit $\theta \in (0, 1)$ 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 (die Schwierigkeit ist, für alle Kompenenten 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*

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

2 NUMERISCHE VERFAHREN

In diesem Kapitel betrachten 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 aufwendiger 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 2.1 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 Nebenbedingung betrachten wir für $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, und $c \in \mathbb{R}^n$ das lineare Optimierungsproblem

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

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

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

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

Satz 3.1 (schwache Dualität). *Für alle $x \in P(A,b)$ und $y \in P^=(A^T,-c)$ ist*

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

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.

Satz 3.2 (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*

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

Der Beweis beruht auf dem Trennungssatz von Hahn–Banach, der auch im Laufe dieser Vorlesung wichtig sein wird.

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

$$a^T x_0 > \alpha \geq a^T x \qquad \text{für alle } x \in M. \tag{3.2}$$

Daraus erhält man als Zwischenschritt von unabhängiger Bedeutung das *Farkas-Lemma.*

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 $d \in \mathbb{R}^n$ mit $A^T d \geq 0$ gilt $b^T d \geq 0$.

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

$$\begin{cases} A\bar{x} \leq b, & \text{(primale Zulässigkeit)} \\ A^T \bar{y} = -c, \quad \bar{y} \geq 0, & \text{(duale Zulässigkeit)} \\ \bar{y}_i (b_i - A_i \bar{x}) = 0 \quad \text{für alle } i = 1, \ldots, m. & \text{(Komplementarität)} \end{cases} \tag{3.3}$$

Die Komplementaritätsbedingung sagt, dass $\bar{y}_i = 0$ oder $b_i = A_i \bar{x}$ für jedes $1 \leq i \leq m$ 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 3.6 (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 $i = 1, \ldots, m$.

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.

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*

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

Ist x ein innerer Punkt von X, so gilt $T_X(x) = \mathbb{R}^n$ (da wir in dem Fall $\{x^k\}_{k \in \mathbb{N}} \subset B_\varepsilon(x) \subset X$ beliebig wählen können). 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 $f : X \to \mathbb{R}$ 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 $\xi^k = \bar{x} + \theta_k(x^k - \bar{x})$ mit $\theta_k \in (0,1)$ 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 $x^k \in X$). 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

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

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.

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 Mengen der speziellen Form

$$X := \left\{x \in \mathbb{R}^n : g_i(x) \leq 0,\ 1 \leq i \leq m, \quad h_j(x) = 0,\ 1 \leq j \leq p\right\} \tag{4.4}$$

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.

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

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

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

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 $\xi_k := x + \theta_k(x^k - x)$ mit $\theta_k \in (0, 1)$. 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*

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

und den *Linearisierungskegel*

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

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. Als Beispiel betrachten wir die Menge

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

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

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

Der Tangentialkegel ist nach Lemma 4.3 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

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

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

$$\nabla g_i(x)^T v < 0 \qquad \text{für alle } i \in \mathcal{A}_X(x), \tag{4.5}$$

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

Beweis. Nach Lemma 4.3 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 $x_t \in X$ gilt, 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 $s := \min\{t_i : 1 \leq i \leq m\} > 0$ mit $x_t \in X$ für alle $t < s$.

Wir setzen nun $t_k := \frac{1}{k} \to 0$ und $x^k := x_{t_k} \to x$. 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 (4.5), 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 4.5. *Sei $x \in X$. Ist die Menge $\{\nabla g_i(x) : i \in \mathcal{A}_X(x)\} \subset \mathbb{R}^n$ linear unabhängig, so ist x regulär.*

Beweis. Unter dieser Voraussetzung hat das lineare Gleichungssystem

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

vollen (Zeilen-)Rang und damit für beliebige $b_i \in \mathbb{R}$ eine (nicht unbedingt eindeutige) Lösung. Wählt man $b_i < 0$ für alle $i \in \mathcal{A}_X(x)$, ist somit (4.5) erfüllt, und die Aussage folgt aus Satz 4.4. □

Wieder sind für konvexe Funktionen stärkere Aussagen möglich.

Folgerung 4.6. *Sei g_i konvex für alle $1 \le i \le m$. Existiert ein $\tilde{x} \in X$ mit*

$$g_i(\tilde{x}) < 0 \qquad \text{für alle } 1 \le i \le m, \tag{4.6}$$

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

Beweis. Für $x = \tilde{x}$ ist wegen (4.6) die aktive Menge $\mathcal{A}_X(x)$ leer und damit (4.5) 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 (4.5) mit $v := \tilde{x} - x$. □

Die Bedingung (4.6) 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 4.7. *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 4.4. Für $x \in X$ beliebig und $d \in L_X(x)$ setze $t_k := \frac{1}{k} \to 0$ und $x^k := x + t_k d \to x$. 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

$$a_i^T x^k = a_i^T x + t_k a_i^T d \le \alpha_i.$$

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

$$a_i^T x^k = a_i^T x + t_k a_i^T d < \alpha_i$$

für t_k hinreichend klein.

Also ist $x^k \in X$ für $k \in \mathbb{N}$ groß genug 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.

$$X = \left\{x \in \mathbb{R}^n : h_j(x) = 0, \quad 1 \leq j \leq p\right\}$$

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

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

Völlig analog zu Lemma 4.3 (nur mit Gleichheit an Stelle der Ungleichung) beweist man nun die folgende Inklusion.

Lemma 4.8. *Für alle $x \in X$ gilt $T_X(x) \subset L_X(x)$.*

Für die umgekehrte Inklusion benötigt man wieder eine Regularitätsbedingung.

Satz 4.9. *Sei $x \in X$. Ist die Menge $\left\{\nabla h_j(x) : 1 \leq j \leq p\right\} \subset \mathbb{R}^n$ linear unabhängig, so ist x regulär.*

Beweis. Der Beweis folgt im Prinzip dem von Satz 4.4; 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 $1 \leq i \leq p$. Zu diesem Zweck definieren wir zunächst die Funktion

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

und konstruieren mit ihrer Hilfe eine Funktion $H : \mathbb{R}^{p+1} \to \mathbb{R}^p$ komponentenweise durch

$$H_j(y, t) := h_j(x + td + \nabla h(x)^T y), \qquad 1 \leq j \leq p,$$

wobei $\nabla h(x)$ die Jacobi-Matrix von h bezeichnet (deren Spalten genau die $\nabla h_i(x)$ sind). 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

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

Nach Voraussetzung hat $\nabla h(x)$ vollen Rang, und damit ist $H_y(0,0)$ invertierbar. 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

$$y'(t) = -H_y(y(t), t)^{-1} H_t(y(t), t) \qquad \text{für alle } t \in (-\varepsilon, \varepsilon).$$

Weiter hat die Funktion $t \mapsto H(y, t)$ für $y \in \mathbb{R}^p$ fest die Ableitung

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

und damit folgt

$$y'(0) = -H_y(0,0)^{-1} H_t(0,0) = -H_y(0,0)^{-1} \nabla h(x) d = 0$$

wegen $\nabla h_i(x)^T d = 0$ für alle $1 \le i \le m$ nach Annahme an $d \in L_X(x)$.

Wir definieren nun die gesuchte Kurve durch

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

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 $x'(0) = d + \nabla h(x)^T y'(0) = d$. Setzen wir also $t_k := \frac{1}{k} \to 0$ und $x^k := x(t_k)$, so gilt $x^k \in X$ für $k \in \mathbb{N}$ groß genug, $x^k \to x(0) = x$ wegen $y(0) = 0$, sowie

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

d. h. $d \in T_X(x)$. □

Wieder ist die Linearität an sich eine Regularitätsbedingung.

Satz 4.10. *Seien alle h_j affin-linear für alle $1 \le i \le p$, 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 4.7: 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,

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

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

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

Die "einfache" Inklusion beweist man wieder völlig analog zu Lemma 4.3, indem man die Nebenbedingungen einzeln betrachtet.

Lemma 4.11. *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 4.12. *Sei $x \in X$. Gilt*

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

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

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

so 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 4.9 konstruiert man nun mit Hilfe von Annahme (i) für $\tilde{d} := d + \alpha v$ eine Kurve $x : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ mit $x(0) = x$, $x'(0) = \tilde{d} = d + \alpha v$ 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

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

Wie im Beweis von Satz 4.4 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 $t_k := \frac{1}{k} \to 0$ und $x^k := x(t_k) \to x$ mit $x^k \in X$ für $k \in \mathbb{N}$ hinreichend groß folgt nun wie zuvor, dass $d + \alpha v \in T_X(x)$ für alle $\alpha > 0$ ist. Aus der Abgeschlossenheit von Tangentialkegeln erhält man nun $d \in T_X(x)$. □

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

Folgerung 4.13. *Sei $x \in X$. Ist die Menge*

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

linear unabhängig, so ist x regulär.

Beweis. Unter dieser Voraussetzung hat das lineare Gleichungssystem

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

vollen (Zeilen-)Rang und damit für beliebige $b_i, c_j \in \mathbb{R}$ eine (nicht unbedingt eindeutige) Lösung. Wählt man $b_i < 0$ und $c_j = 0$, ist somit die MFCQ erfüllt, und die Aussage folgt aus Satz 4.12. □

Ebenso kombiniert man die "globalen" Regularitätsbedingungen.

Folgerung 4.14. *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 $\{b_j : 1 \leq j \leq p\}$ ist linear unabhängig,

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

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

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

Beweis. Wir müssen lediglich noch Annahme (ii) von Satz 4.12 nachprüfen. Dafür betrachten wir wieder für $x \in X$ beliebig die Richtung $v := \tilde{x} - x$. Genau wie im Beweis von Folgerung 4.6 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 4.12. □

Durch Kombination der Beweise von Satz 4.7 und Satz 4.10 erhält man schließlich den folgenden Satz, der erklärt, warum in der linearen Optimierung keine Regularitätsbedingung notwendig war.

Satz 4.15. *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, $X = \{x \in \mathbb{R}^n : h(x) = 0\}$ für $h : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar. Dann ist $L_X(x) = \left\{d \in \mathbb{R}^n : \nabla h(x)^T d = 0\right\} = \ker \nabla h(x)^T$ ein linearer Unterraum (nämlich der *Tangentialraum* von 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 4.16. *Seien $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und $X \subset \mathbb{R}^n$ von der Form (4.4). 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*

$$\begin{cases} \nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{\mu}_i \nabla g_i(\bar{x}) + \sum_{j=1}^{p} \bar{\lambda}_j \nabla h_j(\bar{x}) = 0, \\ \qquad\qquad\qquad\qquad h_j(\bar{x}) = 0, \quad 1 \le j \le p, \\ \bar{\mu}_i \ge 0, \quad g_i(\bar{x}) \le 0, \quad \bar{\mu}_i g_i(\bar{x}) = 0, \quad 1 \le i \le m. \end{cases} \tag{4.7}$$

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

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

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

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

Setzen wir

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

ist dies genau die erste Zeile von (4.7). Nach Definition gilt 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 (4.7) werden *Karush–Kuhn–Tucker-* oder kurz *KKT-Bedingungen* genannt. Die letzte Zeile ist dabei eine *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 4.17. *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 (4.7) 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. □

Für affin-lineare bzw. konvexe Gleichungs- bzw. Ungleichungsrestriktionen sind die KKT-Bedingungen sogar hinreichend. Dafür ist nicht mal eine Slater-Bedingung erforderlich.

Folgerung 4.18. *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 (4.7), so 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

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

Damit folgt aus der Konvexität von f mit Satz 1.2 und der ersten Zeile von (4.7)

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

denn wegen $x, \bar{x} \in X$ gilt $b_j^T x = \beta_j = b_j^T \bar{x}$ für alle $1 \leq j \leq p$. 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 3.5.

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*

$$L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}, \qquad (x, \mu, \lambda) \mapsto f(x) + \sum_{i=1}^{m} \mu_i g_i(x) + \sum_{j=1}^{p} \lambda_j h_j(x) \tag{4.8}$$

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

$$L(x, \mu, \lambda) = f(x) + \sum_{i=1}^{m} \mu_i g_i(x) + \sum_{j=1}^{p} \lambda_j h_j(x) \leq f(x) \tag{4.9}$$

wegen $g_i(x) \leq 0$ und $h_j(x) = 0$. Erfüllen $(\bar{x}, \bar{\mu}, \bar{\lambda})$ die KKT-Bedingungen (4.7), so gilt wegen der Komplementarität sogar

$$L(\bar{x}, \bar{\mu}, \bar{\lambda}) = f(\bar{x}), \tag{4.10}$$

und die erste Bedingung in (4.7) ist äquivalent zu

$$\nabla_x L(\bar{x}, \bar{\mu}, \bar{\lambda}) = \nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{\mu}_i \nabla g_i(\bar{x}) + \sum_{j=1}^{p} \bar{\lambda}_j \nabla h_j(\bar{x}) = 0. \tag{4.11}$$

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

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

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

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

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

Satz 4.19 (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 (4.7) und gilt*

$$d^T \nabla^2_{xx} L(\bar{x}, \bar{\mu}, \bar{\lambda}) d > 0 \qquad \text{für alle } d \in K_X(\bar{x}, \bar{\mu}) \setminus \{0\}, \tag{4.12}$$

so 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 $\{d^k\}_{k\in K}$ mit $d^k \to d$ für ein $d \in \mathbb{R}^n$ mit $\|d\| = 1$, d. h. $d \neq 0$. Wir zeigen nun, dass $d \in L_X(\bar{x})$ gilt. Für beliebige $k \in \mathbb{N}$ folgt aus Satz 1.10 die Existenz eines $\xi^{i,k}$ mit

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

wegen $g_i(x^k) \leq 0$ für $x^k \in X$ und $g_i(\bar{x}) = 0$ für $i \in \mathcal{A}_X(\bar{x})$. Wie üblich folgt aus $x^k \to \bar{x}$ auch $\xi^{i,k} \to \bar{x}$, und Division durch $\|x^k - \bar{x}\|$ und Grenzübergang $K \ni k \to \infty$ liefert

$$\nabla g_i(\bar{x})^T d \leq 0 \qquad \text{für alle } i \in \mathcal{A}_X(\bar{x}). \tag{4.13}$$

Analog folgt für beliebige $1 \leq j \leq p$ wegen $x^k, x \in X$ auch

$$\nabla h_j(\xi^{j,k})^T(x^k - \bar{x}) = h_j(x^k) - h_j(\bar{x}) = 0$$

und damit nach Division durch $\|x^k - \bar{x}\|$ und Grenzübergang

$$\nabla h_j(\bar{x})^T d = 0 \qquad \text{für alle } 1 \leq j \leq p. \tag{4.14}$$

Also ist $d \in L_X(\bar{x}) \setminus \{0\}$.

Wir machen nun eine Fallunterscheidung.

(i) $d \in K_X(\bar{x}, \bar{\mu})$. In diesem Fall zeigen wir, dass d die Bedingung (4.12) verletzt. Nach Konstruktion von $x^k \in X$ folgt aus Satz 1.11 für $x \mapsto L(x, \bar{\mu}, \bar{\lambda})$ zusammen mit (4.9)–(4.11)

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

Wieder gilt $\xi^k \to \bar{x}$ wegen $x^k \to \bar{x}$, und Division durch $\|x^k - \bar{x}\|^2$ und Grenzübergang $K \ni k \to \infty$ liefert

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

im Widerspruch zu (4.12).

(ii) $d \notin K_X(\bar{x}, \bar{\mu})$. Wegen (4.13) muss dann ein $i_+ \in \mathcal{A}_+(\bar{x}, \bar{\mu})$ mit $\nabla g_{i_+}(\bar{x})^T d < 0$ existieren. Analog zu oben folgt mit $f(x^k) \leq f(\bar{x})$ aus

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

durch Division und Grenzübergang zusammen mit $\nabla_x L(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0$ und (4.13), (4.14) die Ungleichung

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

und damit ebenfalls ein Widerspruch. □

Notwendige Optimalitätsbedingungen bedürfen wieder einer Regularitätsbedingung.

Satz 4.20. *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 $\bar{\lambda} \in \mathbb{R}^p$*

$$d^T \nabla^2_{xx} L(\bar{x}, \bar{\mu}, \bar{\lambda}) d \geq 0 \qquad \text{für alle } d \in K_X(\bar{x}, \bar{\mu}). \tag{4.15}$$

Beweis. Aus der LICQ folgt insbesondere die Regularität von $\bar{x}$ und damit aus Satz 4.16 die Existenz von Lagrange-Multiplikatoren, die nach Folgerung 4.17 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 4.12 konstruieren wir nun für dieses d eine zulässige Kurve $x(t) \in X$ so, dass für $t > 0$ klein genug $f(x(t)) < f(\bar{x})$ gilt. Dafür zerlegen wir die „biaktive Menge“ $\mathcal{A}_0(\bar{x}, \bar{\mu})$ weiter in

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

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 4.12 eine Kurve $x : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ konstruieren mit $x(0) = \bar{x}$, $x'(0) = d$, und

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

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 zweimal stetig differenzierbar ist, 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 (4.9)–(4.11) 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 $\theta_t \in (0, t)$ mit

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

Zusammen mit (4.9) erhalten wir also

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

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 (4.9) 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

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

als den *effektiven Definitionsbereich* von q.

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

Beispiel 5.2. Betrachte das lineare Optimierungsproblem

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

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 solche $\mu \geq 0$ angenommen, für die $A^T\mu = -c$ gilt. Damit ist das Lagrange-duale Problem genau (LD),

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

Im allgemeinen ist die explizite Berechnung der dualen Funktion $q(\mu,\lambda)$ wegen der fehlenden Nebenbedingung $x \in X$ zwar einfacher als die Lösung des ursprünglichen (primalen) Minimierungsproblem, aber trotzdem nur in Spezialfällen möglich. Es gelten jedoch stets die folgenden Eigenschaften.

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

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

wegen $(\mu, \lambda), (\tilde{\mu}, \tilde{\lambda}) \in \operatorname{dom} q$. 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* (4.7)*;*

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

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

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 (4.10) und (4.9) 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 (4.11) gilt weiterhin $\nabla_x L(x, \bar{\mu}, \bar{\lambda}) = 0$. 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* Minimimum 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 KKT-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 (4.7) 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 $\min_{x \in X} f(x)$, so hat auch das duale Problem eine Lösung $\bar{\mu} \geq 0, \lambda \in \mathbb{R}^p$, und es gilt*

$$\max_{\mu \geq 0, \lambda} q(\mu, \lambda) = \min_{x \in X} f(x). \tag{5.4}$$

Beweis. Nach Satz 5.1 gilt $\sup_{\mu \geq 0, \lambda} q(\mu, \lambda) \leq \inf_{x \in X} f(x)$, wir müssen also nur noch die umgekehrte Ungleichung zeigen. Nach Annahme und Satz 1.9 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 (4.10) folgt daher

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

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 STRAFVERFAHREN

Wir kommen nun zu Verfahren zur numerischen Lösung von Optimierungsproblemen mit 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.

6.1 QUADRATISCHE STRAFVERFAHREN

Wir betrachten wieder den konkreten Fall

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

für $g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar, 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 $g_i(x) \leq 0$ 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

$$\begin{aligned} P_\alpha(x) &:= f(x) + \frac{\alpha}{2}\sum_{i=1}^{m}|(g_i)^+|^2 + \frac{\alpha}{2}\sum_{j=1}^{p}|h_j(x)|^2 \\ &= f(x) + \frac{\alpha}{2}\|(g(x))^+\|^2 + \frac{\alpha}{2}\|h(x)\|^2, \end{aligned} \tag{6.1}$$

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

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

Wir fragen uns zuerst, wann das penalisierte Problem

$$\min_{x \in \mathbb{R}^n} P_\alpha(x) \tag{6.2}$$

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

Satz 6.1. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ stetig, $X \subset \mathbb{R}^n$ nichtleer und abgeschlossen, und entweder*

(i) f koerziv oder

(ii) X beschränkt, π koerziv, und f nach unten beschränkt.

Dann existiert für alle $\alpha > 0$ eine Lösung $x_\alpha \in \mathbb{R}^n$ von (6.2).

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 6.1 erfüllt sind. Ein Strafverfahren hat dann die Form von

Algorithmus 6.1 : Quadratisches Strafverfahren

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

for $k = 0, \ldots$ **do**
  Berechne x^{k+1} als globalen Minimierer von (6.2) 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 6.2. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ stetig und $X \subset \mathbb{R}^n$ nichtleer. Angenommen, Algorithmus 6.1 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 nun $\pi(x^k) \ge \pi(x^{k+1})$.

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 $h(x^k) \to 0$ gehen. □

Damit können wir nun die Konvergenz zeigen.

Satz 6.3. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ stetig und $X \subset \mathbb{R}^n$ nichtleer. Dann bricht Algorithmus 6.1 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 6.1 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 6.2 (iv)

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

Nach Definition von π impliziert dies $\bar{x} \in X$. Für alle $x \in X$ und $k \in K$ gilt wegen $\pi(x) \geq 0$ 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 (6.2) 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

$$\nabla P_\alpha(x) = \nabla f(x) + \alpha \sum_{i=1}^{m} (g_i(x))^+ \nabla g_i(x) + \alpha \sum_{j=1}^{p} h_j(x) \nabla h_j(x). \tag{6.3}$$

Vergleicht man dies mit (4.8), so erhält man

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

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 6.4. *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 6.1 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 (4.7).

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

$$\mu_i^k = \alpha_k (g_i(x^k))^+ \to 0 =: \bar{\mu}_i \qquad \text{für alle } i \notin \mathcal{A}_X(\bar{x}). \tag{6.4}$$

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 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 (6.3))

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

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 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 erste Relation von (4.7). Aus Satz 6.3 folgt insbesondere $\bar{x} \in X$ und damit $h_j(\bar{x}) = 0$ für $1 \le j \le p$, d. h. die zweite Relation. 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 $1 \le i \le m$. Schließlich folgt aus (6.4) auch die Komplementaritätsbedingung $\bar{\mu}_i g_i(\bar{x}) = 0$ für alle $1 \le i \le m$ und damit die dritte Relation. □

Ein Nachteil der quadratischen Penalisierung ist, dass in der Regel $x_\alpha \notin X$ gilt. Aus (6.3) 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).

6.2 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 für α groß genug in der Tat exakt ist.

Satz 6.5. *Seien $f : \mathbb{R}^n \to \mathbb{R}$ und $g_i : \mathbb{R}^n \to \mathbb{R}$, $1 \le i \le m$, stetig differenzierbar und konvex, und sei $h_j : \mathbb{R}^n \to \mathbb{R}$, $1 \le j \le p$, affin-linear. Es sei weiter die Slater-Bedingung (4.6) 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 (4.7). Insbesondere existieren Lagrange-Multiplikatoren $\bar{\mu} \in \mathbb{R}^m$ und $\bar{\lambda} \in \mathbb{R}^p$. Wir wählen nun

$$\bar{\alpha} := \max\{\bar{\mu}_1, \dots, \bar{\mu}_m, |\bar{\lambda}_1|, \dots, |\bar{\lambda}_p|\}. \tag{6.5}$$

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

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

Also ist $\bar{x}$ globaler Minimierer von $f + \alpha\pi_1$. □

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 in Minimierern, die im Inneren von X liegen).

Satz 6.6. *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

$$\nabla f(\bar{x}) + \alpha \nabla \pi(\bar{x}) = 0. \tag{6.6}$$

Für beliebige $\alpha_1, \alpha_2 \geq \bar{\alpha}$ 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 (6.6) auch $\nabla f(\bar{x}) = 0$, im Widerspruch zur Annahme. □

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

6.3 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.

GLEICHUNGSNEBENBEDINGUNGEN

Wir betrachten das Optimierungsproblem

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{mit } h(x) = 0 \tag{6.7}$$

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 (6.7), so ist dies trotz der stetigen Differenzierbarkeit eine exakte Penalisierung.

Satz 6.7. *Seien f, h zweimal stetig differenzierbar. Erfüllt $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^p$ für (6.7) die hinreichende Optimalitätsbedingungen aus Satz 4.19, 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. Nach Annahme erfüllt $(\bar{x}, \bar{\lambda})$ insbesondere die KKT-Bedingungen. Daraus folgt sofort für alle $\alpha > 0$

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

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

$$d^T \nabla^2_{xx} L(\bar{x}, \bar{\mu}) d > 0 \qquad \text{für alle } d \neq 0 \text{ mit } \nabla h(\bar{x})^T d = 0. \tag{6.8}$$

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

$$d^T \left(\nabla^2_{xx} L(\bar{x}, \bar{\lambda}) + \alpha \nabla h(\bar{x}) \nabla h(\bar{x})^T \right) d > 0 \quad \text{für alle } d \neq 0. \tag{6.9}$$

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

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

Die linke Seite 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 (6.8).

Also erfüllt $\bar{x}$ für $\alpha > 0$ groß genug die hinreichende Optimalitätsbedingung Satz 1.8 für die Funktion $x \mapsto L_\alpha(x, \bar{\lambda})$. □

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 6.1 die Konvergenz zeigen zu können.

Satz 6.8. *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 (6.1).*

Beweis. Aus der Optimalität von x^k und (4.9) folgt für $k \in \mathbb{N}$ beliebig

$$f(x^k) + (\lambda^k)^T h(x^k) + \frac{\alpha_k}{2}\|h(x^k)\|^2 = L_{\alpha_k}(x^k, \lambda^k) \le \inf_{x\in X} L_{\alpha_k}(x, \lambda^k) = \inf_{x\in X} f(x). \tag{6.10}$$

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 (6.10), dass gilt

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

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 (6.1), woraus $f(\bar{x}) \ge \inf_{x\in X} f(x)$ folgt. Daraus erhalten wir

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

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

Analog zu Satz 6.4 haben wir auch ein Konvergenzresultat für stationäre Punkte.

Satz 6.9. *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 (6.1).

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

$$\lambda^k + \alpha_k h(x^k) \to \bar{\lambda} := \left(\nabla h(\bar{x})\nabla h(\bar{x})^T\right)^{-1} \nabla h(\bar{x}) \left(-\nabla f(\bar{x})^T\right). \tag{6.11}$$

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 (6.11) auch

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

und damit wegen der LICQ die KKT-Bedingungen. □

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

Algorithmus 6.2 : 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].

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 schließlich $z_i = s_i^2$ 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 6.2 an. Dabei müssen wir in Schritt 2 den Minimierer bezüglich (x, s) bestimmen, d. h. das Problem

$$\min_{x\in\mathbb{R}^n, s\in\mathbb{R}^m} f(x) + \sum_{i=1}^m \left(\mu_i(g_i(x) + s_i^2) + \frac{\alpha}{2}(g_i(x) + s_i^2)\right)$$

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 Fallunterscheidung.

(i) $\mu_i + \alpha g_i(x) \leq 0$: Dann ist $\hat{z}_i \geq 0$ und damit auch der gewünschte Minimierer. Also ist wegen $g_i(x) + \hat{z}_i = -\frac{\mu_i}{\alpha}$

$$\mu_i(g_i(x) + \hat{z}_i) + \frac{\alpha}{2}(g_i(x) + \hat{z}_i)^2 = -\frac{\mu_i^2}{\alpha} + \frac{\alpha}{2}\frac{\mu_i^2}{\alpha^2} = -\frac{1}{2\alpha}\mu_i^2.$$

(ii) $\mu_i + \alpha g_i(x) > 0$: Dann ist $\hat{z}_i < 0$ und damit wird das Minimum im Randpunkt $z = 0$ angenommen. Also ist mit quadratischer Ergänzung

$$\mu_i g_i(x) + \frac{\alpha}{2} g_i(x)^2 = \frac{1}{2\alpha} \left((\mu_i + \alpha g_i(x)) - \mu_i^2 \right).$$

In beiden Fällen können wir dies schreiben als

$$\mu_i (g_i(x) + \hat{z}_i) + \frac{\alpha}{2} (g_i(x) + \hat{z}_i)^2 = \frac{1}{2\alpha} \left(\max\{0, \mu_i + \alpha g_i(x)\}^2 - \mu_i^2 \right).$$

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 Minimierung vernachlässigt werden.

Betrachten wir nun wie im Beweis von Satz 6.9 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

$$\mu^{k+1} := \max\{0, \mu^k + \alpha_k g(x^k)\} \geq 0.$$

Die Kombination mit Gleichungsnebenbedingungen ist nun trivial, und wir erhalten das volle Multiplikator-Strafverfahren.

Algorithmus 6.3 : Multiplikator-Strafverfahren

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

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

2 Berechne x^{k+1} als stationären Punkt von

$$\min_{x \in \mathbb{R}^n} f(x) + (\lambda^k)^T h(x) + \frac{\alpha_k}{2} \|h(x)\|^2 + \frac{1}{2\alpha_k} \|(\mu^k + \alpha_k g(x))^+\|^2$$

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 Setze $\mu^{k+1} = \max\{0, \mu^k + \alpha_k g(x^{k+1})\}$

8 Wähle $\alpha_{k+1} \geq \alpha_k$

Für die Aktualisierung von α_k muss man nun neben der Verletzung der Gleichungsnebenbedingung auch die der Ungleichungsnebenbedingung betrachten – sowie der für die KKT-Bedingungen notwendigen Komplementarität der Lagrange-Multiplikatoren $\bar{\mu} \geq 0$. 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 auch, wenn gilt

$$\| \min\{-g(x^{k+1}), \mu^{k+1}\} \| \geq \rho \| \min\{-g(x^k), \mu^k\} \|.$$

[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.

7 BARRIERE- UND INNERE-PUNKTE-VERFAHREN

Strafverfahren sind ungeeignet, wenn die Zielfunktion f für $x \notin X$ gar nicht definiert ist. In *Barriereverfahren* wird dagegen 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

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

für $g_i : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar. (Zusätzliche Gleichungsnebenbedingungen werden entweder durch eine Straffunktion oder den im folgenden Kapitel 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

$$\min_{x \in \mathbb{R}^n} B_\alpha(x) \tag{7.1}$$

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 (4.6).

Satz 7.1. *Sei $f : X \to \mathbb{R}$ stetig, $X \subset \mathbb{R}^n$, beschränkt, nichtleer und abgeschlossen. Gilt die Slater-Bedingung (4.6), dann existiert für alle $\alpha > 0$ eine Lösung $x_\alpha \in X$ von (7.1).*

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

$$X_M := \{x \in X : B_\alpha(x) \leq M\}.$$

Offensichtlich muss ein Minimierer von B_α (falls existent) in X_M liegen, d. h. auch Lösung sein von

$$\min_{x \in X_M} B_\alpha(x). \tag{7.2}$$

Es genügt also zu zeigen, dass dieses Problem eine Lösung hat.

Wir zeigen zuerst, dass X_M kompakt ist. Mit X ist auch $X_M \subset X$ beschränkt. Für die Abgeschlossenheit sei $\{x^k\}_{k\in\mathbb{N}} \subset X_M$ eine konvergente Folge mit Grenzwert $x \in X$. 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 (7.2). □

Man kann nun hoffen, dass für $\alpha \to 0$ die Folge $\{x_\alpha\}_{\alpha>0}$ der entsprechenden Minimierer von (7.1) gegen einen Minimierer $\bar{x} \in X$ von f konvergiert. Entsprechend hat das Barriereverfahren nun die Form von

Algorithmus 7.1 : Barriereverfahren

Input : $\alpha_0 > 0$, $x^0 \in X$ strikt zulässig

1 Setze $k = 0$
2 **for** $k = 0, \ldots$ **do**
3 Berechne x^{k+1} als globalen Minimierer von (7.1) mit $\alpha = \alpha_k$ (und Startwert x^k)
4 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 7.2. *Angenommen, Algorithmus 7.1 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 7.1 zeigen zu können, bedarf es einer Art Regularitätsbedingung. Dafür definieren wir das *strikte Innere*

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

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 Innerern, von der konkreten Beschreibung von X durch die g_i ab.)

Satz 7.3. *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 7.1 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 $\{x^k\}_{k\in K} \subset X^-$ eine konvergente Teilfolge mit Grenzwert $\bar{x}$, der wegen $\overline{X^-} = X$ in X liegt. 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 7.2 (i) und der Optimalität von x^k folgt nun

$$f(x^k) + \alpha_k\beta(x^0) \leq f(x^k) + \alpha_k\beta(x^k) \leq f(\hat{x}) + \alpha_k\beta(\hat{x})$$

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 stetig differenzierbare g_i folgt aus $\frac{d}{dt}\ln(t) = \frac{1}{t}$ mit der Summen- und Kettenregel

$$\nabla B_\alpha(x) = \nabla f(x) - \alpha \sum_{i=1}^{m} \frac{\nabla g_i(x)}{g_i(x)}. \tag{7.3}$$

Vergleicht man dies wieder mit (4.8), so erhält man

$$\nabla B_\alpha(x) = \nabla_x L(x, \mu) \qquad \text{für} \quad \mu := \frac{\alpha}{-g(x)},$$

und in der Tat kann man analog zu Satz 6.4 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 (4.7) – 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

$$\text{(LP)} \qquad \begin{cases} \min\limits_{x \in \mathbb{R}^n} c^T x \\ \text{mit } Ax = b, \\ \quad x \ge 0. \end{cases}$$

Wir nehmen in Folge stets an, dass dieses Problem eine zulässige Lösung $\bar{x} \in P^=(A, b)$ besitzt. Da $\bar{x}$ nach Satz 4.15 regulär ist, existieren nach Satz 4.16 Lagrange-Multiplikatoren $s \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}^m$ mit

$$\begin{cases} A^T \lambda + s = c, \\ \quad A\bar{x} = b, \\ \quad \bar{x}_i s_i = 0, \\ \quad \bar{x}, s \ge 0. \end{cases}$$

Wir ersetzen nun diese KKT-Bedingungen durch die KKT-Bedingungen des entsprechenden Barriereproblems

$$\text{(7.4)} \qquad \begin{cases} \min\limits_{x \in \mathbb{R}^n} c^T x - \alpha \sum\limits_{i=1}^{n} \ln(x_i) \\ \text{mit } Ax = b, \end{cases}$$

was ebenfalls ein (sogar strikt) konvexes Optimierungsproblem ist. Wir haben hier nur affin-lineare Gleichungsnebenbedingungen, also existiert nach Satz 4.10 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*

$$\begin{cases} A^T\lambda_\alpha + s_\alpha = c, \\ Ax_\alpha = b, \\ [x_\alpha]_i[s_\alpha]_i = \alpha, \\ x_\alpha, s_\alpha > 0. \end{cases} \tag{7.5}$$

Wir müssen noch zeigen, dass überhaupt Lösungen von (7.5) existieren, was wir analog zu Satz 7.1 machen. (Beachte, dass $P^=(A,b)$ nicht als beschränkt vorausgesetzt war.) Dazu definieren wir die *primal-dual zulässige Menge*

$$Z := \left\{(x,\lambda,s) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n : Ax = b, A^T\lambda + s = c, x, s \geq 0\right\}$$

sowie die *strikt zulässige Menge*

$$Z^+ := \left\{(x,\lambda,s) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n : Ax = b, A^T\lambda + s = c, x, s > 0\right\}.$$

Satz 7.4. *Sei Z^+ nichtleer. Dann existiert für alle $\alpha > 0$ eine Lösung $(x_\alpha, \lambda_\alpha, s_\alpha)$ von (7.5). Dabei sind (x_α, s_α) eindeutig bestimmt. Hat A vollen Rang, ist auch λ_α eindeutig.*

Beweis. Da (7.4) 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

$$X_M := \{x \in \mathbb{R}^n : Ax = b, B_\alpha(x) \leq M\}.$$

Wie im Beweis von Satz 7.1 folgt die Abgeschlossenheit von X_M; wir müssen also nur noch die Beschränktheit zeigen. Zunächst gilt für alle $x \in X_M$ nach Definition

$$\begin{aligned} M &\geq c^Tx - \alpha\sum_{i=1}^n \ln(x_i) \\ &= c^Tx - (Ax-b)^T\hat{\lambda} - \alpha\sum_{i=1}^n \ln(x_i) \\ &= c^Tx - (c-\hat{s})^Tx + b^T\hat{\lambda} - \alpha\sum_{i=1}^n \ln(x_i) \\ &= x^T\hat{s} + b^T\hat{\lambda} - \alpha\sum_{i=1}^n \ln(x_i), \end{aligned}$$

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 $\lim_{t\to 0}(s_i t - \alpha \ln(t)) = \infty$, 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 (7.4) ist. Damit ist auch $s_\alpha := \frac{\alpha}{x_\alpha}$ eindeutig. Dagegen ist der Lagrange-Multiplikator λ_α aus (7.5) 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 (7.5) 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 (7.5) 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 besteht in 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 7.5. *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 $\alpha \mapsto (x_\alpha, \lambda_\alpha, s_\alpha)$ 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

$$(x^{k+1}, \lambda^{k+1}, s^{k+1}) \in Z^\gamma := \left\{(x,\lambda,s) \in Z^+ : x_i s_i \geq \frac{\gamma}{n} x^T s, 1 \leq i \leq n\right\}.$$

Zusammen erhalten wir das folgende zulässige Pfadverfolgungsverfahren.

Algorithmus 7.2 : zulässiges Innere-Punkte-Verfahren

Input : $\gamma \in (0,1)$, $0 < \sigma_{\min} < \sigma_{\max} < 1$, $\varepsilon > 0$, $(x^0, \lambda^0, s^0) \in Z^+$

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

2 **if** $\mu_k := \frac{1}{n}(x^k)^T s^k \leq \varepsilon$ **then return** (x^k, s^k, λ^k)

3 Wähle $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$

4 Löse

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ D_{s^k} & 0 & D_{x^k} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \\ \Delta s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -x^k \odot s^k + \sigma_k \mu_k \mathbb{1} \end{pmatrix} \tag{7.6}$$

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

$$x^{k+1} = x^k + t_k \Delta x, \qquad \lambda^{k+1} = \lambda^k + t_k \Delta\lambda, \qquad s^{k+1} = s^k + t_k \Delta s$$

Wir zeigen nun die Konvergenz von Algorithmus 7.2, wofür wir eine Reihe technischer Lemmas brauchen 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 7.6. *Sei $(\Delta x, \Delta\lambda, \Delta s)$ eine Lösung von (7.6). 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 7.5 durch Multiplikation der ersten Gleichung von (7.6) mit Δx^T und Verwenden der zweiten Gleichung.

Zu (ii): Die dritte Gleichung von (7.6) lautet komponentenweise für $1 \leq i \leq n$

$$s_i^k \Delta x_i + x_i^k \Delta s_i = -x_i^k s_i^k + \sigma_k \mu_k,$$

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 7.7. *Für $u, v \in \mathbb{R}^n$ mit $u^T v \geq 0$ gilt*

$$\|u \odot v\|_2 \leq \frac{1}{2}\|u+v\|_2^2.$$

Beweis. Zunächst gilt für alle $a, b \in \mathbb{R}$

$$\frac{1}{4}(a+b)^2 = \frac{1}{4}(a-b)^2 + ab \geq ab.$$

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 7.8. *Sei $(x^k, \lambda^k, s^k) \in Z^\gamma$ und $(\Delta x, \Delta\lambda, \Delta s)$ Lösung von (7.6). 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 (7.6) mit

$$(D_{x^k} D_{s^k})^{-1/2} := \operatorname{diag}((x_1^k s_1^k)^{-1/2}, \dots (x_n^k s_n^k)^{-1/2})^T$$

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 7.6 (i) folgt aus Lemma 7.7 dann

$$\begin{aligned}
\|\Delta x \odot \Delta s\|_2 \le \frac{1}{2}\|(A^k)^{-1}\Delta x + A^k \Delta s\|_2^2 &= \frac{1}{2}\|(D_{x^k} D_{s^k})^{-1/2}(-x^k \odot s^k + \sigma_k \mu_k \mathbb{1})\|_2^2 \\
&= \frac{1}{2}\sum_{i=1}^n \left(-\sqrt{x_i^k s_i^k} + \sigma_k \mu_k \frac{1}{\sqrt{x_i^k s_i^k}} \right)^2 = \frac{1}{2}\sum_{i=1}^n \left(x_i^k s_i^k - 2\sigma_k \mu_k + \frac{\sigma_k^2 \mu_k^2}{x_i^k s_i^k} \right) \\
&\le \frac{1}{2}\left((x^k)^T s^k - 2n\sigma_k \mu_k + \sigma_k^2 \mu_k^2 \frac{n}{\gamma \mu_k} \right) \\
&\le \frac{1}{2}\left(1 + \frac{1}{\gamma}\right) n \mu_k,
\end{aligned}$$

wobei wir $x_i^k s_i^k \ge \gamma \mu_k$ 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 7.2.

Lemma 7.9. *Sei* $(x^k, \lambda^k, s^k) \in Z^\gamma$ *und* $(\Delta x, \Delta\lambda, \Delta s)$ *Lösung von* (7.6). *Dann gilt*

$$(x^k(t), \lambda^k(t), s^k(t)) \in Z^\gamma \qquad \text{für} \quad 0 \le t \le t_k := 2\gamma \frac{\sigma_k}{n} \frac{1-\gamma}{1+\gamma}.$$

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 (7.6), 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 7.8

$$|\Delta x_i \Delta s_i| \le \|\Delta x \odot \Delta s\|_2 \le \frac{1}{2}\left(1 + \frac{1}{\gamma}\right) n \mu_k.$$

Zusammen ergibt das

$$\begin{aligned} x_i^k(t)s_i^k(t) &= (x_i^k + t\Delta x_i)(s_i^k + t\Delta s_i) \\ &= x_i^k s_i^k + t(x_i^k \Delta s_i + s_i^k \Delta x_i) + t^2 \Delta x_i \Delta s_i \\ &\geq (1-t)x_i^k s_i^k + t\sigma_k \mu_k - t^2 |\Delta x_i \Delta s_i| \\ &\geq (1-t)\gamma\mu_k + t\sigma_k\mu_k - \frac{t^2}{2}\left(1+\frac{1}{\gamma}\right) n\mu_k, \end{aligned}$$

wobei wir im letzten Schritt wieder $x_i^k s_k^i \geq \gamma\mu_k$ für $(x^k, \lambda^k, s^k) \in Z^\gamma$ verwendet haben. Es genügt nach Lemma 7.6 (ii) also zu garantieren, dass gilt

$$(1-t)\gamma\mu_k + t\sigma_k\mu_k - \frac{t^2}{2}\left(1+\frac{1}{\gamma}\right) n\mu_k \geq \gamma\mu_k(t) = \gamma(1-t(1-\sigma_k))\mu_k.$$

Vereinfachen und Auflösen nach $t > 0$ ergibt dann genau $t \leq t_k$.

Es bleibt noch zu zeigen, dass $(x^k(t), \lambda^k(t), s^k(t)) \in Z^+$ für alle $t \leq t_k$ gilt. Da die Gleichungsnebenbedingungen für alle $t \geq 0$ erhalten bleiben, ist nur $x^k(t), s^k(t) > 0$ von Bedeutung. Aber das folgt aus

$$[x^k(t)]_i [s^k(t)]_i \geq \gamma(1-t(1-\sigma_k))\mu_k > 0$$

wegen $\mu_k = (x^k)^T(s^k)/n > 0$ für $x^k, s^k > 0$, $\gamma, \sigma_k \in (0,1)$, und $t \leq t_k < 1$. □

Für die so gewählte maximale Schrittweite können wir das zentrale Konvergenzresultat for Algorithmus 7.2 beweisen.

Satz 7.10. *Sei $\{x^k, \lambda^k, s^k\}_{k\in\mathbb{N}}$ durch Algorithmus 7.2 mit t_k aus Lemma 7.9 erzeugt. Dann existiert ein $\delta > 0$ mit*

$$\mu_{k+1} \leq \left(1 - \frac{\delta}{n}\right)\mu_k \qquad \text{für alle } k \in \mathbb{N}.$$

Beweis. Aus Lemma 7.6 (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

$$\sigma_k(1-\sigma_k) \geq \sigma^* := \min\{\sigma_{\min}(1-\sigma_{\min}), \sigma_{\max}(1-\sigma_{\max})\} > 0$$

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 bekanntermaßen nicht gilt.

Folgerung 7.11. *Sei $\{x^k, \lambda^k, s^k\}_{k\in\mathbb{N}}$ durch Algorithmus 7.2 mit t_k aus Lemma 7.9 erzeugt. Sei $\varepsilon > 0$ und $\kappa > 0$ mit $\mu^0 \le \varepsilon^{-\kappa}$. Dann existiert ein $K = O(n|\log\varepsilon|)$ mit*

$$\mu_k \le \varepsilon \qquad \text{für alle } k \ge K.$$

Beweis. Nach Satz 7.10 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)

$$\log\mu_k \le k \log\left(1 - \frac{\delta}{n}\right) + \kappa \log\tfrac{1}{\varepsilon} \le k\left(-\frac{\delta}{n}\right) + \kappa\log\tfrac{1}{\varepsilon} \tag{7.7}$$

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 (7.7) nicht größer als $\log\varepsilon = -\log\frac{1}{\varepsilon}$ ist; Auflösen nach k ergibt dann für $\varepsilon < 1$

$$k \ge (1+\kappa)\frac{n}{\delta}\log\tfrac{1}{\varepsilon} = \left(\frac{1+\kappa}{\delta}\right) n|\log\varepsilon| := K$$

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 (7.6) 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

$$(x^{k+1}, \lambda^{k+1}, s^{k+1}) \in Z^{\gamma,\beta}$$
$$:= \left\{(x,\lambda,s) : x, s > 0,\ \|(b - Ax, c - s - A^T\lambda)\| \le r^0\beta x^T s,\ x \odot s \ge \frac{\gamma}{n}x^T s\right\}$$

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].

8 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.

8.1 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 8.1. *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 8.1 : Lagrange–Newton-Verfahren

Input : $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^p$

while $k = 0, \dots$ **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 8.2. *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.

8.2 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

$$\begin{pmatrix} \nabla^2_{xx} L(x^k, \lambda^k) & \nabla h(x^k) \\ \nabla h(x^k)^T & 0 \end{pmatrix} \begin{pmatrix} s_x \\ s_\lambda \end{pmatrix} = \begin{pmatrix} -\nabla_x L(x^k, \lambda^k) \\ -h(x^k) \end{pmatrix}. \tag{8.1}$$

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

$$\begin{cases} \min\limits_{s\in\mathbb{R}^n} \nabla f(x^k)^T s + \frac{1}{2} s^T \nabla^2_{xx} L(x^k, \lambda^k) s \\ \text{mit } \; h(x^k) + \nabla h(x^k)^T s = 0. \end{cases} \tag{8.2}$$

Da die Linearität der Nebenbedingungen eine Regularitätsbedingung darstellt (Satz 4.10), 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 (8.1). 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 (8.2). Anstelle eines Updates berechnet man hier also direkt die neue Näherung für den Lagrange-Multiplikator.

Für zusätzliche Ungleichungsnebenbedingungen

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{mit} \quad g(x) \le 0, \quad h(x) = 0, \tag{8.3}$$

betrachtet man analog für $(x^k, \mu^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ das Problem

$$\begin{cases} \min\limits_{s \in \mathbb{R}^n} \nabla f(x^k)^T s + \frac{1}{2} s^T \nabla^2_{xx} L(x^k, \mu^k, \lambda^k) s \\ \text{mit} \;\; g(x^k) + \nabla g(x^k)^T s \le 0, \\ \qquad h(x^k) + \nabla h(x^k)^T s = 0, \end{cases}$$

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 8.2 : 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* (8.3) **then**

3 **return** (x^k, μ^k, λ^k)

4 Berechne $(x^{k+1}, \mu^{k+1}, \lambda^{k+1})$ als KKT-Punkt von

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n} \nabla f(x^k)^T (x - x^k) + \frac{1}{2}(x - x^k)^T \nabla^2_{xx} L(x^k, \mu^k, \lambda^k)(x - x^k) \\ \text{mit} \;\; g(x^k) + \nabla g(x^k)^T (x - x^k) \le 0, \\ \qquad h(x^k) + \nabla h(x^k)^T (x - x^k) = 0, \end{cases} \tag{8.4}$$

  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: Angenommen, zusätzlich zur Regularitätsbedingung gilt im KKT-Punkt $(\bar{x}, \bar{\mu}, \bar{\lambda})$ des nichtlinearen Problems die strikte Komplementarität, d. h. $\bar{\mu}_i = 0$ genau dann, wenn $g_i(\bar{x}) < 0$. Dann sind die KKT-Bedingungen für (8.3) äquivalent zum *Gleichungs*system

$$\begin{cases} \nabla_x L(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0, & \\ h_j(\bar{x}) = 0, & 1 \le j \le p, \\ g_i(\bar{x}) = 0, & i \in \mathcal{A}_X(\bar{x}), \\ \bar{\mu}_i = 0, & i \notin \mathcal{A}_X(\bar{x}). \end{cases}$$

Auf dieses System wird nun das Newton-Verfahren für nichtlineare Gleichungen angesetzt. Da die g_i stetig differenzierbar sind, ist $\mathcal{A}_X(x^k) = \mathcal{A}_X(\bar{x})$ für x^k nahe genug an $\bar{x}$; ebenso gilt $g_i(x^k) + \nabla g_i(x^k)^T(x^k - \bar{x}) < 0$ für alle $i \in \mathcal{A}_X(\bar{x})$ für x^k nahe genug an $\bar{x}$. In diesem Fall kann man wieder zeigen, dass der Newton-Schritt genau den KKT-Bedingungen für (8.4) entspricht. Schließlich zeigt man noch, dass die Newton-Matrix in $(\bar{x}, \bar{\mu}, \bar{\lambda})$ invertierbar ist.

Satz 8.3. *Sei $(\bar{x}, \bar{\mu}, \bar{\lambda})$ ein KKT-Punkt von (8.3), 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 $d \in K_X(\bar{x}, \bar{\mu})$).

Dann konvergiert für alle (x^0, μ^0, λ^0) hinreichend nahe an $(\bar{x}, \bar{\mu}, \bar{\lambda})$ die durch Algorithmus 8.2 erzeugte Folge superlinear.

Beweis. Wir gehen ähnlich wie im Multiplikator-Strafverfahren vor und definieren

$$H(x, \mu, \lambda) := \begin{pmatrix} \nabla_x L(x, \mu, \lambda) \\ h(x) \\ \min\{-g(x), \mu\} \end{pmatrix},$$

wobei die Komplementaritätsfunktion wieder komponentenweise zu verstehen ist. Dann ist $(\bar{x}, \bar{\mu}, \bar{\lambda})$ ein KKT-Punkt von (8.3) genau dann, wenn $H(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0$ gilt. Wir wenden nun auf H ein Newton-Verfahren an. 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

$$(x, \mu, \lambda) \in U_{\varepsilon_1}(\bar{x}, \bar{\mu}, \bar{\lambda}) := \left\{ (x, \mu, \lambda) : \|(x, \mu, \lambda) - (\bar{x}, \bar{\mu}, \bar{\lambda})\| \le \varepsilon_1 \right\}$$

auch gilt

$$-g_i(x) \begin{cases} < \mu_i & \text{für } i \in \mathcal{A}_X(\bar{x}), \\ > \mu_i & \text{für } i \notin \mathcal{A}_X(\bar{x}), \end{cases} \tag{8.5}$$

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 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. Nach Annahmen (i) und (ii) ist nun $H'(\bar{x}, \bar{\mu}, \bar{\lambda})$ invertierbar; 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 8.2 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 (8.5) ein $\varepsilon_3 < \varepsilon_1$ existiert so, dass $(x, \mu) = (x^k, \mu^k)$ ebenfalls (8.5) erfüllt sowie

$$-g_i(x^k) - \nabla g_i(x^k)^T (x - x^k) \begin{cases} < \mu_i^k & \text{für } i \in \mathcal{A}_X(\bar{x}), \\ > \mu_i^k & \text{für } i \notin \mathcal{A}_X(\bar{x}), \end{cases} \tag{8.6}$$

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

$$\begin{aligned} \nabla^2_{xx} L(x^k, \mu^k, \lambda^k)(x^{k+1} - x^k) + \nabla g(x^k)^T (\mu^{k+1} - \mu^k) & \\ + \nabla h(x^k)^T (\lambda^{k+1} - \lambda^k) &= -\nabla_x L(x^k, \mu^k, \lambda^k), \\ \nabla h(x^k)^T (x^{k+1} - x^k) &= -h(x^k), \\ \nabla g_i(x^k)^T (x^{k+1} - x^k) &= -g_i(x^k), && i \in \mathcal{A}_X(\bar{x}), \\ (\mu_i^{k+1} - \mu_i^k) &= -\mu_i^k, && i \notin \mathcal{A}_X(\bar{x}), \end{aligned}$$

wobei wir (8.5) für die rechte Seite der letzten beiden Gleichungen verwendet haben. Nach Definition der Lagrange-Funktion vereinfacht sich nun die erste Zeile zu

$$\nabla f(x^k) + \nabla^2_{xx} L(x^k, \mu^k, s^k)(x^{k+1} - x^k) + (\mu^{k+1})^T \nabla g(x^k) + (\lambda^{k+1})^T \nabla h(x^k) = 0,$$

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 (8.6) und damit insbesondere

$$-g_i(x^k) - \nabla g_i(x^k)^T(x - x^k) > \mu_i^{k+1} = 0 \qquad \text{für } i \notin \mathcal{A}_X(\bar{x}),$$

wobei wir die letzte Zeile,

$$\mu_i^{k+1} = 0 \qquad \text{für } i \notin \mathcal{A}_X(\bar{x}),$$

verwendet haben. Dies sind genau die KKT-Bedingungen für (8.4).

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 (8.4), d. h. erfüllt

$$\begin{aligned}
\nabla f(x^k) + \nabla^2_{xx} L(x^k, \mu^k, s^k)(\tilde{x} - x^k) + \tilde{\mu}^T \nabla g(x^k) + \tilde{\lambda}^T \nabla h(x^k) &= 0, \\
h(x^k) + \nabla h(x^k)^T(\tilde{x} - x^k) &= 0, \\
\min\{-g(x^k) - \nabla g(x^k)^T(\tilde{x} - x^k), \tilde{\mu}\} &= 0,
\end{aligned}$$

wobei wir wieder die Komplementaritätsfunktion verwendet haben. Nach Annahme erfüllen $\tilde{x}$ und $\tilde{\mu}$ ebenfalls (8.6), 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 $(x^{k+1}, \mu^{k+1}, \lambda^{k+1})$ also der nächste KKT-Punkt zu (x^k, μ^k, λ^k), da jeder weitere KKT-Punkt $(\tilde{x}, \tilde{\mu}, \tilde{\lambda}) \notin U_{3\varepsilon_3}(\bar{x}, \bar{\mu}, \bar{\lambda})$ erfüllt. □

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 (6.5) 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 (8.4) 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

$$\begin{aligned} g(x^k) + \nabla g(x^k)^T s &\leq \varepsilon, \\ h(x^k) + \nabla h(x^k)^T s &= \eta, \end{aligned}$$

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.

8.3 AKTIVE-MENGEN-STRATEGIE FÜR QUADRATISCHE PROBLEME

Für die Lösung der SQP-Teilprobleme (8.4) kann man analog zum Konvergenzbeweis die KKT-Bedingung als Gleichungssystem schreiben. Da die linearisierte aktive Menge

$$\mathcal{A}^k_{\text{lin}}(s^k) := \left\{ i : g_i(x^k) + \nabla g_i(x^k)^T s^k = 0 \right\}$$

nicht bekannt ist, geht man iterativ vor: Man wählt eine Startschätzung $\mathcal{A}^0 \subset \mathcal{A}^k_{\text{lin}}(\bar{s})$, 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 (8.4) 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*

$$\mathcal{A}^k := \left\{ i : a_i^T s^k + \alpha_i = 0 \right\}, \qquad \mathcal{I}^k := \{1,\dots,m\} \setminus \mathcal{A}^k.$$

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 Spalten 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 Spalten b_j^T sowie β für den Vektor mit Einträgen β_j, $j = 1, \dots, p$, so haben die KKT-Bedingungen die Form

$$\begin{pmatrix} H & A_k^T & B^T \\ A_k & 0 & 0 \\ B & 0 & 0 \end{pmatrix} \begin{pmatrix} s \\ \mu \\ \lambda \end{pmatrix} = \begin{pmatrix} -c \\ -\alpha_k \\ -\beta \end{pmatrix}. \tag{8.7}$$

Angenommen, wir haben nun eine Lösung $(s^{k+1}, \mu^{k+1}, \lambda^{k+1})$ von (8.7). Gilt dann $s^{k+1} = s^k$ und $\mu^{k+1} \geq 0$, so sind wir fertig: Es ist dann offensichtlich s^{k+1} zulässig für (QP), $\mathcal{A}^{k+1} = \mathcal{A}^k$, und $\bar{\mu}_i := \mu_i^{k+1}$, $i \in \mathcal{A}^{k+1}$, und 0 sonst, erfüllt 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, wählen wir naheliegenderweise diejenige, für die μ_i^{k+1} am kleinsten ist.)

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 $a_i^T d^k < 0$, 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

$$t_k := \min \left\{ -\frac{a_i^T s^k + \alpha_i}{a_i^T d^k} : i \in \mathcal{I}_k,\ a_i^T d^k > 0 \right\}. \tag{8.8}$$

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 8.3 : Aktive-Mengen-Strategie

Input : $s^0 \in \mathbb{R}^n$ zulässig für (QP)

Setze $\mathcal{A}^0 := \left\{ i : a_i^T s^0 + \alpha_i = 0 \right\}$
for $k = 0, \ldots$ **do**
  Berechne $(\tilde{s}^{k+1}, \mu^{k+1}, \lambda^{k+1})$ als Lösung von (8.7)
  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 $r = \arg\min \left\{ \mu_i^{k+1} : i \in \mathcal{A}^k,\ \mu_i^{k+1} < 0 \right\}$
      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 $r := \arg\min \left\{ -\frac{a_i^T s^{k+1} + \alpha_i}{a_i^T d^k} : i \in \mathcal{I}_k,\ a_i^T d^k > 0 \right\}$
      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 (8.7) berechnet werden kann. Dies kann man unter der folgenden Annahme garantieren.

Lemma 8.4. *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 (8.7) 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 (8.7) 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 $\mathcal{A}^k$) als auch $\tilde{s}^{k+1}$ (wegen der zweiten und dritten Gleichung von (8.7)) die Gleichungsnebenbedingungen in (QP_k) erfüllen. Aber dies ist ein Widerspruch. □

Wir können nun eine Konvergenzaussage für Algorithmus 8.3 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. Beachten Sie 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 8.5. *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 8.3:*

(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), 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 8.3 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 8.3 nicht nach endlich vielen Schritten abbricht (was unter den Voraussetzungen nach Lemma 8.4 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. □

Teil III

KONVEXE OPTIMIERUNG

9 KONVEXE UNTERHALBSTETIGE FUNKTIONEN

Wir haben bereits gesehen, dass Konvexität in Optimierungsproblemen besonders starke Aussagen erlaubt. Tatsächlich ist dies eine so starke Eigenschaft, dass sie sogar die Differenzierbarkeit ersetzen kann. Wir beschränken uns hier auf die Theorie im $\mathbb{R}^n$; Algorithmen und die (weitgehend analoge) Theorie in unendlichdimensionalen Räumen bleiben einer vertiefenden Vorlesung vorbehalten.

Sei also $U \subset \mathbb{R}^n$ eine konvexe Funktion und $F : U \to \mathbb{R}$ eine konvexe Menge, und betrachte das Problem

$$\min_{x \in U} F(x)$$

Es ist in Folge hilfreich, die Beschränkung $\bar{x} \in U$ in das Funktional aufzunehmen, indem wir F auf $\mathbb{R}^n$ erweitern, dafür aber den Wert ∞ zulassen. Wir betrachten also

$$\overline{F} : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}, \qquad \overline{F}(x) = \begin{cases} F(x) & x \in U, \\ \infty & x \in \mathbb{R}^n \setminus U. \end{cases}$$

Dabei wird $\overline{\mathbb{R}}$ mit der üblichen Arithmetik versehen, d. h. $t < \infty$ und $t + \infty = \infty$ für alle $t \in \mathbb{R}$; Subtraktion und Multiplikation von negativen Zahlen mit ∞ und insbesondere $F(x) = -\infty$ sind nicht zugelassen. Existiert überhaupt ein $x \in U$, so kann ein Minimierer $\bar{x}$ also nur in U liegen.

Wir betrachten also in Folge Funktionale $F : \mathbb{R}^n \to \overline{\mathbb{R}}$. Die Menge, auf der F endlich ist, bezeichnet man als (*effektiven*) *Definitionsbereich*

$$\operatorname{dom} F := \{x \in \mathbb{R}^n : F(x) < \infty\}.$$

Ist $\operatorname{dom} F \neq \emptyset$, so nennt man F *eigentlich*.

Offensichtlich kann eine Funktion, die den Wert ∞ annimmt, nicht stetig sein. Da wir nur an Minima interessiert sind, brauchen wir aber auch nur eine „einseitige“ Stetigkeit: Man nennt F *unterhalbstetig* in $x \in \mathbb{R}^n$, falls gilt

$$F(x) \leq \liminf_{n \to \infty} F(x_n) \quad \text{für alle Folgen } \{x_n\}_{n \in \mathbb{N}} \subset \mathbb{R}^n \text{ mit } x_n \to x.$$

Offensichtlich ist jede stetige Funktion auch unterhalbstetig; dies gilt insbesondere für

(i) jede Norm $\|\cdot\|$ auf $\mathbb{R}^n$;

(ii) lineare Funktionale der Form

$$F : \mathbb{R}^n \to \mathbb{R}, \qquad x \mapsto \langle x^*, x\rangle := (x^*)^T x,$$

für festes $x^* \in \mathbb{R}^n$.

Damit haben wir alle Begriffe zur Hand, um Satz 1.4 auf Funktionen mit Werten in $\overline{\mathbb{R}}$ zu erweitern. Der Beweis ist völlig analog, hier aber der Vollständigkeit halber komplett angegeben.

Satz 9.1. *Sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ eigentlich, koerziv und unterhalbstetig. Dann hat das Minimierungsproblem*

$$\min_{x\in\mathbb{R}^n} F(x)$$

eine Lösung $\bar{x} \in \operatorname{dom} F$.

Beweis. Der Beweis kann in drei Schritte aufgeteilt werden.

(i) *Zeige, dass eine Minimalfolge existiert.*

Da F eigentlich ist, ist $M := \inf_{x\in\mathbb{R}^n} F(x) < \infty$ (wobei $M = -\infty$ noch nicht ausgeschlossen ist). Wir können also eine Folge $\{y_n\}_{n\in\mathbb{N}} \subset \operatorname{ran} F \setminus \{\infty\} \subset \mathbb{R}$ finden mit $y_n \to M$, d. h. es existiert eine Folge $\{x_n\}_{n\in\mathbb{N}} \subset \mathbb{R}^n$ mit

$$F(x_n) \to M = \inf_{x\in\mathbb{R}^n} F(x).$$

Eine solche Folge wird *Minimalfolge* genannt. Beachten Sie, dass wir aus der Konvergenz von $\{F(x_n)\}_{n\in\mathbb{N}}$ (noch) nicht auf die Konvergenz von $\{x_n\}_{n\in\mathbb{N}}$ schließen können.

(ii) *Zeige, dass die Minimalfolge eine konvergente Teilfolge besitzt.*

Wir zeigen zuerst, dass $\{x_n\}_{n\in\mathbb{N}}$ beschränkt ist. Angenommen, das ist nicht der Fall, d. h. $\|x_n\| \to \infty$ für $n \to \infty$. Aus der Koerzivität von F folgt dann auch $F(x_n) \to \infty$, im Widerspruch zu $F(x_n) \to M < \infty$ nach Definition der Minimalfolge. Also ist $\{x_n\}_{n\in\mathbb{N}}$ beschränkt und enthält daher nach dem Satz von Heine–Borel eine konvergente Teilfolge $\{x_{n_k}\}_{k\in\mathbb{N}}$ mit Grenzwert $\bar{x} \in \mathbb{R}^n$. Dieser Grenzwert ist Kandidat für einen Minimierer.

(iii) *Zeige, dass dieser Grenzwert ein Minimierer ist.*

Aus der Definition der Minimalfolge folgt, dass auch für die Teilfolge $F(x_{n_k}) \to M$ gilt. Mit der Unterhalbstetigkeit von F und der Definition des Infimums erhalten wir daher

$$\inf_{x\in\mathbb{R}^n} F(x) \le F(\bar{x}) \le \liminf_{k\to\infty} F(x_{n_k}) = M = \inf_{x\in\mathbb{R}^n} F(x) < \infty.$$

Daraus folgt $\bar{x} \in \operatorname{dom} F$ sowie $\inf_{x\in\mathbb{R}^n} F(x) = F(\bar{x}) > -\infty$ (da F eigentlich ist). Das Infimum wird also in $\bar{x} \in \operatorname{dom} F$ angenommen, und damit ist $\bar{x}$ der gesuchte Minimierer. □

Ein weiterer Vorzug der Unterhalbstetigkeit ist, dass sie unter bestimmten Operationen erhalten bleibt.

Lemma 9.2. *Sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ unterhalbstetig. Dann sind unterhalbstetig*

(i) αF *für alle* $\alpha \geq 0$*;*

(ii) $F + G$ *für* $G : \mathbb{R}^n \to \overline{\mathbb{R}}$ *unterhalbstetig;*

(iii) $\varphi \circ F$ *für* $\varphi : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ *unterhalbstetig und monoton steigend;*

(iv) $F \circ \Phi$ *für* $\Phi : \mathbb{R}^m \to \mathbb{R}^n$ *stetig;*

(v) $x \mapsto \sup_{i \in I} F_i(x)$ *mit* $F_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ *unterhalbstetig für eine beliebige Menge* I*.*

Beachte, dass Aussage (v) für stetige Funktionen *nicht* gilt!

Beweis. Die Aussagen (i) und (ii) folgen direkt aus den Rechenregeln für den lim inf.

Für (iii) folgt zunächst aus der Unterhalbstetigkeit von F und der Monotonie von φ die Ungleichung

$$\varphi(F(x)) \leq \varphi(\liminf_{n \in \mathbb{N}} F(x_n)).$$

Auf die rechte Seite möchten wir nun die Unterhalbstetigkeit von φ anwenden; das ist aber nicht direkt möglich, da wir nicht voraussetzen können, dass $\{F(x_n)\}_{n \in \mathbb{N}}$ konvergiert. Wir gehen daher einen Umweg und betrachten zunächst die Teilfolge $\{\varphi(F(x_{n_k})\}_{k \in \mathbb{N}}$ für die gilt

$$\liminf_{n \to \infty} \varphi(F(x_n)) = \lim_{k \to \infty} \varphi(F(x_{n_k})).$$

Daraus extrahieren wir eine weitere Teilfolge, der Einfachheit mit k' indexiert, für die gilt $\liminf_{k \to \infty} F(x_{n_k}) = \lim_{k' \to \infty} F(x_{n_{k'}})$. Da der limes inferior einer Teilfolge nicht kleiner sein kann als der der gesamten Folge, erhalten wir aus der Monotonie und Unterhalbstetigkeit von φ

$$\varphi(\liminf_{n \to \infty} F(x_n)) \leq \varphi(\lim_{k' \to \infty} F(x_{n_{k'}})) \leq \liminf_{k' \to \infty} \varphi(F(x_{n_{k'}})) = \liminf_{n \to \infty} \varphi(F(x_n)),$$

wobei wir im letzten Schritt verwendet haben, dass alle Teilfolgen einer konvergenten Folge den gleichen Grenzwert besitzen.

Aussage (iv) folgt direkt aus der Stetigkeit von Φ: Gilt $y_n \to y$, so gilt auch $x_n := \Phi(y_n) \to \Phi(y) =: x$, und aus der Unterhalbstetigkeit von F folgt

$$F(\Phi(y)) \leq \liminf_{n \to \infty} F(\Phi(y_n)).$$

Sei schließlich $\{x_n\}_{n\in\mathbb{N}}$ eine konvergente Folge mit Grenzwert $x \in \mathbb{R}^n$. Dann gilt nach Definition des Supremums

$$F_j(x) \le \liminf_{n\to\infty} F_j(x_n) \le \liminf_{n\to\infty} \sup_{i\in I} F_i(x_n) \qquad \text{für alle } j \in I.$$

Nehmen wir auf beiden Seiten das Supremum über alle $j \in I$, so folgt die Aussage (v). □

Ein weiteres häufig auftretendes Beispiel für unstetige aber unterhalbstetige Funktionale ist die *Indikator-Funktion*[1] einer Menge $U \subset \mathbb{R}^n$, definiert als

$$\delta_U(x) = \begin{cases} 0 & \text{falls } x \in U, \\ \infty & \text{falls } x \in \mathbb{R}^n \setminus U. \end{cases}$$

Der Zweck dieser Definition ist natürlich, die Minimierung eines Funktionals $F : \mathbb{R}^n \to \mathbb{R}$ unter der Nebenbedingung $x \in U$ auf die unbeschränkte Minimierung von $\overline{F} := F + \delta_U$ zurückführen. Für die Existenz von Minimierern ist daher das folgende Resultat wichtig.

Lemma 9.3. *Sei $U \subset \mathbb{R}^n$. Dann ist $\delta_U : \mathbb{R}^n \to \overline{\mathbb{R}}$*

(i) eigentlich, wenn U nichtleer ist;

(ii) unterhalbstetig, wenn U abgeschlossen ist;

(iii) koerziv, wenn U beschränkt ist.

Beweis. Aussage (i) ist klar. Für (ii) betrachte eine konvergente Folge $\{x_n\}_{n\in\mathbb{N}} \subset \mathbb{R}^n$ mit Grenzwert $x \in \mathbb{R}^n$. Ist $x \in U$, dann ist wegen $\delta_U \ge 0$ natürlich

$$\delta_U(x) = 0 \le \liminf_{n\to\infty} \delta_U(x_n).$$

Sei nun $x \notin U$. Da U abgeschlossen ist, muss ein $N \in \mathbb{N}$ existieren mit $x_n \notin U$ für alle $n \ge N$ (sonst könnten wir – durch Übergang zu einer Teilfolge – eine Folge mit $x_n \to x \in U$ konstruieren, im Widerspruch zur Annahme). Also gilt $\delta_U(x_n) = \infty$ für alle $n \ge N$ und damit

$$\delta_U(x) = \infty = \liminf_{n\to\infty} \delta_U(x_n).$$

Für (iii) sei U beschränkt, d. h. es gebe ein $M > 0$ mit $U \subset K_M(0)$. Gilt $\|x_n\| \to \infty$, so existiert ein $N \in \mathbb{N}$ mit $\|x_n\| > M$ für alle $n \ge N$, und damit $x_n \notin K_M(0) \supset U$ für alle $n \ge N$. Also gilt auch $\delta_U(x_n) \to \infty$. □

[1] nicht zu verwechseln mit der *charakteristischen* Funktion $\mathbb{1}_U$ mit $\mathbb{1}_U(x) = 1$ für $x \in U$ und 0 sonst!

Wir betrachten nun konvexe Funktionen auf $\mathbb{R}^n$. Zur Erinnerung: ein eigentliches Funktional $F : X \to \overline{\mathbb{R}}$ heißt *konvex*, wenn für alle $x, y \in X$ und $\lambda \in [0,1]$ gilt

$$F(\lambda x + (1-\lambda)y) \leq \lambda F(x) + (1-\lambda)F(y) \tag{9.1}$$

(dabei ist der Funktionswert ∞ auf beiden Seiten zugelassen).

Eine alternative Charakterisierung der Konvexität eines Funktionals $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ basiert auf ihrem *Epigraph*

$$\operatorname{epi} F := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : F(x) \leq t\}.$$

Lemma 9.4. *Für $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ ist* epi F

(i) nichtleer genau dann, wenn F eigentlich ist;

(ii) konvex genau dann, wenn F konvex ist;

(iii) abgeschlossen genau dann, wenn F unterhalbstetig ist.

Beweis. Aussage (i) folgt direkt aus der Definition: F ist eigentlich genau dann, wenn ein $x \in \mathbb{R}^n$ und ein $t \in \mathbb{R}$ existiert mit $F(x) \leq t < \infty$, d. h. $(x,t) \in \operatorname{epi} F$.

Für (ii) sei F konvex und seien $(x,r), (y,s) \in \operatorname{epi} F$ gegeben. Für beliebige $\lambda \in [0,1]$ folgt dann aus (9.1), dass gilt

$$F(\lambda x + (1-\lambda)y) \leq \lambda F(x) + (1-\lambda)F(y) \leq \lambda r + (1-\lambda)s,$$

d. h. es ist

$$\lambda(x,r) + (1-\lambda)(y,s) = (\lambda x + (1-\lambda)y, \lambda r + (1-\lambda)s) \in \operatorname{epi} F$$

und damit epi F konvex. Sei umgekehrt epi F konvex und $x, y \in \mathbb{R}^n$ beliebig, wobei wir $F(x) < \infty$ und $F(y) < \infty$ annehmen können (ansonsten ist (9.1) trivialerweise erfüllt). Offensichtlich sind $(x, F(x)), (y, F(y)) \in \operatorname{epi} F$. Aus der Konvexität von epi F folgt dann, dass für alle $\lambda \in [0,1]$ gilt

$$(\lambda x + (1-\lambda)y, \lambda F(x) + (1-\lambda)F(y)) = \lambda(x, F(x)) + (1-\lambda)(y, F(y)) \in \operatorname{epi} F$$

und damit nach Definition von epi F auch (9.1).

Nun zu (iii): Sei zuerst F unterhalbstetig und $\{(x_n, t_n)\}_{n\in\mathbb{N}} \subset \operatorname{epi} F$ eine beliebige Folge mit $(x_n, t_n) \to (x,t) \in \mathbb{R}^n \times \mathbb{R}$. Dann gilt

$$F(x) \leq \liminf_{n\to\infty} F(x_n) \leq \limsup_{n\to\infty} t_n = t,$$

d. h $(x,t) \in \operatorname{epi} F$. Sei umgekehrt epi F abgeschlossen und angenommen, F ist eigentlich (sonst ist die Aussage trivial) und nicht unterhalbstetig. Dann existiert eine Folge $\{x_n\}_{n\in\mathbb{N}} \subset \mathbb{R}^n$ mit $x_n \to x \in \mathbb{R}^n$ und

$$F(x) > \liminf_{n\to\infty} F(x_n) =: M \in [-\infty, \infty).$$

Wir machen nun eine Fallunterscheidung:

a) $x \in \operatorname{dom} F$: Dann können wir eine Teilfolge auswählen, die wir wieder mit $\{x_n\}_{n\in\mathbb{N}}$ bezeichnen, so dass ein $\varepsilon > 0$ existiert mit $F(x_n) \leq F(x) - \varepsilon$ und damit $(x_n, F(x) - \varepsilon) \in \operatorname{epi} F$ für alle $n \in \mathbb{N}$. Wegen $x_n \to x$ folgt aus Abgeschlossenheit von $\operatorname{epi} F$ auch $(x, F(x) - \varepsilon) \in \operatorname{epi} F$ und damit $F(x) \leq F(x) - \varepsilon$, im Widerspruch zu $\varepsilon > 0$.

b) $x \notin \operatorname{dom} F$: In diesem Fall argumentiert man analog mit $F(x_n) \leq M + \varepsilon$ für $M > -\infty$ bzw. $F(x_n) \leq \varepsilon$ für $M = -\infty$, um einen Widerspruch zu $F(x) = \infty$ zu erhalten. □

Ebenfalls nützlich für die Betrachtung eines Funktionals $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ sind die zugehörigen *Subniveaumengen*

$$F_t := \{x \in \mathbb{R}^n : F(x) \leq t\}, \qquad t \in \mathbb{R},$$

für die man analog zu Lemma 9.4 die folgenden Eigenschaften zeigt.

Lemma 9.5. *Für $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ gilt:*

(i) Ist F konvex, so ist F_t konvex für alle $t \in \mathbb{R}$ (die Umkehrung gilt aber nicht);

(ii) F ist unterhalbstetig genau dann, wenn F_t abgeschlossen ist für alle $t \in \mathbb{R}$.

Direkt aus der Definition folgt die Konvexität

(i) *affiner Funktionale*, d. h. der Form $x \mapsto \langle x^*, x\rangle - \alpha$ für $x^* \in \mathbb{R}^n$ und $\alpha \in \mathbb{R}$;

(ii) jeder Norm $\|\cdot\|$ auf $\mathbb{R}^n$;

(iii) der Indikatorfunktion δ_C für eine konvexe Menge C.

Weitere Beispiele lassen sich analog zu Lemma 9.2 durch folgende Operationen erzeugen.

Lemma 9.6. *Sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ konvex. Dann sind konvex*

(i) αF für alle $\alpha \geq 0$;

(ii) $F + G$ für $G : \mathbb{R}^n \to \overline{\mathbb{R}}$ konvex (ist F oder G strikt konvex, so auch $F + G$);

(iii) $\varphi \circ F$ für $\varphi : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ konvex und monoton steigend;

(iv) $F \circ A$ für $A : \mathbb{R}^m \to \mathbb{R}^n$ linear;

(v) $x \mapsto \sup_{i\in I} F_i(x)$ mit $F_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ konvex für eine beliebige Menge I.

Nach Lemma 9.6 (v) ist also insbesondere das punktweise Supremum von affinen Funktionalen stets konvex. Tatsächlich lässt sich sogar jedes konvexe Funktional so darstellen. Dafür definieren wir für ein eigentliches Funktional $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ die *konvexe Hülle*

$$F^\Gamma : \mathbb{R}^n \to \overline{\mathbb{R}}, \qquad x \mapsto \sup\{a(x) : a \text{ affin mit } a(\tilde{x}) \leq F(\tilde{x}) \text{ für alle } \tilde{x} \in \mathbb{R}^n\}.$$

Satz 9.7. *Sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ eigentlich. Dann ist F genau dann konvex und unterhalbstetig, wenn gilt $F = F^\Gamma$.*

Beweis. Da affine Funktionale konvex und stetig sind, ist $F = F^\Gamma$ nach Lemma 9.6 (v) und Lemma 9.2 (v) immer konvex und unterhalbstetig.

Für die andere Richtung sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig. Aus der Definition von F^Γ als Supremum ist offensichtlich, dass stets punktweise $F^\Gamma \leq F$ gilt. Angenommen, $F^\Gamma < F$. Dann existiert ein Punkt $x_0 \in \mathbb{R}^n$ und ein $\lambda \in \mathbb{R}$ mit

$$F^\Gamma(x_0) < \lambda < F(x_0).$$

Wir konstruieren nun mit Hilfe des Hahn–Banach-Trennungssatzes ein affines Funktional $a : \mathbb{R}^n \to \mathbb{R}$ mit $a \leq F$ aber $a(x_0) > \lambda > F^\Gamma(x_0)$, was zusammen mit der Definition von F^Γ zum Widerspruch führt. Da F eigentlich, konvex und unterhalbstetig ist, ist epi F nach Lemma 9.4 nichtleer, konvex und abgeschlossen. Weiter ist $\{(x_0, \lambda)\}$ wegen $\lambda < F(x_0)$ kein Element in epi F. Der Satz 3.3 von Hahn–Banach liefert also die Existenz eines $z^* \in (\mathbb{R}^n \times \mathbb{R}) \setminus (0,0)$ und eines $\alpha \in \mathbb{R}$ mit

$$\langle z^*, (x,t)\rangle_{\mathbb{R}^n\times\mathbb{R}} \leq \alpha < \langle z^*, (x_0,\lambda)\rangle_{\mathbb{R}^n\times\mathbb{R}} \qquad \text{für alle } (x,t) \in \operatorname{epi} F.$$

Wir definieren nun ein $x^* \in \mathbb{R}^n$ durch $\langle x^*, x\rangle = \langle z^*, (x,0)\rangle_{\mathbb{R}^n\times\mathbb{R}}$ für alle $x \in \mathbb{R}^n$ und setzen $s := \langle z^*, (0,1)\rangle_{\mathbb{R}^n\times\mathbb{R}} \in \mathbb{R}$. Dann gilt $\langle z^*, (x,t)\rangle_{\mathbb{R}^n\times\mathbb{R}} = \langle x^*, x\rangle + st$ und damit

$$\langle x^*, x\rangle + st \leq \alpha < \langle x^*, x_0\rangle + s\lambda \qquad \text{für alle } (x,t) \in \operatorname{epi} F. \tag{9.2}$$

Nun ist für $(x,t) \in \operatorname{epi} F$ auch $(x,t') \in \operatorname{epi} F$ für alle $t' > t$, und aus der ersten Ungleichung in (9.2) folgt für alle $t' > 0$ groß genug

$$s \leq \frac{\alpha - \langle x^*, x\rangle}{t'} \to 0 \qquad \text{für } t' \to \infty.$$

Also ist $s \leq 0$. Wir machen nun eine Fallunterscheidung.

(i) $s < 0$: Wir setzen

$$a : \mathbb{R}^n \to \mathbb{R}, \qquad x \mapsto \frac{\alpha - \langle x^*, x\rangle}{s}.$$

Dann ist a affin. Für $x \in \operatorname{dom} F$ ist außerdem $(x, F(x)) \in \operatorname{epi} F$, und aus der produktiven Null und der ersten Ungleichung von (9.2) folgt (beachte $s < 0$!)

$$a(x) = \tfrac{1}{s}\left(\alpha - \langle x^*, x\rangle - sF(x)\right) + F(x) \leq F(x).$$

(Für $x \notin \operatorname{dom} F$ ist die Aussage klar.) Aus der zweiten Ungleichung von (9.2) folgt aber

$$a(x_0) = \tfrac{1}{s}\left(\alpha - \langle x^*, x_0\rangle\right) > \lambda.$$

(ii) $s = 0$: Dann folgt $\langle x^*, x\rangle \leq \alpha < \langle x^*, x_0\rangle$ für alle $x \in \operatorname{dom} F$, weshalb $x_0 \notin \operatorname{dom} F$ gelten muss. Allerdings ist F eigentlich, so dass ein $y_0 \in \operatorname{dom} F$ existiert, für das wir analog zu Fall (i) durch Trennung von $\operatorname{epi} F$ und (y_0, μ) für μ klein genug ein affines Funktional $a_0 : \mathbb{R}^n \to \mathbb{R}$ mit punktweise $a_0 \leq F$ konstruieren können. Wir setzen nun für $\rho > 0$

$$a_\rho : \mathbb{R}^n \to \mathbb{R}, \qquad x \mapsto a_0(x) + \rho\left(\langle x^*, x\rangle - \alpha\right).$$

Dann ist auch a_ρ affin, und wegen $\langle x^*, x\rangle \leq \alpha$ gilt $a_\rho(x) \leq a_0(x) \leq F(x)$ für alle $x \in \operatorname{dom} F$ und $\rho > 0$ beliebig. Wegen $\langle x^*, x_0\rangle > \alpha$ existiert aber ein $\rho > 0$ mit $a_\rho(x_0) > \lambda$.

In beiden Fällen muss nach Definition von F^Γ als Supremum also auch $F^\Gamma(x_0) > \lambda$ gelten, im Widerspruch zur Annahme $F^\Gamma(x_0) < \lambda$. □

Zum Abschluß dieses Kapitels zeigen wir, das folgende erstaunliche Resultat: *Jede (lokal) beschränkte konvexe Funktion ist (lokal) Lipschitz-stetig*. Neben seinem Nutzen in späteren Resultaten demonstriert dieses Resultat die Eleganz der konvexen Analysis: eine algebraische, globale Eigenschaft (Konvexität) verknüpft zwei topologische, lokale Eigenschaften (Umgebungen und Stetigkeit). Sei in Folge $B_\rho(x) := \{\tilde{x} \in \mathbb{R}^n : \|\tilde{x} - x\| < \rho\}$ für gegebene $x \in \mathbb{R}^n$ und $\rho > 0$. Zur Erinnerung: eine Funktion $F : \mathbb{R}^n \to \mathbb{R}$ heißt *lokal Lipschitz-stetig* in $x \in \mathbb{R}^n$ mit Konstante $L > 0$, wenn ein $\varepsilon > 0$ existiert mit

$$|F(x_1) - F(x_2)| \leq L\|x_1 - x_2\| \qquad \text{für alle } x_1, x_2 \in B_\varepsilon(x).$$

Satz 9.8. *Sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ konvex. Existiert für $x \in \mathbb{R}^n$ ein $\rho > 0$ so, dass F auf $B_\rho(x)$ nach oben beschränkt ist, so ist F in x lokal Lipschitz-stetig.*

Beweis. Nach Voraussetzung existiert ein $M \in \mathbb{R}$ mit $F(y) \leq M$ für alle $y \in B_\rho(x)$. Wir zeigen zunächst, dass F dann ebenfalls lokal nach unten beschränkt ist. Sei dafür $y \in B_\rho(x)$ beliebig. Dann ist wegen $\|x - y\| < \rho$ auch $z := 2x - y = x - (y - x) \in B_\rho(x)$, und aus der Konvexität von F folgt $F(x) = F\left(\frac{1}{2}y + \frac{1}{2}z\right) \leq \frac{1}{2}F(y) + \frac{1}{2}F(z)$ und damit

$$-F(y) \leq F(z) - 2F(x) \leq M - 2F(x) =: m,$$

d. h. $-m \leq F(y) \leq M$ für alle $y \in B_\rho(x)$.

Wir zeigen nun, dass daraus die Lipschitz-Stetigkeit auf $B_{\frac{\rho}{2}}(x)$ folgt. Seien $y_1, y_2 \in B_{\frac{\rho}{2}}(x)$ mit $y_1 \neq y_2$ und setze

$$z := y_1 + \frac{\rho}{2}\frac{y_1 - y_2}{\|y_1 - y_2\|} \in B_\rho(x)$$

wegen $\|z - x\| \leq \|y_1 - x\| + \frac{\rho}{2} < \rho$. Nach Konstruktion ist nun

$$y_1 = \lambda z + (1 - \lambda)y_2 \quad \text{für} \quad \lambda := \frac{\|y_1 - y_2\|}{\|y_1 - y_2\| + \frac{\rho}{2}} \in (0, 1),$$

so dass wegen der Konvexität von F gilt $F(y_1) \leq \lambda F(z) + (1-\lambda)F(y_2)$. Daraus folgt zusammen mit der Definition von λ sowie $F(z) \leq M$ und $-F(y_2) \leq m = M - 2F(x)$ die Abschätzung

$$\begin{aligned} F(y_1) - F(y_2) \leq \lambda(F(z) - F(y_2)) &\leq \lambda(2M - 2F(x)) \\ &= \frac{2(M-F(x))}{\|y_1 - y_2\| + \frac{\rho}{2}} \|y_1 - y_2\| \\ &\leq \frac{2(M-F(x))}{\rho/2} \|y_1 - y_2\|. \end{aligned}$$

Durch Vertauschung von y_1 und y_2 in der Konstruktion von z erhalten wir damit

$$|F(y_1) - F(y_2)| \leq \frac{2(M-F(x))}{\rho/2} \|y_1 - y_2\| \quad \text{für alle } y_1, y_2 \in B_{\frac{\rho}{2}}(x)$$

und damit die lokale Lipschitz-Stetigkeit mit Konstante $L(x, \rho/2) := 4(M - F(x))/\rho$. □

Daraus folgt die gewünschte Aussage aus rein geometrischen Überlegungen.

Satz 9.9. *Sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ konvex. Dann ist F lokal Lipschitz-stetig auf* $\operatorname{int}(\operatorname{dom} F)$.

Beweis. Sei $x \in \operatorname{int}(\operatorname{dom} F)$. Dann existiert ein $\rho > 0$ mit

$$B_\rho^\infty(x) = \{\tilde{x} \in \mathbb{R}^n : \|\tilde{x} - x\|_\infty < \rho\} \subset \operatorname{dom} F.$$

Sei nun $z := (z_1, \ldots, z_n)^T \in B_\rho^\infty(x)$. Dann können wir mit Hilfe der Einheitsvektoren e_i, $i = 1, \ldots n$, schreiben

$$z = \sum_{i=1}^n z_i e_i = \sum_{i=1}^n \frac{z_i - x_i}{\rho}(x + \rho e_i) + \left(1 - \sum_{i=1}^n \frac{z_i - x_i}{\rho}\right) x.$$

Aus der Konvexität von F folgt nun

$$\begin{aligned} F(z) &\leq \sum_{i=1}^n \frac{z_i - x_i}{\rho} F(x + \rho e_i) + \left(1 - \sum_{i=1}^n \frac{z_i - x_i}{\rho}\right) F(x) \\ &\leq \max\{F(x + \rho e_1), \ldots, F(x + \rho e_n), F(x)\} =: M. \end{aligned}$$

Also ist F nach oben beschränkt auf $B_\rho^\infty(x)$, und die Behauptung folgt aus Satz 9.8 – zunächst speziell bezüglich der Maximumsnorm; wegen der Äquivalenz aller Normen auf $\mathbb{R}^n$ dadurch aber auch bezüglich jeder beliebigen Norm mit gegebenfalls anderer Konstante. □

Wir werden im Laufe der Vorlesung noch weitere Gelegenheiten haben, das unerwartet schöne Verhalten von konvexen Funktionen auf dem Inneren ihres effektiven Definitionsbereichs zu beobachten.

10 DAS KONVEXE SUBDIFFERENTIAL

Wir wenden uns nun der Charakterisierung von Minimierern konvexer Funktionen durch ein Fermatsches Prinzip zu. Dafür benötigen wir einen Ableitungsbegriff, der einerseits allgemein genug ist, um auch für nichtdifferenzierbare Funktionen ein Fermatprinzip zu garantieren, aber andererseits konkret genug ist, um explizite Charakterisierungen und insbesondere Rechenregeln zu erlauben.

Unsere Motivation ist dabei geometrisch: Die klassische Ableitung $f'(t)$ einer skalaren Funktion $f : \mathbb{R} \to \mathbb{R}$ in t kann interpretiert werden als die Steigung der Tangente an f in $(t, f(t))$. Ist die Funktion nicht differenzierbar, existiert keine eindeutige (oder gar keine) Tangente mehr. Die Idee ist nun, in diesem Fall als verallgemeinerte Ableitung die *Menge aller* Tangentensteigungen (die auch leer sein kann!) zu verwenden. Dies führt direkt auf die folgende Definition.

Für $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ und $x \in \operatorname{dom} F$ ist

$$\partial F(x) := \{x^* \in \mathbb{R}^n : \langle x^*, \tilde{x} - x\rangle \leq F(\tilde{x}) - F(x) \quad \text{für alle } \tilde{x} \in \mathbb{R}^n\} \tag{10.1}$$

das *(konvexe) Subdifferential* von F in x. (Beachten Sie, dass $\tilde{x} \notin \operatorname{dom} F$ zugelassen ist, da dann die Ungleichung trivialerweise erfüllt ist.) Für $x \notin \operatorname{dom} F$ setzen wir $\partial F(x) = \emptyset$. Direkt aus der Definition folgt, dass $\partial F(x)$ konvex und abgeschlossen ist. Ein Element $\xi \in \partial F(x)$ heißt *Subgradient*.

Direkt aus der Definition erhalten wir nun ein Fermatsches Prinzip.

Satz 10.1. *Seien $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ und $\bar{x} \in \operatorname{dom} F$. Dann sind äquivalent:*

(i) $0 \in \partial F(\bar{x})$;

(ii) $F(\bar{x}) = \min_{x \in \mathbb{R}^n} F(x)$.

Beweis. Nach Definition ist $0 \in \partial F(\bar{x})$ genau dann, wenn gilt

$$0 = \langle 0, \tilde{x} - \bar{x}\rangle \leq F(\tilde{x}) - F(\bar{x}) \qquad \text{für alle } \tilde{x} \in \mathbb{R}^n,$$

d. h. $F(\bar{x}) \leq F(\tilde{x})$ für alle $\tilde{x} \in \mathbb{R}^n$. □

Dies entspricht auch der geometrischen Anschauung: Für $n = 1$ beschreibt $\tilde{y} := f(\tilde{x}) = f(x) + \xi(\tilde{x} - x)$ mit $\xi \in \partial f(x)$ eine Tangente an $y = f(x)$ mit Steigung ξ; die Bedingung $\xi = 0 \in \partial f(\tilde{x})$ bedeutet also, dass f in $\bar{x}$ eine waagerechte Tangente hat.[1]

Es ist nicht überraschend, dass für konvexe Funktionen stärkere Aussagen über das konvexe Subdifferential möglich sind. Ein zentrales Werkzeug dafür ist eine äquivalente Charakterisierung über Richtungsableitungen. Wir erinnern, dass für $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ die *Richtungsableitung* in $x \in \mathbb{R}^n$ in die feste(!) Richtung $h \in \mathbb{R}^n$ definiert ist als

$$F'(x; h) := \lim_{t \to 0^+} \frac{F(x + th) - F(x)}{t}.$$

Dieser Grenzwert existiert (zumindest in den erweiterten reellen Zahlen) für jede konvexe Funktion.

Lemma 10.2. *Sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ konvex und seien $x \in \operatorname{dom} F$ und $h \in \mathbb{R}^n$ gegeben. Dann gilt:*

(i) Die Funktion

$$\varphi : (0, \infty) \to \overline{\mathbb{R}}, \qquad t \mapsto \frac{F(x + th) - F(x)}{t},$$

ist monoton steigend.

(ii) Der Grenzwert $F'(x; h) = \lim_{t \to 0^+} \varphi(t) \in [-\infty, \infty]$ existiert und erfüllt

$$F'(x; h) \leq F(x + h) - F(x).$$

(iii) Gilt $x \in \operatorname{int}(\operatorname{dom} F)$, dann ist $F'(x; h) \in \mathbb{R}$.

Beweis. Zu (i): Durch Einsetzen und Umformen sieht man, dass für alle $0 < s < t$ die Bedingung $\varphi(s) \leq \varphi(t)$ äquivalent ist zu

$$F(x + sh) \leq \frac{s}{t} F(x + th) + \left(1 - \frac{s}{t}\right) F(x).$$

Dies folgt aber wegen $x + sh = (1 - \frac{s}{t})x + \frac{s}{t}(x + th)$ aus der Konvexität von F.

Aussage (ii) folgt nun sofort aus (i) wegen

$$F'(x; h) = \lim_{t \to 0^+} \varphi(t) = \inf_{t > 0} \varphi(t) \leq \varphi(1) = F(x + h) - F(x).$$

[1] Beachten Sie, dass in Satz 10.1 nirgendwo die Konvexität von F eingeht! Tatsächlich charakterisiert $0 \in \partial F(\bar{x})$ die *globalen* Minimierer jeder Funktion. Nichtkonvexe Funktionen können aber auch lokale Minimierer haben, für die die Subdifferentialinklusion nicht erfüllt ist. Tatsächlich sind (konvexe) Subdifferentiale nichtkonvexer Funktionen in der Regel leer. Dies führt insbesondere für den Beweis einiger Rechenregeln zu Problemen, für die wir in der Tat Konvexität voraussetzen müssen.

Zu (iii): Ist $x \in \operatorname{int}(\operatorname{dom} F)$, so existiert ein $\varepsilon > 0$ mit $x + th \in \operatorname{dom} F$ für alle $t \in (-\varepsilon, \varepsilon)$. Wie in (i) zeigt man $\varphi(s) \leq \varphi(t)$ für alle $s < t < 0$. Schließlich folgt aus $x = \frac{1}{2}(x+th) + \frac{1}{2}(x-th)$ für $t > 0$ mit der Konvexität von F auch

$$\varphi(-t) = \frac{F(x-th) - F(x)}{-t} \leq \frac{F(x+th) - F(x)}{t} = \varphi(t)$$

und damit die Monotonie auf ganz $\mathbb{R} \setminus \{0\}$. Wie in (ii) folgt aus der Wahl von $\varepsilon > 0$ nun

$$-\infty < \varphi(-\varepsilon) \leq F'(x;h) \leq \varphi(\varepsilon) < \infty. \qquad \square$$

Mit Hilfe dieser Eigenschaften können wir die alternative Charakterisierung zeigen.

Lemma 10.3. *Seien $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ konvex und $x \in \operatorname{dom} F$. Dann gilt*

$$\partial F(x) = \{x^* \in \mathbb{R}^n : \langle x^*, h\rangle \leq F'(x;h) \quad \textit{für alle } h \in \mathbb{R}^n\}.$$

Beweis. Da alle $\tilde{x} \in \mathbb{R}^n$ geschrieben werden können als $\tilde{x} = x + h$ für ein $h \in \mathbb{R}^n$ und umgekehrt, genügt zu zeigen, dass für $x^* \in \mathbb{R}^n$ die folgenden Aussagen äquivalent sind:

(i) $\langle x^*, h\rangle \leq F'(x;h)$ für alle $h \in \mathbb{R}^n$;

(ii) $\langle x^*, h\rangle \leq F(x+h) - F(x)$ für alle $h \in \mathbb{R}^n$.

Gilt (i), so folgt direkt aus Lemma 10.2 (ii), dass für alle $h \in \mathbb{R}^n$ gilt

$$\langle x^*, h\rangle \leq F'(x;h) \leq F(x+h) - F(x).$$

Gilt (ii) für alle $h \in \mathbb{R}^n$, so auch für th für alle $h \in \mathbb{R}^n$ und $t > 0$. Division durch t und Grenzübergang liefert dann

$$\langle x^*, h\rangle \leq \lim_{t \to 0^+} \frac{F(x+th) - F(x)}{t} = F'(x;h). \qquad \square$$

Wir betrachten einige Beispiele. Zunächst folgt aus Lemma 10.3, dass das Subdifferential die klassische Ableitung verallgemeinert.

Satz 10.4. *Sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ konvex und differenzierbar in x. Dann ist $\partial F(x) = \{\nabla F(x)\}$.*

Beweis. Ist F in $x \in \mathbb{R}^n$ differenzierbar, so existiert nach Definition die Richtungsableitung für alle $h \in \mathbb{R}^n$ und definiert eine lineare Abbildung $h \mapsto \nabla F(x)^T h$, d. h.

$$\langle \nabla F(x), h\rangle = \nabla F(x)^T h = F'(x;h) \quad \text{für alle } h \in \mathbb{R}^n.$$

Aus Lemma 10.3 folgt nun sofort $\nabla F(x) \in \partial F(x)$.

Umgekehrt folgt aus $\xi \in \partial F(x)$ wieder mit Lemma 10.3, dass gilt

$$\langle \xi, h\rangle \leq F'(x;h) = \langle \nabla F(x), h\rangle \quad \text{für alle } h \in \mathbb{R}^n.$$

Da $h \in \mathbb{R}^n$ beliebig war, ist dies nur für $\xi - \nabla F(x) = 0$ möglich. $\square$

Natürlich möchten wir auch Subdifferentiale von Funktionen haben, die nicht differenzierbar sind. Das kanonische Beispiel ist eine beliebige Norm $\|x\|$ auf $\mathbb{R}^n$, die ja in $x = 0$ nicht differenzierbar ist. Für das folgende Resultat benötigen wir die zugehörige *duale Norm* (oder *Operatornorm*) auf $\mathbb{R}^n$, definiert durch

$$\|x^*\|_* := \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\langle x^*, x\rangle}{\|x\|} = \sup_{\|x\|\le 1} \langle x^*, x\rangle.$$

Aus der linearen Algebra ist bekannt, dass gilt $\|x^*\|_* = \|x^*\|_q$ für $\|x\| = \|x\|_p$ mit $p^{-1}+q^{-1} = 1$ und damit insbesondere

(i) $\|x^*\|_* = \|x\|_2$ für $p = 2$;

(ii) $\|x^*\|_* = \|x^*\|_\infty$ für $p = 1$;

(iii) $\|x^*\|_* = \|x^*\|_1$ für $p = \infty$.

Satz 10.5. *Für $x \in \mathbb{R}^n$ ist*

$$\partial(\|\cdot\|)(x) = \begin{cases} \{x^* \in \mathbb{R}^n : \langle x^*, x\rangle = \|x\| \text{ und } \|x^*\|_* = 1\} & \text{falls } x \neq 0, \\ \{x^* \in \mathbb{R}^n : \|x^*\| \le 1\}_* & \text{falls } x = 0. \end{cases}$$

Beweis. Für $x = 0$ ist nach Definition $\xi \in \partial(\|\cdot\|)(x)$ genau dann, wenn gilt

$$\langle \xi, \tilde{x}\rangle \le \|\tilde{x}\| \qquad \text{für alle } \tilde{x} \in \mathbb{R}^n \setminus \{0\}$$

(für $\tilde{x} = 0$ ist die Ungleichung trivial). Dies ist aber wegen der Definition der Operatornorm äquivalent mit $\|\xi\|_* \le 1$.

Sei nun $x \neq 0$ und betrachte $\xi \in \partial(\|\cdot\|)(x)$. Indem wir nacheinander $\tilde{x} = 0$ und $\tilde{x} = 2x$ in die Definition (10.1) einsetzen, erhalten wir

$$\|x\| \le \langle \xi, x\rangle = \langle \xi, 2x - x\rangle \le \|2x\| - \|x\| = \|x\|,$$

d. h. $\langle \xi, x\rangle = \|x\|$. Analog haben wir für alle $\tilde{x} \in \mathbb{R}^n$, dass gilt

$$\langle \xi, \tilde{x}\rangle = \langle \xi, (\tilde{x} + x) - x\rangle \le \|\tilde{x} + x\| - \|x\| \le \|\tilde{x}\|,$$

woraus wie im Fall $x = 0$ folgt $\|\xi\|_* \le 1$. Für $\tilde{x} = x/\|x\|$ gilt nun

$$\langle \xi, \tilde{x}\rangle = \|x\|^{-1}\langle \xi, x\rangle = \|x\|^{-1}\|x\| = 1.$$

Also ist tatsächlich $\|\xi\|_* = 1$.

Es sei umgekehrt $x^* \in \mathbb{R}^n$ mit $\langle x^*, x\rangle = \|x\|$ und $\|x^*\|_* = 1$. Dann folgt für alle $\tilde{x} \in \mathbb{R}^n \setminus \{0\}$ aus der Definition der Operatornorm

$$1 = \|x^*\|_* = \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\langle x^*, x\rangle}{\|x\|} \ge \frac{\langle x^*, \tilde{x}\rangle}{\|\tilde{x}\|},$$

d. h. $\langle x^*, \tilde{x}\rangle \le \|\tilde{x}\|$. Da diese Ungleichung für $\tilde{x} = 0$ trivialerweise gilt, folgt für alle $\tilde{x} \in \mathbb{R}^n$

$$\langle x^*, \tilde{x} - x\rangle = \langle x^*, \tilde{x}\rangle - \langle x^*, x\rangle \le \|\tilde{x}\| - \|x\|,$$

und daher nach Definition $x^* \in \partial(\|\cdot\|)(x)$ □

Für den Fall $n = 1$ erhalten wir daraus das Subdifferential der Betragsfunktion als

$$\partial(|\cdot|)(t) = \operatorname{sign}(t) := \begin{cases} \{1\} & \text{falls } t > 0, \\ \{-1\} & \text{falls } t < 0, \\ [-1,1] & \text{falls } t = 0. \end{cases}$$

Schließlich haben wir auch eine einfache Darstellung für das Subdifferential der Indikatorfunktion einer konvexen Menge $C \subset \mathbb{R}^n$. Für $x \in C = \operatorname{dom} \delta_C$ gilt nämlich

$$\begin{aligned} x^* \in \partial\delta_C(x) &\Leftrightarrow \langle x^*, \tilde{x} - x\rangle \le \delta_C(\tilde{x}) && \text{für alle } \tilde{x} \in \mathbb{R}^n \\ &\Leftrightarrow \langle x^*, \tilde{x} - x\rangle \le 0 && \text{für alle } \tilde{x} \in C, \end{aligned}$$

da die Ungleichung für alle $\tilde{x} \notin C$ trivialerweise erfüllt ist. Die Menge $\partial\delta_C(x)$ nennt man auch *Normalenkegel* an C in x. Wieder ist das Beispiel $n = 1$ erhellend: Betrachte $C = [-1, 1]$ und $t \in C$. Dann ist $\xi \in \partial\delta_{[-1,1]}(t)$ genau dann, wenn $\xi(\tilde{t} - t) \le 0$ für alle $\tilde{t} \in [-1, 1]$ gilt. Wir machen nun eine Fallunterscheidung:

(i) $t = 1$. Dann ist $\tilde{t} - t \in [-2, 0]$ und damit gilt die Bedingung genau dann, wenn $\xi \ge 0$ ist.

(ii) $t = -1$. Dann ist $\tilde{t} - t \in [0, 2]$ und damit gilt die Bedingung genau dann, wenn $\xi \le 0$ ist.

(iii) $t \in (-1, 1)$. Dann kann $\tilde{t} - t$ sowohl positive als auch negative Werte annehmen, und damit muss $\xi = 0$ sein.

Also ist

$$\partial\delta_{[-1,1]}(t) = \begin{cases} [0, \infty) & \text{falls } t = 1, \\ (-\infty, 0] & \text{falls } t = -1, \\ \{0\} & \text{falls } t \in (-1, 1), \\ \emptyset & \text{falls } t \in \mathbb{R} \setminus [-1, 1]. \end{cases}$$

Dies entspricht den Komplementaritätsbedingungen für Lagrange-Multiplikatoren zu den Ungleichungen $-1 \le t \le 1$; vergleiche Satz 4.16.

Den Fall $n > 1$ kann man aus diesen Beispielen mit Hilfe des folgenden Resultats erhalten.

Satz 10.6. *Sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ konvex und* separabel, *d. h. $F(x) = \sum_{i=1}^n f_i(x_i)$ und $f_i : \mathbb{R} \to \overline{\mathbb{R}}$ konvex. Dann gilt für $x \in \operatorname{dom} F$*

$$\partial F(x) = \{\xi \in \mathbb{R}^n : \xi_i \in \partial f_i(x_i), \quad 1 \le i \le n\}.$$

Beweis. Ist $\xi_i \in \partial f_i(x_i)$ für alle $1 \le i \le n$, so erhalten wir durch Summation über (10.1) für alle i sofort für $\tilde{x} \in \mathbb{R}^n$ beliebig

$$\langle \xi, \tilde{x} - x \rangle = \sum_{i=1}^{n} \xi_i(\tilde{x}_i - x_i) \le \sum_{i=1}^{n} (f_i(\tilde{x}_i) - f_i(x_i)) = F(\tilde{x}) - F(x).$$

Sei umgekehrt $\xi \in \partial F(x)$ und wähle für $1 \le i \le n$ und $t \in \mathbb{R}$ beliebig

$$\tilde{x} := (\tilde{x}_1, \dots, \tilde{x}_n)^T, \qquad \tilde{x}_j := \begin{cases} t & \text{falls } j = i, \\ x_i & \text{falls } j \ne i. \end{cases}$$

Dann ist

$$\xi_i(t - x_i) = \langle \xi, \tilde{x} - x_i \rangle \le F(\tilde{x}) - F(x) = f_i(t) - f_i(x_i),$$

d. h. $\xi_i \in \partial f_i(x_i)$. □

Zusammen mit den obigen Beispielen erhält man daraus die Subdifferentiale der Norm $\|\cdot\|_1 = \sum_i |x_i|$ sowie der Indikatorfunktion $\partial \delta_{\|\cdot\|_\infty \le 1}(x) = \sum_i \partial_{[-1,1]}(x_i)$.

Subdifferentiale weiterer Funktionale erhält man durch Rechenregeln. Es ist naheliegend, dass diese umso aufwändiger zu beweisen sind, je schwächer der Differenzierbarkeitsbegriff ist (d. h. je mehr Funktionen in diesem Sinne differenzierbar sind). Die ersten beiden Regeln folgen noch direkt aus der Definition.

Lemma 10.7. *Für $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ konvex und $x \in \operatorname{dom} F$ gilt*

(i) $\partial(\lambda F)(x) = \lambda(\partial F(x)) := \{\lambda\xi : \xi \in \partial F(x)\}$ *für* $\lambda \ge 0$;

(ii) $\partial F(\cdot + x_0)(x) = \partial F(x + x_0)$ *für* $x_0 \in \mathbb{R}^n$ *mit* $x + x_0 \in \operatorname{dom} F$.

Schon die Summenregel ist deutlich aufwändiger. Wir benötigen dafür den Hahn–Banach-Trennungssatz in der folgenden Form, die als *Satz von Eidelheit* bekannt ist.

Folgerung 10.8. *Seien $A, B \subset \mathbb{R}^n$ konvex und nichtleer. Ist das Innere* $\operatorname{int} A$ *nichtleer und disjunkt zu B, dann existiert ein $x^* \in \mathbb{R}^n \setminus \{0\}$ und ein $\lambda \in \mathbb{R}$ mit*

$$\langle x^*, x_1 \rangle \le \lambda \le \langle x^*, x_2 \rangle \qquad \text{für alle } x_1 \in A, x_2 \in B. \tag{10.2}$$

Satz 10.9 (Summenregel). *Seien $F, G : \mathbb{R}^n \to \overline{\mathbb{R}}$ konvex. Dann gilt für alle $x \in \operatorname{dom} F \cap \operatorname{dom} G$*

$$\partial F(x) + \partial G(x) \subset \partial(F + G)(x).$$

Existiert ein $x_0 \in \operatorname{int}(\operatorname{dom} F) \cap \operatorname{dom} G$, so gilt Gleichheit.

Beweis. Die Inklusion folgt direkt aus der Definition des Subdifferentials durch Addition. Seien daher $x \in \operatorname{dom} F \cap \operatorname{dom} G$ und $\xi \in \partial(F+G)(x)$, erfüllen also

$$\langle \xi, \tilde{x} - x \rangle \leq (F(\tilde{x}) + G(\tilde{x})) - (F(x) + G(x)) \quad \text{für alle } \tilde{x} \in \mathbb{R}^n. \tag{10.3}$$

Unser Ziel ist nun, ähnlich wie im Beweis von Satz 9.7 mit Hilfe der Charakterisierung konvexer Funktionale durch ihren Epigraphen und des Trennungssatzes ein lineares Funktional $\zeta \in \partial G(x) \subset \mathbb{R}^n$ zu finden mit $\xi - \zeta \in \partial F(x)$, d. h.

$$\begin{aligned} F(\tilde{x}) - F(x) - \langle \xi, \tilde{x} - x \rangle &\geq \langle \zeta, x - \tilde{x} \rangle \quad \text{für alle } \tilde{x} \in \operatorname{dom} F, \\ G(x) - G(\tilde{x}) &\leq \langle \zeta, x - \tilde{x} \rangle \quad \text{für alle } \tilde{x} \in \operatorname{dom} G. \end{aligned}$$

Wir definieren dafür die Mengen

$$\begin{aligned} C_1 &:= \{(\tilde{x}, t - (F(x) - \langle \xi, x \rangle)) : \tilde{x} \in \operatorname{dom} F,\ t \geq F(\tilde{x}) - \langle \xi, \tilde{x} \rangle\}, \\ C_2 &:= \{(\tilde{x}, G(x) - t) : \tilde{x} \in \operatorname{dom} G,\ t \geq G(\tilde{x})\}, \end{aligned}$$

d. h.

$$C_1 = \operatorname{epi}(F - \xi) - (0, F(x) - \langle \xi, x \rangle), \qquad C_2 = -(\operatorname{epi} G - (0, G(x))).$$

Um auf diese Mengen Folgerung 10.8 anwenden zu können, müssen wir die Voraussetzungen nachweisen.

(i) Wegen $x \in \operatorname{dom} F \cap \operatorname{dom} G$ sind sowohl C_1 als auch C_2 nichtleer. Weiterhin sind F und G konvex, und damit ist es nicht schwer (wenn auch lästig) mit Hilfe der Definition nachzuweisen, dass C_1 und C_2 konvex sind.

(ii) Der Kernpunkt ist natürlich der Nachweis, dass $\operatorname{int} C_1$ nichtleer ist. Wegen $x_0 \in \operatorname{int}(\operatorname{dom} F)$ ist F nach Satz 9.9 beschränkt in einer offenen Kugel $U \subset \operatorname{int}(\operatorname{dom} F)$ um x_0. Wir können daher ein offenes Intervall $I \subset \mathbb{R}$ finden mit $U \times I \subset C_1$. Dann ist $U \times I$ offen nach Definition der Produkttopologie auf $\mathbb{R}^n \times \mathbb{R}$, und damit ist jedes Paar (x_0, α) mit $\alpha \in I$ ein innerer Punkt von C_1.

(iii) Es bleibt zu zeigen, dass $\operatorname{int} C_1 \cap C_2 = \emptyset$ gilt. Angenommen, es existiert ein $(\tilde{x}, \alpha) \in \operatorname{int} C_1 \cap C_2$. Aber dann folgt aus der Definition der Mengen dass gilt

$$F(\tilde{x}) - F(x) - \langle \xi, \tilde{x} - x \rangle < \alpha \leq G(x) - G(\tilde{x}),$$

im Widerspruch zu (10.3). Also sind $\operatorname{int} C_1$ und C_2 disjunkt

Folgerung 10.8 liefert also ein $(x^*, s) \in (\mathbb{R}^n \times \mathbb{R}) \setminus \{(0,0)\}$ und ein $\lambda \in \mathbb{R}$ mit

$$\langle x^*, \tilde{x} \rangle + s(t - (F(x) - \langle \xi, x \rangle)) \leq \lambda, \quad \text{für alle } \tilde{x} \in \operatorname{dom} F, t \geq F(\tilde{x}) - \langle \xi, \tilde{x} \rangle, \tag{10.4a}$$

$$\langle x^*, \tilde{x} \rangle + s(G(x) - t) \geq \lambda, \quad \text{für alle } \tilde{x} \in \operatorname{dom} G, t \geq G(\tilde{x}). \tag{10.4b}$$

Wir zeigen nun, dass $s < 0$ ist. Für $s = 0$ folgt mit $\tilde{x} = x_0 \in \operatorname{dom} F \cap \operatorname{dom} G$ sofort der Widerspruch

$$\langle x^*, x_0 \rangle < \lambda \leq \langle x^*, x_0 \rangle,$$

da (x_0, α) für α groß genug innerer Punkt von C_1 ist und daher nach Satz 3.3 die Trennung sogar mit strikter Ungleichung erfüllt ist. Gilt $s > 0$, so ist für $t > F(x) - \langle \xi, x \rangle$ die Klammer in (10.4a) positiv, und $t \to \infty$ mit $\tilde{x}$ fest führt zum Widerspruch zur Beschränktheit durch λ.

Also ist $s < 0$, und aus (10.4a) mit $t = F(\tilde{x}) - \langle \xi, \tilde{x} \rangle$ und aus (10.4b) mit $t = G(\tilde{x})$ folgt

$$F(\tilde{x}) - F(x) + \langle \xi, \tilde{x} - x \rangle \geq s^{-1}(\lambda - \langle x^*, \tilde{x} \rangle), \quad \text{für alle } \tilde{x} \in \operatorname{dom} F, \tag{10.5}$$

$$G(x) - G(\tilde{x}) \leq s^{-1}(\lambda - \langle x^*, \tilde{x} \rangle), \quad \text{für alle } \tilde{x} \in \operatorname{dom} G. \tag{10.6}$$

Setzen wir $\tilde{x} = x \in \operatorname{dom} F \cap \operatorname{dom} G$ in beiden Ungleichungen, so folgt sofort $\lambda = \langle x^*, x \rangle$. Damit ist $\zeta = s^{-1} x^*$ das gewünschte Funktional mit $(\xi - \zeta) \in \partial F(x)$ und $\zeta \in \partial G(x)$, d. h. $\xi \in \partial F(x) + \partial G(x)$. □

Daraus erhält man per Induktion Summenregeln für beliebig viele Summanden (wobei x_0 im Inneren aller bis auf eines effektiven Definitionsbereichs liegen muss). Man kann daraus auch eine Kettenregel für lineare Operatoren ableiten.

Satz 10.10 (Kettenregel). *Seien $A \in \mathbb{R}^{m \times n}$ und $F : \mathbb{R}^m \to \overline{\mathbb{R}}$ konvex und unterhalbstetig. Dann gilt für alle $x \in \operatorname{dom}(F \circ A)$*

$$\partial(F \circ A)(x) \supset A^* \partial F(Ax) := \{A^* y^* : y^* \in \partial F(Ax)\}.$$

Existiert ein $x_0 \in \mathbb{R}^n$ mit $Ax_0 \in \operatorname{int}(\operatorname{dom} F)$, so gilt Gleichheit.

Beweis. Die Inklusion folgt wieder direkt aus der Definition: Für $\eta \in \partial F(Ax) \subset \mathbb{R}^m$ gilt insbesondere für alle $\tilde{y} = A\tilde{x} \in \mathbb{R}^m$ mit $\tilde{x} \in \mathbb{R}^n$

$$F(A\tilde{x}) - F(Ax) \geq \langle \eta, A\tilde{x} - Ax \rangle = \langle A^* \eta, \tilde{x} - x \rangle,$$

d. h. $\xi := A^* \eta \in \partial(F \circ A) \subset \mathbb{R}^n$.

Sei nun $x \in \operatorname{dom}(F \circ A)$ und $\xi \in \partial(F \circ A)(x)$, d. h.

$$F(Ax) + \langle \xi, \tilde{x} - x \rangle \leq F(A\tilde{x}) \quad \text{für alle } \tilde{x} \in \mathbb{R}^n.$$

Wir konstruieren nun ein $\eta \in \partial F(Ax)$ mit $\xi = A^* \eta$ durch Anwenden der Summenregel auf

$$H : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}, \qquad H(x, y) := F(y) + \delta_{\operatorname{graph} A}(x, y).$$

Da A linear und stetig ist, ist $\operatorname{graph} A$ konvex und abgeschlossen und damit $\delta_{\operatorname{graph} A}$ konvex und unterhalbstetig. Weiter ist nach Annahme $Ax \in \operatorname{dom} F$ und damit $(x, Ax) \in \operatorname{dom} H$.

Zuerst zeigen wir, dass $\xi \in \partial(F \circ A)(x)$ genau dann gilt, wenn $(\xi, 0) \in \partial H(x, Ax)$ ist. Sei dafür $(\xi, 0) \in \partial H(x, Ax)$. Dann gilt für alle $\tilde{x} \in \mathbb{R}^n$, $\tilde{y} \in \mathbb{R}^m$

$$\langle \xi, \tilde{x} - x \rangle + \langle 0, \tilde{y} - Ax \rangle \leq F(\tilde{y}) - F(Ax) + \delta_{\operatorname{graph} A}(\tilde{x}, \tilde{y}) - \delta_{\operatorname{graph} A}(x, Ax).$$

Insbesondere gilt dies für alle $\tilde{y} \in \operatorname{ran}(A) = \{A\tilde{x} : \tilde{x} \in \mathbb{R}^n\}$. Wegen $\delta_{\operatorname{graph} A}(\tilde{x}, A\tilde{x}) = 0$ ist also

$$\langle \xi, \tilde{x} - x \rangle \leq F(A\tilde{x}) - F(Ax) \quad \text{für alle } \tilde{x} \in \mathbb{R}^n,$$

d. h. $\xi \in \partial(F \circ A)(x)$. Sei umgekehrt $\xi \in \partial(F \circ A)(x)$. Dann ist für alle $\tilde{x} \in \mathbb{R}^n$ und $\tilde{y} \in \mathbb{R}^m$ wegen $\delta_{\operatorname{graph} A}(x, Ax) = 0$ und $\delta_{\operatorname{graph} A}(\tilde{x}, \tilde{y}) \geq 0$

$$\begin{aligned}\langle \xi, \tilde{x} - x \rangle + \langle 0, \tilde{y} - Ax \rangle &= \langle \xi, \tilde{x} - x \rangle \\ &\leq F(A\tilde{x}) - F(Ax) + \delta_{\operatorname{graph} A}(\tilde{x}, \tilde{y}) - \delta_{\operatorname{graph} A}(x, Ax) \\ &= F(\tilde{y}) - F(Ax) + \delta_{\operatorname{graph} A}(\tilde{x}, \tilde{y}) - \delta_{\operatorname{graph} A}(x, Ax),\end{aligned}$$

da für $\tilde{y} \neq A\tilde{x}$ beide Seiten der letzten Gleichung unendlich sind. Also ist $(\xi, 0) \in \partial H(x, Ax)$.

Wir betrachten nun die „geliftete“ Funktion $\tilde{F} : \mathbb{R}^n \times \mathbb{R}^m$, $(x, y) \mapsto F(y)$ sowie $(x_0, Ax_0) \in \operatorname{graph} A = \operatorname{dom} \delta_{\operatorname{graph} A}$. Da nach Annahme $Ax_0 \in \operatorname{int}(\operatorname{dom} F)$ ist, ist auch $(x_0, Ax_0) \in \operatorname{int}(\operatorname{dom} \tilde{F}) = \mathbb{R}^n \times \operatorname{int}(\operatorname{dom} F) \subset \mathbb{R}^n \times \mathbb{R}^m$. Wir können daher Satz 10.9 anwenden und erhalten

$$(\xi, 0) \in \partial H(x, Ax) = \partial F(Ax) + \partial \delta_{\operatorname{graph} A}(x, Ax),$$

d. . $(\xi, 0) = (x^*, y^*) + (w^*, z^*)$ für ein $(x^*, y^*) \in \partial F(Ax)$ und ein $(w^*, z^*) \in \partial \delta_{\operatorname{graph} A}(x, Ax)$.

Nun „kollabieren“ wir diese Subdifferentiale wieder auf die einzelnen Komponenten, um die gewünschte Charakterisierung zu erhalten. Zuerst ist $(x^*, y^*) \in \partial \tilde{F}(Ax)$ genau dann, wenn gilt

$$\langle x^*, \tilde{x} - x \rangle + \langle y^*, \tilde{y} - Ax \rangle \leq F(\tilde{y}) - F(Ax) \quad \text{für alle } \tilde{x} \in \mathbb{R}^n, \tilde{y} \in \mathbb{R}^m.$$

Festhalten von $\tilde{x} = x$ bzw. $\tilde{y} = Ax$ liefert $y^* \in \partial F(Ax)$ und $x^* = 0$. Weiter ist $(w^*, z^*) \in \partial \delta_{\operatorname{graph} A}(x, Ax)$ genau dann, wenn gilt

$$\langle w^*, \tilde{x} - x \rangle + \langle z^*, \tilde{y} - Ax \rangle \leq 0 \quad \text{für alle } (\tilde{x}, \tilde{y}) \in \operatorname{graph} A,$$

d. h. für alle $\tilde{x} \in \mathbb{R}^n$ und $\tilde{y} = A\tilde{x}$. Also ist

$$\langle w^* + A^* z^*, \tilde{x} - x \rangle \leq 0 \quad \text{für alle } \tilde{x} \in \mathbb{R}^n$$

und damit $w^* = -A^* z^*$. Zusammen erhalten wir

$$(\xi, 0) = (0, y^*) + (-A^* z^*, z^*),$$

woraus $y^* = -z^*$ und daher $\xi = -A^* z^* = A^* y^*$ mit $y^* \in \partial F(Ax)$ folgt, was zu zeigen war. □

Eine Kettenregel gilt auch, wenn die *innere* Funktion nichtdifferenzierbar ist.

Satz 10.11. *Sei $F : \mathbb{R}^n \to \mathbb{R}$ konvex und sei $\varphi : \mathbb{R} \to \mathbb{R}$ konvex, monoton wachsend, und differenzierbar. Dann ist $\varphi \circ F$ konvex, und für alle $x \in \mathbb{R}^n$ gilt*

$$\partial[\varphi \circ F](x) = \varphi'(F(x))\partial F(x) = \{\varphi'(F(x))x^* : x^* \in \partial F(x)\}.$$

Beweis. Die Konvexität von $\varphi \circ F$ folgt aus Lemma 9.6 (iii). Für die Charakterisierung des Subdifferentials sei $x \in \mathbb{R}^n$ gegeben. Aus Satz 9.9 folgt dann, dass φ Lipschitz-stetig mit Konstante L um $F(x) \in \operatorname{int}(\operatorname{dom}\varphi) = \mathbb{R}$ ist. Also gilt für alle $h \in \mathbb{R}^n$

$$\begin{aligned}
(\varphi \circ F)'(x;h) &= \lim_{t\to 0^+} \frac{[\varphi \circ F](x+th) - [\varphi \circ F](x)}{t} \\
&= \lim_{t\to 0^+} \frac{\varphi(F(x+th)) - \varphi(F(x) + tF'(x;h))}{t} \\
&\quad + \lim_{t\to 0^+} \frac{\varphi(F(x) + tF'(x;h)) - \varphi(F(x))}{t} \\
&\le \lim_{t\to 0^+} L \left| \frac{F(x+th) - F(x)}{t} - F'(x;h) \right| + \varphi'(F(x);F'(x;h)) \\
&= \varphi'(F(x);F'(x;h)),
\end{aligned}$$

wobei wir im letzten Schritt die Richtungsdifferenzierbarkeit von F in $x \in \operatorname{int}(\operatorname{dom} F) = \mathbb{R}^n$ aus Lemma 10.2 verwendet haben. Analog zeigt man die umgekehrte Ungleichung unter Verwendung der Lipschitz-Stetigkeit in der Form $\varphi(t_1) - \varphi(t_2) \ge -L|t_1 - t_2|$. Also ist

$$[\varphi \circ F]'(x;h) = \varphi'(F(x);F'(x;h)) = \varphi'(F(x))F'(x;h)$$

wegen der Differenzierbarkeit von φ.

Damit ist nach Lemma 10.3

$$\partial(\varphi \circ F)(x) = \{z^* \in \mathbb{R}^n : \langle z^*, h\rangle \le \varphi'(F(x))F'(x;h) \text{ für alle } h \in \mathbb{R}^n\}.$$

Wegen der Monotonie und Differenzierbarkeit von $\varphi : \mathbb{R} \to \mathbb{R}$ ist nun $\varphi'(F(x)) \ge 0$ (vergleiche Satz 1.2). Ist $\varphi'(F(x)) \ne 0$, dann können wir $x^* := \varphi'(F(x))^{-1}z^*$ definieren; ansonsten ist $z^* = 0$ der einzige Subgradient. In beiden Fällen können wir äquivalent schreiben

$$\partial(\varphi \circ F)(x) = \{\varphi'(F(x))x^* : \langle x^*, h\rangle \le F'(x;h) \text{ für alle } h \in \mathbb{R}^n\},$$

und die Aussage folgt aus Lemma 10.3. □

Zusammenfassend erhalten wir eine Charakterisierung von Minimierern konvexer Funktionen unter (konvexen) Nebenbedingungen.

Folgerung 10.12. *Sei $U \subset \mathbb{R}^n$ nichtleer, konvex und abgeschlossen, und sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig. Existiert ein $x_0 \in \operatorname{int} U \cap \operatorname{dom} F$, so ist $\bar{x} \in U$ Lösung von*

$$\min_{x \in U} F(x)$$

genau dann, wenn ein $\xi \in \mathbb{R}^n$ existiert mit

$$\begin{cases} -\xi \in \partial F(\bar{x}), \\ \langle \xi, \tilde{x} - x \rangle \leq 0 \quad \text{für alle } \tilde{x} \in U. \end{cases} \tag{10.7}$$

Beweis. Aufgrund der Voraussetzungen an F und U können wir Satz 10.1 auf $J := F + \delta_U$ anwenden; und da $x_0 \in \operatorname{int} U = \operatorname{int}(\operatorname{dom} \delta_U)$ ist, können wir auch die Summenregel anwenden. Also hat F in $\bar{x}$ ein Minimum genau dann, wenn gilt

$$0 \in \partial J(\bar{x}) = \partial F(\bar{x}) + \partial \delta_U(\bar{x}).$$

Zusammen mit der Charakterisierung des Subdifferentials der Indikatorfunktion als Normalenkegel erhält man damit (10.7). □

Für eine differenzierbare Funktion $F : \mathbb{R}^n \to \mathbb{R}$ und $U = \{x \in \mathbb{R}^n : g(x) \leq 0\}$ ergibt (10.7) die *Karush–Kuhn–Tucker-Bedingungen* aus Satz 4.16; die Existenz des inneren Punktes $x_0 \in \operatorname{int} U$ entspricht dabei genau der *Slater-Bedingung* aus Folgerung 4.6.

11 FENCHEL-DUALITÄT

Ein Grund für die Nützlichkeit des konvexen Subdifferentials ist, wie wir sehen werden, seine Verbindung mit der Fenchel–Legendre-Transformation. Sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ eigentlich. Dann ist die *Fenchel-Konjugierte* zu F definiert als

$$F^* : \mathbb{R}^n \to \overline{\mathbb{R}}, \qquad F^*(x^*) = \sup_{x \in \mathbb{R}^n} \langle x^*, x \rangle - F(x).$$

(Da $\operatorname{dom} F \neq \emptyset$ angenommen ist, gilt $F^*(x^*) > -\infty$ für alle $x^* \in \mathbb{R}^n$, also ist die Definition sinnvoll.) Aus Lemma 9.6 (v) und Lemma 9.2 (v) folgt sofort, dass F^* für eigentliche F stets konvex und unterhalbstetig ist. Ist F nach unten durch ein affin-lineares Funktional beschränkt (was nach Satz 9.7 für konvexe und unterhalbstetige Funktionen immer der Fall ist), so ist auch F^* eigentlich. Die Definition liefert außerdem sofort die *Fenchel–Young-Ungleichung*

$$\langle x^*, x \rangle \leq F(x) + F^*(x^*) \qquad \text{für alle } x \in \mathbb{R}^n, x^* \in \mathbb{R}^n. \tag{11.1}$$

Anschaulich ist $F^*(x^*)$ der (negative) affine Anteil der Tangente an F (im Punkt x, in dem das Supremum angenommen wird) mit der Steigung x^*. Konjugieren wir ein weiteres Mal, erhalten wir die *Bikonjugierte* $F^{**} := (F^*)^*$, die selbst für nichtkonvexe und unterhalbstetige Funktionen wieder konvex und unterhalbstetig ist. Anschaulich ist F^{**} die konvexe Hülle von F, die für konvexe Funktionen nach Satz 9.7 ja mit F übereinstimmt.

Satz 11.1 (Fenchel–Moreau–Rockafellar). *Sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ eigentlich. Dann gilt*

(i) $F^{**} \leq F$;

(ii) $F^{**} = F^{\Gamma}$;

(iii) $F^{**} = F$ *genau dann, wenn F konvex und unterhalbstetig ist.*

Beweis. Für Aussage (i) nehmen wir in der Fenchel–Young-Ungleichung (11.1) das Supremum über alle $x^* \in \mathbb{R}^n$ und erhalten

$$F(x) \geq \sup_{x^* \in \mathbb{R}^n} \langle x^*, x \rangle - F^*(x^*) = F^{**}(x).$$

Für (ii) stellen wir zuerst fest, dass F^{**} nach Definition der Fenchel-Konjugierten konvex und unterhalbstetig und wegen (i) auch eigentlich ist. Also gilt nach Satz 9.7

$$F^{**}(x) = (F^{**})^{\Gamma}(x) = \sup\{a(x) : a : \mathbb{R}^n \to \mathbb{R} \text{ affin mit } a \leq F^{**}\}.$$

Wir zeigen nun, dass wir auf der rechten Seite F^{**} durch F ersetzen können. Sei dafür $a(x) = \langle x^*, x\rangle - \alpha$ für $x^* \in \mathbb{R}^n$ und $\alpha \in \mathbb{R}$ beliebig. Gilt $a \leq F^{**}$, so folgt aus (i) sofort $a \leq F$. Gilt umgekehrt $a \leq F$, so ist $\langle x^*, x\rangle - F(x) \leq \alpha$ für alle $x \in \mathbb{R}^n$, und durch Supremum über alle $x \in \mathbb{R}^n$ erhalten wir $\alpha \geq F^*(x^*)$. Daraus folgt nun nach Definition von F^{**}

$$a(x) = \langle x^*, x\rangle - \alpha \leq \langle x^*, x\rangle - F^*(x^*) \leq F^{**}(x) \quad \text{für alle } x \in \mathbb{R}^n,$$

d. h. $a \leq F^{**}$.

Aussage (iii) folgt nun sofort aus (ii) und Satz 9.7. □

Wir betrachten wieder relevante Beispiele.

Satz 11.2. *Sei $B := \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ die Einheitskugel bezüglich der Norm $\|\cdot\|$ und setze $F = \delta_B$. Dann ist $F^*(x^*) = \|x^*\|_*$ für die zugehörige duale Norm.*

Beweis. Nach Definition der dualen Norm gilt für alle $x^* \in \mathbb{R}^n$

$$(\delta_B)^*(x^*) = \sup_{x\in\mathbb{R}^n} \langle x^*, x\rangle - \delta_B(x) = \sup_{\|x\|\leq 1} \langle x^*, x\rangle = \|x^*\|_*. \qquad \square$$

Satz 11.3. *Sei $F(x) = \|x\|$ und $\|\cdot\|_*$ die zugehörige duale Norm mit Einheitskugel $B^* := \{x^* \in \mathbb{R}^n : \|x^*\|_* \leq 1\}$. Dann ist $F^*(x^*) = \partial\delta_{B^*}(x^*)$.*

Beweis. Für $x^* \in \mathbb{R}^n$ machen wir die Fallunterscheidung

(i) $\|x^*\|_* \leq 1$. Dann folgt wie im Beweis von Satz 10.5 für alle $x \in \mathbb{R}^n$, dass gilt $\langle x^*, x\rangle - \|x\| \leq 0$. Weiterhin ist $\langle x^*, 0\rangle = 0 = \|0\|$. Also gilt

$$F^*(x^*) = \sup_{x\in\mathbb{R}^n} \langle x^*, x\rangle - \|x\| = 0.$$

(ii) $\|x^*\|_* > 1$. Nach Definition der dualen Norm existiert dann ein $x_0 \in \mathbb{R}^n$ mit $\langle x^*, x_0\rangle > \|x_0\|$. Lassen wir daher $t \to \infty$ gehen in

$$0 < t(\langle x^*, x_0\rangle - \|x_0\|) = \langle x^*, tx_0\rangle - \|tx_0\| \leq F^*(x^*),$$

so erhalten wir $F^*(x^*) = \infty$. □

Analog zu Satz 10.6 kann man zeigen, dass separable Funktionen komponentenweise konjugiert werden können.

Satz 11.4. *Sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ eigentlich und separabel, d. h. $F(x) = \sum_{i=1}^n f_i(x_i)$ für $f_i : \mathbb{R} \to \overline{\mathbb{R}}$ eigentlich. Dann ist*

$$F^*(x^*) = \sum_{i=1}^n f_i^*(x_i^*) \qquad \text{für alle } x^* = (x_1^*, \dots, x_n^*)^T \in \mathbb{R}^n.$$

Beweis. Direkt aus der Definition folgt

$$\begin{aligned} F^*(x^*) &= \sup_{x\in\mathbb{R}^n} \langle x^*, x\rangle - F(x) = \sup_{x\in\mathbb{R}^n} \sum_{i=1}^n x_i^* x_i - f_i(x_i) = \sum_{i=1}^n \sup_{x_i\in\mathbb{R}} \left[x_i^* x_i - f_i(x_i)\right] \\ &= \sum_{i=1}^n f_i^*(x_i^*), \end{aligned}$$

da wegen der Separabilität jeder Summand separat maximiert werden kann. □

Wir notieren noch einige nützliche Rechenregeln.

Lemma 11.5. *Sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ eigentlich. Dann ist*

(i) $(\alpha F)^* = \alpha F^* \circ (\alpha^{-1}\,\mathrm{Id})$ *für $\alpha > 0$;*

(ii) $(F(\cdot + x_0) + \langle x_0^*, \cdot\rangle)^* = F^*(\cdot - x_0^*) - \langle \cdot - x_0^*, x_0\rangle$ *für alle $x_0, x_0^* \in \mathbb{R}^n$;*

(iii) $(F \circ A)^* = F^* \circ A^{-*}$ *für $A \in \mathbb{R}^{n\times m}$ invertierbar und $A^{-*} := (A^{-1})^*$.*

Beweis. Die Regeln folgen direkt aus den Eigenschaften des Supremums.

Aussage (i) gilt wegen $\alpha > 0$ und

$$(\alpha F)^*(x^*) = \sup_{x\in\mathbb{R}^n} \left(\alpha\langle \alpha^{-1}x^*, x\rangle - \alpha F(x)\right) = \alpha \sup_{x\in\mathbb{R}^n} \left(\langle \alpha^{-1}x^*, x\rangle - F(x)\right) = \alpha F^*(\alpha^{-1}x^*).$$

Aussage (ii) gilt wegen $\{x + x_0 : x \in \mathbb{R}^n\} = \mathbb{R}^n$ und

$$\begin{aligned} (F(\cdot + x_0) + \langle x_0^*, \cdot\rangle)^*(x^*) &= \sup_{x\in\mathbb{R}^n} \langle x^*, x\rangle - F(x + x_0) - \langle x_0^*, x\rangle \\ &= \sup_{x\in\mathbb{R}^n} \left(\langle x^* - x_0^*, x + x_0\rangle - F(x + x_0)\right) - \langle x^* - x_0^*, x_0\rangle \\ &= \sup_{\tilde{x}=x+x_0, x\in\mathbb{R}^n} \left(\langle x^* - x_0^*, \tilde{x}\rangle - F(\tilde{x})\right) - \langle x^* - x_0^*, x_0\rangle \\ &= F^*(x^* - x_0^*) - \langle x^* - x_0^*, x_0\rangle. \end{aligned}$$

Aussage (iii) gilt wegen $\operatorname{ran} A = \mathbb{R}^n$ und

$$\begin{aligned} (F \circ A)^*(y^*) &= \sup_{y\in\mathbb{R}^n} \langle y^*, A^{-1}Ay\rangle - F(Ay) \\ &= \sup_{x=Ay, y\in\mathbb{R}^n} \langle A^{-*}y^*, x\rangle - F(x) = F^*(A^{-*}y^*). \end{aligned}$$

□

Die Definition der Fenchel-Konjugierten ist auf besondere Weise verträglich mit der des Subdifferentials.

Satz 11.6. *Sei $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig. Dann sind äquivalent für $x \in \mathbb{R}^n$ und $x^* \in \mathbb{R}^n$:*

(i) $\langle x^*, x\rangle = F(x) + F^*(x^*)$;

(ii) $x^* \in \partial F(x)$;

(iii) $x \in \partial F^*(x^*)$.

Beweis. Gilt (i), so folgt aus der Definition von F^* als Supremum

$$\langle x^*, x\rangle - F(x) = F^*(x^*) \geq \langle x^*, \tilde{x}\rangle - F(\tilde{x}) \qquad \text{für alle } \tilde{x} \in \mathbb{R}^n, \tag{11.2}$$

was nach Definition äquivalent ist zu $x^* \in \partial F(x)$. Umgekehrt ergibt Supremum über alle $\tilde{x} \in \mathbb{R}^n$ auf beiden Seiten von (11.2)

$$\langle x^*, x\rangle \geq F(x) + F^*(x^*),$$

und zusammen mit der Fenchel–Young-Ungleichung (11.1) folgt (i).

Analog erhält man aus (i) zusammen mit Satz 11.1 für alle $\tilde{x}^* \in \mathbb{R}^n$ die Ungleichung

$$\langle x^*, x\rangle - F^*(x^*) = F(x) = F^{**}(x) \geq \langle \tilde{x}^*, x\rangle - F^*(\tilde{x}^*),$$

woraus wie oben die Äquivalenz von (i) und (iii) folgt. □

Satz 11.6 spielt die Rolle des „Satzes von der konvexen Umkehrfunktion“. Damit kann man insbesondere das Subdifferential einer komplizierten Norm durch das (einfachere) der konjugierten Indikatorfunktion ersetzen. Hat man zum Beispiel ein Problem der Form

$$\inf_{x\in\mathbb{R}^n} F(x) + G(Ax) \tag{P}$$

für $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ und $G : \mathbb{R}^m \to \overline{\mathbb{R}}$ eigentlich, konvex, und unterhalbstetig sowie $A \in \mathbb{R}^{m\times n}$, so können wir G mit Hilfe von Satz 11.1 ersetzen durch die Definition von G^{**} und erhalten

$$\inf_{x\in\mathbb{R}^n} \sup_{y^*\in\mathbb{R}^m} F(x) + \langle y^*, Ax\rangle - G^*(y^*). \tag{11.3}$$

Dürften wir nun inf und sup vertauschen, so könnten wir schreiben (mit $\inf F = -\sup(-F)$)

$$\begin{aligned}\inf_{x\in\mathbb{R}^n} \sup_{y^*\in y^*} F(x) + \langle y^*, Ax\rangle - G^*(y^*) &= \sup_{y^*\in\mathbb{R}^m} \inf_{x\in\mathbb{R}^n} F(x) + \langle y^*, Ax\rangle - G^*(y^*)\\ &= \sup_{y^*\in\mathbb{R}^m} -\left(\sup_{x\in\mathbb{R}^n} -F(x) + \langle -A^*y^*, x\rangle\right) - G^*(y^*).\end{aligned}$$

Einsetzen der Definition von F^* ergibt dann das *duale Problem*

$$\sup_{y^* \in \mathbb{R}^m} -F^*(-A^* y^*) - G^*(y^*). \tag{D}$$

Als Nebeneffekt haben wir den Operator A zwischen den Funktionalen verschoben.

Der folgende Satz nutzt auf elegante Weise das Fermat-Prinzip, Summen- und Kettenregel, sowie die Fenchel–Young-Gleichung, um hinreichende Bedingungen für die Vertauschbarkeit von inf und sup zu geben.

Satz 11.7 (Fenchel–Rockafellar). *Seien $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ und $G : \mathbb{R}^m \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig und sei $A \in \mathbb{R}^{m \times n}$. Gelte weiterhin:*

(i) das primale Problem (P) *hat eine Lösung $\bar{x} \in \mathbb{R}^n$;*

(ii) es existiert ein $x_0 \in \operatorname{dom} F \cap \operatorname{dom}(G \circ A)$ mit $Ax_0 \in \operatorname{int}(\operatorname{dom} G)$.

Dann hat das duale Problem (D) eine Lösung $\bar{y}^* \in \mathbb{R}^m$ und es gilt

$$\min_{x \in \mathbb{R}^n} F(x) + G(Ax) = \max_{y^* \in \mathbb{R}^m} -F^*(-A^* y^*) - G^*(y^*). \tag{11.4}$$

Weiterhin sind $\bar{x}$ und $\bar{y}^*$ Lösungen von (P) bzw. (D) genau dann, wenn die Fenchel-Extremalitätsbedingungen

$$\begin{cases} -A^* \bar{y}^* \in \partial F(\bar{x}), \\ \quad \bar{y}^* \in \partial G(A\bar{x}), \end{cases} \tag{E}$$

erfüllt sind.

Beweis. Sei zuerst $\bar{x} \in \mathbb{R}^n$ eine Lösung von (P). Wegen Voraussetzung (ii) sind Satz 10.9 (wegen $x^0 \in \operatorname{int} \operatorname{dom}(G \circ A)$ aufgrund der Linearität von A) und Satz 10.10 anwendbar; aus Satz 10.1 folgt daher

$$0 \in \partial(F + G \circ A)(\bar{x}) = \partial F(\bar{x}) + A^* \partial G(A\bar{x})$$

und damit die Existenz eines $\bar{y}^* \in \partial G(A\bar{x})$ mit $-A^* \bar{y}^* \in \partial F(\bar{x})$, d. h. (E) ist erfüllt.

Es gelte umgekehrt (E) für $\bar{x} \in \mathbb{R}^n$ und $\bar{y}^* \in \mathbb{R}^m$. Dann ist wiederum aufgrund der Sätze 10.1, 10.9 und 10.10 $\bar{x}$ Lösung von (P). Weiter folgt aus (E) mit Satz 11.6 die Gleichheit in den Fenchel–Youngschen Ungleichungen für F und G, d. h.

$$\begin{cases} \langle -A^* \bar{y}^*, \bar{x} \rangle = F(\bar{x}) + F^*(-A^* \bar{y}^*), \\ \quad \langle \bar{y}^*, A\bar{x} \rangle = G(A\bar{x}) + G^*(\bar{y}^*). \end{cases} \tag{11.5}$$

Durch Summieren beider Gleichungen erhalten wir

$$F(\bar{x}) + G(A\bar{x}) = -F^*(-A^* \bar{y}^*) - G^*(\bar{y}^*). \tag{11.6}$$

Es bleibt zu zeigen, dass $\bar{y}^*$ Lösung von (D) ist. Setze dafür

$$L : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}, \qquad L(x, y^*) = F(x) + \langle y^*, Ax\rangle - G^*(y^*).$$

Dann gilt für alle $\tilde{x} \in \mathbb{R}^n$ und $\tilde{y}^* \in \mathbb{R}^m$ stets

$$\sup_{y^* \in \mathbb{R}^m} L(\tilde{x}, y^*) \geq L(\tilde{x}, \tilde{y}^*) \geq \inf_{x \in \mathbb{R}^n} L(x, \tilde{y}^*),$$

und damit (Infimum über alle $\tilde{x}$ in der ersten und Supremum über alle $\tilde{y}^*$ in der zweiten Ungleichung)

$$\inf_{x \in \mathbb{R}^n} \sup_{y^* \in \mathbb{R}^m} L(x, y^*) \geq \sup_{y^* \in \mathbb{R}^m} \inf_{x \in \mathbb{R}^n} L(x, y^*).$$

Also ist

$$\begin{aligned} F(\bar{x}) + G(A\bar{x}) &= \inf_{x \in \mathbb{R}^n} \sup_{y^* \in \mathbb{R}^m} F(x) + \langle y^*, Ax\rangle - G^*(y^*) \\ &\geq \sup_{y^* \in \mathbb{R}^m} \inf_{x \in \mathbb{R}^n} F(x) + \langle y^*, Ax\rangle - G^*(y^*) \\ &= \sup_{y^* \in \mathbb{R}^m} -F^*(-A^* y^*) - G^*(y^*). \end{aligned}$$

Zusammen mit der Gleichung (11.6) erhalten wir damit

$$-F^*(-A^*\bar{y}^*) - G(\bar{y}^*) = F(\bar{x}) + G(A\bar{x}) \geq \sup_{y^* \in \mathbb{R}^m} -F^*(-A^* y^*) - G^*(y^*),$$

d. h. $\bar{y}^*$ ist Lösung von (D), woraus insbesondere die behauptete Existenz einer Lösung folgt.

Da alle Lösungen von (D) nach Definition den selben (maximalen) Funktionalwert haben, folgt aus (11.6) auch (11.4).

Sind schließlich $\bar{x} \in \mathbb{R}^n$ und $\bar{y}^* \in \mathbb{R}^m$ Lösungen von (P) bzw. (D), so folgt aus der gerade gezeigten starken Dualität die Gleichung (11.6). Zusammen mit der produktiven Null erhalten wir daraus

$$0 = \left[F(\bar{x}) + F^*(-A^*\bar{y}^*) - \langle -A^*\bar{y}^*, \bar{x}\rangle\right] + \left[G(A\bar{x}) + G^*(\bar{y}^*) - \langle \bar{y}^*, A\bar{x}\rangle\right].$$

Da beide Klammern wegen der Fenchel–Young-Ungleichung nicht-negativ sind, müssen sie einzeln verschwinden. Also gilt (11.5), und aus Satz 11.6 folgt (E). □

Mit Hilfe von Satz 11.6 können wir aus den Fenchel-Extremalitätsbedingungen weitere äquivalente Optimalitätsbedingungen erzeugen, indem wir eine oder beide Subdifferentialinklusionen invertieren. Dies kann man ausnutzen, um praktisch durchführbare Algorithmen für die Lösung von Optimierungsproblemen dieser Form herzuleiten. Der Satz erlaubt auch, den Subgradienten $\bar{y}^*$ aus der Kettenregel als Minimierer eines konvexen Funktionals zu charakterisieren. Ist zum Beispiel F^* oder G^* strikt konvex, so ist $\bar{y}^*$ eindeutig; unter dieser Voraussetzung sind oft stärkere Aussagen über die Stabilität von Minimierern von (P) oder die Konvergenz von Verfahren zu deren Berechnung möglich.

12 SUBGRADIENTENBASIERTE VERFAHREN

Wir betrachten zum Abschluss zwei einfache Verfahren für nichtdifferenzierbare konvexe Optimierungsprobleme, die auf der Verwendung von Subgradienten anstelle von Gradienten basieren. Wir werden sehen, dass dies funktioniert, wenn auch nur mit Einschränkungen. Anders als in der Theorie ist es hier vorteilhafter, etwaige Nebenbedingungen separat zu behandeln. Wir betrachten daher für eine nichtleere, konvexe, und abgeschlossene Menge $X \subset \mathbb{R}^n$ und eine konvexe unterhalbstetige Funktion $F : \mathbb{R}^n \to \mathbb{R}$ das Problem

$$\text{(P)} \qquad \min_{x \in X} F(x),$$

von dem wir wie üblich annehmen, dass es eine Lösung $\bar{x} \in X$ hat.

12.1 SUBGRADIENTENVERFAHREN

Die erste Idee ist, im Gradientenverfahren den Gradienten einfach durch einen (beliebigen!) Subgradienten zu ersetzen; die Nebenbedingungen werden dann einfach durch eine Projektion der Iterierten auf die zulässige Menge X behandelt, die unter den Voraussetzungen an X wohldefiniert ist. Wir erinnern an den *Projektionssatz* aus Optimierung 1 (den man nun auch leicht aus Folgerung 10.12 herleitet):

Satz 12.1. *Sei $X \subset \mathbb{R}^n$ nichtleer, konvex, und abgeschlossen und $z \in \mathbb{R}^n$ beliebig. Dann hat das Problem*

$$\min_{x \in X} \|x - z\|_2^2$$

eine eindeutige Lösung $\bar{x} =: \operatorname{proj}_X(z) \in X$, genannt Projektion von z auf X. *Weiter ist $x = \operatorname{proj}_X(z)$ genau dann, wenn gilt*

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

Die Schwierigkeit ist nun, dass im Allgemeinen ein Subgradient *keine* Abstiegsrichtung ist (dies gilt nur für den Subgradienten mit minimaler Norm). Daraus folgt, dass

(i) die üblichen Verfahren zur Schrittweitenbestimmung nicht funktionieren, und

(ii) die erzeugte Folge der Funktionswerte nicht monoton fallend ist.

Man muss daher in der Regel mit einer a priori gewählten Schrittweitenfolge arbeiten sowie sich den minimalen bisher erreichten Funktionswert merken. Dies führt auf das folgende Verfahren.

Algorithmus 12.1 : Subgradientenverfahren

Wähle einen Startpunkt $x^0 \in \mathbb{R}^n$, berechne $m_0 := F(x^0)$
for $k = 0, \dots$ **do**
  Wähle $\xi^k \in \partial F(x^k)$
  if $\xi^k = 0$ **then**
    return x^k
  else
    setze $d^k := -\xi^k / \|\xi^k\|_2$
  Setze $x^{k+1} = \mathrm{proj}_X(x^k + t_k d^k)$ für ein $t_k > 0$
  Setze $m_{k+1} = \min\{f(x^{k+1}), m_k\}$

Wir untersuchen nun die Konvergenzeigenschaften des Subgradientenverfahrens. Dazu zeigen wir zuerst, dass die Projektion auf eine konvexe Menge *nichtexpansiv* ist.

Lemma 12.2. *Sei $X \subset \mathbb{R}^n$ nichtleer, konvex, und abgeschlossen. Dann gilt*

$$\|\mathrm{proj}_X(x) - \mathrm{proj}_X(y)\|_2 \leq \|x - y\|_2 \qquad \textit{für alle } x, y \in \mathbb{R}^n.$$

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

$$x - y = \mathrm{proj}_X(x) - \mathrm{proj}_X(y) + (x - \mathrm{proj}_X(x)) + (\mathrm{proj}_X(y) - y).$$

Aus dem Satz von Pythagoras folgt dann

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

Aus Satz 12.1 folgt aber für $z = x$ und $\tilde{x} = \mathrm{proj}_X(y) \in X$ bzw. $z = y$ und $\tilde{x} = \mathrm{proj}_X(x) \in X$

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

und damit

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

Damit können wir nun die Konvergenz zumindest der (minimalen) Funktions*werte* zeigen, falls die Schrittweiten nicht zu schnell gegen Null gehen.

Satz 12.3. *Sei $\{m_k\}_{k\in\mathbb{N}}$ die durch Algorithmus 12.1 erzeugte Folge für Schrittweiten $\{t_k\}_{k\in\mathbb{N}}$ mit*

$$\lim_{k\to\infty} t_k = 0, \qquad \sum_{k=0}^{\infty} t_k = \infty.$$

Dann gilt

$$\lim_{k\to\infty} m_k = F^* := \min\{F(x) : x \in X\}.$$

Beweis. Nach Konstruktion ist $\{m_k\}_{k\in\mathbb{N}} \subset \mathbb{R}$ monoton fallend sowie nach unten durch F^* beschränkt und daher konvergent mit Grenzwert $m_* \geq F^*$. Angenommen, es gilt $m_* > F^*$. Dann können wir ein $\alpha \in \mathbb{R}$ wählen mit $F^* < \alpha < m_*$; nach Definition von F^* existiert daher ein $\hat{x} \in X$ mit $F(\hat{x}) < \alpha$. Nun ist $F : \mathbb{R}^n \to \mathbb{R}$ konvex und daher nach Satz 9.9 stetig; es existiert daher ein $\delta > 0$ mit $F(x) \leq \alpha$ für alle $\|x - \hat{x}\|_2 \leq \delta$. Insbesondere gilt das für

$$z^k := \hat{x} + \delta\frac{\xi^k}{\|\xi^k\|_2} \qquad \text{für alle } k \in \mathbb{N}$$

mit $\xi^k \in \partial F(x^k)$ aus Algorithmus 12.1. Aus der Definition des Subgradienten und $F(x^k) \geq m_k \geq m_* > \alpha$ folgt dann

$$\langle \xi^k, z^k - x^k\rangle \leq F(z^k) - F(x^k) \leq \alpha - m_k < 0.$$

Für $d^k = -\xi^k/\|\xi^k\|_2$ (falls $\xi^k \neq 0$; sonst ist x^k nach Satz 10.1 bereits Minimierer von F) und $z^k = \hat{x} - \delta d_k$ gilt daher

$$0 < \langle d^k, z^k - x^k\rangle = \langle d^k, \hat{x} - x^k\rangle - \delta\langle d^k, d^k\rangle$$

und damit

$$\langle d^k, x^k - \hat{x}\rangle < -\delta \qquad \text{für alle } k \in \mathbb{N}.$$

Aus Lemma 12.2 folgt wegen $\hat{x} \in X$ nun

$$\begin{aligned}\|x^{k+1} - \hat{x}\|_2^2 &= \|\mathrm{proj}_X(x^k + t_k d^k) - \hat{x}\|_2^2\\ &\leq \|x^k + t_k d^k - \hat{x}\|_2^2\\ &= \|x^k - \hat{x}\|_2^2 + 2\langle d^k, x^k - \hat{x}\rangle + t_k^2\\ &< \|x^k - \hat{x}\|_2^2 + t_k(t_k - 2\delta).\end{aligned}$$

Wegen $0 < t_k \to 0$ existiert aber ein $k_0 \in \mathbb{N}$ mit $t_k \leq \delta$ und damit

$$\|x^{k+1} - \hat{x}\|_2^2 < \|x^k - \hat{x}\|_2^2 - \delta t_k \quad \text{für alle } k \geq k_0.$$

Aufsummieren über alle $k = k_0, \ldots, K$ für $K > k_0$ beliebig und Verwenden der Teleskopsumme ergibt dann

$$\delta\sum_{k=k_0}^{K} t_k \leq \|x^{k_0} - \hat{x}\|_2^2 - \|x^{K+1} - \hat{x}\|_2^2 \leq \|x^{k_0} - \hat{x}\|_2^2,$$

im Widerspruch zur Divergenz der linken Seite für $K \to \infty$ nach Voraussetzung. □

Eine mögliche zulässige Wahl für die Schrittweiten ist z. B. $t_k = \frac{1}{k+1}$, was allerdings zu einer äußerst langsamen Konvergenz führt. Ist jedoch zumindest der optimale Funktionswert bekannt, kann man bessere Schrittweiten wählen.

Lemma 12.4. *Sei $\bar{x} \in X$ Lösung von (P) und $\{x^k\}_{k\in\mathbb{N}}$ die durch Algorithmus 12.1 erzeugte Folge für Schrittweiten $\{t_k\}_{k\in\mathbb{N}}$ mit*

$$0 < t_k < 2\frac{F(x^k) - F(\bar{x})}{\|\xi^k\|_2} \qquad \text{für alle } k \in \mathbb{N}.$$

Dann gilt

$$\|x^{k+1} - \bar{x}\|_2 < \|x^k - \bar{x}\|_2 \qquad \text{für alle } k \in \mathbb{N}.$$

Beweis. Aus der Definition des Subgradienten gilt für $\xi^k \in \partial F(x^k)$ wieder

$$\langle \xi^k, \bar{x} - x^k \rangle \leq F(\bar{x}) - F(x^k) \qquad \text{für alle } k \in \mathbb{N}.$$

Für $d^k = -\xi^k / \|\xi^k\|_2$ und $x^{k+1} = x^k + t_k d_k$ ist daher

$$\begin{aligned}
\|x^k + t_k d^k - \bar{x}\|_2^2 &\leq \|x^k - \bar{x}\|_2^2 + 2t_k \langle d^k, x^k - \bar{x}\rangle + t_k^2 \|d^k\|_2^2 \\
&= \|x^k - \bar{x}\|_2^2 + 2\frac{t_k}{\|\xi^k\|_2}\langle \xi^k, \bar{x} - x^k\rangle + t_k^2 \\
&\leq \|x^k - \bar{x}\|_2^2 + 2\frac{t_k}{\|\xi^k\|_2}(F(\bar{x}) - F(x^k)) + t_k^2 \\
&= \|x^k - \bar{x}\|_2^2 + t_k\left(-\frac{2}{\|s_k\|_2}(F(x^k) - F(\bar{x})) + t_k\right) \\
&< \|x^k - \bar{x}\|_2^2
\end{aligned}$$

nach Annahme an t_k.

Aus Lemma 12.2 und $\bar{x} \in X$ folgt damit

$$\|x^{k+1} - \bar{x}\|_2 = \|\mathrm{proj}_X(x^k + t_k d^k) - \mathrm{proj}_X(\bar{x})\|_2 \leq \|x^k + t_k d^k - \bar{x}\|_2 < \|x^k - \bar{x}\|_2. \qquad \square$$

Dies legt die Wahl $t_k = (F(x^k) - F(\bar{x})/\|\xi^k\|_2$ nahe – wenn $F(\bar{x})$ oder zumindest eine hinreichend gute obere Schranke bekannt ist. Schwerwiegender ist allerdings, dass für das Subgradientenverfahren kein praktikables Abbruchkriterium zur Verfügung steht, da für nichtdifferenzierbare Funktionen $\|\xi^k\| \to 0$ nicht zu erwarten ist (betrachte z. B. $f(x) = |x|$ mit $|\xi^k| = 1$ für alle $x^k \neq \bar{x} = 0$).

12.2 SCHNITTEBENENVERFAHREN

Das Problem ist, dass für nichtdifferenzierbare Funktionen das Subdifferential zwar globale Aussagen erlaubt, aber im Gegensatz zu (stetigen) klassischen Ableitungen keine Nachbarschaftsinformationen enthält. Praktikable Verfahren verwenden daher *mehrere* Subgradienten aus dem Subdifferential an *verschiedenen* Punkten. Ein Prototyp ist das in Folge hergeleitete Verfahren. Angenommen, wir haben bereits für $j = 0, \dots, k$ Punkte x^j und zugehörige Subgradienten $\xi^j \in \partial F(x^j)$ bestimmt. Dann folgt aus der Definition des Subdifferentials für alle $x \in X$

$$F(x) \geq F(x^j) + \langle \xi^j, x - x^j \rangle \qquad \text{für } j = 0, \dots, k.$$

Also gilt auch

$$F(x) \geq \max_{j=0,\dots,k} F(x^j) + \langle \xi^j, x - x^j \rangle =: F_k(x),$$

d. h. $F_k \leq F = F^{\Gamma}$ ist eine *endliche* Approximation der konvexen Hülle, die für konvexe Funktionen ja mit der Funktion selber übereinstimmt. Die Idee ist nun, die Näherung zu verbessern, indem wir x^{k+1} als Minimierer von F_k über X zu wählen (da dort der größte Unterschied zwischen F und F_k zu erwarten ist). Dabei stellt sich natürlich die Frage, ob dieses Problem wesentlich einfacher als das ursprüngliche ist. Die folgende Reformulierung zeigt, dass das der Fall ist.

Lemma 12.5. *Sei $\hat{x} \in X$ und $\hat{\eta} = F_k(\hat{x})$. Dann ist $\hat{x}$ Lösung von $\min_{x \in X} F_k(x)$ genau dann, wenn $(\hat{x}, \hat{\eta})$ Lösung ist von*

$$(\mathrm{P}_k) \qquad \begin{cases} \min\limits_{x \in X, \eta \in \mathbb{R}} \eta \\ \quad \textit{mit} \quad F(x^j) + \langle \xi^j, x - x^j \rangle \leq \eta, \quad j = 0, \dots, k. \end{cases}$$

Beweis. Ist $\hat{x} \in X$ Lösung von $\min_{x \in X} F_k(x)$ und $\hat{\eta} = F_k(x^{k+1})$, dann ist offensichtlich $(\hat{x}, \hat{\eta})$ zulässig für (P_k). Angenommen, $(\hat{x}, \hat{\eta})$ ist nicht optimal. Dann existiert ein zulässiger Punkt $(\tilde{x}, \tilde{\eta})$ mit $\tilde{\eta} < \hat{\eta}$. Nach Definition von F_k ist dann aber

$$F_k(\tilde{x}) = \max_{j=0,\dots,k} F(x^j) + \langle \xi^j, \tilde{x} - x^j \rangle \leq \tilde{\eta} < \hat{\eta} = F_k(\hat{x}),$$

im Widerspruch zur Annahme, dass $\hat{x}$ Minimierer von F_k ist.

Sei nun umgekehrt $(\hat{x}, \hat{\eta})$ eine Lösung von (P_k). Dann ist insbesondere $\hat{x} \in X$. Angenommen, $\hat{x}$ ist kein Minimierer von F_k, d. h. es existiert ein $\tilde{x} \in X$ mit

$$\tilde{\eta} := F_k(\tilde{x}) < F_k(\hat{x}) = \max_{j=0,\dots,k} F(x^j) + \langle \xi^j, \hat{x} - x^j \rangle \leq \hat{\eta}.$$

Also ist auch $(\tilde{x}, \tilde{\eta})$ zulässig für (P_k) mit $\tilde{\eta} < \hat{\eta}$, im Widerspruch zur Annahme, dass $\hat{\eta}$ minimal ist. □

Wir können x^{k+1} also als Minimierer einer *differenzierbaren* Funktion unter zusätzlichen linearen Nebenbedingungen berechnen; ist X selber beschrieben durch endlich viele lineare Gleichungen und Ungleichungen, liegt sogar ein lineares Optimierungsproblem vor, das mit Simplex- oder Innere-Punkte-Verfahren gelöst werden kann. Insbesondere unterscheidet sich (P_{k+1}) von (P_k) nur durch die Zunahme einer neuen Restriktion

$$F(x^{k+1}) + \langle \xi^{k+1}, x - x^{k+1} \rangle \leq \eta$$

mit $\xi^{k+1} \in \partial F(x^{k+1})$. Erfüllt nun die Lösung (x^{k+1}, η^{k+1}) von (P_k) bereits diese Ungleichung, so folgt daraus nach Lemma 12.5

$$F(x^{k+1}) \leq \eta^{k+1} = F_k(x^{k+1}) \leq F_k(x) \leq F^\Gamma(x) = F(x) \qquad \text{für alle } x \in X,$$

d. h. x^{k+1} ist Lösung von (P). Anders formuliert: Ist x^{k+1} noch *nicht* der gesuchte Minimierer, schneidet die zusätzliche Ungleichung x^{k+1} vom zulässigen Bereich ab, so dass der nächste Kandidat woanders gesucht wird. Daher bezeichnet man diesen Ansatz als *Schnittebenenverfahren* (Englisch: "cutting plane method").

Algorithmus 12.2 : Schnittebenenverfahren

```
1 Wähle einen Startpunkt x^0 ∈ X
2 for k = 0, ... do
3     Wähle ξ^k ∈ ∂F(x^k)
4     if (x^k, F_k(x^k)) zulässig für (P_k) then return x^k
5     Berechne (x^{k+1}, η^{k+1}) als Lösung von (P_k)
```

Wir zeigen nun die Konvergenz von Algorithmus 12.2.

Satz 12.6. *Sei $F : \mathbb{R}^n \to \mathbb{R}$ eine konvexe unterhalbstetige Funktion und $X \subset \mathbb{R}^n$ eine nichtleere, konvexe, und abgeschlossene Menge. Dann ist jeder Häufungspunkt einer durch Algorithmus 12.2 erzeugte Folge $\{x^k\}_{k\in\mathbb{N}}$ eine Lösung von (P).*

Beweis. Sei x^* ein Häufungspunkt und $\{x^k\}_{k\in\mathbb{N}}$ eine (nicht von der Notation unterschiedene) Teilfolge mit $x^k \to x^*$. Wegen $x^k \in X$ für alle $k \in \mathbb{N}$ und der Abgeschlossenheit von X ist dann auch $x^* \in X$.

Sei nun $k \in \mathbb{N}$ beliebig. Dann gilt nach Konstruktion für alle $x \in X$ und alle $j = 0, \ldots, k$

$$F(x) = F^\Gamma(x) \geq F_k(x) \geq F_k(x^{k+1}) \geq F(x^j) + \langle \xi^j, x^{k+1} - x^j \rangle.$$

Grenzübergang $k \to \infty$ ergibt dann

$$F(x) \geq F(x^j) + \langle \xi^j, x^* - x^j \rangle \qquad \text{für alle } j \in \mathbb{N}.$$

Wir möchten nun auch $j \to \infty$ (entlang der ursprünglich gewählten Teilfolge) gehen lassen. Dazu verwenden wir, dass $F : \mathbb{R}^n \to \mathbb{R}$ nach Satz 9.9 lokal Lipschitz-stetig ist. Aus Lemma 10.3 folgt daher für $j \in \mathbb{N}$ groß genug

$$\langle \xi^j, x^* - x^j \rangle \leq F'(x^j; x^* - x^j) = \lim_{t \to 0} \frac{F(x^j + t(x^* - x^j)) - F(x^j)}{t} \leq L\|x^* - x^j\|$$

und analog $\langle \xi^j, x^j - x^* \rangle \leq L\|x^j - x^*\|$. Also gilt

$$0 \leq |\langle \xi^j, x^j - x^* \rangle| \leq L\|x^j - x^*\| \to 0$$

und daher (ebenfalls wegen der Stetigkeit von F)

$$F(x) \geq F(x^j) + \langle \xi^j, x^* - x^j \rangle \to F(x^*),$$

d. h. x^* ist Minimierer von F über X. □

Im Gegensatz zum Subgradientenverfahren können wir hier ein praktikables Abbruchkriterium finden. Nach Definition von $\xi^j \in \partial F(x^j)$ gilt für die Lösung $\bar{x} \in X$ von (P_k) nämlich

$$F(x^j) + \langle \xi^j, \bar{x} - x^j \rangle \leq F(\bar{x}) \qquad \text{für alle } j = 0, \ldots, k.$$

Also ist $(\bar{x}, F(\bar{x}))$ zulässig für (P_k), so dass nach Konstruktion gilt $\eta^{k+1} \leq F(\bar{x})$ und damit

$$F(x^{k+1}) - F(\bar{x}) \leq F(x^{k+1}) - \eta^{k+1}.$$

Um zu garantieren, dass wir höchstens $\varepsilon > 0$ über dem optimalen Funktionswert liegen, reicht es zu prüfen, ob $F(x^{k+1}) - \eta^{k+1} < \varepsilon$ gilt.

Trotzdem hat Algorithmus 12.2 noch einige Nachteile:

(i) Das Teilproblem (P_k) kann keine Lösung haben, wenn X und damit die zulässige Menge unbeschränkt ist. Dies kann man aber behandeln, indem man die Zielfunktion durch Hinzufügen eines quadratischen Terms koerziv macht; man betrachtet statt F_k daher

$$F_k^\gamma(x) := F_k(x) + \frac{1}{2\gamma_k}\|x - x^k\|_2^2$$

für $\gamma_k > 0$ geeignet gewählt. Dies bezeichnet man als *Proximal-Schnittebenenverfahren*.

(ii) Die Anzahl der Nebenbedingungen in (P_k) wächst im Laufe der Iteration, so dass die Teilprobleme immer komplizierter zu lösen sind. Tatsächlich verwendete Verfahren enthalten daher auch Schritte, in denen (geeignet gewählte) Ungleichungen auch wieder entfernt werden.

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. [Harcourt Brace Jovanovich, Publishers], New York-London.

S. Boyd & L. Vandenberghe (2004), *Convex Optimization*, Cambridge University Press, Cambridge, doi: 10.1017/cbo9780511804441.

C. Clason & T. Valkonen (2020), Introduction to Nonsmooth Analysis and Optimization, arxiv: 2020.00216.

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

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