

Julius-Maximilians-Universität Würzburg

Ein semismoothes SQP-Verfahren zur
Bestimmung normalisierter
Nash-Gleichgewichte

Diplomarbeit im Fach Mathematik

Vorgelegt von

Alexandra Väth

Würzburg, den 10.07.2008

Betreuer:

Prof. Dr. Christian Kanzow

Lehrstuhl für Angewandte Mathematik II

Inhaltsverzeichnis

1	Einleitung	5
1.1	Spieltheoretischer Hintergrund	7
1.2	Anwendungsbeispiel: Energiekontrolle in Digital Subscriber Lines	10
2	Stetige Differenzierbarkeit der Zielfunktion $V_\gamma(x)$	15
2.1	Ein Resultat aus der konvexen Optimierung	15
2.2	Der Satz von Danskin	17
2.3	Punkt-Menge-Abbildungen	18
2.4	Stetige Differenzierbarkeit von $V_\gamma(x)$	22
3	Semismoothness der Ableitung der Zielfunktion $V_\gamma(x)$	27
3.1	Semismoothe und strongly semismoothe Funktionen	27
3.2	Satz über implizite Abbildungen für (strongly) semismoothe Funktionen	31
3.3	(Strongly) Semismoothness der Ableitung von $V_\gamma(x)$	38
4	Konvergenzeigenschaften des semismoothen SQP-Verfahrens	45
4.1	Lokal superlineare Konvergenz des semismoothen SQP-Verfahrens	45
4.2	Lokal quadratische Konvergenz des semismoothen SQP-Verfahrens	57
4.3	Hinreichende Bedingungen für die strong second-order sufficiency condition	59
5	Numerische Umsetzung des SQP-Verfahrens	69
5.1	Zwei Varianten des SQP-Verfahrens	69
5.2	Numerische Illustration	74
	Anhang	83
A.1	Tabellen	83
A.2	Quellcode	87

Kapitel 1

Einleitung

Das Thema dieser Arbeit ist ein numerisches Verfahren zur Bestimmung normalisierter Nash-Gleichgewichte von verallgemeinerten Nash-Gleichgewichtsproblemen. Eine genaue Definition der auftretenden spieltheoretischen Begriffe findet sich in Abschnitt 1.1. Unter einem verallgemeinerten Nash-Gleichgewichtsproblem versteht man eine Situation, in der mehrere Parteien Entscheidungen treffen müssen, um ihren Nutzen zu maximieren, und sich durch ihre Entscheidungen gegenseitig beeinflussen. So kann es sein, dass der erzielbare Nutzen einer Partei von den Entscheidungen der anderen Parteien abhängt. Ein Beispiel hierfür ist ein Anbieterduopol, bei dem der erzielbare Marktpreis davon abhängt, wieviele Mengeneinheiten des Gutes von beiden Anbietern insgesamt auf dem Markt angeboten wird. Alternativ kann es auch passieren, dass eine Partei durch ihre Entscheidung die Entscheidungsmöglichkeiten einer anderen Partei beeinflusst. Dieses Problem tritt zum Beispiel dann auf, wenn mehrere Fischer in dem selben See fischen und entscheiden müssen, wieviele Fische sie fangen wollen. Wie man sieht, treten solche Problemstellungen häufig im Bereich der Wirtschaftswissenschaften auf; einige kleine Beispiele hierzu finden sich im Abschnitt 5.2. Dass spieltheoretische Problemstellungen allerdings auch in ganz anderen Bereichen auftreten können, zeigt das Anwendungsbeispiel in Abschnitt 1.2. In diesem Abschnitt betrachten wir ein aktuelles Problem im Zusammenhang mit Digital Subscriber Lines (DSL).

Um solche Problemstellungen zu lösen, sind grundsätzlich zwei Lösungsansätze denkbar: Zum einen könnten alle Parteien kooperieren und so versuchen, den gesamten Nutzen aller Parteien zu maximieren. Hierbei tritt allerdings häufig das Problem auf, dass einzelne Parteien Anreize haben, von dem gemeinsamen Optimum abzuweichen, weil sie dadurch ihren Nutzen (evtl. auf Kosten der anderen) weiter steigern können. Dieses Problem ist in den Wirtschaftswissenschaften unter anderem als „free-rider“-Problem bekannt. Es tritt zum Beispiel dann auf, wenn zwei Personen A, B an einem gemeinsamen Projekt arbeiten sollen, und es nicht möglich ist, den Arbeitseinsatz der beiden Personen zu beobachten. Der Nutzen der beiden Personen ergibt sich jeweils durch ihren Anteil am Erfolg des Projektes abzüglich ihres Arbeitsleids. Daher steht jeder der beiden vor der Entscheidung: shirking oder non-shirking, d.h. lohnt es sich mit vollem Einsatz zu arbeiten, oder ist es eventuell besser Leistung zurückzubehalten. Hierbei hängt der Nutzen der beiden über den Projekterfolg auch von der Entscheidung des jeweils anderen ab; mögliche Nutzenwerte finden sich in der Tabelle 1.1.

Wie man sieht, arbeiten beide Personen im gemeinsamen Optimum mit vollem Ar-

		A	
		non-shirking	shirking
B	non-shirking	3,3	1,4
	shirking	4,1	2,2

Tabelle 1.1: Das free-rider-Problem

beitseinsatz. Solange es jedoch nicht möglich ist, diesen Arbeitseinsatz zu kontrollieren, haben beide den Anreiz von dem gemeinsamen Optimum abzuweichen und werden schlussendlich beide Leistung zurückhalten. Die Strategiekombination (shirking,shirking) ist daher das Nash-Gleichgewicht dieses Problems.

Allgemein definiert man als Nash-Gleichgewicht eine Strategiekombination, bei der sich keine Partei besserstellen kann, indem sie alleine von diesem Gleichgewicht abweicht. Dieser Definition liegt der zweite Lösungsansatz für solche Problemstellungen zugrunde. Bei diesem Ansatz geht man davon aus, dass alle N Parteien egoistisch versuchen ihren Nutzen zu maximieren. Hierbei hängen die Nutzenfunktionen und die Strategiemengen der einzelnen Parteien allerdings von den Entscheidungen der anderen $N - 1$ Parteien ab. Mathematisch bedeutet dies, dass man N restringierte Optimierungsprobleme lösen muss, deren Daten jeweils von den Lösungen der anderen $N - 1$ Optimierungsprobleme abhängen.

In Abschnitt 1.1 werden wir jedoch zeigen, dass man normalisierte Nash-Gleichgewichte als globale Minima *eines* restringierten Optimierungsproblems darstellen kann. Dieses Optimierungsproblem hat unter geeigneten Voraussetzungen zwar eine stetig differenzierbare Zielfunktion, vergleiche Kapitel 2, aber diese ist im Allgemeinen nicht zweimal stetig differenzierbar. Um bei einem numerischen Verfahren zur Lösung dieses Optimierungsproblems dennoch lokal schnelle Konvergenz garantieren zu können, werden wir in Kapitel 3 mit Hilfe eines Satzes über implizite Abbildungen für semismoothe Funktionen beweisen, dass die erste Ableitung der Zielfunktion unter gewissen Voraussetzungen wenigstens semismooth bzw. strongly semismooth ist. Daher können wir ein semismoothes SQP-Verfahren anwenden, um das restringierte Optimierungsproblem zu lösen. Kapitel 4 beschäftigt sich mit den lokalen Konvergenzeigenschaften eines solchen Verfahrens. Schließlich wollen wir in Kapitel 5 ein semismoothes SQP-Verfahren an einigen Nash-Gleichgewichtsproblemen numerisch testen.

Doch bevor wir mit der Definition des spieltheoretischen Problems anfangen können, müssen wir noch einige Notationsdetails klären:

Die Norm, die in dieser Arbeit verwendet wird, ist die euklidische Norm, d.h. für $x \in \mathbb{R}^n$ ist

$$\|x\|_2 := \sqrt{\sum_{i=1}^n x_i^2}.$$

Damit definieren wir Kugelumgebungen $K_\delta(x^0)$ als

$$K_\delta(x^0) := \{x \in \mathbb{R}^n \mid \|x - x^0\|_2 < \delta\}.$$

Zwei Teilmengen des \mathbb{R}^n , für die wir eine eigene Bezeichnung einführen wollen, sind

$$\mathbb{R}_+^n := \{x \in \mathbb{R}^n \mid x_i \geq 0 \quad \forall i = 1, \dots, n\},$$

$$\mathbb{R}_{++}^n := \{x \in \mathbb{R}^n \mid x_i > 0 \quad \forall i = 1, \dots, n\}.$$

Mit $\operatorname{argmax}_{x \in X} f(x)$ wollen wir die, eventuell leere, Menge aller $x \in X$ bezeichnen, in denen das restringierte Maximum des Problems

$$\max f(x) \quad \text{u.d.N.} \quad x \in X$$

angenommen wird.

Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine differenzierbare Funktion. Dann bezeichnet $f'(x) \in \mathbb{R}^{m \times n}$ oder auch $Df(x)$ die Ableitung bzw. die Jacobi-Matrix von f im Punkt x . Im Gegensatz dazu bezeichnen wir mit $\nabla f(x) \in \mathbb{R}^{n \times m}$ die transponierte Jacobi-Matrix, d.h. es gilt $\nabla f(x)^T = f'(x)$.

Manchmal wollen wir auch nur die Ableitung von f nach gewissen Blockkomponenten x^ν von x betrachten. Hierfür schreiben wir dann $\nabla_{x^\nu} f(x)$ und meinen damit die transponierten partiellen Ableitungen von f bezüglich der Komponenten von x^ν . Analog schreiben wir bei der zweiten Ableitung $\nabla_{x^\nu x^\mu}^2 f(x)$, was bedeuten soll, dass wir f zunächst bezüglich x^ν und anschließend nochmal bezüglich x^μ differenzieren.

Ist $A \in \mathbb{R}^{m \times n}$ eine Matrix und $I \subseteq \{1, \dots, m\}$ sowie $J \subseteq \{1, \dots, n\}$, so bezeichnet $A_{I,J} \in \mathbb{R}^{|I| \times |J|}$ die Teilmatrix von A , die nur aus den Elementen $a_{i,j}$, $(i, j) \in I \times J$, von A besteht.

Die Landau-Symbole o und O sind folgendermaßen definiert: Seien $\{a_k\}_k \subset \mathbb{R}^m$, $\{b_k\}_k \subset \mathbb{R}^n$ zwei Folgen. Dann sagen wir $\|a_k\| = o(\|b_k\|)$, wenn gilt

$$\lim_{k \rightarrow \infty} \frac{\|a_k\|}{\|b_k\|} = 0$$

und $\|a_k\| = O(\|b_k\|)$, wenn gilt

$$\limsup_{k \rightarrow \infty} \frac{\|a_k\|}{\|b_k\|} < \infty.$$

Damit können wir nun auch erklären, wie wir lineare, superlineare und quadratische Konvergenz definieren wollen. Sei $\{x^k\}_k \subset \mathbb{R}^n$ eine gegen ein $x^0 \in \mathbb{R}^n$ konvergente Folge. Dann konvergiert $\{x^k\}_k$ *linear* gegen x^0 , wenn es eine Konstante $0 < c < 1$ gibt, so dass für alle $k \in \mathbb{N}$ gilt

$$\|x^{k+1} - x^0\| \leq c \|x^k - x^0\|,$$

superlinear gegen x^0 , wenn gilt

$$\|x^{k+1} - x^0\| = o(\|x^k - x^0\|),$$

und *quadratisch* gegen x^0 , wenn gilt

$$\|x^{k+1} - x^0\| = O(\|x^k - x^0\|^2).$$

1.1 Spieltheoretischer Hintergrund

Eine häufige Problemstellung in der Spieltheorie sind sogenannte verallgemeinerte Nash-Gleichgewichts-Probleme (eng. generalized Nash-equilibrium problem, kurz GNEP). Hierbei geht es darum, dass N Spieler jeweils eine Variable $x^\nu \in \mathbb{R}^{n^\nu}$ kontrollieren können

und versuchen hierdurch ihren Verlust zu minimieren.

Sei $x = (x^1, \dots, x^N) \in \mathbb{R}^n$ mit $n := n_1 + \dots + n_N$. Weiter bezeichnen wir mit $X \subseteq \mathbb{R}^n$ die gemeinsame Strategiemenge aller Spieler, d.h. für eine zulässige Strategie muss immer $x \in X$ gelten. Wenn wir besonders die Entscheidungsvariable des ν -ten Spielers betonen wollen, schreiben wir $x = (x^\nu, x^{-\nu})$ und fassen die Entscheidungsvariablen aller anderen Spieler unter $x^{-\nu}$ zusammen.

Bezeichnen wir die Verlustfunktion des ν -ten Spielers mit $\theta_\nu : \mathbb{R}^n \rightarrow \mathbb{R}$, so versucht der ν -te Spieler das restringierte Optimierungsproblem

$$\min_{x^\nu} \theta_\nu(x^\nu, x^{*, -\nu}) \quad \text{u.d.N.} \quad (x^\nu, x^{*, -\nu}) \in X \quad (1.1)$$

zu lösen, wobei $x^{*, -\nu}$ die von den anderen Spielern gewählte Strategie bezeichnet. Ein Vektor $x^* = (x^{*,1}, \dots, x^{*,N}) \in X$ heißt *verallgemeinertes Nash-Gleichgewicht* oder *Lösung des GNEPs*, wenn $x^{*,\nu}$ für alle $\nu = 1, \dots, N$ das Problem (1.1) löst.

Wir wollen uns aber in dieser Arbeit nicht mit verallgemeinerten Nash-Gleichgewichten, sondern mit einem Spezialfall davon, nämlich den *normalisierten Nash-Gleichgewichten* beschäftigen. Ein Vektor $x^* \in X$ heißt *normalisiertes Nash-Gleichgewicht*, wenn gilt:

$$\sup_{y \in X} \Psi(x^*, y) = 0.$$

Hierbei bezeichnet $\Psi(x, y)$ die *Nikaido-Isoda-Funktion*

$$\Psi(x, y) := \sum_{\nu=1}^N [\theta_\nu(x^\nu, x^{-\nu}) - \theta_\nu(y^\nu, x^{-\nu})].$$

Man kann zeigen, dass jedes normalisierte Nash-Gleichgewicht auch ein verallgemeinertes Nash-Gleichgewicht ist. Die Umkehrung gilt allerdings im Allgemeinen nicht.

In dieser Arbeit gehen wir davon aus, dass die Strategiemenge X dargestellt werden kann als

$$X = \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}$$

mit gewissen Funktionen $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ und $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$, und dass die Funktionen θ_ν, g, h den folgenden Bedingungen genügen:

Voraussetzungen 1.1.1. a) Die Funktionen θ_ν sind zweimal stetig differenzierbar und als Funktionen von x^ν konvex.

b) Die Funktion g ist zweimal stetig differenzierbar und ihre Komponenten g_i sind konvex in x .

c) Die Funktion h ist affin-linear.

d) Die durch die Funktionen g und h definierte Strategiemenge

$$X = \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}$$

ist nichtleer.

Aus diesen Voraussetzungen folgt sofort, dass die Strategiemenge X abgeschlossen und konvex ist. Sie ist allerdings nicht notwendig beschränkt. Außerdem ist die Funktion h zweimal stetig differenzierbar mit lokal Lipschitz-stetiger zweiter Ableitung. Manchmal werden wir auch die folgenden, etwas stärkeren Bedingungen stellen:

Voraussetzungen 1.1.2. a) Die Funktionen θ_ν sind zweimal stetig differenzierbar und als Funktionen von x^ν konvex. Weiter seien die zweiten Ableitungen der Funktionen θ_ν noch lokal Lipschitz-stetig.

b) Die Funktion g ist zweimal stetig differenzierbar und ihre Komponenten g_i sind konvex in x mit lokal Lipschitz-stetigen zweiten Ableitungen.

c) Die Funktion h ist affin-linear.

d) Die durch die Funktionen g und h definierte Strategiemenge

$$X = \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}$$

ist nichtleer.

Um ein normalisiertes Nash-Gleichgewicht zu finden, verwenden wir die *regularisierte Nikaido-Isoda-Funktion*

$$\Psi_\gamma(x, y) := \sum_{\nu=1}^N [\theta_\nu(x^\nu, x^{-\nu}) - \theta_\nu(y^\nu, x^{-\nu})] - \frac{\gamma}{2} \|x - y\|_2^2$$

mit einer Konstanten $\gamma > 0$ und definieren für $x \in X$

$$V_\gamma(x) := \sup_{y \in X} \Psi_\gamma(x, y).$$

Unter den Voraussetzungen 1.1.1, und damit auch unter den stärkeren Voraussetzungen 1.1.2, gelten dann die folgenden Aussagen:

Satz 1.1.3. a) $V_\gamma(x) \geq 0 \quad \forall \quad x \in X$.

b) x^* ist ein normalisiertes Nash-Gleichgewicht genau dann, wenn gilt

$$x^* \in X \quad \text{und} \quad V_\gamma(x^*) = 0.$$

c) Für alle $x \in X$ gibt es genau einen Vektor $y_\gamma(x) \in X$, so dass gilt

$$\operatorname{argmax}_{y \in X} \Psi_\gamma(x, y) = y_\gamma(x)$$

und die Abbildung $x \mapsto y_\gamma(x)$ ist stetig.

d) Die Funktion V_γ ist stetig differenzierbar für alle $x \in X$ mit

$$\nabla V_\gamma(x) = \nabla_x \Psi_\gamma(x, y) \Big|_{y=y_\gamma(x)}.$$

e) x^* ist ein normalisiertes Nash-Gleichgewicht genau dann, wenn $x^* = y_\gamma(x^*)$ gilt.

Diese Aussagen werden in Kapitel 2 bewiesen. Nach dem obigen Satz sind also die folgenden beiden Probleme äquivalent:

- Finde ein normalisiertes Nash-Gleichgewicht x^* des GNEPs.
- Finde eine Lösung x^* von

$$\min V_\gamma(x) \quad \text{u.d.N.} \quad x \in X \quad (1.2)$$

mit $V_\gamma(x^*) = 0$.

Hierbei ist das zweite Problem zwar stetig differenzierbar, aber im Allgemeinen leider nicht zweimal stetig differenzierbar. Wie in der Einleitung erwähnt, werden wir uns deshalb in Kapitel 3 mit (strongly) semismooth Funktionen und einem Satz über implizite Funktionen beschäftigen, um am Ende dieses Kapitels beweisen zu können, dass ∇V_γ unter den gegebenen Voraussetzungen zumindest eine (strongly) semismoothe Funktion ist.

1.2 Anwendungsbeispiel: Energiekontrolle in Digital Subscriber Lines

Doch zunächst wollen wir mit dem folgenden Beispiel motivieren, warum es auch für die Anwendung interessant ist Nash-Gleichgewichtsprobleme zu betrachten. Weitere Informationen zu diesem Beispiel finden sich in [23], [24], [10] und [1].

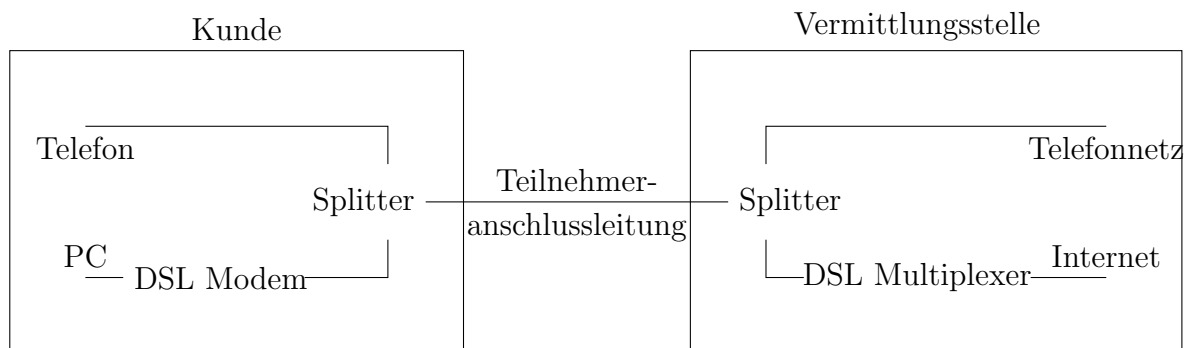


Abbildung 1.1: Die Grundstruktur eines DSL-Anschlusses

Die Abbildung 1.1 gibt etwas vereinfacht die grundlegende Struktur eines DSL-Anschlusses wieder. Der große Vorteil von DSL ist, dass auf diese Art auch Nutzern mit einem normalen Telefonanschluss über Kupferkabel ein schneller Internetzugang ermöglicht werden kann. Bei einem DSL-Anschluss über Telefonleitung werden die Internetdaten ebenfalls über die Teilnehmeranschlussleitung gesendet, allerdings in einem Frequenzbereich, der oberhalb des für Sprachtelefonie oder ISDN genutzten Frequenzbandes liegt. Bei dem in Deutschland weit verbreiteten ADSL (Asymmetric DSL) ist der genutzte Frequenzbereich wie folgt aufgeteilt: 0-138 kHz für Telefon und ISDN, 138-275 kHz für den Upstream und 275-1104 kHz für den Downstream. Hierbei teilt man Up- und Downstream nochmals in jeweils 32 bzw. 192 Frequenzbänder der Breite 4,3125 kHz auf. In der nur wenige

Kilometer vom Telefonanschluss entfernt liegenden Vermittlungsstelle wird das analoge DSL-Signal in ein digitales gewandelt und von dort über Glasfaserkabel weiterübertragen. Aus der Nutzung der schon vorhandenen Anschlüsse resultieren allerdings auch die Probleme, mit denen DSL zu kämpfen hat. Die Teilnehmeranschlussleitungen sind in der Regel Kupferkabel, die nie für die Übertragung von Signalen in diesem hohen Frequenzbereich gedacht waren. Die tatsächlich realisierbare Geschwindigkeit des Internetanschlusses hängt daher stark von der Qualität und dem Durchmesser des genutzten Kabelpaares, sowie dem Abstand der Kundenanschlusses vom DSL-Multiplexer ab. Ein weiteres Problem sind elektromagnetische Interferenzen, ähnlich der Störgeräusche, wie man sie aus den Anfängen der Telefonie kennt. Die Hauptursache solcher Interferenzen ist das sogenannte Übersprechen (engl. crosstalk). Ein über ein Kupferkabel gesendetes Signal induziert ein elektromagnetisches Feld um dieses Kabel. So wird dieses Signal in die anderen, im selben Kabelstrang verlaufenden, Adern eingekoppelt und stört so die in diesen Adern übertragenen Signale. Man unterscheidet hierbei zwischen NEXT (near-end-crosstalk), also einem Störsignal, das auf der Seite der Störsenders in einer anderen Leitung empfangen wird, und FEXT (far-end-crosstalk), einem Störsignal, das auf der Seite des Empfängers in einer anderen Leitung empfangen wird, vergleiche hierzu auch Abbildung 1.2.

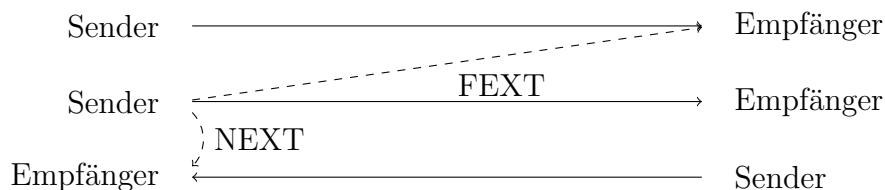


Abbildung 1.2: near-end-crosstalk und far-end-crosstalk

Im Allgemeinen ist NEXT ein größeres Problem als FEXT, da bei NEXT das Störsignal im Vergleich zu dem ankommenden, durch die lange Leitung schon gedämpften, Signal stärker ist. Bei FEXT hingegen wird das Störsignal durch die lange Leitung schon etwas gedämpft und es ist daher unwahrscheinlicher, dass es das gewünschte Signal überdeckt. Eine Ausnahme hiervon ist das sogenannte near-far-Problem, vergleiche Abbildung 1.3.

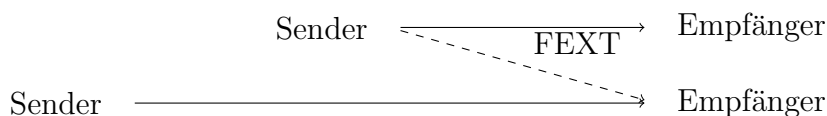


Abbildung 1.3: near-far-Problem

Hier betrachtet man zwei Anschlüsse in einem Kabelstrang, von denen einer deutlich weiter von der Vermittlungsstelle entfernt ist als der andere. Dadurch kann es passieren, dass das durch FEXT entstehende Störsignal des nähergelegenen Anschlusses das durch die lange Leitung gedämpfte Signal des weiter entfernten Anschlusses überdeckt. Weitere Ursachen für Störgeräusche sind Funksignale, da die Kupferadern der Teilnehmeranschlussleitungen wie Antennen Funksignale auffangen, und elektromagnetische Impulse in der Nähe des Kabelstrangs.

Um trotz der Störgeräusche eine akzeptable Bit-Fehler-Rate garantieren zu können, untersucht man vor allem die drei folgenden Lösungsansätze:

Der traditionelle Ansatz ist ein statisches Spektrum-Management, bei dem allen Anschlüssen eine identische „spectral mask“ vorgegeben wird, d.h. Bestimmungen, mit wieviel Energie auf welchen Frequenzen gesendet werden darf. Hierdurch wird auch im schlimmsten möglichen Fall die gewünschte Bit-Fehler-Rate garantiert. Im Allgemeinen sind diese Masken daher zu restriktiv und benachteiligen die weiter vom DSL-Multiplexer entfernten Anschlüsse, vgl. near-far-Problem.

Ein zweiter Ansatz ist die Einführung eines zentralen Spektrum-Managements, das versucht die Summe aller Nutzer-Raten zu maximieren. Dieser Ansatz hat jedoch mehrere Nachteile: Zum einen ist das entstehende Optimierungsproblem nicht konvex und besitzt mehrere lokale Maxima. Zum anderen setzt dieser Ansatz voraus, dass es eine zentrale Stelle gibt, die die Kontrolle über alle Anschlüsse hat. Dies ist heutzutage nicht mehr realistisch, da oft mehrere (konkurrierende) Anbieter Anschlüsse in dem gleichen Kabelstrang versorgen.

Der dritte Ansatz, den wir im Folgenden genauer betrachten wollen, interpretiert die Situation als ein nichtkooperatives Spiel, in dem jeder Nutzer versucht, seine Bit-Rate zu maximieren. Sei N die Anzahl der Nutzer in einem Kabelstrang und F die Anzahl der ihnen zur Verfügung stehenden Frequenzen. Weiter bezeichne $x^\nu \in \mathbb{R}^F$ die Energieverteilung des Nutzers ν auf die möglichen Frequenzen. Hierbei ist zu beachten, dass nur solche Vektoren x^ν zulässig sind, für die $x^\nu \in X_\nu$ gilt, wobei

$$X_\nu = \left\{ x^\nu \in \mathbb{R}^F \mid 0 \leq x_f^\nu \leq \text{CAP}_f^\nu \quad \forall f = 1, \dots, F, \quad \sum_{f=1}^F x_f^\nu \leq P_{\max}^\nu \right\}$$

das Energiebudget des Nutzers ν beschreibe. Die Konstanten $0 < \text{CAP}_f^\nu \leq \infty$ bzw. $0 < P_{\max}^\nu$ geben hierbei an, wieviel Energie der Nutzer ν maximal auf die Frequenzen $f = 1, \dots, F$ verteilen darf, bzw. wieviel Energie er insgesamt höchstens verteilen darf. Bezeichnen wir mit $\alpha_f^{\mu\nu} \geq 0$ die Koeffizienten für das Übersprechen von Nutzer μ auf Nutzer ν auf der Frequenz f und mit $\sigma_f^\nu > 0$ die Störgeräusche, die den Nutzer ν auf der Frequenz f beeinträchtigen, so ist

$$\text{SNR}_f^\nu := \frac{x_f^\nu}{\sigma_f^\nu + \sum_{\mu \neq \nu} \alpha_f^{\mu\nu} x_f^\mu}$$

der signal-to-noise-ratio der Nutzers ν auf der Frequenz f , d.h. der Quotient aus dem eigentlich gewünschten Signal und den Störgeräuschen auf der Frequenz f . Betrachtet man alle Interferenzen als Störgeräusche, so ergibt sich unter in der Literatur gängigen Annahmen die folgende Formel für die erreichbare Bit-Rate des Nutzers ν auf der Frequenz f :

$$b_f^\nu = \frac{1}{2} \log_2 \left(1 + \frac{\text{SNR}_f^\nu}{\Gamma} \right),$$

wobei $\Gamma \geq 1$ den SNR-Gap bezeichnet. Im Idealfall $\Gamma = 1$ spricht man statt von der erreichbaren Bit-Rate auch von der Kanalkapazität. In der Realität gilt allerdings im Allgemeinen $\Gamma > 1$, der genaue Wert hängt unter anderem von der für die Übertragung verwendeten Kodierung und der gewünschten Bit-Fehler-Rate ab. Im Folgenden wollen wir dennoch von dem Idealfall $\Gamma = 1$ ausgehen, anderenfalls kann man Γ mit den Konstanten $\alpha_f^{\mu\nu}$ und σ_f^ν verrechnen. Damit können wir nun die vom Nutzer ν erreichbare

durchschnittliche Bit-Rate bestimmen:

$$b_\nu = \frac{1}{F} \sum_{f=1}^F b_f^\nu = \frac{1}{2F} \sum_{f=1}^F \log_2 \left(1 + \frac{x_f^\nu}{\sigma_f^\nu + \sum_{\mu \neq \nu} \alpha_f^{\mu\nu} x_f^\mu} \right).$$

Wir definieren daher als Nutzenfunktion des Nutzers ν

$$\theta_\nu(x) := \sum_{f=1}^F \log_2 \left(1 + \frac{x_f^\nu}{\sigma_f^\nu + \sum_{\mu \neq \nu} \alpha_f^{\mu\nu} x_f^\mu} \right).$$

Folglich muss jeder Nutzer für festes $x^{-\nu}$ das Optimierungsproblem

$$\max_{x^\nu} \theta_\nu(x) \quad \text{u.d.N.} \quad x^\nu \in X_\nu$$

lösen. Setzt man $X = X_1 \times \dots \times X_N$, so sieht man, dass dieses Problem ein (verallgemeinertes) Nash-Gleichgewichtsproblem ist. Genaugenommen handelt es sich hierbei sogar um ein Standard-Nash-Gleichgewichtsproblem, da die zulässige Menge X_ν des Spielers ν nicht von den gewählten Strategien der anderen Spieler abhängt.

Kapitel 2

Stetige Differenzierbarkeit der Zielfunktion $V_\gamma(x)$

In diesem Kapitel wollen wir beweisen, dass die Funktion $V_\gamma(x)$ unter den Voraussetzungen 1.1.1 stetig differenzierbar ist. Im Abschnitt 2.1 werden wir zeigen, dass Maximierungsprobleme mit gleichmäßig konkaver Zielfunktion über konvexen Mengen immer genau eine Lösung haben. Dies garantiert die Wohldefiniertheit von $y_\gamma(x)$. Im Abschnitt 2.2 werden wir dann den Satz von Danskin beweisen, der dafür sorgt, dass $V_\gamma(x)$ stetig differenzierbar ist. Um eine im Satz von Danskin auftretende Voraussetzung genauer zu untersuchen, machen wir im Abschnitt 2.3 einen Ausflug in die Theorie mengenwertiger Abbildungen. Im Abschnitt 2.4 setzen wir schließlich alle Ergebnisse zusammen und beweisen die Aussagen aus Satz 1.1.3.

2.1 Ein Resultat aus der konvexen Optimierung

Wir wollen in diesem Abschnitt ein Resultat aus der konvexen Optimierung beweisen, mit dem wir später zeigen können, dass die Abbildung $x \mapsto y_\gamma(x)$ aus Satz 1.1.3 wohldefiniert ist. Dafür benötigen wir allerdings zunächst noch zwei Eigenschaften konvexer Funktionen.

Satz 2.1.1. *Sei $X \subseteq \mathbb{R}^n$ konvex und $f : X \rightarrow \mathbb{R}$ konvex. Dann ist f im Inneren von X lokal Lipschitz-stetig.*

Einen Beweis dieser Aussage findet man in [12], Satz 6.14.

Satz 2.1.2. *Seien $X \subseteq \mathbb{R}^n$ nichtleer, offen und konvex, $f : X \rightarrow \mathbb{R}$ konvex und $x \in X$. Dann ist das konvexe Subdifferential*

$$\partial f(x) := \{s \in \mathbb{R}^n \mid f(y) \geq f(x) + s^T(y - x) \quad \forall y \in X\}$$

nichtleer.

Den Beweis für diesen Satz findet man zum Beispiel in [12], Satz 6.17. Das hier definierte konvexe Subdifferential sollte nicht mit dem in Abschnitt 3.1 definierten Clarke'schen Subdifferential verwechselt werden. Allerdings stimmen für eine konvexe und

lokal Lipschitz-stetige Funktion das konvexe Subdifferential und das nach Clarke definierte überein, vergleiche Proposition 2.2.7 in [2]. Mit diesen Aussagen können wir jetzt den gewünschten Satz beweisen.

Satz 2.1.3. *Es sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ eine konvexe Funktion und $X \subseteq \mathbb{R}^n$ nichtleer, abgeschlossen und konvex. Ist f zusätzlich gleichmäßig konvex auf X , so besitzt das Optimierungsproblem*

$$\min f(x) \quad \text{u.d.N.} \quad x \in X$$

genau eine Lösung.

Beweis. Existenz einer Lösung:

Sei $f^* = \inf_{x \in X} f(x)$. Da f auf X wohldefiniert ist, ist $f^* < +\infty$. Dann sind alle auf X eingeschränkten Levelmengen

$$\mathcal{L}(c; X) := \{x \in X \mid f(x) \leq c\}$$

mit $c > f^*$ nichtleer. Da f auf \mathbb{R}^n konvex, also nach Satz 2.1.1 stetig ist, sind die Levelmengen

$$\mathcal{L}(c) := \{x \in \mathbb{R}^n \mid f(x) \leq c\}$$

abgeschlossen. Da auch X nach Voraussetzung abgeschlossen ist, sind auch die eingeschränkten Levelmengen

$$\mathcal{L}(c; X) = \mathcal{L}(c) \cap X$$

abgeschlossen.

Sei nun $c \geq f^*$ beliebig. Wir wollen zeigen, dass $\mathcal{L}(c; X)$ dann auch beschränkt ist. Sei dafür $x \in X$ beliebig, aber fest gewählt. Da f auf \mathbb{R}^n konvex ist, ist nach Satz 2.1.2 das konvexe Subdifferential $\partial f(x) \neq \emptyset$. d.h. es gibt ein $s \in \partial f(x)$.

Annahme: Es gibt eine Folge $\{x^k\}_k \subseteq \mathcal{L}(c; X)$ mit $\|x^k\|_k \rightarrow +\infty$. Da f auf X gleichmäßig konvex ist, gilt mit einem gewissen Modulus $\mu > 0$ für alle $k \in \mathbb{N}$ und alle $\lambda \in (0, 1)$

$$f(\lambda x + (1 - \lambda)x^k) + \mu\lambda(1 - \lambda)\|x - x^k\|_2^2 \leq \lambda f(x) + (1 - \lambda)f(x^k).$$

Speziell mit $\lambda = \frac{1}{2}$ folgt für alle $k \in \mathbb{N}$

$$c \geq f(x^k) \geq 2f\left(\frac{1}{2}x + \frac{1}{2}x^k\right) + \frac{\mu}{2}\|x - x^k\|_2^2 - f(x).$$

Wegen $s \in \partial f(x)$ gilt für alle $y \in X$

$$f(y) \geq f(x) + s^T(y - x).$$

Mit der Cauchy-Schwarzschen Ungleichung folgt daraus für alle $k \in \mathbb{N}$

$$f\left(\frac{1}{2}x + \frac{1}{2}x^k\right) \geq f(x) + s^T\left(\frac{1}{2}x^k - \frac{1}{2}x\right) \geq f(x) - \frac{1}{2}\|s\|_2\|x - x^k\|_2.$$

Setzt man die zweite Ungleichung in die erste ein, so erhält man

$$\begin{aligned} c &\geq 2\left(f(x) - \frac{1}{2}\|s\|_2\|x - x^k\|_2\right) + \frac{\mu}{2}\|x - x^k\|_2^2 - f(x) \\ &= f(x) + \frac{1}{2}(\mu\|x - x^k\|_2 - \|s\|_2)\|x - x^k\|_2. \end{aligned}$$

Da wegen $\|x^k\|_k \rightarrow +\infty$ auch $\|x - x^k\|_k \rightarrow +\infty$ gilt und $\mu > 0$ ist, ist für alle hinreichend großen $k \in \mathbb{N}$ der Faktor $\frac{1}{2}(\mu\|x - x^k\|_2 - \|s\|_2) > 0$. Daraus folgt aber $c \geq +\infty$, ein Widerspruch.

Also ist $\mathcal{L}(c; X)$ beschränkt. Insgesamt wissen wir daher, dass $\mathcal{L}(c; X)$ kompakt ist. Nach Konstruktion von $\mathcal{L}(c; X)$ gilt außerdem

$$\min f(x) \quad \text{u.d.N.} \quad x \in X$$

ist äquivalent zu

$$\min f(x) \quad \text{u.d.N.} \quad x \in \mathcal{L}(c; X).$$

Auf der kompakten Menge $\mathcal{L}(c; X)$ nimmt die stetige Funktion f aber mindestens ein globales Minimum an, d.h. das Optimierungsproblem hat mindestens eine Lösung.

Eindeutigkeit der Lösung:

Da f auf X gleichmäßig konvex ist, ist f dort insbesondere auch strikt konvex. Annahme: Es gibt zwei verschiedene Lösungen $x^1 \neq x^2$ des Optimierungsproblems. Da X konvex ist, ist dann auch $\frac{1}{2}(x^1 + x^2) \in X$ und es gilt

$$f\left(\frac{1}{2}(x^1 + x^2)\right) < \frac{1}{2}(f(x^1) + f(x^2)) = f^*.$$

Dies widerspricht aber der Definition $f^* = \inf_{x \in X} f(x)$. Also hat das gegebene Optimierungsproblem genau eine Lösung $x^* \in X$. \square

2.2 Der Satz von Danskin

In diesem Abschnitt wollen wir den Satz von Danskin beweisen, der uns später die Differenzierbarkeit der Zielfunktion $V_\gamma(x)$ garantieren wird.

Satz 2.2.1. (Satz von Danskin) Seien $X \subseteq \mathbb{R}^n$ nichtleer, abgeschlossen und konvex sowie $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ stetig differenzierbar und als Funktion der zweiten Variablen $f(x, \cdot)$ gleichmäßig konkav. Für beliebiges $x \in \mathbb{R}^n$ sei $p(x) \in X$ die eindeutige Lösung des Optimierungsproblems

$$\max f(x, y) \quad \text{u.d.N.} \quad y \in X.$$

Ist die hierdurch definierte Abbildung $x \mapsto p(x)$ stetig, so ist die durch

$$g(x) := \max_{y \in X} f(x, y) = f(x, p(x))$$

definierte Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}$ stetig differenzierbar mit dem Gradienten

$$\nabla g(x) = \nabla_x f(x, y)|_{y=p(x)}.$$

Beweis. Da $f(x, \cdot)$ für festes $x \in \mathbb{R}^n$ gleichmäßig konkav ist, ist nach Satz 2.1.3 $p(x)$ wohldefiniert. Wir wollen zunächst zeigen, dass g in jedem Punkt richtungsdifferenzierbar ist, indem wir die Richtungsableitung berechnen. Sei dafür $x \in \mathbb{R}^n$ beliebig und $d \in \mathbb{R}^n$ eine beliebige Richtung. Aus den Definitionen von g und $p(x)$ folgt dann

$$g(x + td) - g(x) = f(x + td, p(x + td)) - f(x, p(x)) \geq f(x + td, p(x)) - f(x, p(x))$$

für jedes $t > 0$. Division durch $t > 0$ und anschließender Grenzübergang $t \rightarrow 0$ ergeben

$$\liminf_{t \rightarrow 0} \frac{g(x + td) - g(x)}{t} \geq \nabla_x f(x, p(x))^T d,$$

da f stetig differenzierbar, insbesondere also richtungsdifferenzierbar ist.

Analog zu oben erhält man die Ungleichung

$$g(x + td) - g(x) = f(x + td, p(x + td)) - f(x, p(x)) \leq f(x + td, p(x + td)) - f(x, p(x + td))$$

Nach dem Mittelwertsatz der Differentialrechnung gibt es für jedes $t > 0$ einen Zwischenpunkt $\xi_t \in (x, x + td)$ mit

$$f(x + td, p(x + td)) - f(x, p(x + td)) = \nabla_x f(\xi_t, p(x + td))^T td.$$

Für dieses ξ_t gilt dann mit der obigen Ungleichung

$$g(x + td) - g(x) \leq \nabla_x f(\xi_t, p(x + td))^T td.$$

Division durch $t > 0$ und anschließender Grenzübergang $t \rightarrow 0$ ergeben hier

$$\limsup_{t \rightarrow 0} \frac{g(x + td) - g(x)}{t} \leq \nabla_x f(x, p(x))^T d.$$

Hierbei haben wir die Stetigkeit von $p(x)$ und $\nabla_x f$ ausgenutzt, sowie $\xi_t \rightarrow x$ für $t \rightarrow 0$. Insgesamt folgt also die Richtungs-differenzierbarkeit von g im Punkt x in Richtung d mit

$$g'(x; d) = \nabla_x f(x, p(x))^T d.$$

Da dies für alle $d \in \mathbb{R}^n$ gilt, folgt

$$\nabla g(x) = \nabla_x f(x, p(x)).$$

Aus der Stetigkeit von $\nabla_x f$ und $p(x)$ ergibt sich die noch zu zeigende Stetigkeit von ∇g . \square

2.3 Punkt-Menge-Abbildungen

Im Satz von Danskin wird die Stetigkeit der Funktion $x \mapsto p(x)$ mit

$$p(x) = \operatorname{argmax}_{y \in X} f(x, y)$$

vorausgesetzt. Wir wollen in diesem Abschnitt genauer beleuchten, unter welchen Bedingungen diese Voraussetzung erfüllt ist. Die folgenden Aussagen findet man zum Beispiel auch in [8].

Zunächst müssen wir einige Begriffe definieren, mit denen wir Eigenschaften von Punkt-Menge-Abbildungen beschreiben können.

Definition 2.3.1. Eine Abbildung $G : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^m)$, wobei $\mathcal{P}(\mathbb{R}^m)$ die Potenzmenge von \mathbb{R}^m bezeichne, heißt

- offen in $x^* \in \mathbb{R}^n$, wenn es für alle Folgen $\{x^k\}_k \subset \mathbb{R}^n$ mit $x^k \rightarrow x^*$ und alle $y^* \in G(x^*)$ eine natürliche Zahl l und eine Folge $\{y^k\}_k \subset \mathbb{R}^m$ gibt mit $y^k \in G(x^k)$ für alle $k \geq l$ und $y^k \rightarrow y^*$,
- abgeschlossen in $x^* \in \mathbb{R}^n$, wenn für alle Folgen $\{x^k\}_k \subset \mathbb{R}^n$ mit $x^k \rightarrow x^*$ und alle konvergenten Folgen $y^k \in G(x^k)$ mit $y^k \rightarrow y^*$ gilt: $y^* \in G(x^*)$,
- stetig in $x^* \in \mathbb{R}^n$, wenn sie in $x^* \in \mathbb{R}^n$ offen und abgeschlossen ist,
- gleichmäßig kompakt in $x^* \in \mathbb{R}^n$, wenn es eine offene Umgebung $U(x^*)$ gibt, so dass der Abschluss der Menge $\bigcup_{x \in U(x^*)} G(x)$ kompakt ist.

Wir wollen diese Definitionen nun an einem kleinen Beispiel nachvollziehen.

Beispiel 2.3.2. Sei $G : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^m)$ mit $G(x) \equiv X$, wobei $X \subseteq \mathbb{R}^m$ nichtleer und abgeschlossen sei. Wir wollen zeigen, dass G dann stetig auf \mathbb{R}^n ist.

G ist offen:

Sei $x^* \in \mathbb{R}^n$ beliebig. Sei weiter $\{x^k\}_k \subset \mathbb{R}^n$ eine beliebige Folge mit $x^k \rightarrow x^*$ und sei $y^* \in G(x^*) = X$. Setzt man $y^k := y^*$, so gilt $y^k \in X = G(x^k)$ für alle $k \in \mathbb{N}$ und $y^k \rightarrow y^*$. Also ist G offen.

G ist abgeschlossen:

Sei wieder $x^* \in \mathbb{R}^n$ beliebig, $\{x^k\}_k \subset \mathbb{R}^n$ eine beliebige Folge mit $x^k \rightarrow x^*$. Sei außerdem $y^k \in G(x^k) = X$ eine Folge mit $y^k \rightarrow y^*$. Da X abgeschlossen ist, gilt $y^* \in X = G(x^*)$, folglich ist G abgeschlossen.

Nun definieren wir noch zwei Abbildungen, wie sie in ähnlicher Form auch schon im Satz von Danskin auftauchen.

Definition 2.3.3. Seien $G : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^m)$ und $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ beliebig. Dann definieren wir

$$g(x) := \sup_{y \in G(x)} f(x, y)$$

und

$$P(x) := \{y \in G(x) \mid f(x, y) = g(x)\}.$$

Auch hierzu wollen wir wieder ein Beispiel betrachten.

Beispiel 2.3.4. Sei $G : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^m)$ mit $G(x) \equiv X$, wobei $X \subseteq \mathbb{R}^m$ nichtleer, abgeschlossen und konvex sei. Weiter sei $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ eine in der zweiten Variablen gleichmäßig konkave Funktion. Dann lauten die oben definierten Funktionen

$$g(x) := \sup_{y \in X} f(x, y)$$

und

$$P(x) := \{y \in X \mid f(x, y) = g(x)\}.$$

Nach Satz 2.1.3 hat das Maximierungsproblem

$$\max f(x, y) \quad \text{u.d.N.} \quad y \in X$$

unter den gegebenen Voraussetzungen für jedes feste $x \in \mathbb{R}^n$ genau eine Lösung. Daher besteht in diesem Fall $P(x)$ immer aus genau einem Element. Folglich nimmt g immer einen endlichen Wert an und man könnte hier statt des Supremums auch ein Maximum schreiben.

Da wir die folgenden Resultate immer nur unter den in Beispiel 2.3.4 gegebenen Voraussetzungen anwenden wollen, werden wir in den folgenden Beweisen immer nur den Fall zeigen, in dem für alle $x \in \mathbb{R}^n$ die Mengen $G(x)$ und $P(x)$ nichtleer sind und $g(x)$ einen endlichen Wert annimmt. Die Beweise für die Spezialfälle, in denen dies nicht erfüllt ist, kann man in [8] nachlesen.

Lemma 2.3.5. *Sei $G : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^m)$ offen in $x^* \in \mathbb{R}^n$ und $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ unterhalbstetig in $x^* \times G(x^*)$, d.h für alle Folgen $\{(x^k, y^k)\}_k \subset \mathbb{R}^n \times \mathbb{R}^m$ mit $(x^k, y^k) \rightarrow (x^*, y^*) \in x^* \times G(x^*)$ gilt $\liminf_{k \rightarrow \infty} f(x^k, y^k) \geq f(x^*, y^*)$. Dann ist auch g unterhalbstetig in x^* .*

Beweis. Sei $\varepsilon > 0$ beliebig und $\{x^k\}_k \subset \mathbb{R}^n$ eine beliebige Folge mit $x^k \rightarrow x^*$. Wähle $y^* \in G(x^*)$ mit $g(x^*) \leq f(x^*, y^*) + \varepsilon$, zum Beispiel kann man $y^* \in P(x^*)$ wählen, da $P(x^*)$ als nichtleer vorausgesetzt war. Da G in x^* offen ist, gibt es eine Folge $y^k \in G(x^k)$ mit $y^k \rightarrow y^*$. Weil f in $x^* \times G(x^*)$ unterhalbstetig ist, folgt damit

$$g(x^*) - \varepsilon \leq f(x^*, y^*) \leq \liminf_{k \rightarrow \infty} f(x^k, y^k) \leq \liminf_{k \rightarrow \infty} g(x^k).$$

Da $\varepsilon > 0$ und $x^k \rightarrow x^*$ beliebig gewählt waren, folgt die Unterhalbstetigkeit von g in x^* . \square

Satz 2.3.6. *Sei $G : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^m)$ stetig in $x^* \in \mathbb{R}^n$ und $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ stetig auf $\mathbb{R}^n \times \mathbb{R}^m$ sowie in der zweiten Variablen konkav für jedes feste $x \in \mathbb{R}^n$. Weiter sei $G(x)$ konvex für alle x in einer offenen Umgebung von x^* und $P(x^*)$ sei nichtleer und beschränkt. Dann ist P gleichmäßig kompakt in x^* .*

Beweis. Annahme: P ist nicht gleichmäßig kompakt in x^* , d.h. es gibt keine offene Umgebung $U(x^*)$, so dass der Abschluss der Menge $\bigcup_{x \in U(x^*)} P(x)$ kompakt ist. Folglich muss für alle offenen Umgebungen $U(x^*)$ der Abschluss der Menge $\bigcup_{x \in U(x^*)} P(x)$ unbeschränkt sein und damit auch die Menge $\bigcup_{x \in U(x^*)} P(x)$ selbst unbeschränkt sein. Da $P(x^*)$ beschränkt ist, muss es dann eine Folge $x^k \rightarrow x^*$ geben, so dass eine Folge $y_1^k \in P(x^k)$ mit $\|y_1^k\| \rightarrow \infty$ existiert.

Sei $y^* \in P(x^*) \subset G(x^*)$ beliebig. Da G in x^* offen ist, gibt es eine Folge $y_2^k \in G(x^k)$ mit $y_2^k \rightarrow y^*$. Damit folgt für alle $k \in \mathbb{N}$

$$f(x^k, y_2^k) \leq g(x^k) = f(x^k, y_1^k).$$

Definiere nun

$$y_3^k := \lambda_k y_2^k + (1 - \lambda_k) y_1^k$$

mit $\lambda_k \in [0, 1]$. Aus der Konvexität von $G(x^k)$ folgt $y_3^k \in G(x^k)$ für alle $k \in \mathbb{N}$.

Da f für jedes x^k konkav in der zweiten Variablen ist, gilt

$$f(x^k, y_3^k) \geq \lambda_k f(x^k, y_2^k) + (1 - \lambda_k) f(x^k, y_1^k) \geq f(x^k, y_2^k).$$

Sei nun $c > 0$ beliebig. Wählt man nun

$$\lambda_k = \begin{cases} 1 - \frac{c}{\|y_1^k\|} & \text{für } 1 - \frac{c}{\|y_1^k\|} > 0, \\ 0 & \text{sonst,} \end{cases}$$

so gilt wegen $\|y_1^k\| \rightarrow \infty$ für alle hinreichend großen $k \in \mathbb{N}$

$$\|y_3^k\| \leq \lambda_k \|y_2^k\| + (1 - \lambda_k) \|y_1^k\| \leq \|y_2^k\| + c.$$

Wegen $y_2^k \rightarrow y^*$ ist dann die Folge $\{y_3^k\}_k$ beschränkt, hat also mindestens einen Häufungspunkt. Wir können also ohne Beschränkung der Allgemeinheit annehmen, dass die gesamte Folge gegen einen Punkt $\hat{y} \in \mathbb{R}^m$ konvergiert. Da G in x^* abgeschlossen ist, folgt aus $y_3^k \in G(x^k)$ auch $\hat{y} \in G(x^*)$. Für diesen Grenzwert gilt wegen der Stetigkeit von f

$$f(x^*, \hat{y}) = \lim_{k \rightarrow \infty} f(x^k, y_3^k) \geq \lim_{k \rightarrow \infty} f(x^k, y_2^k) = f(x^*, y^*) = g(x^*).$$

Daraus folgt $\hat{y} \in P(x^*)$. Andererseits gilt

$$\|\hat{y}\| \geq \lim_{k \rightarrow \infty} [(1 - \lambda_k) \|y_1^k\| - \lambda_k \|y_2^k\|] = c - \|y^*\|.$$

Durch geeignete Wahl von $c > 0$ kann also $\|\hat{y}\|$ beliebig vergrößert werden, was der Beschränktheit von $P(x^*)$ widerspricht. Also war die Annahme falsch und P ist gleichmäßig kompakt in x^* . \square

Satz 2.3.7. *Sei $G : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^m)$ stetig in $x^* \in \mathbb{R}^n$ und $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ stetig auf $\mathbb{R}^n \times \mathbb{R}^m$. Sei weiter P in x^* gleichmäßig kompakt, $P(x)$ in einer Umgebung von x^* nichtleer und $P(x^*)$ einelementig. Dann ist P stetig in x^* .*

Beweis. P ist abgeschlossen:

Seien $x^k \rightarrow x^*$ und $y^k \in P(x^k)$ mit $y^k \rightarrow y^*$ beliebig gewählt. Da G in x^* abgeschlossen ist, gilt dann wegen $P(x^k) \subset G(x^k)$ auch $y^* \in G(x^*)$.

Da G in x^* auch offen ist und f auf $\mathbb{R}^n \times \mathbb{R}^m$ stetig ist, ist g nach Lemma 2.3.5 in x^* unterhalbstetig. Mit der Stetigkeit von f folgt daraus

$$f(x^*, y^*) = \lim_{k \rightarrow \infty} f(x^k, y^k) = \lim_{k \rightarrow \infty} g(x^k) \geq g(x^*).$$

Nach Definition gilt daher $y^* \in P(x^*)$, d.h. P ist abgeschlossen in x^* .

P ist offen:

Seien $x^k \rightarrow x^*$, $y^* \in P(x^*)$ beliebig gewählt. Weiter wählen wir auch die Folge $y^k \in P(x^k)$ beliebig.

Annahme: y^k konvergiert nicht gegen y^* . Da P in x^* gleichmäßig kompakt ist, gibt es eine offene Umgebung $U(x^*)$, so dass der Abschluss der Menge $\bigcup_{x \in U(x^*)} P(x)$ kompakt ist, also insbesondere die Menge $\bigcup_{x \in U(x^*)} P(x)$ beschränkt ist. Für alle hinreichend großen $k \in \mathbb{N}$ gilt $x^k \in U(x^*)$. Daher muss die Folge y^k einen Häufungspunkt $\hat{y} \neq y^*$ haben. Da P wie oben gezeigt in x^* abgeschlossen ist, muss auch $\hat{y} \in P(x^*)$ gelten. Das widerspricht aber der Voraussetzung, dass $P(x^*)$ einelementig ist. Folglich ist $P(x^*)$ offen. \square

Die eigentlich gewünschte Aussage erhalten wir jetzt als Korollar aus den vorherigen Sätzen.

Korollar 2.3.8. *Sei $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ stetig und für jedes feste $x \in \mathbb{R}^n$ in der zweiten Variablen gleichmäßig konkav. Sei weiter $X \subseteq \mathbb{R}^n$ nichtleer, abgeschlossen und konvex. Für beliebiges $x \in \mathbb{R}^n$ sei $p(x) \in X$ die eindeutige Lösung des Optimierungsproblems*

$$\max f(x, y) \quad \text{u.d.N.} \quad y \in X.$$

Dann ist die Abbildung $x \mapsto p(x)$ stetig.

Beweis. Nach Satz 2.1.3 ist die Funktion p wohldefiniert. Definieren wir nun $G : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^n)$ mit $G(x) \equiv X$, so ist G nach Beispiel 2.3.2 stetig auf \mathbb{R}^n . Nach Definition 2.3.3 gilt damit

$$g(x) := \sup_{y \in X} f(x, y)$$

und

$$P(x) = \{y \in X \mid f(x, y) = g(x)\} = \{p(x)\}.$$

Folglich ist für alle $x \in \mathbb{R}^n$ die Menge $P(x)$ nichtleer und beschränkt. Da $G(x) = X$ nach Voraussetzung konvex ist, sind alle Voraussetzungen für Satz 2.3.6 erfüllt und P ist gleichmäßig kompakt für alle $x \in \mathbb{R}^n$.

Da $P(x)$ für alle $x \in \mathbb{R}^n$ einelementig ist, sind damit auch alle Voraussetzungen für Satz 2.3.7 erfüllt. Daher ist P stetig für alle $x \in \mathbb{R}^n$.

Sei nun $x^* \in \mathbb{R}^n$ beliebig und $x^k \rightarrow x^*$ eine beliebige Folge. Da P in x^* offen ist, gibt es zu $p(x^*) \in P(x^*)$ eine Folge $y^k \in P(x^k)$ mit $y^k \rightarrow p(x^*)$. Wie oben gezeigt, ist für alle $k \in \mathbb{R}^n$ die Menge $P(x^k) = \{p(x^k)\}$ einelementig. Folglich gilt $p(x^k) \rightarrow p(x^*)$. Da die Folge $x^k \rightarrow x^*$ beliebig gewählt war, ist p stetig in x^* . Da auch dieser Punkt beliebig war, folgt die Behauptung. \square

2.4 Stetige Differenzierbarkeit von $V_\gamma(x)$

Sei wieder für $x, y \in X$

$$\Psi_\gamma(x, y) = \sum_{\nu=1}^N [\theta_\nu(x^\nu, x^{-\nu}) - \theta_\nu(y^\nu, x^{-\nu})] - \frac{\gamma}{2} \|x - y\|_2^2$$

die regularisierte Nikaido-Isoda-Funktion mit einer Konstanten $\gamma > 0$ und für $x \in X$

$$V_\gamma(x) = \sup_{y \in X} \Psi_\gamma(x, y).$$

In diesem Kapitel können wir die Voraussetzungen 1.1.1 etwas abschwächen:

Voraussetzungen 2.4.1. a) Die Funktionen θ_ν seien stetig differenzierbar und als Funktionen von x^ν konvex.

b) Die zulässige Menge X sei nichtleer, abgeschlossen und konvex.

Dann können wir nun die ersten drei Aussagen aus Satz 1.1.3 beweisen. Alle Beweise in diesem Abschnitt findet man auch in [6].

Satz 2.4.2. a) $V_\gamma(x) \geq 0 \quad \forall \quad x \in X$.

b) x^* ist ein normalisiertes Nash-Gleichgewicht genau dann, wenn gilt

$$x^* \in X \quad \text{und} \quad V_\gamma(x^*) = 0.$$

c) Für alle $x \in X$ gibt es genau einen Vektor $y_\gamma(x) \in X$, so dass gilt

$$\operatorname{argmax}_{y \in X} \Psi_\gamma(x, y) = y_\gamma(x)$$

und die Abbildung $x \mapsto y_\gamma(x)$ ist stetig.

Beweis. a) Für jedes $x \in X$ gilt

$$V_\gamma(x) = \sup_{y \in X} \Psi_\gamma(x, y) \geq \Psi_\gamma(x, x) = 0.$$

b) „ \Rightarrow “ Sei zunächst x^* ein normalisiertes Nash-Gleichgewicht. Dann gilt nach Definition $x^* \in X$ und $\sup_{y \in X} \Psi(x^*, y) = 0$. Folglich ist $\Psi(x^*, y) \leq 0$ für alle $y \in X$. Damit ergibt sich für alle $y \in X$

$$\Psi_\gamma(x^*, y) = \Psi(x^*, y) - \frac{\gamma}{2} \|x - y\|_2^2 \leq \Psi(x^*, y) \leq 0.$$

Daher ist auch

$$V_\gamma(x^*) = \sup_{y \in X} \Psi_\gamma(x^*, y) \leq 0.$$

Zusammen mit Teil a) folgt $V_\gamma(x^*) = 0$.

„ \Leftarrow “ Sei nun umgekehrt $x^* \in X$ und $V_\gamma(x^*) = 0$. Dann folgt $\Psi_\gamma(x^*, y) \leq 0$ für alle $y \in X$. Wegen $x^* \in X$ ist $\sup_{y \in X} \Psi(x^*, y) \geq 0$. Wir wollen nun zeigen, dass hier Gleichheit gelten muss.

Annahme: Es gibt ein $y^* \in X$ mit $\Psi(x^*, y^*) > 0$. Da X nach Voraussetzung konvex ist, gilt für alle $\lambda \in (0, 1)$

$$\lambda x^* + (1 - \lambda)y^* \in X.$$

Da die Nikaido-Isoda-Funktion in y konkav ist, folgt für alle $\lambda \in (0, 1)$

$$\Psi(x^*, \lambda x^* + (1 - \lambda)y^*) \geq \lambda \Psi(x^*, x^*) + (1 - \lambda)\Psi(x^*, y^*) = (1 - \lambda)\Psi(x^*, y^*) > 0.$$

Damit erhalten wir

$$\begin{aligned} \Psi_\gamma(x^*, \lambda x^* + (1 - \lambda)y^*) &= \Psi(x^*, \lambda x^* + (1 - \lambda)y^*) - \frac{\gamma}{2} \|x^* - \lambda x^* - (1 - \lambda)y^*\|_2^2 \\ &= \Psi(x^*, \lambda x^* + (1 - \lambda)y^*) - \frac{\gamma}{2} (1 - \lambda)^2 \|x^* - y^*\|_2^2 \\ &\geq (1 - \lambda)\Psi(x^*, y^*) - \frac{\gamma}{2} (1 - \lambda)^2 \|x^* - y^*\|_2^2 \\ &> 0 \end{aligned}$$

für alle λ hinreichend nahe bei 1. Dies ist aber ein Widerspruch dazu, dass für alle $y \in X$ gilt $\Psi_\gamma(x^*, y) \leq 0$.

Also war die Annahme falsch und x^* ist ein normalisiertes Nash-Gleichgewicht.

c) Da die regularisierte Nikaido-Isoda-Funktion in y gleichmäßig konkav ist und X nach Voraussetzung nichtleer, abgeschlossen und konvex ist, folgt aus Satz 2.1.3 die Wohldefiniertheit von $y_\gamma(x) \in X$. Nach Korollar 2.3.8 ist die Abbildung $x \mapsto y_\gamma(x)$ stetig für alle $x \in X$. □

Als Korollar hieraus erhalten wir die Fixpunkt-Charakterisierung normalisierter Nash-Gleichgewichte.

Korollar 2.4.3. Sei $y_\gamma(x)$ wie in Satz 2.4.2 definiert. x^* ist genau dann ein normalisiertes Nash-Gleichgewicht, wenn x^* ein Fixpunkt der Abbildung $x \mapsto y_\gamma(x)$ ist, d.h. $x^* = y_\gamma(x^*)$.

Beweis. „ \Rightarrow “ Sei zunächst x^* ein normalisiertes Nash-Gleichgewicht. Dann gilt $x^* \in X$ und nach Satz 2.4.2 außerdem $V_\gamma(x^*) = 0$. Damit folgt

$$0 = V_\gamma(x^*) = \sup_{y \in X} \Psi_\gamma(x^*, y) = \Psi_\gamma(x^*, y_\gamma(x^*)).$$

Da andererseits auch $0 = \Psi_\gamma(x^*, x^*)$ gilt, folgt aus Satz 2.4.2 c) $x^* = y_\gamma(x^*)$.

„ \Leftarrow “ Sei nun umgekehrt x^* ein Fixpunkt der Abbildung $x \mapsto y_\gamma(x)$, d.h. es gilt $x^* = y_\gamma(x^*) \in X$. Außerdem folgt daraus

$$V_\gamma(x^*) = \Psi_\gamma(x^*, y_\gamma(x^*)) = \Psi_\gamma(x^*, x^*) = 0.$$

Mit Satz 2.4.2 folgt dann, dass x^* ein normalisiertes Nash-Gleichgewicht ist. \square

Schließlich wollen wir nun noch die stetige Differenzierbarkeit von $V_\gamma(x)$ beweisen.

Satz 2.4.4. *Die Funktion $V_\gamma(x)$ ist stetig differenzierbar für jedes $x \in X$ und ihr Gradient ist gegeben durch*

$$\begin{aligned} \nabla V_\gamma(x) &= \nabla_x \Psi_\gamma(x, y) \Big|_{y=y_\gamma(x)} \\ &= \sum_{\nu=1}^N [\nabla \theta_\nu(x^\nu, x^{-\nu}) - \nabla \theta_\nu(y_\gamma^\nu(x), x^{-\nu})] \\ &\quad + \begin{pmatrix} \nabla_{x^1} \theta_1(y_\gamma^1(x), x^{-1}) \\ \vdots \\ \nabla_{x^N} \theta_N(y_\gamma^N(x), x^{-N}) \end{pmatrix} - \gamma(x - y_\gamma(x)), \end{aligned}$$

wobei $y_\gamma(x)$ den Vektor aus Satz 2.4.2 bezeichne.

Beweis. Nach Voraussetzung ist X nichtleer, abgeschlossen und konvex. Nach Definition ist

$$V_\gamma(x) = \sup_{y \in X} \Psi_\gamma(x, y)$$

und die regularisierte Nikaido-Isoda-Funktion ist für jedes feste $x \in X$ in y gleichmäßig konkav. Da die Abbildung $x \mapsto y_\gamma(x)$ stetig ist, folgt mit dem Satz von Danskin, vgl. Satz 2.2.1, dass $V_\gamma(x)$ stetig differenzierbar ist mit dem Gradienten

$$\nabla V_\gamma(x) = \nabla_x \Psi_\gamma(x, y) \Big|_{y=y_\gamma(x)}.$$

Setzen wir die Definition von

$$\Psi_\gamma(x, y) = \sum_{\nu=1}^N \left[\theta_\nu(x^\nu, x^{-\nu}) - \theta_\nu(y^\nu, x^{-\nu}) - \frac{\gamma}{2} \|x^\nu - y^\nu\|_2^2 \right]$$

ein, so folgt

$$\begin{aligned} \nabla_{x^\mu} \Psi_\gamma(x, y) &= \sum_{\nu=1}^N \nabla_{x^\mu} \theta_\nu(x^\nu, x^{-\nu}) - \sum_{\substack{\nu=1 \\ \nu \neq \mu}}^N \nabla_{x^\mu} \theta_\nu(y^\nu, x^{-\nu}) - \gamma(x^\mu - y^\mu) \\ &= \sum_{\nu=1}^N [\nabla_{x^\mu} \theta_\nu(x^\nu, x^{-\nu}) - \nabla_{x^\mu} \theta_\nu(y^\nu, x^{-\nu})] + \nabla_{x^\mu} \theta_\mu(y^\mu, x^{-\mu}) - \gamma(x^\mu - y^\mu). \end{aligned}$$

Fügt man alle partiellen Ableitungen zum Gradienten zusammen, so erhält man

$$\nabla_x \Psi_\gamma(x, y) = \sum_{\nu=1}^N [\nabla \theta_\nu(x^\nu, x^{-\nu}) - \nabla \theta_\nu(y^\nu, x^{-\nu})] + \begin{pmatrix} \nabla_{x^1} \theta_1(y^1, x^{-1}) \\ \vdots \\ \nabla_{x^N} \theta_N(y^N, x^{-N}) \end{pmatrix} - \gamma(x - y).$$

Wenn wir nun noch den Punkt $y = y_\gamma(x)$ einsetzen, so erhalten wir die gesuchte Formel für den Gradienten. □

Kapitel 3

Semismoothness der Ableitung der Zielfunktion $V_\gamma(x)$

Wir wollen uns in dieser Arbeit mit einem Verfahren zur Lösung des Problems

$$\min_x V_\gamma(x) \quad \text{u.d.N.} \quad g(x) \leq 0, h(x) = 0,$$

wie es in Abschnitt 1.1 definiert war, beschäftigen. Wie in Kapitel 2 bewiesen, ist die Zielfunktion $V_\gamma(x)$ unter den Voraussetzungen 1.1.1 bzw 1.1.2 zwar stetig differenzierbar. Sie ist aber im Allgemeinen nicht zweimal stetig differenzierbar. Wir wollen in diesem Kapitel, genauer in Abschnitt 3.3, zeigen, dass $\nabla V_\gamma(x)$ unter den obigen Voraussetzungen zumindest noch semismooth bzw. strongly semismooth ist. Um dies beweisen zu können benötigen wir allerdings gewisse Eigenschaften semismoothter Funktionen, wie sie in Abschnitt 3.1 gesammelt sind, und einen Satz über implizite Abbildungen bei semismoothen Funktionen, den wir in Abschnitt 3.2 herleiten werden.

3.1 Semismoothe und strongly semismoothe Funktionen

In diesem Abschnitt wollen wir definieren, was wir unter semismoothen bzw. strongly semismoothen Funktionen verstehen und grundlegende Eigenschaften dieser Funktionen darstellen.

Ist $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine lokal Lipschitz-stetige Funktion, so besagt ein Satz von Rademacher, vergleiche [19], dass f fast überall differenzierbar ist. Sei $D_f \subseteq \mathbb{R}^n$ die Menge aller Punkte, in denen f differenzierbar ist. Dann können wir die folgenden Mengen definieren:

Definition 3.1.1. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine lokal Lipschitz-stetige Funktion. Dann heißt für alle $x \in \mathbb{R}^n$

$$\partial_B f(x) := \{H \in \mathbb{R}^{m \times n} \mid \exists \{x^k\}_k \subseteq D_f : x^k \rightarrow x, f'(x^k) \rightarrow H\}$$

das B-Subdifferential von f in x . Seine konvexe Hülle

$$\partial f(x) := \text{conv } \partial_B f(x) \subseteq \mathbb{R}^{m \times n}$$

ist Clarkes verallgemeinerte Jacobi-Matrix von f in x . Weiter heißt

$$\partial_C f(x) := (\partial f_1(x)^T \times \dots \times \partial f_m(x)^T)^T \subseteq \mathbb{R}^{m \times n}$$

das C-Subdifferential von f in x . Analog zu $\partial_C f(x)$ definieren wir das b-Subdifferential

$$\partial_b f(x) := (\partial_B f_1(x)^T \times \dots \times \partial_B f_m(x)^T)^T \subseteq \mathbb{R}^{m \times n}.$$

Nach [2], Proposition 2.6.2, gelten immer die folgenden Inklusionen

$$\partial_B f(x) \subseteq \partial f(x) \subseteq \partial_C f(x).$$

Analog gilt auch

$$\partial_B f(x) \subseteq \partial_b f(x) \subseteq \partial_C f(x).$$

Nach [11], Satz 4.3 bis 4.5, gelten außerdem die folgenden Aussagen:

Satz 3.1.2. *Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine lokal Lipschitz-stetige Funktion und $x \in \mathbb{R}^n$. Dann gelten:*

- a) $\partial f(x)$ ist nichtleer, konvex und kompakt. $\partial_B f(x)$ ist nichtleer und kompakt.
- b) Die Abbildung $x \mapsto \partial f(x)$ ist abgeschlossen, d.h. für alle Folgen $\{x^k\}_k \rightarrow x$ und alle konvergenten Folgen $\{H_k\}_k \rightarrow H$ mit $H_k \in \partial f(x^k)$ gilt $H \in \partial f(x)$. Die Abbildung $x \mapsto \partial_B f(x)$ ist ebenfalls abgeschlossen.
- c) Die Abbildung $x \mapsto \partial f(x)$ ist oberhalbstetig, d.h. für alle $\varepsilon > 0$ gibt es ein $\delta > 0$, so dass für alle $y \in K_\delta(x)$ gilt

$$\partial f(y) \subseteq \partial f(x) + K_\varepsilon(0).$$

Die Abbildung $x \mapsto \partial_B f(x)$ ist ebenfalls oberhalbstetig.

Die Abbildungen $x \mapsto \partial_C f(x)$ bzw. $x \mapsto \partial_b f(x)$ erben jeweils die drei obengenannten Eigenschaften ihrer Komponentenfunktionen $\partial f_i(x)$ bzw. $\partial_B f_i(x)$.

Für die Bestimmung der verallgemeinerten Jacobi-Matrix einer Funktion sind die folgenden zwei Resultate hilfreich.

Lemma 3.1.3. *Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine lokal Lipschitz-stetige Funktion. Sei weiter $S \subset \mathbb{R}^n$ eine (Lebesgue-)Nullmenge und $x \in \mathbb{R}^n$ beliebig. Dann gelten:*

Ist $m = 1$, so ist

$$\partial f(x) = \text{conv}\{H \in \mathbb{R}^n \mid \exists \{x^k\}_k \subseteq D_f \setminus S : x^k \rightarrow x, f'(x^k) \rightarrow H\}.$$

Ist $m > 1$, so gilt zumindest für alle $d \in \mathbb{R}^n$

$$\partial f(x) \cdot d = \text{conv}\{H \in \mathbb{R}^{m \times n} \mid \exists \{x^k\}_k \subseteq D_f \setminus S : x^k \rightarrow x, f'(x^k) \rightarrow H\} \cdot d.$$

Den Beweis dieser beiden Aussagen findet man in [2], Theorem 2.5.1 und Proposition 2.6.4.

Für Verknüpfungen lokal Lipschitz-stetiger Funktionen gelten die folgenden Kettenregeln:

Satz 3.1.4. Seien $F : \mathbb{R}^a \rightarrow \mathbb{R}^b$ und $g : \mathbb{R}^b \rightarrow \mathbb{R}$ lokal Lipschitz-stetig. Dann ist auch $g \circ F$ lokal Lipschitz-stetig und es gilt

$$\partial(g \circ F)(x) \subseteq \text{conv}\{\partial g(F(x)) \cdot \partial F(x)\}.$$

Ist g sogar stetig differenzierbar, so gilt

$$\partial(g \circ F)(x) = g'(F(x)) \cdot \partial F(x).$$

Seien $F : \mathbb{R}^a \rightarrow \mathbb{R}^b$ und $G : \mathbb{R}^b \rightarrow \mathbb{R}^c$ lokal Lipschitz-stetig. Dann ist auch $G \circ F$ lokal Lipschitz-stetig und es gilt für alle Vektoren $d \in \mathbb{R}^a$

$$\partial(G \circ F)(x) \cdot d \subseteq \text{conv}\{\partial G(F(x)) \cdot \partial F(x) \cdot d\}.$$

Ist G sogar stetig differenzierbar, so gilt

$$\partial(G \circ F)(x) \cdot d = G'(F(x)) \cdot \partial F(x) \cdot d.$$

Den Beweis dieses Satzes findet man in [2], Theorem 2.6.6 und nachfolgendes Korollar sowie dem Korollar nach Proposition 2.2.1.

Mit Hilfe der verallgemeinerten Jacobi-Matrix können wir jetzt (strongly) semismoothe Funktionen definieren.

Definition 3.1.5. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine lokal Lipschitz-stetige Funktion. f heißt semismooth in $x \in \mathbb{R}^n$, wenn f in x richtungsdifferenzierbar ist und für alle $d \rightarrow 0$ und alle $H \in \partial f(x + d)$ gilt

$$\|Hd - f'(x; d)\| = o(\|d\|).$$

f heißt strongly semismooth in $x \in \mathbb{R}^n$, wenn f in x richtungsdifferenzierbar ist und für alle $d \rightarrow 0$ und alle $H \in \partial f(x + d)$ gilt

$$\|Hd - f'(x; d)\| = O(\|d\|^2).$$

Offensichtlich ist jede strongly semismoothe Funktion semismooth. Abschließend wollen wir in diesem Abschnitt noch einige Eigenschaften von (strongly) semismoothen Funktionen sammeln, die sich später als sehr hilfreich herausstellen werden.

Satz 3.1.6. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine lokal Lipschitz-stetige und richtungsdifferenzierbare Funktion und $x \in \mathbb{R}^n$. Dann gelten:

a) f ist B-differenzierbar in x , d.h. für alle $d \rightarrow 0$ gilt

$$\|f(x + d) - f(x) - f'(x; d)\| = o(\|d\|).$$

Insbesondere ist daher jede semismoothe Funktion B-differenzierbar in x .

b) Ist f in \mathbb{R}^n stetig differenzierbar, so ist f dort semismooth.

c) Kompositionen semismoother Funktionen sind wieder semismooth.

d) Die folgenden Aussagen sind äquivalent:

(1) f ist semismooth in x , d.h.

$$\|Hd - f'(x; d)\| = o(\|d\|) \text{ für alle } d \rightarrow 0 \text{ und alle } H \in \partial f(x + d).$$

(2) f ist semismooth in x in jeder Komponente f_i , d.h.

$$\|h_i d - f'_i(x; d)\| = o(\|d\|) \text{ für alle } d \rightarrow 0 \text{ und alle } h_i \in \partial f_i(x + d)$$

gilt für alle $i = 1, \dots, m$.

(3) Für alle $d \rightarrow 0$ gilt

$$\|f'(x; d) - f'(x + d; d)\| = o(\|d\|).$$

(4) Für alle $d \rightarrow 0$ gilt

$$\|f(x + d) - f(x) - f'(x + d; d)\| = o(\|d\|).$$

(5) Für alle $d \rightarrow 0$ und alle $H \in \partial f(x + d)$ gilt

$$\|f(x + d) - f(x) - Hd\| = o(\|d\|).$$

Die Aussagen a) bis c) sind aus [11], vergleiche Lemma 4.12, Lemma 4.16, Korollar 4.23 und Satz 4.22, bekannt ebenso wie die Äquivalenz der Aussagen (1), (2), (3) und (5) in d). Mit Hilfe von a) lässt sich außerdem leicht die Äquivalenz der (3) und (4) Aussagen in d) beweisen. Nun wollen wir ein ähnliches Resultat für strongly semismoothe Funktionen formulieren.

Satz 3.1.7. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine lokal Lipschitz-stetige und richtungsdifferenzierbare Funktion und $x \in \mathbb{R}^n$.

a) Ist f strongly semismooth in x , so ist f dort B -differenzierbar zweiten Grades, d.h. für alle $d \rightarrow 0$ gilt

$$\|f(x + d) - f(x) - f'(x; d)\| = O(\|d\|^2).$$

b) Ist f in \mathbb{R}^n stetig differenzierbar und f' dort lokal Lipschitz-stetig, so ist f dort strongly semismooth.

c) Kompositionen strongly semismoothe Funktionen sind wieder strongly semismooth.

d) Die folgenden Aussagen sind äquivalent:

(1) f ist strongly semismooth in x , d.h.

$$\|Hd - f'(x; d)\| = O(\|d\|^2) \text{ für alle } d \rightarrow 0 \text{ und alle } H \in \partial f(x + d).$$

(2) f ist strongly semismooth in x in jeder Komponente f_i , d.h.

$$\|h_i d - f'_i(x; d)\| = O(\|d\|^2) \text{ für alle } d \rightarrow 0 \text{ und alle } h_i \in \partial f_i(x + d)$$

gilt für alle $i = 1, \dots, m$.

(3) Für alle $d \rightarrow 0$ gilt

$$\|f'(x; d) - f'(x + d; d)\| = O(\|d\|^2).$$

e) Ist f strongly semismooth in x , so gilt für alle $d \rightarrow 0$

$$\|f(x + d) - f(x) - f'(x + d; d)\| = O(\|d\|^2).$$

Den Beweis von Aussage a) findet man in [18], Lemma 2.3. Die Aussagen b) und c) sind aus [11] bekannt, siehe Lemma 4.16, Korollar 4.23 und Satz 4.22, ebenso wie die Äquivalenz von (1) und (2) in Aussage d). Die Äquivalenz von (1) und (3) in Aussage d) wird ebenfalls in [18], Lemma 2.3 bewiesen. Die noch fehlende Aussage e) erhält man, indem man Aussage a) mit Aussage (3) kombiniert.

Bemerkung 3.1.8. Die Minimum-Funktion $\varphi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ definiert durch $\varphi(a, b) = \min\{a, b\}$ ist strongly semismooth in $\mathbb{R} \times \mathbb{R}$. Nach Satz 3.1.7, Aussage b), gilt dies für alle $(a, b) \neq (0, 0)$. Für den Fall $(a, b) = (0, 0)$ zeigt man die Behauptung, indem man die Definition von strongly semismooth nachrechnet.

3.2 Satz über implizite Abbildungen für (strongly) semismoothe Funktionen

Um zu beweisen, dass die Zielfunktion unseres Ausgangsproblems (1.2) unter den gegebenen Voraussetzungen eine (strongly) semismoothe Ableitung besitzt, benötigen wir einen Satz über implizite Abbildungen. Einen solchen wollen wir in diesem Abschnitt herleiten. Doch zunächst wollen wir einen Begriff definieren, der im folgenden häufiger auftauchen wird.

Definition 3.2.1. Seien $U, V \subseteq \mathbb{R}^n$ offen und $f : U \rightarrow V$ eine Abbildung.

- a) f heißt lokaler Lipschitz-Homeomorphismus in $x_0 \in U$, wenn offene Umgebungen $X \subseteq U$ von x_0 und $Y \subseteq V$ von $f(x_0)$ existieren, so dass $f : X \rightarrow Y$ invertierbar ist mit einer lokalen Inversen $f_{x_0}^{-1} : Y \rightarrow X$ und die beiden Funktionen f und $f_{x_0}^{-1}$ auf X bzw. Y Lipschitz-stetig sind.
- b) f heißt lokaler Lipschitz-Homeomorphismus in U , wenn f ein lokaler Lipschitz-Homeomorphismus in allen $x_0 \in U$ ist.
- c) f heißt globaler Lipschitz-Homeomorphismus in U , wenn f auf ganz U invertierbar ist und f sowie die Inverse f^{-1} Lipschitz-stetig auf ganz U bzw. V sind.

Eine in diesem Zusammenhang nützliche Tatsache findet man in [22], Lemma 3.2.2:

Lemma 3.2.2. Sei $U \subseteq \mathbb{R}^n$ eine offene Menge und $f : U \rightarrow \mathbb{R}^n$ lokal Lipschitz-stetig in U . Dann ist f genau dann ein lokaler Lipschitz-Homeomorphismus in $x_0 \in U$, wenn es eine Umgebung $X \subseteq U$ von x_0 und eine Konstante $l > 0$ gibt, so dass für alle $x_1, x_2 \in X$ gilt

$$\|f(x_1) - f(x_2)\| \geq l\|x_1 - x_2\|.$$

Ein weiteres, zumindest hinreichendes, Kriterium stammt aus [2], Theorem 7.1.1:

Satz 3.2.3. *Sei $U \subseteq \mathbb{R}^n$ eine offene Menge und $f : U \rightarrow \mathbb{R}^n$ lokal Lipschitz-stetig. Sind alle Elemente der verallgemeinerten Jacobi-Matrix $\partial f(x_0)$ von f in $x_0 \in U$ regulär, so ist f ein lokaler Lipschitz-Homeomorphismus in x_0 .*

In diesem Zusammenhang werden wir später auch die folgende, aus [11], Lemma 4.26, bekannte Aussage verwenden:

Lemma 3.2.4. *Sei $U \subseteq \mathbb{R}^n$ eine offene Menge und $f : U \rightarrow \mathbb{R}^n$ lokal Lipschitz-stetig. Sind alle Matrizen des B-Subdifferentials $\partial_B f(x_0)$ von f in $x_0 \in U$ regulär, so gibt es Konstanten $\delta, c > 0$, so dass für alle $x \in K_\delta(x_0) \cap U$ gilt: Alle Matrizen $H \in \partial_B f(x)$ sind ebenfalls regulär und es gilt $\|H^{-1}\| \leq c$.*

Eine analoge Aussage gilt auch, falls alle Matrizen in $\partial f(x_0)$, $\partial_C f(x_0)$ oder $\partial_b f(x_0)$ regulär sind.

Genaugenommen wurde diese Aussage in [11] nur für das B-Subdifferential $\partial_B f(x)$ bewiesen, der Beweis für die anderen Subgradienten geht allerdings völlig analog unter Verwendung von Satz 3.1.2.

Da wir auch in diesem Abschnitt wieder mit Richtungsableitungen arbeiten werden, wollen wir an dieser Stelle einige wichtige Eigenschaften wiederholen.

Lemma 3.2.5. *Sei $U \subseteq \mathbb{R}^n$ eine offene Menge und $f : U \rightarrow \mathbb{R}^n$ richtungsdifferenzierbar in $x_0 \in U$.*

- a) *Die Abbildung $d \mapsto f'(x_0; d)$ ist positiv homogen, d.h. für alle $t > 0$ und alle $d \in \mathbb{R}^n$ gilt $f'(x_0; td) = tf'(x_0; d)$.*
- b) *Ist f lokal Lipschitz-stetig in x_0 mit einer Lipschitz-Konstanten $L > 0$, so ist $d \mapsto f'(x_0; d)$ global Lipschitz-stetig mit der gleichen Lipschitz-Konstanten L .*
- c) *Ist die Abbildung $d \mapsto f'(x_0; d)$ lokal invertierbar im Ursprung, so ist sie invertierbar auf \mathbb{R}^n und ihre Inverse ist ebenfalls positiv homogen.*
- d) *Ist die Abbildung $d \mapsto f'(x_0; d)$ ein lokaler Lipschitz-Homeomorphismus im Ursprung, so ist sie ein globaler Lipschitz-Homeomorphismus.*

Die Aussagen a) und b) sind aus [11], Lemma 4.16, bekannt, die Aussagen c) und d) findet man als Proposition 3.2.1 in [22]. Bevor wir den Satz über inverse Abbildungen für (strongly) semismoothe Funktionen beweisen können, benötigen wir noch den folgenden Satz über die Richtungs-differenzierbarkeit von inversen Abbildungen.

Satz 3.2.6. *Sei $U \subseteq \mathbb{R}^n$ eine offene Menge und $f : U \rightarrow \mathbb{R}^n$ richtungsdifferenzierbar in $x_0 \in U$. Ist f ein lokaler Lipschitz-Homeomorphismus in x_0 , so ist die Richtungsableitung $f'(x_0, \cdot)$ ein globaler Lipschitz-Homeomorphismus. Außerdem ist auch die lokale Inverse $g := f_{x_0}^{-1}$ richtungsdifferenzierbar in $f(x_0)$ und ihre Richtungsableitung $g'(f(x_0), \cdot)$ ist die Inverse von $f'(x_0, \cdot)$.*

Beweis. Eine etwas allgemeinere Version des obigen Satzes findet man in [22], Theorem 3.2.3. Wir wollen den Satz aber an dieser Stelle trotzdem beweisen.

Nach Lemma 3.2.5 b) ist $f'(x_0, \cdot)$ global Lipschitz-stetig, insbesondere also Lipschitz-stetig in einer Umgebung des Ursprungs. Da f ein lokaler Lipschitz-Homeomorphismus in x_0 ist, gibt es nach Lemma 3.2.2 eine Umgebung $X \subseteq U$ von x_0 und eine Konstante $l > 0$, so dass für alle $x_1, x_2 \in X$ gilt

$$\|f(x_1) - f(x_2)\| \geq l\|x_1 - x_2\|.$$

Also gilt für alle $d_1, d_2 \in \mathbb{R}^n$

$$\begin{aligned} \|f'(x_0; d_1) - f'(x_0; d_2)\| &= \left\| \lim_{t \rightarrow 0^+} \frac{f(x_0 + td_1) - f(x_0)}{t} - \lim_{t \rightarrow 0^+} \frac{f(x_0 + td_2) - f(x_0)}{t} \right\| \\ &= \left\| \lim_{t \rightarrow 0^+} \frac{f(x_0 + td_1) - f(x_0 + td_2)}{t} \right\| \\ &\geq \lim_{t \rightarrow 0^+} l \left\| \frac{t(d_1 - d_2)}{t} \right\| \\ &= l\|d_1 - d_2\|. \end{aligned}$$

Nach Lemma 3.2.2 ist $f'(x_0, \cdot)$ daher ein globaler Lipschitz-Homeomorphismus. Nach Lemma 3.2.5 Teil a) ist auch die Inverse dieser Abbildung positiv homogen.

Um die zweite Behauptung zu beweisen, definieren wir $h(d) := f'(x_0; d)$ und wählen $L > 0$ als gemeinsame Lipschitz-Konstante von h^{-1} in \mathbb{R}^n und von f^{-1} in einer Umgebung Y von $f(x_0)$. Dies ist möglich, da h ein globaler Lipschitz-Homeomorphismus und f ein lokaler Lipschitz-Homeomorphismus in x_0 ist. Sei weiter $d \in \mathbb{R}^n$ eine beliebige Richtung mit o.B.d.A $\|d\| = 1$. Dann gilt für hinreichend kleines $t > 0$ auch $f(x_0) + td \in Y$ und wir können die folgende Hilfsfunktion definieren:

$$y(t) := f^{-1}(f(x_0) + td) - f^{-1}(f(x_0)) = f^{-1}(f(x_0) + td) - x_0.$$

Dann folgt sofort

$$td = f(x_0 + y(t)) - f(x_0)$$

und wegen der lokalen Lipschitz-Stetigkeit von f^{-1}

$$\|y(t)\| = \|f^{-1}(f(x_0) + td) - f^{-1}(f(x_0))\| \leq L\|td\| = Lt.$$

Insbesondere folgt daraus

$$\lim_{t \rightarrow 0^+} \|y(t)\| = 0.$$

Unter Verwendung der positiven Homogenität von h^{-1} und der Definition von L erhalten

wir dann die folgenden Abschätzungen für alle $t > 0$:

$$\begin{aligned}
0 &\leq \left\| \frac{f^{-1}(f(x_0) + td) - f^{-1}(f(x_0))}{t} - h^{-1}(d) \right\| \\
&= \left\| \frac{f^{-1}(f(x_0) + td) - f^{-1}(f(x_0)) - h^{-1}(td)}{t} \right\| \\
&= \left\| \frac{h^{-1}(h(y(t))) - h^{-1}(td)}{t} \right\| \\
&\leq L \frac{\|h(y(t)) - td\|}{t} \\
&= L \frac{\|h(y(t)) - f(x_0 + y(t)) + f(x_0)\|}{t} \\
&\leq L^2 \frac{\|f(x_0 + y(t)) - f(x_0) - f'(x_0; y(t))\|}{\|y(t)\|}.
\end{aligned}$$

Da f nach Satz 3.1.6 a) B-differenzierbar in x_0 ist und $\lim_{t \rightarrow 0^+} \|y(t)\| = 0$ gilt, folgt aus der obigen Abschätzung

$$\begin{aligned}
0 &\leq \lim_{t \rightarrow 0^+} \left\| \frac{f^{-1}(f(x_0) + td) - f^{-1}(f(x_0))}{t} - h^{-1}(d) \right\| \\
&\leq \lim_{t \rightarrow 0^+} L^2 \frac{\|f(x_0 + y(t)) - f(x_0) - f'(x_0; y(t))\|}{\|y(t)\|} \\
&= 0.
\end{aligned}$$

Also ist $g := f_{x_0}^{-1}$ richtungsdifferenzierbar in $f(x_0)$ und es gilt $g'(f(x_0); d) = h^{-1}(d)$ für alle $d \in \mathbb{R}^n$. Dies zeigt die zweite Behauptung. \square

Damit können wir jetzt den Satz über inverse Abbildungen für (strongly) semismoothe Funktionen beweisen, aus dem wir dann das gewünschte Resultat für implizite Funktionen herleiten können.

Satz 3.2.7. *Sei $U \subseteq \mathbb{R}^n$ offen und $f : U \rightarrow \mathbb{R}^n$ eine lokal Lipschitz-stetige und richtungsdifferenzierbare Funktion. Sei weiter $x_0 \in U$ und alle Matrizen in $\partial f(x_0)$ seien regulär. Dann gelten:*

- a) *Es existieren Umgebungen X von x_0 und Y von $f(x_0)$, so dass $f : X \rightarrow Y$ ein Lipschitz-Homeomorphismus ist und dass die Inverse f^{-1} richtungsdifferenzierbar in allen $y = f(x) \in Y$ ist, wobei die Richtungsableitungen durch die Inverse von $f'(x; \cdot)$ gegeben sind. Hierbei können X und Y so gewählt werden, dass für alle $x \in X$ ebenfalls alle Matrizen in $\partial f(x)$ regulär sind.*
- b) *Ist f zusätzlich semismooth in X , so ist f^{-1} semismooth in Y .*
- c) *Ist f sogar strongly semismooth in X , so ist auch f^{-1} strongly semismooth in Y .*

Beweis. In dem Beweis dieses Satzes werden einige Ideen aus dem Beweis von Theorem 2 in [4] verwendet.

- a) Nach Satz 3.2.3 gibt es Umgebungen X von x_0 und Y von $f(x_0)$, so dass $f : X \rightarrow Y$ ein Lipschitz-Homeomorphismus mit inverser Funktion $g : Y \rightarrow X$ ist. Wegen Lemma 3.2.4 können X und Y sogar so gewählt werden, dass für alle $x \in X$ ebenfalls alle Matrizen in $\partial f(x)$ regulär sind. Diese Eigenschaft werden wir für Teil b) und c) benötigen. Insbesondere ist f in jedem $x \in X$ ein lokaler Lipschitz-Homeomorphismus und die lokalen Inversen f_x^{-1} stimmen auf Y mit g überein. Nach Satz 3.2.6 ist g daher in allen $y = f(x) \in Y$ richtungsdifferenzierbar und die Richtungsableitungen $g'(f(x); \cdot)$ sind die Inversen von $f'(x; \cdot)$.
- b) Nach Teil a) ist $g := f^{-1}$ Lipschitz-stetig und richtungsdifferenzierbar in Y . Nach Satz 3.1.6 d) genügt es zu zeigen, dass für alle $y \in Y$ und alle Folgen $d^k \rightarrow 0$ gilt

$$\|g(y + d^k) - g(y) - g'(y + d^k; d^k)\| = o(\|d^k\|).$$

Sei also $y \in Y$ beliebig und $d^k \rightarrow 0$ eine beliebige Folge, von der wir o.B.d.A. voraussetzen können, dass immer $d^k \neq 0$ und $y^k := y + d^k \in Y$ gilt. Hierzu definieren wir $x^k := g(y^k)$ und $x := g(y)$. Dann gilt $x^k - x = g(y + d^k) - g(y) \rightarrow 0$ und, wie wir gleich zeigen werden,

$$m := \inf_{k \in \mathbb{N}} \frac{\|y^k - y\|}{\|x^k - x\|} > 0.$$

Anderenfalls würde, evtl. auf einer Teilfolge, gelten

$$0 = \lim_{k \rightarrow \infty} \frac{y^k - y}{\|x^k - x\|} = \lim_{k \rightarrow \infty} \frac{f(x^k) - f(x)}{\|x^k - x\|}.$$

Nach Satz 3.1.6 d), Aussage (5), gilt aufgrund der Semismoothness von f aber auch für alle $H_k \in \partial f(x^k)$

$$0 = \lim_{k \rightarrow \infty} \frac{f(x^k) - f(x) - H_k(x^k - x)}{\|x^k - x\|}.$$

Zusammen mit dem obigen Grenzwert folgt daraus

$$0 = \lim_{k \rightarrow \infty} H_k \frac{(x^k - x)}{\|x^k - x\|}.$$

Da die Abbildung $x \mapsto \partial f(x)$ abgeschlossen und oberhalbstetig mit kompakten Bildern $\partial f(x)$ ist, konvergiert die Folge $\{H_k\}_k$, evtl. auf einer Teilfolge, gegen eine Element $H \in \partial f(x)$. Andererseits konvergiert auch die Folge $\left\{ \frac{x^k - x}{\|x^k - x\|} \right\}_k$, evtl. auf einer weiteren Teilfolge, gegen einen Vektor $d \neq 0$ und wir erhalten

$$0 = \lim_{k \rightarrow \infty} H_k \frac{(x^k - x)}{\|x^k - x\|} = Hd.$$

Dies widerspricht aber der Voraussetzung, dass alle Matrizen in $\partial f(x)$ regulär sind. Sei weiter $L > 0$ eine Lipschitz-Konstante für g auf Y . Wegen Lemma 3.2.5 b) ist

L dann auch eine globale Lipschitz-Konstante von $g'(y^k; \cdot)$ für alle $k \in \mathbb{N}$ und von $g'(y; \cdot)$. Mit der Semismoothness von f und Satz 3.2.6 folgt dann die Behauptung:

$$\begin{aligned}
& \lim_{k \rightarrow \infty} \frac{\|g(y + d^k) - g(y) - g'(y + d^k; d^k)\|}{\|d^k\|} = \\
& \lim_{k \rightarrow \infty} \frac{\|g(y^k) - g(y) - g'(y^k; y^k - y)\|}{\|y^k - y\|} \leq \\
& \lim_{k \rightarrow \infty} \frac{1}{m} \frac{\|x^k - x - g'(y^k; f(x^k) - f(x))\|}{\|x^k - x\|} = \\
& \lim_{k \rightarrow \infty} \frac{1}{m} \frac{\|g'(y^k; f'(x^k; x^k - x)) - g'(y^k; f(x^k) - f(x))\|}{\|x^k - x\|} \leq \\
& \lim_{k \rightarrow \infty} \frac{L}{m} \frac{\|f'(x^k; x^k - x) - f(x^k) + f(x)\|}{\|x^k - x\|} = 0.
\end{aligned}$$

c) Nach Teil a) ist $g := f^{-1}$ Lipschitz-stetig und richtungsdifferenzierbar in Y . Nach Satz 3.1.7 d) genügt es zu zeigen, dass für alle $y \in Y$ und alle Folgen $d^k \rightarrow 0$ gilt

$$\|g'(y; d^k) - g'(y + d^k; d^k)\| = O(\|d^k\|^2).$$

Sei also $y \in Y$ beliebig und $d^k \rightarrow 0$ eine beliebige Folge, von der wir o.B.d.A. voraussetzen können, dass immer $d^k \neq 0$ und $y^k := y + d^k \in Y$ gilt. Hierzu definieren wir $x^k := g(y^k)$ und $x := g(y)$. Dann gilt $x^k - x = g(y + d^k) - g(y) \rightarrow 0$ und, wie in Teil b),

$$m := \inf_{k \in \mathbb{N}} \frac{\|y^k - y\|}{\|x^k - x\|} > 0.$$

Wie in Teil b) kann man zeigen, dass gilt

$$\begin{aligned}
& \lim_{k \rightarrow \infty} \frac{\|g(y + d^k) - g(y) - g'(y + d^k; d^k)\|}{\|d^k\|^2} \leq \\
& \lim_{k \rightarrow \infty} \frac{L}{m^2} \frac{\|f'(x^k; x^k - x) - f(x^k) + f(x)\|}{\|x^k - x\|^2}
\end{aligned}$$

und

$$\begin{aligned}
& \lim_{k \rightarrow \infty} \frac{\|g(y + d^k) - g(y) - g'(y; d^k)\|}{\|d^k\|^2} \leq \\
& \lim_{k \rightarrow \infty} \frac{1}{m} \frac{\|x^k - x - g'(y; f(x^k) - f(x))\|}{\|x^k - x\|^2} = \\
& \lim_{k \rightarrow \infty} \frac{1}{m} \frac{\|g'(y; f'(x; x^k - x)) - g'(y; f(x^k) - f(x))\|}{\|x^k - x\|^2} \leq \\
& \lim_{k \rightarrow \infty} \frac{L}{m^2} \frac{\|f'(x; x^k - x) - f(x^k) + f(x)\|}{\|x^k - x\|^2}.
\end{aligned}$$

Da f strongly semismooth in X ist, gilt nach Satz 3.1.7 e) für alle $k \in \mathbb{N}$

$$\|f(x^k) - f(x) - f'(x^k; x^k - x)\| = O(\|x^k - x\|^2)$$

und nach Satz 3.1.7 a) gilt außerdem

$$\|f(x^k) - f(x) - f'(x; x^k - x)\| = O(\|x^k - x\|^2).$$

Mit den obigen Abschätzungen folgt daraus

$$\begin{aligned} \|g(y + d^k) - g(y) - g'(y + d^k; d^k)\| &= O(\|d^k\|^2), \\ \|g(y + d^k) - g(y) - g'(y; d^k)\| &= O(\|d^k\|^2). \end{aligned}$$

Kombiniert man beide Ergebnisse, so folgt die Behauptung wegen

$$\|g'(y; d^k) - g'(y + d^k; d^k)\| \leq \|g(y + d^k) - g(y) - g'(y + d^k; d^k)\| + \|g(y + d^k) - g(y) - g'(y; d^k)\|.$$

□

Zum Abschluss dieses Abschnitts erhalten wir nun den versprochenen Satz über implizite Abbildungen für (strongly) semismoothe Funktionen, für den wir nur noch die folgende Definition brauchen.

Definition 3.2.8. Sei $H : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^n$, $(x, y) \mapsto H(x, y)$ eine lokal Lipschitz-stetige Funktion. Dann bezeichnen wir mit $\pi_y \partial H(x, y)$ die Menge aller Matrizen $M \in \mathbb{R}^{n \times n}$, für die eine Matrix $N \in \mathbb{R}^{n \times m}$ existiert, so dass $[N, M] \in \partial H(x, y)$ gilt. Analog definieren wir die Menge $\pi_x \partial H(x, y) \subseteq \mathbb{R}^{n \times m}$.

Satz 3.2.9. Sei $H : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ eine lokal Lipschitz-stetige Funktion derart, dass alle Matrizen in $\pi_y \partial H(x_0, y_0)$ regulär sind, wobei (x_0, y_0) eine Lösung von $H(x, y) = 0$ ist. Dann gelten:

- a) Ist H semismooth in einer Umgebung von (x_0, y_0) , so gibt es eine offene Umgebung von x_0 und eine lokal Lipschitz-stetige, semismoothe Funktion $g : X \rightarrow \mathbb{R}^n$, so dass $g(x_0) = y_0$ gilt und $H(x, g(x)) = 0$ für alle $x \in X$.
- b) Ist H strongly semismooth in einer Umgebung von (x_0, y_0) , so gibt es eine offene Umgebung von x_0 und eine lokal Lipschitz-stetige, strongly semismoothe Funktion $g : X \rightarrow \mathbb{R}^n$, so dass $g(x_0) = y_0$ gilt und $H(x, g(x)) = 0$ für alle $x \in X$.

Beweis. Den Beweis von Teil a) findet man in [7], Theorem 2.3. Da er außerdem völlig analog zu dem Beweis von Teil b) verläuft, wollen wir hier nur die Aussage für strongly semismoothe Funktionen beweisen.

Wir betrachten die Hilfsfunktion $F : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^{m+n}$ definiert durch

$$F(x, y) = \begin{pmatrix} x - x_0 \\ H(x, y) \end{pmatrix}.$$

Wegen der Inklusion $\partial F(x, y) \subseteq \partial_C F(x, y)$ gilt dann

$$\partial F(x_0, y_0) \subseteq \begin{pmatrix} I_{m \times m} & 0 \\ \pi_x \partial H(x_0, y_0) & \pi_y \partial H(x_0, y_0) \end{pmatrix}.$$

Deswegen sind unter den gegebenen Voraussetzungen alle Matrizen in $\partial F(x_0, y_0)$ regulär. Nach Satz 3.1.7 ist F außerdem strongly semismooth in einer Umgebung von (x_0, y_0) ,

da alle Komponentenfunktionen von F dort strongly semismooth sind. Hierbei verwenden wir, dass eine stetig differenzierbare Funktion mit lokal Lipschitz-stetiger Ableitung strongly semismooth ist, insbesondere sind alle lineare Funktionen strongly semismooth. Deswegen können wir Satz 3.2.7 auf F anwenden und erhalten offene Umgebungen U von (x_0, y_0) und W von $(0, 0) = F(x_0, y_0)$, so dass $F : U \rightarrow W$ ein lokaler Lipschitz-Homeomorphismus mit strongly semismoother Inverser $G : W \rightarrow U$ ist. Da W eine offene Menge in $\mathbb{R}^m \times \mathbb{R}^n$ ist, ist die Menge

$$X := \{x \in \mathbb{R}^m \mid (x - x_0, 0) \in W\}$$

eine nichtleere offene Menge in \mathbb{R}^m . Wir zeigen nun, dass X alle gewünschten Eigenschaften hat. Sei dafür $x \in X$ beliebig. Dann ist $(x - x_0, 0) \in W$ und es gibt wegen der Bijektivität von F einen eindeutigen Vektor y , so dass $(x, y) \in U$ und $H(x, y) = 0$ gilt. Setzen wir $g(x) := y$, so haben wir eine Abbildung $g : X \rightarrow \mathbb{R}^n$ definiert, die $H(x, g(x)) \equiv 0$ erfüllt. Wir müssen also nur noch zeigen, dass g lokal Lipschitz-stetig und strongly semismooth ist. Wie oben erwähnt gilt für alle $x \in X$:

$$F(x, g(x)) = \begin{pmatrix} x - x_0 \\ H(x, g(x)) \end{pmatrix} = \begin{pmatrix} x - x_0 \\ 0 \end{pmatrix}$$

Wenden wir nun auf beide Seiten die inverse Abbildung G an, so erhalten wir für alle $x \in X$

$$\begin{pmatrix} x \\ g(x) \end{pmatrix} = G(x - x_0, 0).$$

Mit G ist auch die Abbildung $x \mapsto G(x - x_0, 0)$ lokal Lipschitz-stetig und strongly semismooth, folglich ist sie lokal Lipschitz-stetig und strongly semismooth in all ihren Komponentenfunktionen. Daher hat die Funktion g alle gewünschten Eigenschaften. \square

3.3 (Strongly) Semismoothness der Ableitung von $V_\gamma(x)$

Wie im Abschnitt 1.1 erläutert, wollen wir das folgende Problem lösen, vergleiche Problem (1.2):

$$\min V_\gamma(x) \quad \text{u.d.N.} \quad x \in X$$

Hierbei ist $X = \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}$ und

$$V_\gamma(x) = \sup_{y \in X} \Psi_\gamma(x, y) = \Psi_\gamma(x, y_\gamma(x))$$

wobei $\Psi_\gamma(x, y)$ definiert war als

$$\Psi_\gamma(x, y) := \sum_{\nu=1}^N [\theta_\nu(x^\nu, x^{-\nu}) - \theta_\nu(y^\nu, x^{-\nu})] - \frac{\gamma}{2} \|x - y\|_2^2.$$

Aus Satz 1.1.3 wissen wir, dass die Funktion V_γ stetig differenzierbar für alle $x \in X$ ist mit

$$\nabla V_\gamma(x) = \nabla_x \Psi_\gamma(x, y) \Big|_{y=y_\gamma(x)}. \quad (3.1)$$

Wir wollen in diesem Abschnitt zeigen, dass ∇V_γ unter den Voraussetzungen 1.1.1 und LICQ sogar semismooth ist und unter den Voraussetzungen 1.1.2 und LICQ strongly semismooth. Da unter den gegebenen Voraussetzungen $\nabla_x \Psi_\gamma(x, y)$ stetig differenzierbar ist, ist ∇V_γ semismooth, falls $y_\gamma(x)$ semismooth ist. Setzen wir sogar voraus, dass die zweite Ableitung aller θ_ν lokal Lipschitz-stetig ist, ist ∇V_γ strongly semismooth, falls $y_\gamma(x)$ strongly semismooth ist. Wir wollen also unter Verwendung des im letzten Abschnitt bewiesenen Satzes über implizite Funktionen zeigen, dass $y_\gamma(x)$ (strongly) semismooth ist. (Einen sehr ähnlichen Ansatz, der allerdings nur Ungleichungsrestriktionen in der Menge X betrachtet, findet man in [7].) Hierfür erinnern wir uns an die Definition von $V_\gamma(x) := \sup_{y \in X} \Psi_\gamma(x, y) = \Psi_\gamma(x, y_\gamma(x))$. Für alle $x \in X$ ist $y_\gamma(x)$ also die eindeutige Lösung eines Optimierungsproblems der Form

$$\min_y f(x, y) \quad \text{u.d.N.} \quad g(y) \leq 0, h(y) = 0 \quad (3.2)$$

mit einer zweimal stetig differenzierbaren und in y gleichmäßig konvexen Funktion $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, einer zweimal stetig differenzierbaren und in allen Komponenten konvexen Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ und einer affin linearen Funktion $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$. Wir wollen daher im Folgenden dieses Problem genauer untersuchen. Sei ab jetzt $x \in X$ beliebig, aber fest gewählt und $y = y(x)$ die zugehörige eindeutige Lösung von Problem (3.2). Setzen wir voraus, dass in y LICQ gilt, d.h. dass die Gradienten aller aktiven Ungleichungsrestriktionen $\nabla g_i(y)$ ($i : g_i(y) = 0$) und die Gradienten aller Gleichungsrestriktionen $\nabla h_i(y)$ ($i = 1, \dots, p$) linear unabhängig sind, so gibt es eindeutige Lagrange-Multiplikatoren $\lambda = \lambda(x) \in \mathbb{R}^m$ und $\mu = \mu(x) \in \mathbb{R}^p$, so dass (y, λ, μ) das KKT-System

$$\begin{aligned} \nabla_y L(x, y, \lambda, \mu) &= \nabla_y f(x, y) + \nabla g(y)\lambda + \nabla h(y)\mu = 0, \\ g(y) &\leq 0, \quad \lambda \geq 0, \quad g(y)^T \lambda = 0, \\ h(y) &= 0 \end{aligned}$$

lösen. Hierbei ist $L(x, y, \lambda, \mu)$ die Lagrange-Funktion, definiert durch

$$L(x, y, \lambda, \mu) = f(x, y) + \lambda^T g(y) + \mu^T h(y).$$

Mit Hilfe der Minimum-Funktion $\varphi(a, b) := \min\{a, b\}$ können wir dieses System äquivalent umformen in $\tilde{\Phi}(x, y, \lambda, \mu) = 0$ mit

$$\tilde{\Phi}(x, y, \lambda, \mu) := \begin{pmatrix} \nabla_y L(x, y, \lambda, \mu) \\ \phi(-g(y), \lambda) \\ h(y) \end{pmatrix},$$

wobei $\phi(-g(y), \lambda)$ definiert ist durch

$$\phi(-g(y), \lambda) = (\varphi(-g_1(y), \lambda_1), \dots, \varphi(-g_m(y), \lambda_m))^T \in \mathbb{R}^m.$$

Im Folgenden werden wir öfter die Abkürzung $w := (x, y, \lambda, \mu)$ verwenden. Das nächste Lemma gibt Aufschluss über die Struktur der Elemente von $\partial \tilde{\Phi}(w)$.

Lemma 3.3.1. *Seien f, g und h zweimal stetig differenzierbar und $w := (x, y, \lambda, \mu)$. Dann lässt sich jedes Element $H \in \partial \tilde{\Phi}(w)$ schreiben als*

$$H = \begin{pmatrix} \nabla_{yx}^2 L(x, y, \lambda, \mu) & \nabla_{yy}^2 L(x, y, \lambda, \mu)^T & \nabla g(y) & \nabla h(y) \\ 0 & -D_a(y, \lambda) \nabla g(y)^T & D_b(y, \lambda) & 0 \\ 0 & \nabla h(y)^T & 0 & 0 \end{pmatrix}$$

wobei $D_a = \text{diag}(a_1(y, \lambda), \dots, a_m(y, \lambda)) \in \mathbb{R}^{m \times m}$ und $D_b = \text{diag}(b_1(y, \lambda), \dots, b_m(y, \lambda)) \in \mathbb{R}^{m \times m}$ Diagonalmatrizen mit den folgenden Einträgen

$$a_i(y, \lambda) = \begin{cases} 1, & \text{falls } -g_i(y) < \lambda_i, \\ 0, & \text{falls } -g_i(y) > \lambda_i, \\ \alpha_i, & \text{falls } -g_i(y) = \lambda_i, \end{cases} \quad b_i(y, \lambda) = \begin{cases} 0, & \text{falls } -g_i(y) < \lambda_i, \\ 1, & \text{falls } -g_i(y) > \lambda_i, \\ 1 - \alpha_i, & \text{falls } -g_i(y) = \lambda_i, \end{cases}$$

mit $\alpha_i \in [0, 1]$ sind. Man erhält $H \in \partial\tilde{\Phi}(w)$ durch geeignete Wahl der α_i .

Beweis. Wir verwenden hier die Inklusion $\partial\tilde{\Phi}(w) \subseteq \partial_C\tilde{\Phi}(w)$. Da unter den gegebenen Voraussetzungen die Funktionen $\nabla_y L(x, y, \lambda, \mu)$ und $h(y)$ in w stetig differenzierbar sind, ergeben sich die Einträge in den ersten n und den letzten p Zeilen von H wie oben angegeben. Gilt $-g_i(y) \neq \lambda_i$, so ist auch φ in $(-g_i(y), \lambda_i)$ stetig differenzierbar und die zugehörigen Zeilen von H haben die oben angegebene Form. Im Fall $-g_i(y) = \lambda_i$ gilt

$$\partial_B\varphi(-g_i(y), \lambda_i) = \{(0, -\nabla g_i(y)^T, 0, 0), (0, 0, e_i^T, 0)\},$$

wobei $e_i \in \mathbb{R}^m$ den i -ten Einheitsvektor bezeichne. Damit folgt

$$\partial\varphi(-g_i(y), \lambda_i) = \text{conv}\partial_B\varphi(-g_i(y), \lambda_i) = \{(0, -\alpha_i \nabla g_i(y)^T, (1 - \alpha_i)e_i^T, 0) \mid \alpha_i \in [0, 1]\}.$$

Hieraus folgt die Darstellung der noch fehlenden Zeilen von H . □

Um den Satz über implizite Funktionen auf $\tilde{\Phi}$ anwenden zu können, müssen wir sicherstellen, dass alle Matrizen in $\pi_{(y, \lambda, \mu)}\partial\tilde{\Phi}(w^*)$ regulär sind, wobei w^* eine Lösung von $\tilde{\Phi}(w) = 0$ sei. Dafür benötigen wir noch genauere Aussagen über die Struktur der Elemente von $\partial\tilde{\Phi}(w^*)$. Hierfür teilen wir die Indexmenge $\{1, \dots, m\}$ der Ungleichungsrestriktionen folgendermaßen auf:

$$I_0 := \{i \mid g_i(y^*) = 0\} \quad \text{und} \quad I_- := \{i \mid g_i(y^*) < 0\}$$

Die Menge I_0 teilen wir weiter auf in

$$I_{00} := \{i \in I_0 \mid \lambda_i^* = 0\} \quad \text{und} \quad I_{0+} := \{i \in I_0 \mid \lambda_i^* > 0\}.$$

Schließlich splitten wir noch die Menge I_{00} auf in

$$I_{000} := \{i \in I_{00} \mid \alpha_i = 0\} \quad \text{und} \quad I_{00\alpha} := \{i \in I_{00} \mid \alpha_i \in (0, 1)\} \quad \text{und} \quad I_{001} := \{i \in I_{00} \mid \alpha_i = 1\}.$$

Man beachte, dass diese drei Indexmengen von dem jeweils gewählten Element aus $\partial\tilde{\Phi}(w^*)$ abhängen. Nach Lemma 3.3.1 hat jedes Element $V \in \pi_{(y, \lambda, \mu)}\partial\tilde{\Phi}(w^*)$ wegen $\nabla_{yy}^2 L(x, y, \lambda, \mu) = \nabla_{yy}^2 L(x, y, \lambda, \mu)^T$ die Gestalt

$$V = \begin{pmatrix} \nabla_{yy}^2 L(x^*, y^*, \lambda^*, \mu^*) & \nabla g(y^*) & \nabla h(y^*) \\ -D_a(y^*, \lambda^*) \nabla g(y^*)^T & D_b(y^*, \lambda^*) & 0 \\ \nabla h(y^*)^T & 0 & 0 \end{pmatrix}.$$

Verwenden wir nun die oben eingeführten Indexmengen, so erhalten wir durch geeignetes Umsortieren der Komponenten von g und unter Verwendung der Definition der Matrizen

$D_a(y, \lambda)$ und $D_b(y, \lambda)$ die folgende Struktur:

$$V = \begin{pmatrix} \nabla_{yy}^2 L & \nabla g_{0+} & \nabla g_{001} & \nabla g_{00\alpha} & \nabla g_{000} & \nabla g_- & \nabla h \\ -\nabla g_{0+}^T & 0 & 0 & 0 & 0 & 0 & 0 \\ -\nabla g_{001}^T & 0 & 0 & 0 & 0 & 0 & 0 \\ -(D_a)_{00\alpha} \nabla g_{00\alpha}^T & 0 & 0 & (D_b)_{00\alpha} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 \\ \nabla h^T & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

In der obigen Darstellung haben wir der Übersichtlichkeit wegen die Variablen $w^* := (x^*, y^*, \lambda^*, \mu^*)$ unterdrückt und z.B. statt $\nabla g_{I_{0+}}$ nur ∇g_{0+} geschrieben. Gemäß der Definition von $I_{00\alpha}$ sind die Matrizen $(D_a)_{00\alpha}$ und $(D_b)_{00\alpha}$ beide positiv definit. Mit dieser Darstellung von V können wir jetzt die gewünschte Regularitätsaussage beweisen.

Satz 3.3.2. *Betrachte das Problem (3.2) mit f, g und h zweimal stetig differenzierbar sowie f gleichmäßig konvex in y , g konvex in allen Komponentenfunktionen und h affin linear. Sei $w^* = (x^*, y^*, \lambda^*, \mu^*)$ eine Lösung von $\tilde{\Phi}(w) = 0$ und es gelte LICQ in y^* . Dann sind alle Matrizen $V \in \pi_{(y, \lambda, \mu)} \partial \tilde{\Phi}(w^*)$ regulär.*

Beweis. Da w^* eine Lösung von $\tilde{\Phi}(w) = 0$ ist, gilt $\lambda^* \geq 0$. Daher ist die Lagrange-Funktion $L(x^*, y, \lambda^*, \mu^*)$ gleichmäßig konvex in y . Daher ist die Matrix $\nabla_{yy}^2 L(x^*, y^*, \lambda^*, \mu^*)$ positiv definit. Sei nun $V \in \pi_{(y, \lambda, \mu)} \partial \tilde{\Phi}(w^*)$ beliebig und $v = (v_1^T, v_2^T, \dots, v_7^T)^T \in \mathbb{R}^{n+m+p}$ ein Vektor im Kern von V , den wir passend zu der obigen Darstellung von V partitioniert haben. Dann löst v das folgende lineare Gleichungssystem:

$$\begin{aligned} \nabla_{yy}^2 L v_1 + \nabla g_{0+} v_2 + \nabla g_{001} v_3 + \nabla g_{00\alpha} v_4 + \nabla g_{000} v_5 + \nabla g_- v_6 + \nabla h v_7 &= 0 \\ -\nabla g_{0+}^T v_1 &= 0 \\ -\nabla g_{001}^T v_1 &= 0 \\ -(D_a)_{00\alpha} \nabla g_{00\alpha}^T v_1 + (D_b)_{00\alpha} v_4 &= 0 \\ v_5 &= 0 \\ v_6 &= 0 \\ \nabla h^T v_1 &= 0 \end{aligned}$$

Daraus folgt

$$\begin{aligned} v_1^T \nabla_{yy}^2 L v_1 + v_1^T \nabla g_{0+} v_2 + v_1^T \nabla g_{001} v_3 + v_1^T \nabla g_{00\alpha} v_4 + v_1^T \nabla h v_7 &= 0 \\ -v_2^T \nabla g_{0+}^T v_1 &= 0 \\ -v_3^T \nabla g_{001}^T v_1 &= 0 \\ -((D_a)_{00\alpha}^{-1} v_4)^T (D_a)_{00\alpha} \nabla g_{00\alpha}^T v_1 + ((D_a)_{00\alpha}^{-1} v_4)^T (D_b)_{00\alpha} v_4 &= 0 \\ -v_7^T \nabla h^T v_1 &= 0 \end{aligned}$$

Addiert man diese 5 Gleichungen, so erhält man die folgende Gleichung:

$$v_1^T \nabla_{yy}^2 L v_1 + v_4^T (D_b)_{00\alpha} (D_a)_{00\alpha}^{-1} v_4 = 0$$

Da unter den gegebenen Voraussetzungen aber die Matrizen $\nabla_{yy}^2 L$ und $(D_b)_{00\alpha}(D_a)_{00\alpha}^{-1}$ positiv definit sind, folgt $v_1 = 0$ und $v_4 = 0$. Daraus ergibt sich die folgende Bestimmungsgleichung für die übrigen Komponenten von v :

$$\nabla g_{0+} v_2 + \nabla g_{001} v_3 + \nabla h v_7 = 0$$

Da I_{0+} und I_{001} Teilmengen der Menge der aktiven Ungleichungen I_0 sind, folgt wegen LICQ auch $v_2 = 0$, $v_3 = 0$ und $v_7 = 0$. Insgesamt folgt also $v = 0$, d.h. V ist regulär. \square

Nun wollen wir zeigen, dass unter den obigen Voraussetzungen y in einer Umgebung von x^* als (strongly) semismoothe Funktion von x dargestellt werden kann.

Satz 3.3.3. *Betrachte das Problem (3.2) mit f, g und h zweimal stetig differenzierbar sowie f gleichmäßig konvex in y , g konvex in allen Komponentenfunktionen und h affin linear. Sei $w^* = (x^*, y^*, \lambda^*, \mu^*)$ eine Lösung von $\tilde{\Phi}(w) = 0$ und es gelte LICQ in y^* . Dann gibt es eine offene Umgebung U von x^* und eine semismoothe Funktion $G : U \rightarrow \mathbb{R}^{n+m+p}$, $x \mapsto (y(x), \lambda(x), \mu(x))$, so dass $\tilde{\Phi}(x, G(x)) = 0$ für alle $x \in U$ gilt. Insbesondere ist die Abbildung $x \mapsto y(x)$ semismooth.*

Sind die zweiten Ableitungen von f und g sogar lokal Lipschitz-stetig, so gibt es eine strongly semismoothe Funktion G mit den oben genannten Eigenschaften. Insbesondere ist die Abbildung $x \mapsto y(x)$ dann strongly semismooth in U .

Beweis. Da der Beweis für die erste Aussage völlig analog zu dem Beweis der zweiten verläuft, werden wir hier nur den zweiten Teil beweisen. Unter den gegebenen Voraussetzungen ist $\tilde{\Phi} : \mathbb{R}^n \times \mathbb{R}^{n+m+p} \rightarrow \mathbb{R}^{n+m+p}$,

$$\tilde{\Phi}(x, y, \lambda, \mu) := \begin{pmatrix} \nabla_y f(x, y) + \nabla g(y)\lambda + \nabla h(y)\mu \\ \phi(-g(y), \lambda) \\ h(y) \end{pmatrix}$$

strongly semismooth, da alle Komponentenfunktionen von $\tilde{\Phi}$ diese Eigenschaft haben. Hierbei geht ein, dass die Minimum-Funktion strongly semismooth ist, Kompositionen strongly semismoother Funktionen wieder strongly semismooth sind und stetig differenzierbare Funktionen mit lokal Lipschitz-stetiger Ableitung strongly semismooth sind. Desweiteren sind nach Satz 3.3.2 alle Elemente von $\pi_{(y, \lambda, \mu)} \partial \tilde{\Phi}(w^*)$ regulär, wobei $w^* = (x^*, y^*, \lambda^*, \mu^*)$ eine Lösung von $\tilde{\Phi}(w) = 0$ ist. Deswegen gibt es nach Satz 3.2.9 eine Umgebung U von x^* und eine strongly semismoothe Funktion $G : U \rightarrow \mathbb{R}^{n+m+p}$, $x \mapsto (y(x), \lambda(x), \mu(x))$, mit $\tilde{\Phi}(x, G(x)) = 0$ für alle $x \in U$. Da sich die Abbildung $x \mapsto y(x)$ durch Komponentenfunktionen von G darstellen lässt, ist sie in U ebenfalls strongly semismooth. \square

Als Korollar aus diesem Satz erhalten wir jetzt das zentrale Resultat des Kapitels 3.

Korollar 3.3.4. *Sei x^* ein normalisiertes Nash-Gleichgewicht eines GNEPs, das den Voraussetzungen 1.1.1 genügt. Gilt in x^* LICQ, so ist die Abbildung $x \mapsto y_\gamma(x)$ in einer Umgebung semismooth. Daher ist auch $\nabla V_\gamma(x)$ in einer Umgebung von x^* semismooth. Genügt das GNEP sogar den Voraussetzungen 1.1.2, so sind $y_\gamma(x)$ und $\nabla V_\gamma(x)$ strongly semismooth in einer Umgebung von x^* .*

Beweis. Wir zeigen hier wieder nur die Aussage für Voraussetzung 1.1.2. Da x^* ein normalisiertes Nash-Gleichgewicht ist, löst x^* das Problem

$$\min V_\gamma(x) \quad \text{u.d.N.} \quad g(x) \leq 0, h(x) = 0$$

und es gilt $x^* = y_\gamma(x^*)$. Der Vektor $y_\gamma(x^*)$ wiederum löst das Problem

$$\min -\Psi_\gamma(x^*, y) \quad \text{u.d.N.} \quad g(y) \leq 0, h(y) = 0.$$

Gilt in x^* also LICQ bezüglich des ersten Problems, so gilt auch in $y_\gamma(x^*)$ LICQ bezüglich des zweiten Problems. Setzen wir $f(x, y) := -\Psi_\gamma(x, y)$, so ist f gleichmäßig konvex und zweimal stetig differenzierbar mit lokal Lipschitz-stetiger zweiter Ableitung. Hierbei geht ein, dass die zweite Ableitung von $\|x - y\|_2^2$ konstant und daher insbesondere lokal Lipschitz-stetig ist. Nach Voraussetzung 1.1.2 sind auch g und h zweimal stetig differenzierbar und haben eine lokal Lipschitz-stetige zweite Ableitung; g ist konvex in allen Komponenten und h ist affin linear. Da in $y_\gamma(x^*)$ LICQ gilt, gibt es Lagrange-Multiplikatoren λ^* und μ^* , so dass $\tilde{\Phi}(x^*, y_\gamma(x^*), \lambda^*, \mu^*) = 0$ gilt. Nach Satz 3.3.3 gibt es daher eine Umgebung U von x^* und eine strongly semismoothe Funktion $G : U \rightarrow \mathbb{R}^{n+m+p}$, $x \mapsto (y(x), \lambda(x), \mu(x))$, so dass $\Phi(x, G(x)) = 0$ für alle $x \in U$ gilt. Es gibt also eine strongly semismoothe Funktion $y(x)$, so dass $y(x)$ für alle $x \in U$ das Problem

$$\min -\Psi_\gamma(x, y) \quad \text{u.d.N.} \quad g(y) \leq 0, h(y) = 0$$

löst. Da unter den gegebenen Voraussetzungen dieses Problem für jedes $x \in U \subseteq X$ die eindeutige Lösung $y_\gamma(x)$ hat, gilt $y(x) = y_\gamma(x)$ für alle $x \in U$. Also ist die Abbildung $x \mapsto y_\gamma(x)$ strongly semismooth in U .

Wegen $\nabla V_\gamma(x) = \nabla_x \Psi_\gamma(x, y_\gamma(x))$ ist dann auch $\nabla V_\gamma(x)$ als Komposition strongly semismoother Funktionen strongly semismooth in einer Umgebung U von x^* . \square

Kapitel 4

Konvergenzeigenschaften des semismoothen SQP-Verfahrens

Wie wir in Abschnitt 3.3 gezeigt haben, ist die erste Ableitung der Funktion $V_\gamma(x)$ unter den Voraussetzungen 1.1.1 semismooth. Wir können daher zur Lösung des Problems

$$\min_x V_\gamma(x) \quad \text{u.d.N.} \quad g(x) \leq 0, h(x) = 0$$

ein semismoothes SQP-Verfahren verwenden. In diesem Kapitel wollen wir die lokalen Konvergenzeigenschaften eines solchen Verfahrens untersuchen. In den Abschnitten 4.1 und 4.2 zeigen wir, dass ein solches Verfahren unter gewissen Voraussetzungen lokal superlinear bzw. lokal quadratisch konvergiert. Der Abschnitt 4.3 ist der strong second-order sufficiency condition gewidmet, einer Bedingung, die in den Konvergenzsätzen aus den ersten zwei Abschnitten dieses Kapitels auftritt. Wir werden einige hinreichende Bedingungen herleiten, unter denen die strong second-order sufficiency condition erfüllt ist.

4.1 Lokal superlineare Konvergenz des semismoothen SQP-Verfahrens

Dieser Abschnitt ist eine Zusammenfassung der Arbeit [5], wobei wir manche Stellen etwas genauer ausführen werden.

Definition 4.1.1. *Eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt SC^1 -Funktion, wenn sie stetig differenzierbar ist und ihr Gradient $\nabla f(x)$ semismooth ist. Ist $\nabla f(x)$ sogar strongly semismooth, so nennen wir f eine sSC^1 -Funktion.*

Wir betrachten im Folgenden Probleme der Gestalt

$$\min f(x) \quad \text{u.d.N.} \quad g(x) \leq 0, h(x) = 0 \tag{4.1}$$

mit SC^1 -Funktionen $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ und $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$. Ein Weg, ein solches Problem zu lösen, ist das sogenannte SQP-Verfahren. Hierbei sucht man in jeder Iteration als neue Iterierte einen KKT-Punkt $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ des folgenden quadratischen Problems

$$\begin{aligned} \min \quad & \nabla f(x^k)^T(x - x^k) + \frac{1}{2}(x - x^k)^T G_k(x - x^k) \\ \text{u.d.N.} \quad & g(x^k) + \nabla g(x^k)^T(x - x^k) \leq 0 \\ & h(x^k) + \nabla h(x^k)^T(x - x^k) = 0 \end{aligned} \tag{4.2}$$

mit einer zunächst beliebigen symmetrischen Matrix $G_k \in \mathbb{R}^{n \times n}$. Meistens wird man für G_k eine Approximation von $\nabla_{xx}^2 L(x^k, \lambda^k, \mu^k)$ wählen, wobei L die Lagrange-Funktion des Problems (4.1) bezeichnet. Doch zunächst wollen wir uns weiter mit Problem (4.1) beschäftigen. Die zugehörige Lagrange-Funktion lautet

$$L(x, \lambda, \mu) = f(x) + g(x)^T \lambda + h(x)^T \mu$$

und die KKT-Bedingungen

$$\begin{aligned} \nabla_x L(x, \lambda, \mu) &= \nabla f(x) + \nabla g(x) \lambda + \nabla h(x) \mu = 0, \\ g(x) &\leq 0, \lambda \geq 0, g(x)^T \lambda = 0, \\ h(x) &= 0. \end{aligned}$$

Mittels der Minimum-Funktion $\varphi(a, b) = \min\{a, b\}$ können wir die KKT-Bedingungen wieder in ein nichtlineares Gleichungssystem der Gestalt

$$\Phi(z) = \begin{pmatrix} \Phi_1(z) \\ \Phi_2(z) \\ \Phi_3(z) \end{pmatrix} = 0$$

umformen. Hierbei steht z als Abkürzung für $z = (x, \lambda, \mu)$ und die Komponenten von Φ sind wie folgt definiert:

$$\begin{aligned} \Phi_1(z) &:= \nabla f(x) + \nabla g(x) \lambda + \nabla h(x) \mu, \\ [\Phi_2(z)]_i &:= \varphi(-g_i(x), \lambda_i) \quad \forall \quad i = 1, \dots, m, \\ \Phi_3(z) &:= h(x). \end{aligned}$$

Setzen wir voraus, dass in der Lösung x^* des Problems (4.1) LICQ gilt, so existieren eindeutige Lagrange-Multiplikatoren λ^* und μ^* , so dass $z^* = (x^*, \lambda^*, \mu^*)$ eine Lösung von $\Phi(z) = 0$ ist.

Wir wollen nun das folgende Subdifferential von $\Phi(z)$ genauer betrachten:

$$\partial_Q \Phi(z) := \partial \Phi_1(z) \times \partial_b \Phi_2(z) \times \{(\nabla \Phi_3(z))^T, 0, 0\}.$$

Wegen Satz 3.1.2 und der stetigen Differenzierbarkeit von $\Phi_3(z)$ ist auch die Abbildung $z \mapsto \partial_Q \Phi(z)$ oberhalbstetig und abgeschlossen und ihre Bilder $\partial_Q \Phi(z)$ sind nichtleer und kompakt. Offensichtlich gilt immer $\partial_Q \Phi(z) \subseteq \partial_C \Phi(z)$. Ausgeschrieben gilt

$$\partial_Q \Phi(z) = \partial \Phi_1(z) \times \partial_B \varphi(-g_1(x), \lambda_1) \times \dots \times \partial_B \varphi(-g_m(x), \lambda_m) \times \{(\nabla \Phi_3(z))^T, 0, 0\}$$

Hierbei gilt wieder

$$\partial_B \varphi(-g_i(x), \lambda_i) = \begin{cases} \{(-\nabla g_i(x)^T, 0, 0)\} & \text{falls } -g_i(x) < \lambda_i, \\ \{(-\alpha_i \nabla g_i(x)^T, (1 - \alpha_i) e_i^T, 0) \mid \alpha_i \in \{0, 1\}\} & \text{falls } -g_i(x) = \lambda_i, \\ \{(0, e_i^T, 0)\} & \text{falls } -g_i(x) > \lambda_i. \end{cases}$$

Da Φ_1 bezüglich λ und μ stetig differenzierbar ist, hat jedes Element $M \in \partial \Phi_1(z)$ die Form

$$M = (H, \nabla g(x), \nabla h(x))$$

mit einer Matrix $H \in \mathbb{R}^{n \times n}$. Deshalb definieren wir wieder

$$\pi_x \partial \Phi_1(z) := \{H \in \mathbb{R}^{n \times n} \mid (H, \nabla g(x), \nabla h(x)) \in \partial \Phi_1(z)\}.$$

Um die Struktur der Elemente von $\partial_Q \Phi(z)$ noch genauer aufzuschlüsseln, definieren wir wie im letzten Abschnitt für eine Lösung $z^* = (x^*, \lambda^*, \mu^*)$ von $\Phi(z) = 0$ die Mengen

$$I_0 := \{i \in \{1, \dots, m\} \mid g_i(y^*) = 0\} \quad \text{und} \quad I_- := \{i \in \{1, \dots, m\} \mid g_i(y^*) < 0\},$$

$$I_{00} := \{i \in I_0 \mid \lambda_i^* = 0\} \quad \text{und} \quad I_{0+} := \{i \in I_0 \mid \lambda_i^* > 0\},$$

$$I_{000} := \{i \in I_{00} \mid \alpha_i = 0\} \quad \text{und} \quad I_{001} := \{i \in I_{00} \mid \alpha_i = 1\}.$$

Auch hier hängen die Mengen I_{000} und I_{001} wieder von dem jeweils betrachteten Element des Subgradienten $\partial_Q \Phi(z^*)$ ab. Die Menge $I_{00\alpha}$ benötigen wir hier nicht, da wir statt $\partial \varphi(-g_i(x^*, \lambda_i^*))$ hier $\partial_B \varphi(-g_i(x^*, \lambda_i^*))$ betrachten. Analog zu Lemma 3.3.1 folgt auch hier, dass jedes Element $V \in \partial_Q \Phi(z^*)$ die Gestalt

$$V = \begin{pmatrix} H & \nabla g_{0+} & \nabla g_{001} & \nabla g_{000} & \nabla g_- & \nabla h \\ -\nabla g_{0+}^T & 0 & 0 & 0 & 0 & 0 \\ -\nabla g_{001}^T & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 \\ \nabla h^T & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

hat, wobei I jeweils eine Einheitsmatrix passender Größe bezeichne und $H \in \pi_x \partial \Phi_1(z^*)$ gilt.

Bevor wir beweisen, dass unter gewissen Voraussetzungen alle Matrizen in $\partial_Q \Phi(z^*)$ regulär sind, benötigen wir noch die folgende Definition.

Definition 4.1.2. Sei $z = (x, \lambda, \mu)$ und $G(z)$ definiert durch

$$G(z) := \{d \in \mathbb{R}^n \mid \nabla f(x)^T d = 0, \nabla g_i(x)^T d = 0 \forall i \in I_{0+}, \nabla h_i(x)^T d = 0 \forall i = 1, \dots, p\}.$$

Ein Punkt $z^* = (x^*, \lambda^*, \mu^*)$ genügt der strong second-order sufficiency condition für Problem (4.1), wenn er $\Phi(z^*) = 0$ erfüllt und für alle $d \in G(z^*) \setminus \{0\}$ und alle $H \in \pi_x \partial \Phi_1(z^*)$ gilt $d^T H d > 0$.

Satz 4.1.3. Der Punkt $z^* = (x^*, \lambda^*, \mu^*)$ genüge der strong second-order sufficiency condition für Problem (4.1) und LICQ. Dann sind alle Elemente von $\partial_Q \Phi(z^*)$ regulär.

Beweis. Sei $V \in \partial_Q \Phi(z^*)$ beliebig. Dann wissen wir, dass V die Gestalt

$$V = \begin{pmatrix} H & \nabla g_{0+} & \nabla g_{001} & \nabla g_{000} & \nabla g_- & \nabla h \\ -\nabla g_{0+}^T & 0 & 0 & 0 & 0 & 0 \\ -\nabla g_{001}^T & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 \\ \nabla h^T & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

mit einer Matrix $H \in \pi_x \partial \Phi_1(z^*)$ hat. Sei $v = (v_1^T, \dots, v_6^T)^T$ ein entsprechend partitionierter Vektor im Kern von V . Dann löst v das folgende lineare Gleichungssystem:

$$\begin{aligned} H v_1 + \nabla g_{0+} v_2 + \nabla g_{001} v_3 + \nabla g_{000} v_4 + \nabla g_- v_5 + \nabla h v_6 &= 0, \\ -\nabla g_{0+}^T v_1 &= 0, \\ -\nabla g_{001}^T v_1 &= 0, \\ v_4 &= 0, \\ v_5 &= 0, \\ \nabla h^T v_1 &= 0. \end{aligned}$$

Hieraus folgt wie im Beweis von Satz 3.3.2 $v_1^T H v_1 = 0$. Wenn wir nun zeigen können, dass $v_1 \in G(z^*)$ gilt, so folgt wegen der strong second-order sufficiency condition $v_1 = 0$. Aus der zweiten und sechsten Gleichung in dem obigen Gleichungssystem ergibt sich $\nabla g_i(x)^T v_1 = 0$ für alle $i \in I_{0+}$ und $\nabla h_i(x)^T v_1 = 0$ für alle $i = 1, \dots, p$. Da nach Voraussetzung $\Phi(z^*) = 0$ gilt, gilt auch $\Phi_1(z^*)^T v_1 = 0$. Ausgeschrieben heißt das

$$\nabla f^T v_1 + (\lambda_{0+}^*)^T \nabla g_{0+}^T v_1 + (\lambda_{001}^*)^T \nabla g_{001}^T v_1 + (\lambda_{000}^*)^T \nabla g_{000}^T v_1 + (\lambda_-^*)^T \nabla g_-^T v_1 + (\mu^*)^T \nabla h^T v_1 = 0.$$

Nach Konstruktion der Indexmengen I_- , I_{000} und I_{001} gilt $\lambda_-^* = 0$, $\lambda_{000}^* = 0$ und $\lambda_{001}^* = 0$. Zusammen mit der zweiten und sechsten Gleichung des obigen Gleichungssystems folgt daraus $\nabla f(x^*)^T v_1 = 0$. Also ist $v_1 \in G(z^*)$ und es folgt $v_1 = 0$. Damit vereinfacht sich die erste Gleichung in dem obigen System zu

$$\nabla g_{0+} v_2 + \nabla g_{001} v_3 + \nabla h v_6 = 0.$$

Da nach Konstruktion I_{0+} und I_{001} Teilmengen der Menge der aktiven Ungleichungen I_0 sind, folgt mit LICQ, dass auch $v_2 = 0$, $v_3 = 0$ und $v_6 = 0$ gelten muss. Also ist $v = 0$ und daher V regulär. \square

Da Φ unter den gegebenen Voraussetzungen lokal Lipschitz-stetig ist und $\partial_Q \Phi(z)$ ähnliche Eigenschaften wie $\partial_B \Phi(z)$ hat, folgt mit Satz 3.2.4 aus dem obigen Satz das folgende Korollar.

Korollar 4.1.4. *Der Punkt $z^* = (x^*, \lambda^*, \mu^*)$ genüge der strong second-order sufficiency condition für Problem (4.1) und LICQ. Dann gibt es Konstanten $\delta, c > 0$, so dass für alle $z \in K_\delta(z^*)$ gilt: Alle Elemente $V \in \partial_Q \Phi(z)$ sind regulär und es gilt $\|V^{-1}\| \leq c$.*

Anstatt direkt lokal superlineare Konvergenz des SQP-Verfahrens zu beweisen, zeigen Han und Sun in [5] die superlineare Konvergenz einer sogenannten „generalized approximate Newton method“, kurz GANM, und zeigen dann, dass dieses Verfahren lokal mit dem SQP-Verfahren übereinstimmt. Ziel der GANM ist es, eine Lösung $z^* = (x^*, \lambda^*, \mu^*)$ von $\Phi(z) = 0$ zu bestimmen.

Algorithmus 4.1.5. (GANM)

(S.0) Wähle einen Startvektor $z^0 = (x^0, \lambda^0, \mu^0) \in \mathbb{R}^{n+m+p}$, setze $k = 0$.

(S.1) Ist ein geeignetes Abbruchkriterium erfüllt: STOPP

(S.2) Wähle $I \in \mathcal{P}(I_{00}(z^*))$ und setze

$$z^{k+1} = z^k - B_{(I)k}^{-1} \Phi^{(I)}(z^k)$$

mit $B_{(I)k} = \nabla \Phi^{(I)k}(z^k)^T$ und $G^k \in \mathbb{R}^{n \times n}$. Setze $k \leftarrow k + 1$ und gehe zu (S.1).

Die hierbei verwendeten Funktionen Φ^I und $\Phi^{(I)k}$ sind folgendermaßen definiert: Seien $z^* = (x^*, \lambda^*, \mu^*)$ eine Lösung von $\Phi(z) = 0$ und $I_-(z^*)$, $I_{0+}(z^*)$ und $I_{00}(z^*)$ wie oben. Sei weiter $I \in \mathcal{P}(I_{00}(z^*))$ ein Element der Potenzmenge von $I_{00}(z^*)$. Dann ist

$$\Phi^{(I)}(z) := \begin{pmatrix} \nabla f(x) + \nabla g(x)\lambda + \nabla h(x)\mu \\ p^{(I)}(z) \\ h(x) \end{pmatrix},$$

wobei $p^{(I)}(z)$ definiert ist durch

$$p^{(I)}(z) := \begin{cases} -g_i(x) & \text{falls } i \in I_{0+}(z^*), \\ -g_i(x) & \text{falls } i \in I, \\ \lambda_i & \text{falls } i \in I_{00}(z^*) \setminus I, \\ \lambda_i & \text{falls } i \in I_-(z^*). \end{cases}$$

Die Funktion $\Phi^{(I)k}$ ist analog definiert als

$$\Phi^{(I)k}(z) := \begin{pmatrix} \nabla f(x^k) + \nabla g(x^k)\lambda + \nabla h(x^k)\mu + G_k(x - x^k) \\ p^{(I)k}(z) \\ h(x^k) + \nabla h(x^k)^T(x - x^k) \end{pmatrix},$$

wobei $p^{(I)k}(z)$ definiert ist durch

$$p^{(I)k}(z) := \begin{cases} -g_i(x^k) - \nabla g_i(x^k)(x - x^k) & \text{falls } i \in I_{0+}(z^*), \\ -g_i(x^k) - \nabla g_i(x^k)(x - x^k) & \text{falls } i \in I, \\ \lambda_i & \text{falls } i \in I_{00}(z^*) \setminus I, \\ \lambda_i & \text{falls } i \in I_-(z^*). \end{cases}$$

$\Phi^{(I)k}$ ist also so etwas ähnliches wie eine lineare Approximation von $\Phi^{(I)}$. Die Funktion $\Phi^{(I)}$ wiederum ist eine Approximation von Φ im Punkt z^* . Dies spiegelt sich auch im Beweis des folgenden Lemmas wieder.

Lemma 4.1.6. *Der Punkt $z^* = (x^*, \lambda^*, \mu^*)$ genüge der strong second-order sufficiency condition für Problem (4.1) und LICQ. Dann gibt es Konstanten $\delta, c > 0$, so dass für alle $z \in K_\delta(z^*)$ und alle $I \in \mathcal{P}(I_{00}(z^*))$ gilt: Alle Elemente $V \in \partial_Q \Phi^{(I)}(z)$ sind regulär und es gilt $\|V^{-1}\| \leq c$.*

Beweis. Für beliebiges $I \in \mathcal{P}(I_{00}(z^*))$ ist

$$\partial_Q \Phi^{(I)}(z^*) = \partial \Phi_1(z^*) \times \partial_{BP_1^{(I)}}(z^*) \times \dots \times \partial_{BP_m^{(I)}}(z^*) \times \{(\nabla \Phi_3(z^*)^T, 0, 0)\}$$

mit

$$\partial_{BP_i^{(I)}}(z^*) = \begin{cases} (-\nabla g_i(x^*)^T, 0, 0) & \text{falls } i \in I_{0+}(z^*), \\ (-\nabla g_i(x^*)^T, 0, 0) & \text{falls } i \in I, \\ (0, e_i^T, 0) & \text{falls } i \in I_{00}(z^*) \setminus I, \\ (0, e_i^T, 0) & \text{falls } i \in I_-(z^*). \end{cases}$$

Es gilt also $\partial_Q \Phi^{(I)}(z^*) \subseteq \partial_Q \Phi(z^*)$. Nach Satz 4.1.3 sind deshalb alle Matrizen in $\partial_Q \Phi^{(I)}(z^*)$ regulär. Nach Lemma 3.2.4 gibt es daher Konstanten $\delta_I, c_I > 0$, so dass für alle $z \in K_{\delta_I}(z^*)$ gilt: Alle Matrizen $V \in \partial_Q \Phi^{(I)}(z)$ sind regulär und es gilt $\|V^{-1}\| \leq c_I$. Da es höchstens $2^{|I_{00}(z^*)|}$ verschiedene Funktionen $\Phi^{(I)}$ gibt, haben die Konstanten $\delta := \min\{\delta_I : I \in \mathcal{P}(I_{00}(z^*))\}$ und $c := \max\{c_I : I \in \mathcal{P}(I_{00}(z^*))\}$ die gewünschten Eigenschaften. \square

Der folgende Satz beschäftigt sich mit den Konvergenzeigenschaften der GANM. Obwohl die GANM offensichtlich ein rein theoretisches Verfahren ist, schließlich setzt es die Kenntnis von $I_-(z^*)$, $I_{00}(z^*)$ und $I_{0+}(z^*)$ voraus, sind die Konvergenzeigenschaften dieses Verfahrens für uns interessant, da wir im Anschluss an den folgenden Satz zeigen werden, dass das SQP-Verfahren lokal ein Spezialfall der GANM ist.

Satz 4.1.7. *Der Punkt $z^* = (x^*, \lambda^*, \mu^*)$ genüge der strong second-order sufficiency condition für Problem (4.1) und LICQ. Hierbei seien f, g und h SC^1 -Funktionen. Weiter seien $\delta, c > 0$ die Konstanten aus Lemma 4.1.6. Existiert für alle $k \in \mathbb{N}$ eine Matrix $H_k \in \pi_x \partial \Phi_1(z^k)$, so dass*

$$\|G_k - H_k\|_2 \leq \frac{1}{4c}$$

gilt, so ist GANM wohldefiniert und konvergiert lokal linear. Gilt zusätzlich

$$\lim_{k \rightarrow \infty} \frac{\|(G_k - H_k)(x^{k+1} - x^k)\|_2}{\|z^{k+1} - z^k\|_2} = 0,$$

so konvergiert das Verfahren lokal superlinear.

Beweis. Da f, g und h SC^1 -Funktionen sind, sind Φ und $\Phi^{(I)}$, $I \in \mathcal{P}(I_{00}(z^*))$, semismooth in z^* .

Partitionieren wir $\Phi^{(I)}$ und $\Phi^{(I)k}$ genauso wie Φ , so gilt für $\Phi^{(I)}$

$$\begin{aligned} \partial \Phi_1^{(I)}(z^k) &= \partial \Phi_1(z^k) = (H, \nabla g(x^k), \nabla h(x^k)) \quad \text{mit } H \in \pi_x \partial \Phi_1(z^k), \\ \partial_B \left[\Phi_2^{(I)}(z^k) \right]_i &= \begin{cases} (-\nabla g_i(x^k)^T, 0, 0) & \text{falls } i \in I_- \cup I, \\ (0, e_i^T, 0) & \text{falls } i \in I_{0+} \cup (I_{00} \setminus I), \end{cases} \\ \nabla \Phi_3^{(I)}(z^k)^T &= \nabla \Phi_3(z^k)^T. \end{aligned}$$

Für $\Phi^{(I)k}$ gilt analog

$$\begin{aligned} \partial \Phi_1^{(I)k}(z^k) &= (G_k, \nabla g(x^k), \nabla h(x^k)), \\ \partial_B \left[\Phi_2^{(I)k}(z^k) \right]_i &= \begin{cases} (-\nabla g_i(x^k)^T, 0, 0) & \text{falls } i \in I_- \cup I, \\ (0, e_i^T, 0) & \text{falls } i \in I_{0+} \cup (I_{00} \setminus I), \end{cases} \\ \nabla \Phi_3^{(I)k}(z^k)^T &= \nabla \Phi_3(z^k)^T. \end{aligned}$$

Wählen wir also das Element $W_{(I)k} \in \partial_Q \Phi^{(I)}(z^k)$, für das $H = H_k$ gilt, so gilt für alle $z = (x, \lambda, \mu)$ wegen $B_{(I)k} = \nabla \Phi^{(I)k}(z^k)^T$

$$\|(B_{(I)k} - W_{(I)k})z\|_2 = \|(H_k - G_k)x\|_2.$$

Insbesondere folgt daraus für alle $k \in \mathbb{N}$

$$\|B_{(I)k} - W_{(I)k}\|_2 = \sup_{z \neq 0} \frac{\|(B_{(I)k} - W_{(I)k})z\|_2}{\|z\|_2} \leq \sup_{z \neq 0} \frac{\|H_k - G_k\|_2 \|x\|_2}{\|z\|_2} \leq \|H_k - G_k\|_2 \leq \frac{1}{4c}.$$

Gilt $\|z^k - z^*\|_2 \leq \delta$, so ist $W_{(I)k}$ nach Lemma 4.1.6 regulär und es gilt $\|W_{(I)k}^{-1}\|_2 \leq c$. Nach dem Perturbation Lemma 2.3.3 von Ortega und Rheinboldt, vergleiche [16], ist daher auch $B_{(I)k}$ für alle $k \in \mathbb{N}$ mit $\|z^k - z^*\|_2 \leq \delta$ regulär und es gilt

$$\|B_{(I)k}^{-1}\|_2 \leq \frac{c}{1 - c\frac{1}{4c}} = \frac{4}{3}c.$$

Da die Funktion $\Phi^{(I)}$ in z^* semismooth ist, ist sie nach Satz 3.1.6 dort auch in jeder Komponente $\Phi_j^{(I)}$, $j = 1, \dots, n + m + p$ semismooth. Wegen $W_{(I)k} \in \partial_Q \Phi^{(I)}(z^k) \subseteq \partial_C \Phi^{(I)}(z^k)$ gilt für alle $j = 1, \dots, n + m + p$ auch $(W_{(I)k})_j \in \partial \Phi_j^{(I)}(z^k)$. Hierbei bezeichne $(W_{(I)k})_j$ die j -te Zeile von $W_{(I)k}$. Da es nur endlich viele Mengen $I \subseteq I_{00}(z^*)$ gibt, folgt aus Satz 3.1.6 a) und d2) die Existenz einer Konstanten $\delta_1 > 0$, so dass für alle $k \in \mathbb{N}$ mit $\|z^k - z^*\|_2 \leq \delta_1$ und alle $I \subseteq I_{00}(z^*)$ gilt

$$\begin{aligned} \|\Phi^{(I)}(z^k) - \Phi^{(I)}(z^*) - W_{(I)k}(z^k - z^*)\|_2 &\leq \sum_{j=1}^{n+m+p} \left| \Phi_j^{(I)}(z^k) - \Phi_j^{(I)}(z^*) - (W_{(I)k})_j(z^k - z^*) \right| \\ &\leq \frac{1}{8c} \|z^k - z^*\|_2. \end{aligned}$$

Setzen wir nun $\delta^* := \min\{\delta, \delta_1\}$, so folgt für alle $k \in \mathbb{N}$ mit $\|z^k - z^*\|_2 \leq \delta^*$ wegen $\Phi^{(I)}(z^*) = 0$

$$\begin{aligned} \|z^{k+1} - z^*\|_2 &= \|z^k - B_{(I)k}^{-1} \Phi^{(I)}(z^k) - z^*\|_2 \\ &\leq \|B_{(I)k}^{-1}\|_2 \|\Phi^{(I)}(z^k) - \Phi^{(I)}(z^*) - B_{(I)k}(z^k - z^*)\|_2 \\ &\leq \|B_{(I)k}^{-1}\|_2 (\|\Phi^{(I)}(z^k) - \Phi^{(I)}(z^*) - W_{(I)k}(z^k - z^*)\|_2 \\ &\quad + \|(W_{(I)k} - B_{(I)k})(z^k - z^*)\|_2) \\ &\leq \frac{4}{3}c \left(\frac{1}{8c} \|z^k - z^*\|_2 + \frac{1}{4c} \|z^k - z^*\|_2 \right) \\ &= \frac{1}{2} \|z^k - z^*\|_2. \end{aligned}$$

Für $z^0 \in K_{\delta^*}(z^*)$ ist die GANM also wohldefiniert und konvergiert mindestens linear. Gilt zusätzlich noch die zweite Bedingung aus dem Satz, d.h.

$$\|(G_k - H_k)(x^{k+1} - x^k)\|_2 = o(\|z^{k+1} - z^k\|_2),$$

so folgt analog zu oben

$$\begin{aligned} \|z^{k+1} - z^*\|_2 &\leq \|B_{(I)k}^{-1}\|_2 (\|\Phi^{(I)}(z^k) - \Phi^{(I)}(z^*) - W_{(I)k}(z^k - z^*)\|_2 + \\ &\quad \|(W_{(I)k} - B_{(I)k})(z^{k+1} - z^k)\|_2 + \|(W_{(I)k} - B_{(I)k})(z^{k+1} - z^*)\|_2) \\ &\leq \frac{4}{3}c \left(o(\|z^k - z^*\|_2) + \|(H_k - G_k)(x^{k+1} - x^k)\|_2 + \frac{1}{4c} \|z^{k+1} - z^*\|_2 \right) \\ &= o(\|z^k - z^*\|_2) + o(\|z^{k+1} - z^k\|_2) + \frac{1}{3} \|z^{k+1} - z^*\|_2 \end{aligned}$$

Wegen der oben bewiesenen Konvergenz impliziert $o(\|z^{k+1} - z^k\|_2)$ insbesondere auch $o(\|z^k - z^*\|_2)$. Damit folgt aus der obigen Gleichung die superlineare Konvergenz des Verfahrens für alle $z^0 \in K_{\delta^*}(z^*)$ wegen

$$\|z^{k+1} - z^*\|_2 = o(\|z^k - z^*\|_2).$$

□

Bemerkung 4.1.8. Wählen wir $G_k \in \pi_x \partial \Phi_1(z^k)$, so sind die beiden Bedingungen aus Satz 4.1.7 trivialerweise erfüllt.

Wie oben schon erwähnt, interessieren wir uns ja eigentlich für die Konvergenzeigenschaften des SQP-Verfahrens.

Algorithmus 4.1.9. (SQP-Verfahren)

(S.0) Wähle einen Startvektor $z^0 = (x^0, \lambda^0, \mu^0) \in \mathbb{R}^{n+m+p}$, setze $k = 0$.

(S.1) Ist ein geeignetes Abbruchkriterium erfüllt: STOPP

(S.2) Finde einen KKT-Punkt $z^{k+1} = (x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ des Problems

$$\begin{aligned} \min \quad & \nabla f(x^k)^T(x - x^k) + \frac{1}{2}(x - x^k)^T G_k(x - x^k) \\ \text{u.d.N.} \quad & g(x^k) + \nabla g(x^k)^T(x - x^k) \leq 0 \\ & h(x^k) + \nabla h(x^k)^T(x - x^k) = 0 \end{aligned} \quad (4.3)$$

Ist z^{k+1} nicht eindeutig, so wähle denjenigen KKT-Punkt von (4.3), der den Abstand $\|z^{k+1} - z^k\|$ zu z^k minimiert. Setze $k \leftarrow k + 1$, gehe zu(S.1).

Mit dem Satz über die Konvergenzeigenschaften der GANM können wir nun auch beweisen, dass das SQP-Verfahren lokal superlinear konvergiert.

Satz 4.1.10. *Der Punkt $z^* = (x^*, \lambda^*, \mu^*)$ genüge der strong second-order sufficiency condition für Problem (4.1) und LICQ. Hierbei seien f, g und h SC^1 -Funktionen. Weiter seien $\delta^*, c > 0$ die Konstanten aus Satz 4.1.7 und Lemma 4.1.6. Dann gibt es eine Konstante $\varepsilon > 0$, so dass gilt:*

Existiert für alle $k \in \mathbb{N}$ eine Matrix $H_k \in \pi_x \partial \Phi_1(z^k)$, so dass

$$\|G_k - H_k\|_2 \leq \min\left\{\varepsilon, \frac{1}{4c}\right\}$$

gilt, so ist das SQP-Verfahren wohldefiniert und konvergiert lokal linear. Gilt zusätzlich

$$\lim_{k \rightarrow \infty} \frac{\|(G_k - H_k)(x^{k+1} - x^k)\|_2}{\|z^{k+1} - z^k\|_2} = 0,$$

so konvergiert das Verfahren lokal superlinear.

Beweis. Um diesen Satz zu beweisen, müssen wir nur noch zeigen, dass das SQP-Verfahren in einer Umgebung von z^* ein Spezialfall der GANM ist. Mit Satz 4.1.7 folgen dann die

Behauptungen.

Zunächst definieren wir für $z = (x, \lambda, \mu)$ die Indexmengen

$$\alpha(z) := \{i : -g_i(x) < \lambda_i\} \quad \beta(z) := \{i : -g_i(x) = \lambda_i\} \quad \gamma(z) := \{i : -g_i(x) > \lambda_i\}$$

Offensichtlich gilt $\alpha(z^*) = I_{0+}(z^*)$, $\beta(z^*) = I_{00}(z^*)$ und $\gamma(z^*) = I_-(z^*)$.

Dann wählen wir $\delta_2 \in (0, \delta^*/3]$, so dass für alle $z, z^k \in K_{3\delta_2}(z^*)$ gilt

$$\begin{aligned} -g_i(x^k) - \nabla g_i(x^k)^T(x - x^k) &< \lambda_i^k \quad \text{für alle } i \in \alpha(z^*), \\ -g_i(x^k) - \nabla g_i(x^k)^T(x - x^k) &> \lambda_i^k \quad \text{für alle } i \in \gamma(z^*). \end{aligned} \quad (4.4)$$

Dies geht, da die Funktionen $-g_i(y) - \nabla g_i(y)^T(x - y) - \lambda_i$ stetig in allen Variablen sind und $-g_i(x^*) < \lambda_i^*$ für alle $i \in \alpha(z^*)$ sowie $-g_i(x^*) > \lambda_i^*$ für alle $i \in \gamma(z^*)$ gilt. Setzen wir $x = x^k$, so folgt daraus für alle $x^k \in K_{3\delta_2}$

$$\alpha(z^*) \subseteq \alpha(z^k), \quad \beta(z^*) \supseteq \beta(z^k), \quad \gamma(z^*) \subseteq \gamma(z^k).$$

Die KKT-Bedingungen des in jeder Iteration des SQP-Verfahrens auftauchenden quadratischen Problems (4.3) können wir wieder als nichtlineares Gleichungssystem

$$\Phi^k(z) = 0 \quad (4.5)$$

formulieren mit

$$\Phi^k(z) := \begin{pmatrix} \nabla f(x^k) + \nabla g(x^k)\lambda + \nabla h(x^k)\mu - G_k(x - x^k) \\ \min\{-g_1(x^k) - \nabla g_1(x^k)^T(x - x^k), \lambda_1\} \\ \vdots \\ \min\{-g_m(x^k) - \nabla g_m(x^k)^T(x - x^k), \lambda_m\} \\ h(x^k) + \nabla h(x^k)^T(x - x^k) \end{pmatrix}.$$

Um zu zeigen, dass das SQP-Verfahren wohldefiniert ist, müssen wir zeigen, dass die Gleichung (4.5) immer eine Lösung hat, falls δ_2 hinreichend klein ist. Hierzu betrachten wir zunächst die Matrix

$$A(z^*) := \begin{pmatrix} H_* & \nabla g_{\alpha(z^*)}(x^*) & \nabla h(x^*) \\ -\nabla g_{\alpha(z^*)}(x^*)^T & 0 & 0 \\ \nabla h(x^*)^T & 0 & 0 \end{pmatrix}$$

mit $H_* \in \pi_x \partial \Phi_1(z^*)$ beliebig. Wegen $\alpha(z^*) = I_{0+}(z^*)$, LICQ und der strong second-order sufficiency condition kann man analog zum Beweis von Satz 4.1.3 zeigen, dass $A(z^*)$ regulär ist.

Setzen wir

$$C(z^*) := \begin{pmatrix} \nabla g_{\beta(z^*)}(x^*) \\ 0 \\ 0 \end{pmatrix},$$

so ist die Matrix

$$B(z^*) := C(z^*)^T A(z^*)^{-1} C(z^*)$$

eine P-Matrix, d.h. eine Matrix mit positiven Hauptminoren. Der Beweis hierfür ist ähnlich zu dem Beweis von Theorem 4.1 in [20]. Definieren wir die Matrizen M_1, M_2 und M_3 durch

$$B(z^*) = C(z^*)^T (A(z^*)^{-1} C(z^*)) =: C(z^*)^T \begin{pmatrix} M_1 \\ M_2 \\ M_3 \end{pmatrix},$$

so gilt

$$\begin{pmatrix} \nabla g_{\beta(z^*)}(x^*) \\ 0 \\ 0 \end{pmatrix} = C(z^*) = A(z^*) \begin{pmatrix} M_1 \\ M_2 \\ M_3 \end{pmatrix} = \begin{pmatrix} H_* M_1 + \nabla g_{\alpha(z^*)}(x^*) M_2 + \nabla h(x^*) M_3 \\ -\nabla g_{\alpha(z^*)}(x^*)^T M_1 \\ \nabla h(x^*)^T M_1 \end{pmatrix}.$$

Folglich ist

$$B(z^*) = \nabla g_{\beta(z^*)}(x^*)^T M_1 = M_1^T H_*^T M_1 + M_2^T \nabla g_{\alpha(z^*)}(x^*)^T M_1 + M_3^T \nabla h(x^*)^T M_1 = M_1^T H_*^T M_1.$$

Wegen $H_* \in \pi_x \partial \Phi_1(z^*)$ und $\Phi_1(z) = \nabla_x L(x, \lambda, \mu)$ ist H_* symmetrisch, vergleiche [16], Theorem 3.3.4. Daher ist auch $B(z^*)$ eine symmetrische Matrix und als solche genau dann eine P-Matrix, wenn sie positiv definit ist. Sei also v ein Vektor mit $v^T B(z^*) v \leq 0$. Definieren wir $d := M_1 v$, so gilt auch $d^T H_* d \leq 0$. Nach dem obigen Gleichungssystem gilt aber außerdem

$$\begin{aligned} \nabla g_{\alpha(z^*)}(x^*)^T d &= \nabla g_{\alpha(z^*)}(x^*)^T M_1 v = 0, \\ \nabla h(x^*)^T d &= \nabla h(x^*)^T M_1 v = 0. \end{aligned}$$

Wie im Beweis von Satz 4.1.3 folgt daraus wegen $\alpha(z^*) = I_{0+}(z^*)$ auch

$$\nabla f(x^*)^T d = 0.$$

Also ist $d \in G(z^*)$ und mit der strong second-order sufficiency condition folgt $M_1 v = d = 0$. Damit folgt aus der ersten Gleichung in dem obigen Gleichungssystem

$$\nabla g_{\beta(z^*)}(x^*) v = \nabla g_{\alpha(z^*)}(x^*) M_2 v + \nabla h(x^*) M_3 v.$$

Wegen $\alpha(z^*) = I_{0+}(z^*)$, $\beta(z^*) = I_{00}(z^*)$ und LICQ folgt daraus

$$v = 0, \quad M_2 v = 0, \quad M_3 v = 0.$$

Wegen $v = 0$ ist $B(z^*)$ symmetrisch positiv definit, also eine P-Matrix.

Da die Abbildung $z \mapsto \partial \Phi_1(z)$ oberhalbstetig ist, gibt es zu jedem $\varepsilon > 0$ ein $\delta_3(\varepsilon) > 0$, so dass für alle $z^k \in K_{\delta_3(\varepsilon)}(z^*)$ gilt

$$\pi_x \partial \Phi_1(z^k) \subseteq \pi_x \partial \Phi_1(z^*) + K_\varepsilon(0).$$

Wegen $H_k \in \pi_x \partial \Phi_1(z^k)$ und $\|G_k - H_k\| \leq \varepsilon$ können wir daher, indem wir ε und δ_2 hinreichend verkleinern ($\delta_2 \leq \delta_3(\varepsilon)$), erreichen, dass für alle $z^k \in K_{\delta_2}(z^*)$ gilt: Die Matrix

$$A(z^k) := \begin{pmatrix} G_k & \nabla g_{\alpha(z^*)}(x^k) & \nabla h(x^k) \\ -\nabla g_{\alpha(z^*)}(x^k)^T & 0 & 0 \\ \nabla h(x^k)^T & 0 & 0 \end{pmatrix}$$

ist regulär und

$$B(z^k) := C(z^k)^T A(z^k)^{-1} C(z^k)$$

mit

$$C(z^k) := \begin{pmatrix} \nabla g_{\beta(z^*)}(x^k) \\ 0 \\ 0 \end{pmatrix}$$

ist eine P-Matrix. Man beachte, dass bei $A(z^k)$, $B(z^k)$ und $C(z^k)$ die Indexmengen $\alpha(z^*)$ und $\beta(z^*)$ immer noch von z^* und nicht von z^k abhängen.

Um die Lösbarkeit von $\Phi^k(z) = 0$ zu zeigen, untersuchen wir zunächst die Lösbarkeit des folgenden Systems:

Finde einen Vektor $d^k = (d_x^k, d_\lambda^k, d_\mu^k)$, der das folgende System löst:

$$\begin{aligned} \nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k + G_k d_x + \nabla g(x^k)d_\lambda + \nabla h(x^k)d_\mu &= 0, \\ -g_i(x^k) - \nabla g_i(x^k)^T d_x &= 0 \quad \text{für alle } i \in \alpha(z^*), \\ \min\{\lambda_i^k + (d_\lambda)_i, -g_i(x^k) - \nabla g_i(x^k)^T d_x\} &= 0 \quad \text{für alle } i \in \beta(z^*), \\ \lambda_i^k + (d_\lambda)_i &= 0 \quad \text{für alle } i \in \gamma(z^*), \\ h(x^k) + \nabla h(x^k)^T d_\mu &= 0. \end{aligned} \tag{4.6}$$

Gilt nämlich $z^k, z^k + d^k \in K_{3\delta_2}(z^*)$, so ist $z^k + d^k$ wegen (4.4) eine Lösung von $\Phi^k(z) = 0$. Für alle $i \in \gamma(z^*)$ ist folglich $(d_\lambda)_i = -\lambda_i^k$. Damit und mit den oben definierten Matrizen können wir das System (4.6) äquivalent in das folgende System umformen:

$$\begin{aligned} \begin{pmatrix} \nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k - \nabla g_\gamma(x^k)\lambda_\gamma^k \\ -g_\alpha(x^k) \\ h(x^k) \end{pmatrix} + A(z^k) \begin{pmatrix} d_x \\ (d_\lambda)_\alpha \\ d_\mu \end{pmatrix} + C(z^k)(d_\lambda)_\beta &= 0, \\ -g_\beta(x^k) - C(z^k)^T \begin{pmatrix} d_x \\ (d_\lambda)_\alpha \\ d_\mu \end{pmatrix} &\geq 0, \\ \lambda_\beta^k + (d_\lambda)_\beta &\geq 0, \\ \left[-g_\beta(x^k) - C(z^k)^T \begin{pmatrix} d_x \\ (d_\lambda)_\alpha \\ d_\mu \end{pmatrix} \right]^T (\lambda_\beta^k + (d_\lambda)_\beta) &= 0. \end{aligned} \tag{4.7}$$

Die erste Gleichung in (4.7) entspricht der ersten, zweiten, vierten und fünften Gleichung in (4.6) und die restlichen drei Gleichungen in (4.7) entsprechen der dritten Gleichung in (4.6). Der Übersichtlichkeit wegen stehen in (4.7) α, β, γ für die Indexmengen $\alpha(z^*), \beta(z^*), \gamma(z^*)$.

Für alle $z^k \in K_{\delta_2}(z^*)$ ist $(d_x, (d_\lambda)_\alpha, d_\mu)$ eindeutig bestimmt durch

$$\begin{pmatrix} d_x \\ (d_\lambda)_\alpha \\ d_\mu \end{pmatrix} = -A(z^k)^{-1} \begin{pmatrix} \nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k - \nabla g_\gamma(x^k)\lambda_\gamma^k \\ -g_\alpha(x^k) \\ h(x^k) \end{pmatrix} - A(z^k)^{-1} C(z^k)(d_\lambda)_\beta,$$

d.h. $(d_x, (d_\lambda)_\alpha, d_\mu)$ ist eindeutig bestimmt, falls $(d_\lambda)_\beta$ eindeutig bestimmt ist. Mit den Abkürzungen

$$w := -g_\beta(x^k) - C(z^k)^T \begin{pmatrix} d_x \\ (d_\lambda)_\alpha \\ d_\mu \end{pmatrix},$$

$$q := -g_\beta(x^k) + C(z^k)^T A(z^k)^{-1} \left[\begin{pmatrix} \nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k - \nabla g_\gamma(x^k)\lambda_\gamma^k \\ -g_\alpha(x^k) \\ h(x^k) \end{pmatrix} - C(z^k)\lambda_\beta^k \right]$$

können wir das System (4.7) umschreiben als

$$\begin{aligned} w - C(z^k)^T A(z^k)^{-1} C(z^k)(\lambda_\beta^k + (d_\lambda)_\beta) &= q, \\ w &\geq 0, \\ (\lambda_\beta^k + (d_\lambda)_\beta) &\geq 0, \\ w^T(\lambda_\beta^k + (d_\lambda)_\beta) &= 0. \end{aligned} \tag{4.8}$$

Nach Theorem 3.13 in [15] hat das lineare Komplementaritätsproblem (4.8) eindeutige Lösungen w und $\lambda_\beta^k + (d_\lambda)_\beta$ für alle $z^k \in K_{\delta_2}(z^*)$, da $B(z^k) := C(z^k)^T A(z^k)^{-1} C(z^k)$ für solche z^k eine P-Matrix ist. Wie oben schon erwähnt, hat dann auch das System (4.7) für alle $z^k \in K_{\delta_2}(z^*)$ eindeutige Lösungen $(d_x, (d_\lambda)_\alpha, d_\mu)$ und $(d_\lambda)_\beta$. Folglich hat das System (4.6) ebenfalls für alle $z^k \in K_{\delta_2}(z^*)$ eine eindeutige Lösung. Wir nennen diese Lösung

$$d^k = (d_x^k, d_\lambda^k, d_\mu^k).$$

Setzen wir

$$I := \{i \in \beta(z^*) : -g_i(x^k) - \nabla g_i(x^k)^T (d_\lambda^k) \leq \lambda_i^k + (d_\lambda^k)_i\},$$

so gilt wegen $\alpha(z^*) = I_{0+}(z^*)$, $\beta(z^*) = I_{00}(z^*)$ und $\gamma(z^*) = I_-(z^*)$ nach (4.6)

$$\begin{aligned} &\Phi^{(I)}(z^k) + B_{(I)k} d^k \\ &= \begin{pmatrix} \nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k \\ -g_{i \in \alpha(z^*)}(x^k) \\ -g_{i \in I}(x^k) \\ \lambda_{i \in \beta(z^*) \setminus I}^k \\ \lambda_{i \in \gamma(z^*)}^k \end{pmatrix} + \begin{pmatrix} G_k & \nabla g(x^k) & \nabla h(x^k) \\ \nabla g_{i \in \alpha(z^*)}(x^k)^T & 0 & 0 \\ \nabla g_{i \in I}(x^k)^T & 0 & 0 \\ 0 & e_{i \in \beta(z^*) \setminus I}^T & 0 \\ 0 & e_{i \in \gamma(z^*)}^T & 0 \end{pmatrix} \begin{pmatrix} d_x^k \\ d_\lambda^k \\ d_\mu^k \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \end{aligned}$$

Wegen $d^k = -B_{(I)k}^{-1} \Phi^{(I)}(z^k)$ gilt für alle $z^k \in K_{\delta_2}(z^*)$ nach Satz 4.1.7

$$\|(z^k + d^k) - z^*\|_2 \leq \frac{1}{2} \|z^k - z^*\|_2,$$

d.h. für alle $z^k \in K_{\delta_2}(z^*)$ gilt auch $z^{k+1} := z^k + d^k \in K_{\delta_2}(z^*)$. Wie nach System (4.6) schon erwähnt, ist $z^{k+1} := z^k + d^k$ daher eine Lösung von $\Phi^k(z) = 0$, d.h. ein KKT-Punkt des Problems (4.3).

Wir zeigen zum Abschluss noch, dass z^{k+1} diejenige Lösung von $\Phi^k(z) = 0$ ist, welche $\|z^{k+1} - z^k\|_2$ minimiert. Angenommen, es gibt eine weitere Lösung \bar{z} mit $\|\bar{z} - z^k\|_2 \leq \|z^{k+1} - z^k\|_2$. Wegen $z^k, z^{k+1} \in K_{\delta_2}(z^*)$ muss dann $\bar{z} \in K_{3\delta_2}(z^*)$ gelten. Nach (4.4) ist dann aber auch $\bar{z} - z^k$ eine Lösung des Systems (4.6). Da dieses System eindeutig lösbar war, folgt $\bar{z} = z^{k+1}$.

Für alle $z^0 \in K_{\delta_2}(z^*)$ ist das SQP-Verfahren also ein Spezialfall der GANM und erbt daher auch ihre Konvergenzeigenschaften. \square

4.2 Lokal quadratische Konvergenz des semismoothen SQP-Verfahrens

Auch in diesem Abschnitt betrachten wir wieder das Problem

$$\min f(x) \quad \text{u.d.N.} \quad g(x) \leq 0, h(x) = 0, \quad (4.9)$$

allerdings diesmal mit sSC^1 -Funktionen f, g und h . Da diese Voraussetzungen die im letzten Abschnitt gegebenen implizieren, können wir alle dort bewiesenen Ergebnisse auch hier anwenden. Wir wollen in diesem Abschnitt zeigen, dass unter den hier gegebenen Voraussetzungen die GANM und das SQP-Verfahren lokal quadratisch konvergieren. Diese Resultate finden sich nicht mehr in [5], lassen sich aber daraus leicht herleiten. Hierfür verwenden wir die Notation aus dem letzten Abschnitt.

Satz 4.2.1. *Der Punkt $z^* = (x^*, \lambda^*, \mu^*)$ genüge der strong second-order sufficiency condition für Problem (4.1) und LICQ. Hierbei seien f, g und h sSC^1 -Funktionen. Weiter seien $\delta, c > 0$ die Konstanten aus Lemma 4.1.6. Existiert für alle $k \in \mathbb{N}$ eine Matrix $H_k \in \pi_x \partial \Phi_1(z^k)$, so dass*

$$\frac{\|G_k - H_k\|_2}{\|z^k - z^*\|_2} \leq \frac{1}{2c}$$

gilt, so ist GANM wohldefiniert und konvergiert lokal quadratisch.

Beweis. Da f, g und h sSC^1 -Funktionen sind, sind Φ und $\Phi^{(I)}$, $I \in \mathcal{P}(I_{00}(z^*))$, strongly semismooth in z^* .

Auch hier wählen wir wieder das Element $W_{(I)k} \in \partial_Q \Phi^{(I)}(z^k)$ so, dass für alle $z = (x, \lambda, \mu)$ gilt

$$\|(B_{(I)k} - W_{(I)k})z\|_2 = \|(H_k - G_k)x\|_2.$$

Insbesondere folgt daraus für alle $k \in \mathbb{N}$ mit $\|z^k - z^*\|_2 \leq 1$

$$\begin{aligned} \|B_{(I)k} - W_{(I)k}\|_2 &= \sup_{z \neq 0} \frac{\|(B_{(I)k} - W_{(I)k})z\|_2}{\|z\|_2} \leq \sup_{z \neq 0} \frac{\|H_k - G_k\|_2 \|x\|_2}{\|z\|_2} \\ &\leq \|H_k - G_k\|_2 \leq \frac{1}{2c} \|z^k - z^*\|_2 \leq \frac{1}{2c}. \end{aligned}$$

Gilt $\|z^k - z^*\|_2 \leq \delta$, so ist $W_{(I)k}$ nach Lemma 4.1.6 regulär und es gilt $\|W_{(I)k}^{-1}\|_2 \leq c$. Nach dem Perturbation Lemma 2.3.3 von Ortega und Rheinboldt, vergleiche [16], ist daher auch $B_{(I)k}$ für alle $k \in \mathbb{N}$ mit $\|z^k - z^*\|_2 \leq \min\{\delta, 1\}$ regulär und es gilt

$$\|B_{(I)k}^{-1}\|_2 \leq \frac{c}{1 - c\frac{1}{2c}} = 2c.$$

Da die Funktion $\Phi^{(I)}$ in z^* strongly semismooth ist, ist sie nach Satz 3.1.7 dort auch in jeder Komponente $\Phi_j^{(I)}$, $j = 1, \dots, n + m + p$ strongly semismooth. Wegen $W_{(I)k} \in \partial_Q \Phi^{(I)}(z^k) \subseteq \partial_C \Phi^{(I)}(z^k)$ gilt für alle $j = 1, \dots, n + m + p$ auch $(W_{(I)k})_j \in \partial \Phi_j^{(I)}(z^k)$. Hierbei bezeichne $(W_{(I)k})_j$ die j -te Zeile von $W_{(I)k}$. Da es nur endlich viele Mengen $I \subseteq I_{00}(z^*)$

gibt, folgt aus Satz 3.1.7 die Existenz von Konstanten $C, \delta_1 > 0$, so dass für alle $k \in \mathbb{N}$ mit $\|z^k - z^*\|_2 \leq \delta_1$ und alle $I \subseteq I_{00}(z^*)$ gilt

$$\begin{aligned} \|\Phi^{(I)}(z^k) - \Phi^{(I)}(z^*) - W_{(I)k}(z^k - z^*)\|_2 &\leq \sum_{j=1}^{n+m+p} \left| \Phi_j^{(I)}(z^k) - \Phi_j^{(I)}(z^*) - (W_{(I)k})_j(z^k - z^*) \right| \\ &\leq C \|z^k - z^*\|_2^2. \end{aligned}$$

Setzen wir nun $\delta_2 := \min\{\delta, \delta_1, 1\}$, so folgt für alle $k \in \mathbb{N}$ mit $\|z^k - z^*\| \leq \delta_2$ wegen $\Phi^{(I)}(z^*) = 0$

$$\begin{aligned} \|z^{k+1} - z^*\|_2 &= \|z^k - B_{(I)k}^{-1} \Phi^{(I)}(z^k) - z^*\|_2 \\ &\leq \|B_{(I)k}^{-1}\|_2 \|\Phi^{(I)}(z^k) - \Phi^{(I)}(z^*) - B_{(I)k}(z^k - z^*)\|_2 \\ &\leq \|B_{(I)k}^{-1}\|_2 (\|\Phi^{(I)}(z^k) - \Phi^{(I)}(z^*) - W_{(I)k}(z^k - z^*)\|_2 \\ &\quad + \|(W_{(I)k} - B_{(I)k})(z^k - z^*)\|_2) \\ &\leq 2c \left(C \|z^k - z^*\|_2^2 + \frac{1}{2c} \|z^k - z^*\|_2^2 \right) \\ &= (2cC + 1) \|z^k - z^*\|_2^2. \end{aligned}$$

Für z^0 mit $\|z^0 - z^*\|_2 \leq \delta^* := \min\{\delta_2, \frac{1}{2(2cC+1)}\}$ ist die GANM also wohldefiniert und konvergiert quadratisch. \square

Bemerkung 4.2.2. Wählen wir $G_k \in \pi_x \partial \Phi_1(z^k)$, so ist die Bedingung aus Satz 4.2.1 trivialerweise erfüllt.

Bemerkung 4.2.3. Wegen der gerade bewiesenen Konvergenz impliziert die Bedingung, dass für alle $k \in \mathbb{N}$ eine Matrix $H_k \in \pi_x \partial \Phi_1(z^k)$ existiert, so dass

$$\frac{\|G_k - H_k\|_2}{\|z^k - z^*\|_2} \leq \frac{1}{2c}$$

gilt, insbesondere, dass es ein $K \in \mathbb{N}$ gibt, so dass für alle $k \geq K$ gilt

$$\|G_k - H_k\|_2 \leq \frac{1}{4c}.$$

Außerdem gilt

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\|(G_k - H_k)(x^{k+1} - x^k)\|_2}{\|z^{k+1} - z^k\|_2} &\leq \lim_{k \rightarrow \infty} \frac{\|G_k - H_k\|_2 \|x^{k+1} - x^k\|_2}{\|z^{k+1} - z^k\|_2} \\ &\leq \lim_{k \rightarrow \infty} \|G_k - H_k\|_2 \\ &\leq \lim_{k \rightarrow \infty} \frac{1}{2c} \|z^k - z^*\|_2 = 0. \end{aligned}$$

Die sind aber gerade die zwei Bedingungen, die in Satz 4.1.7 für lineare bzw. superlineare Konvergenz gestellt wurden.

Die lokal quadratische Konvergenz des SQP-Verfahrens unter den Bedingungen in diesem Abschnitt folgt aus dem nächsten Satz, dem Pendant zu Satz 4.1.10.

Satz 4.2.4. *Der Punkt $z^* = (x^*, \lambda^*, \mu^*)$ genüge der strong second-order sufficiency condition für Problem (4.1) und LICQ. Hierbei seien f, g und h sSC^1 -Funktionen. Weiter seien $\delta^*, c > 0$ die Konstanten aus Satz 4.2.1 und Lemma 4.1.6. Dann gibt es eine Konstante $\varepsilon > 0$, so dass gilt:*

Existiert für alle $k \in \mathbb{N}$ eine Matrix $H_k \in \pi_x \partial \Phi_1(z^k)$, so dass

$$\|G_k - H_k\|_2 \leq \varepsilon$$

und

$$\frac{\|G_k - H_k\|_2}{\|z^k - z^*\|_2} \leq \frac{1}{2c}$$

gilt, so ist das SQP-Verfahren wohldefiniert und konvergiert lokal quadratisch.

Der Beweis dieses Satzes geht völlig analog zu dem Beweis von Satz 4.1.10, deswegen wird er an dieser Stelle nicht noch einmal wiederholt.

4.3 Hinreichende Bedingungen für die strong second-order sufficiency condition

Wie in Kapitel 1.1 erläutert, sind die folgenden beiden Probleme äquivalent:

- Finde ein normalisiertes Nash-Gleichgewicht x^* des GNEPs.
- Finde eine Lösung x^* von

$$\min V_\gamma(x) \quad \text{u.d.N.} \quad g(x) \leq 0, h(x) = 0 \quad (4.10)$$

mit $V_\gamma(x^*) = 0$.

Aus Korollar 3.3.4 wissen wir, dass gilt:

Sei x^* ein normalisiertes Nash-Gleichgewicht eines GNEPs, das den Voraussetzungen 1.1.1 bzw. 1.1.2 genügt. Gilt in x^* LICQ, so ist die Abbildung $x \mapsto y_\gamma(x)$ in einer Umgebung von x^* (strongly) semismooth. Daher ist auch $\nabla V_\gamma(x)$ in einer Umgebung von x^* semismooth.

Sind die zweiten Ableitungen aller Funktionen θ_ν sowie die zweite Ableitung von g zusätzlich lokal Lipschitz-stetig, so sind $y_\gamma(x)$ und $\nabla V_\gamma(x)$ strongly semismooth in einer Umgebung von x^* .

Deswegen wollen wir ein semismoothes SQP-Verfahren anwenden, um das Problem (4.10) zu lösen. Da wir in x^* LICQ voraussetzen, gibt es eindeutige Lagrange-Multiplikatoren λ^*, μ^* , so dass $z^* = (x^*, \lambda^*, \mu^*)$ ein KKT-Punkt von (4.10) ist. In den Konvergenzsätzen 4.1.10 und 4.2.4 war jedoch zusätzlich vorausgesetzt, dass in z^* die strong second-order sufficiency condition erfüllt ist. Mit

$$\Phi_1(z) = \nabla V_\gamma(x) + \nabla g(x)\lambda + \nabla h(x)\mu,$$

$$I_{0+} = \{i : g_i(x^*) = 0, \lambda_i^* > 0\}$$

und

$$G(z^*) := \{d \in \mathbb{R}^n \mid \nabla V_\gamma(x^*)^T d = 0, \nabla g_i(x^*)^T d = 0 \forall i \in I_{0+}, \nabla h_i(x^*)^T d = 0 \forall i = 1, \dots, p\}$$

bedeutet dies, dass für alle $d \in G(z^*) \setminus \{0\}$ und alle $H \in \pi_x \partial \Phi_1(z^*)$ gelten muss

$$d^T H d > 0.$$

Um zu untersuchen, unter welchen Bedingungen die strong second-order sufficiency condition in z^* erfüllt ist, wollen wir zunächst die Struktur der Elemente $H \in \pi_x \partial_B \Phi_1(z^*)$ genauer betrachten.

Nach Satz 1.1.3 ist $V_\gamma(x)$ unter den gegebenen Voraussetzungen stetig differenzierbar und es gilt für alle $x \in \mathbb{R}^n$

$$\nabla V_\gamma(x) = \nabla_x \Psi(x, y) \Big|_{y=y_\gamma(x)}.$$

Wir können daher schreiben $\Phi_1 = H_2 \circ H_1$ mit einer (strongly) semismoothen Funktion

$$H_1(x, \lambda, \mu) = \begin{pmatrix} x \\ y_\gamma(x) \\ \lambda \\ \mu \end{pmatrix}$$

und einer stetig differenzierbaren Funktion

$$H_2(x, y, \lambda, \mu) = \nabla_x \Psi_\gamma(x, y) + \nabla g(x) \lambda + \nabla h(x) \mu.$$

Mit Satz 3.1.4 folgt daraus für alle $\tilde{d} \in \mathbb{R}^{n+m+p}$

$$\partial \Phi_1(x, \lambda, \mu) \tilde{d} = H_2'(x, y_\gamma(x), \lambda, \mu) \cdot \partial H_1(x, \lambda, \mu) \tilde{d}.$$

Da die Funktion h affin linear ist, gilt

$$H_2'(x, y_\gamma(x), \lambda, \mu) = (\nabla_{xx}^2 \Psi_\gamma(x, y_\gamma(x)) + \sum_{i=1}^m \lambda_i \nabla^2 g_i(x), \quad \nabla_{xy}^2 \Psi_\gamma(x, y_\gamma(x)), \quad \nabla g(x), \quad \nabla h(x)).$$

Über die Struktur von $\partial H_1(x, \lambda, \mu)$ gibt das folgende Lemma Aufschluss.

Lemma 4.3.1. *Es gilt*

$$\partial_B H_1(x, \lambda, \mu) = \begin{pmatrix} I & 0 & 0 \\ \partial_B y_\gamma(x) & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \text{ und } \partial H_1(x, \lambda, \mu) = \begin{pmatrix} I & 0 & 0 \\ \partial y_\gamma(x) & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}.$$

Beweis. Wir zeigen zuerst die Aussage über das B-Subdifferential für festes (x, λ, μ) .

Sei zunächst $M \in \partial_B y_\gamma(x)$ beliebig. Dann gibt es eine Folge $\{x^k\}_k \subseteq D_{y_\gamma}$, so dass $x^k \rightarrow x$ und $y'_\gamma(x^k) \rightarrow M$ gilt. Dann ist offensichtlich $\{(x^k, \lambda, \mu)\}_k \subseteq D_{H_1}$ und es gilt

$$\lim_{k \rightarrow \infty} H_1'(x^k, \lambda, \mu) = \lim_{k \rightarrow \infty} \begin{pmatrix} I & 0 & 0 \\ y'_\gamma(x^k) & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} = \begin{pmatrix} I & 0 & 0 \\ M & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} =: H.$$

Folglich gilt $H \in \partial_B H_1(x, \lambda, \mu)$. Dies zeigt die eine Inklusion.

Betrachten wir nun umgekehrt ein beliebiges Element $H \in \partial_B H_1(x, \lambda, \mu)$. Dann gibt es

wiederum eine Folge $\{x^k, \lambda^k, \mu^k\}_k \subseteq D_{H_1}$, so dass $(x^k, \lambda^k, \mu^k) \rightarrow (x, \lambda, \mu)$ und außerdem $H'_1(x^k, \lambda^k, \mu^k) \rightarrow H$ gilt. Folglich ist H von der Gestalt

$$H = \begin{pmatrix} I & 0 & 0 \\ M & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}$$

mit einer passenden Matrix M . Da die Funktion H_1 in den Punkten (x^k, λ^k, μ^k) differenzierbar ist, ist sie dort auch in jeder Komponentenfunktion differenzierbar. Also gilt insbesondere $\{x^k\}_k \subseteq D_{y_\gamma}$. Nach Definition von M gilt außerdem

$$\lim_{k \rightarrow \infty} y'_\gamma(x^k) = M.$$

Daraus folgt $M \in \partial_B y_\gamma(x)$. Dies zeigt die noch fehlende Inklusion.

Die Aussage über den verallgemeinerten Gradienten folgt aus den beiden Identitäten $\partial H_1(x, \lambda, \mu) = \text{conv}\{\partial_B H_1(x, \lambda, \mu)\}$ und $\partial y_\gamma(x) = \text{conv}\{\partial_B y_\gamma(x)\}$. \square

Insgesamt folgt daher für die Struktur von $\partial \Phi_1(x^*, \lambda^*, \mu^*)$:

$$\begin{aligned} \partial \Phi_1(x^*, \lambda^*, \mu^*) \cdot \tilde{d} = & H'_2(x^*, y_\gamma(x^*), \lambda^*, \mu^*) \cdot \partial H_1(x^*, \lambda^*, \mu^*) \cdot \tilde{d} = \\ & \left(\nabla_{xx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) + \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x^*) + \nabla_{xy}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) \cdot \partial y_\gamma(x^*), \nabla g(x^*), \nabla h(x^*) \right) \cdot \tilde{d} \end{aligned}$$

für alle $\tilde{d} \in \mathbb{R}^{n+m+p}$. Eigentlich interessiert uns allerdings weniger $\partial \Phi_1(x^*, \lambda^*, \mu^*)$ als $\pi_x \partial \Phi_1(x^*, \lambda^*, \mu^*)$. Aber auch diese Menge können wir nun wie folgt angeben:

$$\pi_x \partial \Phi_1(x^*, \lambda^*, \mu^*) \cdot d = \nabla_{xx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) \cdot d + \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x^*) \cdot d + \nabla_{xy}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) \cdot \partial y_\gamma(x^*) \cdot d$$

für beliebiges $d \in \mathbb{R}^n$.

Wir wollen ja untersuchen, wann für alle $H \in \pi_x \partial \Phi_1(x^*, \lambda^*, \mu^*)$ und alle $d \in G(z^*) \setminus \{0\}$ gilt $d^T H d > 0$. Da die Funktionen g_i als konvex vorausgesetzt waren, sind die Matrizen $\nabla^2 g_i(x^*)$ alle positiv semidefinit. Desweiteren gilt $\lambda_i^* \geq 0$ für alle $i = 1, \dots, m$, da z^* ein KKT-Punkt von (4.10) ist. Also ist auch die Matrix $\sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x^*)$ positiv semidefinit. Daher genügt es zu zeigen, dass für alle $d \in G(z^*) \setminus \{0\}$ und alle $M \in \partial y_\gamma(x^*)$ gilt

$$d^T \nabla_{xx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) d + d^T \nabla_{xy}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) M d > 0. \quad (4.11)$$

Betrachten wir zum Vergleich $\partial(\nabla V_\gamma(x^*))$, so folgt mit der Kettenregel 3.1.4 für alle $d \in \mathbb{R}^n$

$$\begin{aligned} \partial(\nabla V_\gamma(x^*)) \cdot d &= \partial(\nabla_x \Psi_\gamma(x^*, y_\gamma(x^*))) \cdot d \\ &= \left(\nabla_{xx}^2 \Psi(x^*, y_\gamma(x^*)) \quad \nabla_{xy}^2 \Psi(x^*, y_\gamma(x^*)) \right) \cdot \begin{pmatrix} I \\ \partial y_\gamma(x^*) \end{pmatrix} \cdot d \\ &= \nabla_{xx}^2 \Psi(x^*, y_\gamma(x^*)) \cdot d + \nabla_{xy}^2 \Psi(x^*, y_\gamma(x^*)) \partial y_\gamma(x^*) \cdot d. \end{aligned}$$

Also können wir äquivalent zu (4.11) zeigen, dass für alle Elemente $H \in \partial(\nabla V_\gamma(x^*))$ und alle $d \in G(z^*) \setminus \{0\}$ gilt $d^T H d > 0$.

Dies führt zu einer ersten hinreichenden Bedingung, die die strong second-order sufficiency condition garantiert.

Satz 4.3.2. Sei x^* ein normalisiertes Nash-Gleichgewicht eines GNEPs, das den Voraussetzungen 1.1.1 und LICQ genügt. Ist die Funktion $\Psi_\gamma(x, y)$ gleichmäßig gleichmäßig konvex bezüglich x , d.h. gibt es eine Konstante $\beta > 0$, so dass für alle Punkte $x_1, x_2, y \in X$ und alle $\alpha \in (0, 1)$ gilt

$$\Psi_\gamma(\alpha x_1 + (1 - \alpha)x_2, y) \leq \alpha \Psi_\gamma(x_1, y) + (1 - \alpha) \Psi_\gamma(x_2, y) - \beta \alpha (1 - \alpha) \|x_1 - x_2\|_2^2,$$

so ist die strong second-order sufficiency condition in x^* erfüllt.

Beweis. Wir werden in diesem Beweis zeigen, dass unter den gegebenen Voraussetzungen die Funktion V_γ gleichmäßig konvex ist. Da V_γ stetig differenzierbar ist, folgt dann, dass ∇V_γ eine gleichmäßig monotone Funktion ist. Proposition 2.3 aus [9] garantiert uns dann, dass alle Elemente von $\partial(\nabla V_\gamma(x^*))$ positiv definit sind. Wie wir oben gezeigt haben, ist dies aber eine hinreichende Bedingung dafür, dass die strong second-order sufficiency condition in x^* erfüllt ist.

Seien also $x_1, x_2 \in X$ und $\alpha \in (0, 1)$ beliebig. Dann gilt

$$\begin{aligned} V_\gamma(\alpha x_1 + (1 - \alpha)x_2) &= \Psi_\gamma(\alpha x_1 + (1 - \alpha)x_2, y_\gamma(\alpha x_1 + (1 - \alpha)x_2)) \\ &\leq \alpha \Psi_\gamma(x_1, y_\gamma(\alpha x_1 + (1 - \alpha)x_2)) + (1 - \alpha) \Psi_\gamma(x_2, y_\gamma(\alpha x_1 + (1 - \alpha)x_2)) \\ &\quad - \beta \alpha (1 - \alpha) \|x_1 - x_2\|_2^2 \\ &\leq \alpha \Psi_\gamma(x_1, y_\gamma(x_1)) + (1 - \alpha) \Psi_\gamma(x_2, y_\gamma(x_2)) - \beta \alpha (1 - \alpha) \|x_1 - x_2\|_2^2 \\ &= \alpha V_\gamma(x_1) + (1 - \alpha) V_\gamma(x_2) - \beta \alpha (1 - \alpha) \|x_1 - x_2\|_2^2. \end{aligned}$$

Also ist V_γ unter den gegebenen Bedingungen gleichmäßig konvex mit Modulus $\beta > 0$. \square

Im Folgenden wollen wir zeigen, dass man diese hinreichende Bedingung noch weiter abschwächen kann. Hierzu werden wir den in (4.11) auftretenden Ausdruck Md mit $M \in \partial y_\gamma(x^*)$ und $d \in \mathbb{R}^n$ beliebig genauer untersuchen.

Hierfür betrachten wir wieder die Funktion

$$\tilde{\Phi}(x, y, \lambda, \mu) = \begin{pmatrix} -\nabla_y \Psi_\gamma(x, y) + \nabla g(y)\lambda + \nabla h(y)\mu \\ \min\{-g_i(y), \lambda_i\}_{i=1, \dots, m} \\ h(y) \end{pmatrix}.$$

Setzen wir voraus, dass in $x^* = y_\gamma(x^*)$ zusätzlich zu den Voraussetzungen 1.1.1 bzw 1.1.2 auch LICQ gilt, so gibt es, wie in Kapitel 3.3 bewiesen, eine offene Umgebung $U(x^*)$ von x^* , auf der es eine (strongly) semismoothe Abbildung $G(x) = (y_\gamma(x), \lambda(x), \mu(x))$ mit den folgenden Eigenschaften gibt: Für alle $x \in U(x^*)$ gilt $\tilde{\Phi}(x, G(x)) = 0$ und alle Elemente von $\pi_{(y, \lambda, \mu)} \partial \tilde{\Phi}(x, y_\gamma(x), \lambda(x), \mu(x))$ sind regulär.

Da die Funktionen $\lambda(x)$ und $\mu(x)$ in $U(x^*)$ semismooth sind, sind nach dem Satz von Rademacher die Mengen $U(x^*) \setminus D_\lambda$ und $U(x^*) \setminus D_\mu$ Nullmengen. Daher gilt mit Lemma 3.1.3 für alle $d \in \mathbb{R}^n$

$$\begin{aligned} \partial y_\gamma(x^*) \cdot d &= \text{conv}\{M \in \mathbb{R}^{n \times n} \mid \exists \{x^k\}_k \subseteq D_{y_\gamma} \cap D_\lambda \cap D_\mu : x^k \rightarrow x^*, y'_\gamma(x^k) \rightarrow M\} \cdot d \\ &=: \text{conv}\{S\} \cdot d. \end{aligned}$$

Wählen wir nun eine beliebige Matrix $M \in S$, so gibt es eine Folge $\{x^k\}_k \subseteq D_{y_\gamma} \cap D_\lambda \cap D_\mu$ mit $x^k \rightarrow x^*$ und $y'_\gamma(x^k) \rightarrow M$. O.B.d.A können wir $\{x^k\}_k \subseteq U(x^*)$ annehmen. Wegen $\tilde{\Phi}(x, G(x)) = 0$ für alle $x \in U(x^*)$ dann gilt für alle $k \in \mathbb{N}$

$$0 = D_x \tilde{\Phi}(x, y_\gamma(x), \lambda(x), \mu(x)) \big|_{x=x^k}$$

und folglich auch für beliebiges, aber festes $d \in \mathbb{R}^n$

$$0 = D_x \tilde{\Phi}(x, y_\gamma(x), \lambda(x), \mu(x))|_{x=x^k} \cdot d = \tilde{\Phi}'(x, y_\gamma(x), \lambda(x), \mu(x); d)|_{x=x^k}.$$

Nun betrachten wir die einzelnen Komponenten der Funktion $\tilde{\Phi}$ unter Berücksichtigung von $\{x^k\}_k \subseteq D_{y_\gamma} \cap D_\lambda \cap D_\mu$:

$$\begin{aligned} \tilde{\Phi}'_1(x, y_\gamma(x), \lambda(x), \mu(x); d)|_{x=x^k} &= -\nabla_{yx}^2 \Psi_\gamma(x^k, y_\gamma(x^k)) \cdot d - \nabla_{yy}^2 \Psi_\gamma(x^k, y_\gamma(x^k)) \cdot y'_\gamma(x^k) \cdot d \\ &\quad + \sum_{i=1}^m \lambda_i(x^k) \nabla^2 g_i(y_\gamma(x^k)) \cdot y'_\gamma(x^k) \cdot d \\ &\quad + \nabla g(y_\gamma(x^k)) \cdot \lambda'(x^k) \cdot d + \nabla h(y_\gamma(x^k)) \cdot \mu'(x^k) \cdot d \\ \tilde{\Phi}'_3(x, y_\gamma(x), \lambda(x), \mu(x); d)|_{x=x^k} &= \nabla h(y_\gamma(x^k))^T \cdot y'_\gamma(x^k) \cdot d \end{aligned}$$

Um die Komponenten $\tilde{\Phi}_2$ genauer untersuchen zu können, erinnern wir uns daran, dass für alle $k \in \mathbb{N}$ wegen $x^k \in U(x^*)$ gilt $g(y_\gamma(x^k)) \leq 0$, $\lambda(x^k) \geq 0$ und $g(y_\gamma(x^k))^T \lambda(x^k) = 0$. Daher können wir wieder die folgenden Indexmengen definieren

$$I_-(x^k) = \{i : g_i(y_\gamma(x^k)) < 0\}, \quad I_0(x^k) = \{i : g_i(y_\gamma(x^k)) = 0\},$$

$$I_{0+}(x^k) = \{i \in I_0(x^k) : \lambda_i(x^k) > 0\}, \quad I_{00}(x^k) = \{i \in I_0(x^k) : \lambda_i(x^k) = 0\}.$$

Damit gilt für alle $i \in I_{0+}(x^k)$ wegen der Stetigkeit von $\lambda(x)$

$$(\tilde{\Phi}'_2)_i(x, y_\gamma(x), \lambda(x), \mu(x); d)|_{x=x^k} = -g'_i(y_\gamma(x); d)|_{x=x^k} = -\nabla g_i(y_\gamma(x^k))^T \cdot y'_\gamma(x^k) \cdot d$$

und für alle $i \in I_-(x^k)$ wegen der Stetigkeit von $g(y_\gamma(x))$

$$(\tilde{\Phi}'_2)_i(x, y_\gamma(x), \lambda(x), \mu(x); d)|_{x=x^k} = \lambda'_i(x^k; d) = e_i^T \cdot \lambda'(x^k) \cdot d$$

mit dem i -ten Einheitsvektor $e_i \in \mathbb{R}^m$.

Um $(\tilde{\Phi}'_2)_i(x, y_\gamma(x), \lambda(x), \mu(x); d)|_{x=x^k}$ für die noch verbleibenden $i \in I_{00}(x^k)$ zu bestimmen, greifen wir auf die Definition der Richtungsableitung zurück:

$$\begin{aligned} 0 &= (\tilde{\Phi}'_2)_i(x, y_\gamma(x), \lambda(x), \mu(x); d)|_{x=x^k} \\ &= \lim_{t \downarrow 0} \frac{\min\{-g_i(y_\gamma(x^k + td)), \lambda_i(x^k + td)\} - \min\{-g_i(y_\gamma(x^k)), \lambda_i(x^k)\}}{t}. \end{aligned}$$

Da die Richtungsableitung existiert, genügt es, jeweils eine bestimmte Folge $\{t_j\}_j \subset \mathbb{R}_{++}$ mit $t_j \rightarrow 0$ zu betrachten. Hierbei können wir o.B.d.A. davon ausgehen, dass t_j für alle $j \in \mathbb{N}$ so klein ist, dass $x^k + t_j d \in U(x^*)$ gilt. Dann kann für alle $j \in \mathbb{N}$ nur einer der folgenden drei Fälle auftreten:

- 1.) $g(y_\gamma(x^k + t_j d)) = 0$ und $\lambda(x^k + t_j d) > 0$
- 2.) $g(y_\gamma(x^k + t_j d)) < 0$ und $\lambda(x^k + t_j d) = 0$
- 3.) $g(y_\gamma(x^k + t_j d)) = 0$ und $\lambda(x^k + t_j d) = 0$

Offensichtlich muss mindestens einer dieser drei Fälle für unendlich viele $j \in \mathbb{N}$ eintreten. Gibt es eine Folge $\{t_j\}_j$, so dass für alle $j \in \mathbb{N}$ der erste Fall eintritt, so folgt mit der Definition der Richtungsableitung

$$\begin{aligned} 0 &= (\tilde{\Phi}'_2)_i(x, y_\gamma(x), \lambda(x), \mu(x); d) |_{x=x^k} = \lim_{j \rightarrow \infty} \frac{-g_i(y_\gamma(x^k + t_j d)) - (-g_i(y_\gamma(x^k)))}{t_j} \\ &= -g'_i(y_\gamma(x); d) |_{x=x^k} \\ &= -\nabla g_i(y_\gamma(x^k))^T \cdot y'_\gamma(x^k) \cdot d. \end{aligned}$$

Wir definieren dann $I_{001}(x^k)$ als die Menge aller Indizes i , für die es ein $\varepsilon > 0$ gibt, so dass für alle $0 < t < \varepsilon$ immer der erste Fall eintritt.

Gibt es eine Folge $\{t_j\}_j$, so dass für alle $j \in \mathbb{N}$ der zweite Fall eintritt, so folgt analog zu oben

$$0 = (\tilde{\Phi}'_2)_i(x, y_\gamma(x), \lambda(x), \mu(x); d) |_{x=x^k} = \lambda'_i(x^k; d) = e_i^T \cdot \lambda'(x^k) \cdot d$$

und wir definieren $I_{000}(x^k)$ als die Menge aller Indizes i , für die es ein $\varepsilon > 0$ gibt, so dass für alle $0 < t < \varepsilon$ immer der zweite Fall eintritt.

Gibt es eine Folge $\{t_j\}_j$, so dass für alle $j \in \mathbb{N}$ immer der dritte Fall eintritt, so gilt $g'_i(y_\gamma(x); d) |_{x=x^k} = 0$ und $\lambda'_i(x^k; d) = 0$. Wir können daher mit einem beliebigen $\alpha_i \in \{0, 1\}$ schreiben

$$\begin{aligned} 0 &= (\tilde{\Phi}'_2)_i(x, y_\gamma(x), \lambda(x), \mu(x); d) |_{x=x^k} \\ &= -\alpha_i \underbrace{\nabla g_i(y_\gamma(x^k))^T \cdot y'_\gamma(x^k) \cdot d}_{=0} + (1 - \alpha_i) \underbrace{e_i^T \cdot \lambda'(x^k) \cdot d}_{=0}. \end{aligned}$$

Gibt es Folgen $\{t_j\}_j$ und $\{t_i\}_i$, so dass für alle t_j der erste Fall eintritt und für alle t_i der zweite Fall, so folgt aus den Überlegungen zu diesen Fällen ebenfalls, dass $g'_i(y_\gamma(x); d) |_{x=x^k} = 0$ und $\lambda'_i(x^k; d) = 0$ gelten muss. Wir setzen daher $I_{00\alpha}(x^k) = I_{00}(x^k) \setminus (I_{001}(x^k) \cup I_{000}(x^k))$ und erhalten so für alle $i \in I_{00\alpha}(x^k)$ mit einem beliebigen $\alpha_i \in \{0, 1\}$ die Darstellung

$$\begin{aligned} 0 &= (\tilde{\Phi}'_2)_i(x, y_\gamma(x), \lambda(x), \mu(x); d) |_{x=x^k} \\ &= -\alpha_i \nabla g_i(y_\gamma(x^k))^T \cdot y'_\gamma(x^k) \cdot d + (1 - \alpha_i) e_i^T \cdot \lambda'(x^k) \cdot d. \end{aligned}$$

Es gibt also ein Element $H_k \in \partial_b \Phi(x^k, y_\gamma(x^k), \lambda(x^k), \mu(x^k))$, so dass

$$0 = H_k \cdot \begin{pmatrix} I \\ y'_\gamma(x^k) \\ \lambda'(x^k) \\ \mu'(x^k) \end{pmatrix} \cdot d$$

gilt, wobei H_k von der Gestalt

$$H_k = \begin{pmatrix} -\nabla_{yy}^2 \Psi_\gamma(x^k, y^k) & -\nabla_{yy}^2 \Psi_\gamma(x^k, y^k) + \sum_{i=1}^m \lambda_i(x^k) \nabla^2 g_i(y^k) & \nabla g_{I_1}(y^k) & \nabla g_{I_2}(y^k) & \nabla h(y^k) \\ 0 & -\nabla g_{I_1}(y^k)^T & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & \nabla h(y^k)^T & 0 & 0 & 0 \end{pmatrix}$$

mit $y^k = y_\gamma(x^k)$, $I_1 = I_{0+} \cup I_{001}$ und $I_2 = I_- \cup I_{000} \cup I_{00\alpha}$ ist. Auf Grund der Struktur von H_k können wir die obige Gleichung umschreiben in

$$\begin{pmatrix} \nabla_{yx}^2 \Psi_\gamma(x^k, y_\gamma(x^k)) \\ 0 \\ 0 \\ 0 \end{pmatrix} \cdot d = V_k \cdot \begin{pmatrix} y'_\gamma(x^k) \\ \lambda'(x^k) \\ \mu'(x^k) \end{pmatrix} \cdot d,$$

wobei $V_k := (H_k)_{\{1, \dots, n+m+p\} \times \{n+1, \dots, 2n+m+p\}} \in \pi_{(y, \lambda, \mu)} \partial_b \tilde{\Phi}(x^k, y_\gamma(x^k), \lambda(x^k), \mu(x^k))$ gilt. Da das b-Subdifferential oberhalbstetig und abgeschlossen ist und kompakte Bildmengen $\partial_b \Phi(x, y_\gamma(x), \lambda(x), \mu(x))$ hat, und $G(x) = (y_\gamma(x), \lambda(x), \mu(x))$ stetig ist, gilt (evtl. auf einer Teilfolge)

$$\lim_{k \rightarrow \infty} V_k = V \in \pi_{(y, \lambda, \mu)} \partial_b \Phi(x^*, y_\gamma(x^*), \lambda(x^*), \mu(x^*)).$$

Da das B-Subdifferential die gleichen Eigenschaften hat, folgt analog, dass (evtl. auf weiteren Teilfolgen) gilt

$$\lim_{k \rightarrow \infty} \lambda'(x^k) = M_\lambda \in \partial_B \lambda(x^*)$$

und

$$\lim_{k \rightarrow \infty} \mu'(x^k) = M_\mu \in \partial_B \mu(x^*).$$

Mit der Stetigkeit von $\nabla_{yx} \Psi_\gamma(x, y_\gamma(x))$ folgt daraus

$$\begin{pmatrix} \nabla_{yx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) \\ 0 \\ 0 \\ 0 \end{pmatrix} \cdot d = V \cdot \begin{pmatrix} M \\ M_\lambda \\ M_\mu \end{pmatrix} \cdot d.$$

Analog zu Satz 3.3.2 kann man beweisen, dass unter den gegebenen Voraussetzungen alle Elemente von $\pi_{(y, \lambda, \mu)} \partial_b \tilde{\Phi}(x^k, y_\gamma(x^k), \lambda(x^k), \mu(x^k))$ regulär sind, insbesondere also ist V regulär. Damit folgt

$$M \cdot d = \begin{pmatrix} I & 0 & 0 \end{pmatrix} V^{-1} \begin{pmatrix} \nabla_{yx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) \\ 0 \\ 0 \\ 0 \end{pmatrix} \cdot d.$$

In (4.11) tritt der Ausdruck Md allerdings in der Form $d^T \nabla_{xy}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) Md$ auf. Mit der obigen Gleichung folgt

$$d^T \nabla_{xy}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) Md = d^T \begin{pmatrix} \nabla_{yx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) \\ 0 \\ 0 \\ 0 \end{pmatrix}^T V^{-1} \begin{pmatrix} \nabla_{yx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) \\ 0 \\ 0 \\ 0 \end{pmatrix} \cdot d,$$

wobei man beachten muss, dass die Matrix V der Konstruktion nach von der betrachteten Richtung d abhängt. Mit dieser Vorarbeit können wir nun das folgende Lemma beweisen:

Lemma 4.3.3. *Sei x^* ein normalisiertes Nash-Gleichgewicht eines GNEPs, das den Voraussetzungen 1.1.1 oder 1.1.2 und LICQ genügt. Dann gilt für alle $M \in \partial y_\gamma(x^*)$ und alle $d \in \mathbb{R}^n$*

$$d^T \nabla_{xy}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) M d \geq 0.$$

Beweis. Wie oben schon erwähnt, gilt für alle $d \in \mathbb{R}^n$

$$\begin{aligned} \partial y_\gamma(x^*) \cdot d &= \text{conv}\{M \in \mathbb{R}^{n \times n} \mid \exists \{x^k\}_k \subseteq D_{y_\gamma} \cap D_\lambda \cap D_\mu : x^k \rightarrow x^*, y'_\gamma(x^k) \rightarrow M\} \cdot d \\ &=: \text{conv}\{S\} \cdot d. \end{aligned}$$

Weiter haben wir schon gezeigt, dass es zu jeder Matrix $M \in S$ und jeder beliebigen, aber festen, Richtung $d \in \mathbb{R}^n$ eine reguläre Matrix $V \in \pi_{(y, \lambda, \mu)} \partial_b \tilde{\Phi}(x^*, y_\gamma(x^*), \lambda(x^*), \mu(x^*))$ gibt, so dass

$$d^T \nabla_{xy}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) M d = d^T \begin{pmatrix} \nabla_{yx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) \\ 0 \\ 0 \\ 0 \end{pmatrix}^T V^{-1} \begin{pmatrix} \nabla_{yx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) \\ 0 \\ 0 \\ 0 \end{pmatrix} \cdot d$$

gilt. Wir zeigen nun, dass die Matrix auf der rechten Seite dieser Gleichung positiv semidefinit ist. Damit folgt die Behauptung für alle Matrizen $M \in S$ und alle Richtungen $d \in \mathbb{R}^n$. Wegen $\partial y_\gamma(x^*) \cdot d = \text{conv}\{S\} \cdot d$ für alle $d \in \mathbb{R}^n$, folgt die Behauptung dann auch für alle Matrizen $M \in \partial y_\gamma(x^*)$.

Definieren wir nun

$$\begin{pmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{pmatrix} := V^{-1} \cdot \begin{pmatrix} \nabla_{yx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

so gilt

$$\begin{pmatrix} \nabla_{yx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) \\ 0 \\ 0 \\ 0 \end{pmatrix} = V \cdot \begin{pmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{pmatrix}.$$

Dies führt auf das folgende lineare Gleichungssystem

$$\begin{aligned} \nabla_{yx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) &= -\nabla_{yy}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) N_1 + \sum_{i=1}^m \lambda_i(x^*) \nabla^2 g_i(y_\gamma(x^*)) N_1 \\ &\quad + \nabla g_{I_1}(y_\gamma(x^*)) N_2 + \nabla g_{I_2}(y_\gamma(x^*)) N_3 + \nabla h(y_\gamma(x^*)) N_4, \\ 0 &= -\nabla g_{I_1}(y_\gamma(x^*))^T N_1, \\ 0 &= N_3, \\ 0 &= \nabla h(y_\gamma(x^*))^T N_1. \end{aligned}$$

Mit diesen vier Gleichungen erhalten wir

$$\begin{aligned}
\begin{pmatrix} \nabla_{yx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) \\ 0 \\ 0 \\ 0 \end{pmatrix}^T V^{-1} \begin{pmatrix} \nabla_{yx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) \\ 0 \\ 0 \\ 0 \end{pmatrix} &= \nabla_{yx}^2 \Psi_\gamma(x^*, y_\gamma(x^*))^T N_1 \\
&= -N_1^T \nabla_{yy}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) N_1 \\
&\quad + \sum_{i=1}^m \lambda_i(x^*) N_1^T \nabla^2 g_i(y_\gamma(x^*)) N_1.
\end{aligned}$$

Wegen $\tilde{\Phi}(x^*, y_\gamma(x^*), \lambda(x^*), \mu(x^*)) = 0$ gilt $\lambda_i(x^*) \geq 0$ für alle $i = 1, \dots, m$. Nach Voraussetzung sind die Funktionen g_i konvex, daher sind ihre Hessematrizen positiv semidefinit. Die Funktion $\Psi_\gamma(x, y)$ ist nach Konstruktion gleichmäßig konkav in y , daher ist die Matrix $\nabla_{yy}^2 \Psi_\gamma(x^*, y_\gamma(x^*))$ negativ definit. Insgesamt folgt, dass die Matrix $-N_1^T \nabla_{yy}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) N_1 + \sum_{i=1}^m \lambda_i(x^*) N_1^T \nabla^2 g_i(y_\gamma(x^*)) N_1$ positiv semidefinit ist und positiv definit genau dann, wenn die Matrix N_1 regulär ist. \square

Dieses Lemma motiviert die folgende hinreichende Bedingung für die strong second-order sufficiency condition, die offensichtlich schwächer als die aus Satz 4.3.2 ist.

Satz 4.3.4. *Sei x^* ein normalisiertes Nash-Gleichgewicht eines GNEPs, das den Voraussetzungen 1.1.1 und LICQ genügt. Ist die Funktion $\Psi_\gamma(\cdot, y_\gamma(x^*))$ gleichmäßig konvex in X , so ist die strong second-order sufficiency condition in x^* erfüllt.*

Beweis. Mit dem obigen Lemma vereinfacht sich die hinreichende Bedingung aus (4.11) zu

$$d^T \nabla_{xx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) d > 0$$

für alle $d \in G(z^*) \setminus \{0\}$. Wenn aber die Funktion $\Psi_\gamma(\cdot, y_\gamma(x^*))$ gleichmäßig konvex in X ist, so ist die Hesse-Matrix $\nabla_{xx}^2 \Psi_\gamma(x^*, y_\gamma(x^*))$ positiv definit. Dann ist die obige Bedingung für alle $d \in \mathbb{R}^n \setminus \{0\}$ erfüllt, insbesondere also für alle $d \in G(z^*) \setminus \{0\}$. \square

Zum Abschluss dieses Abschnitts wollen wir das Optimierungsproblem

$$\min_x \Psi_\gamma(x, y_\gamma^*) \quad \text{u.d.N.} \quad g(x) \leq 0, h(x) = 0 \quad (4.12)$$

betrachten, wobei x^* ein normalisiertes Nash-Gleichgewicht eines GNEPs sei, das den Voraussetzungen 1.1.1 oder 1.1.2 und LICQ genügt, und $y_\gamma^* = y_\gamma(x^*)$. Offensichtlich ist dieses Problem eng verwandt mit dem bisher betrachteten Problem

$$\min_x V_\gamma(x) = \Psi_\gamma(x, y_\gamma(x)) \quad \text{u.d.N.} \quad g(x) \leq 0, h(x) = 0. \quad (4.13)$$

Dies motiviert den folgenden Satz, der die Voraussetzungen aus Satz 4.3.4 noch etwas weiter abschwächt.

Satz 4.3.5. *Sei x^* ein normalisiertes Nash-Gleichgewicht eines GNEPs, das den Voraussetzungen 1.1.1 und LICQ genügt. Erfüllt (x^*, λ^*, μ^*) die strong second-order sufficiency*

condition bezüglich des Problems (4.12), d.h ist (x^*, λ^*, μ^*) ein KKT-Punkt des Problems (4.12) und gilt mit

$$\begin{aligned} I_{0+} &= \{i : g_i(x^*) = 0, \lambda_i^* > 0\}, \\ G(x^*, \lambda^*, \mu^*) &:= \{d \in \mathbb{R}^n \mid \nabla_x \Psi_\gamma(x^*, y_\gamma^*)^T d = 0, \nabla g_i(x^*)^T d = 0 \forall i \in I_{0+}, \nabla h(x^*)^T d = 0\}, \\ \Phi_1(x, \lambda, \mu) &= \nabla_x \Psi_\gamma(x, y_\gamma) + \nabla g(x)\lambda + \nabla h(x)\mu \end{aligned}$$

für alle $d \in G(x^*, \lambda^*, \mu^*) \setminus \{0\}$ und für alle $H \in \pi_x \partial \Phi_1(x^*, \lambda^*, \mu^*)$ die Ungleichung $d^T H d > 0$, so erfüllt (x^*, λ^*, μ^*) auch die strong second-order sufficiency condition bezüglich des Problems (4.13), d.h (x^*, λ^*, μ^*) ist ein KKT-Punkt des Problems (4.13) und mit

$$\begin{aligned} \bar{I}_{0+} &= \{i : g_i(x^*) = 0, \lambda_i^* > 0\}, \\ \bar{G}(x^*, \lambda^*, \mu^*) &:= \{d \in \mathbb{R}^n \mid \nabla V_\gamma(x^*, y_\gamma^*)^T d = 0, \nabla g_i(x^*)^T d = 0 \forall i \in \bar{I}_{0+}, \nabla h(x^*)^T d = 0\}, \\ \bar{\Phi}_1(x, \lambda, \mu) &= \nabla V_\gamma(x) + \nabla g(x)\lambda + \nabla h(x)\mu \end{aligned}$$

gilt für alle $d \in \bar{G}(x^*, \lambda^*, \mu^*) \setminus \{0\}$ und für alle $H \in \pi_x \partial \bar{\Phi}_1(x^*, \lambda^*, \mu^*)$ die Ungleichung $d^T H d > 0$.

Beweis. Erfüllt (x^*, λ^*, μ^*) die strong second-order sufficiency condition bezüglich des Problems (4.12), so ist (x^*, λ^*, μ^*) ein KKT-Punkt von (4.12), d.h. es gilt

$$\Phi(x^*, \lambda^*, \mu^*) = \begin{pmatrix} \nabla_x \Psi_\gamma(x^*, y_\gamma^*) + \nabla g(x^*)\lambda^* + \nabla h(x^*)\mu^* \\ \min\{\lambda_i^*, -g_i(x^*)\}_{i=1, \dots, m} \\ h(x^*) \end{pmatrix} = 0.$$

Wegen $\nabla V_\gamma(x) = \nabla_x \Psi_\gamma(x, y_\gamma(x))$ ist (x^*, λ^*, μ^*) also genau dann ein KKT-Punkt von (4.12), wenn es ein KKT-Punkt von (4.13) ist.

Betrachten wir nun den zweiten Teil der Definition der strong second-order sufficiency condition für Problem (4.12). Hier ist wegen der affinen Linearität der Funktion h

$$\pi_x \partial \Phi_1(x^*, \lambda^*, \mu^*) = \{\nabla_{xx}^2 \Psi_\gamma(x^*, y_\gamma^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x^*)\}.$$

Da (x^*, λ^*, μ^*) der strong second-order sufficiency condition bezüglich Problem (4.12) genügt, gilt für alle $d \in G(x^*, \lambda^*, \mu^*) \setminus \{0\}$ die Ungleichung

$$d^T \left(\nabla_{xx}^2 \Psi_\gamma(x^*, y_\gamma^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x^*) \right) d > 0. \quad (4.14)$$

Damit (x^*, λ^*, μ^*) der strong second-order sufficiency condition bezüglich Problem (4.13) genügt, muss für alle $d \in \bar{G}(x^*, \lambda^*, \mu^*) \setminus \{0\}$ und alle $H \in \pi_x \partial \bar{\Phi}_1(x^*, \lambda^*, \mu^*)$ gelten $d^T H d > 0$.

Offensichtlich ist $\bar{I}_{0+} = I_{0+}$ und wegen $\nabla V_\gamma(x^*) = \nabla_x \Psi_\gamma(x^*, y_\gamma^*)$ daher auch $\bar{G}(x^*, \lambda^*, \mu^*) = G(x^*, \lambda^*, \mu^*)$. Wie wir schon gezeigt haben, gibt es für alle $d \in \bar{G}(x^*, \lambda^*, \mu^*) \setminus \{0\}$ und alle $H \in \pi_x \partial \bar{\Phi}_1((x^*, \lambda^*, \mu^*))$ eine Matrix $M \in \partial y_\gamma(x^*)$, so dass gilt

$$d^T H d = d^T \left(\nabla_{xx}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) + \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(x^*) + \nabla_{xy}^2 \Psi_\gamma(x^*, y_\gamma(x^*)) \cdot M \right) d.$$

Mit (4.14) und Lemma 4.3.3 folgt $d^T H d > 0$. Also genügt (x^*, λ^*, μ^*) der strong second-order sufficiency condition bezüglich Problem (4.13). \square

Kapitel 5

Numerische Umsetzung des SQP-Verfahrens

In diesem Kapitel wollen wir das semismoothe SQP-Verfahren, dessen Konvergenzeigenschaften wir in Kapitel 4 untersucht haben, an ein paar numerischen Beispielen testen. In Abschnitt 5.1 werden zwei mögliche Algorithmen vorgestellt und Implementationsdetails geklärt. Der erste Algorithmus verwendet die in Abschnitt 4.3 hergeleitete Darstellung von Elementen aus $\partial y_\gamma(x)$ um die für das SQP-Verfahren benötigte Matrix $G_k \in \pi_x \partial \Phi_1(x^k, \lambda^k, \mu^k)$ zu bestimmen. Im zweiten Algorithmus wird dieses Problem umgangen, indem die Matrix G_k durch eine modifizierte BFGS-Aufdatierung H_k approximiert wird. Diese zwei Algorithmen werden schließlich in Abschnitt 5.2 an einigen Beispielen getestet. Die numerischen Berechnungen wurden mit MATLAB 7.4.0 auf einem Mac mit Intel-Chipsatz durchgeführt.

5.1 Zwei Varianten des SQP-Verfahrens

Der erste Algorithmus zur Lösung des Problems

$$\min_x V_\gamma(x) \quad \text{u.d.N.} \quad g(x) \leq 0, h(x) = 0,$$

den wir in diesem Abschnitt betrachten wollen, verwendet als Matrix in der quadratischen Zielfunktion der Teilprobleme ein Element $G_k \in \pi_x \partial \Phi_1(x^k, \lambda^k, \mu^k)$.

Algorithmus 5.1.1. (Ein semismoothes SQP-Verfahren)

(S.0) Wähle $\gamma, \varepsilon, c > 0$, $\alpha_0 \in \mathbb{R}_{++}^{m+p}$, $\beta, \sigma \in (0, 1)$ und Startvektoren $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}_+^m$ und $\mu^0 \in \mathbb{R}^p$, setze $k = 0$.

(S.1) Ist $|V_\gamma(x^k)| \leq \varepsilon$: STOPP

(S.2) Bestimme ein Element $G_k \in \pi_x \partial \Phi_1(x^k, \lambda^k, \mu^k)$.

(S.3) Finde einen KKT-Punkt $(x^{neu}, \lambda^{k+1}, \mu^{k+1})$ von

$$\begin{aligned} \min \quad & \nabla V_\gamma(x^k)^T (x - x^k) + \frac{1}{2} (x - x^k)^T G_k (x - x^k) \\ \text{u.d.N.} \quad & g(x^k) + \nabla g(x^k)^T (x - x^k) \leq 0 \\ & h(x^k) + \nabla h(x^k)^T (x - x^k) = 0. \end{aligned}$$

Gibt es mehrere KKT-Punkte, so wähle denjenigen, der $\|x^{neu} - x^k\|$ minimiert.

(S.4) Bestimme die Schrittweite $t_k := \max\{\beta^l : l = 0, 1, 2, \dots\}$ mit

$$P_1(x^k + t_k(x^{neu} - x^k); \alpha^k) \leq P_1(x^k) + t_k \sigma P_1'(x^k; x^{neu} - x^k; \alpha^k),$$

wobei P_1 die exakte l_1 -Penalty-Funktion bezeichne.

(S.5) Setze $x^{k+1} = x^k + t_k(x^{neu} - x^k)$, datiere α_k auf. Setze $k \leftarrow k + 1$ und gehe zu (S.1).

Die zur Implementation benötigten Ableitungen der Funktion

$$\Psi_\gamma(x, y) = \sum_{\nu=1}^N [\theta_\nu(x^\nu, x^{-\nu}) - \theta_\nu(y^\nu, x^{-\nu})] - \frac{\gamma}{2} \|x - y\|_2^2$$

sind in dem folgenden Lemma zusammengefasst. Auf den Beweis wird hier verzichtet, da man die Ableitungen durch Anwendung der üblichen Ableitungsregeln erhält, bzw. in [7], Lemma 4.1, findet.

Lemma 5.1.2. *Sind die Funktionen θ_ν zweimal stetig differenzierbar, so ist auch die Funktion $\Psi_\gamma(x, y)$ zweimal stetig differenzierbar. Sie hat dann die folgenden (partiellen) Ableitungen*

$$\nabla_x \Psi_\gamma(x, y) = \sum_{\nu=1}^N [\nabla \theta_\nu(x^\nu, x^{-\nu}) - \nabla \theta_\nu(y^\nu, x^{-\nu})] + \begin{pmatrix} \nabla_{x^1} \theta_1(y^1, x^{-1}) \\ \vdots \\ \nabla_{x^N} \theta_N(y^N, x^{-N}) \end{pmatrix} - \gamma(x - y),$$

$$\nabla_y \Psi_\gamma(x, y) = - \begin{pmatrix} \nabla_{x^1} \theta_1(y^1, x^{-1}) \\ \vdots \\ \nabla_{x^N} \theta_N(y^N, x^{-N}) \end{pmatrix} + \gamma(x - y),$$

$$\begin{aligned} \nabla_{xx}^2 \Psi_\gamma(x, y) &= \sum_{\nu=1}^N [\nabla^2 \theta_\nu(x^\nu, x^{-\nu}) - \nabla^2 \theta_\nu(y^\nu, x^{-\nu})] \\ &+ \begin{pmatrix} \nabla_{x^1 x^1}^2 \theta_1(y^1, x^{-1}) & \cdots & \nabla_{x^1 x^N}^2 \theta_N(y^N, x^{-N}) \\ \vdots & \ddots & \vdots \\ \nabla_{x^N x^1}^2 \theta_1(y^1, x^{-1}) & \cdots & \nabla_{x^N x^N}^2 \theta_N(y^N, x^{-N}) \end{pmatrix} \\ &+ \begin{pmatrix} \nabla_{x^1 x^1}^2 \theta_1(y^1, x^{-1}) & \cdots & \nabla_{x^1 x^N}^2 \theta_1(y^1, x^{-1}) \\ \vdots & \ddots & \vdots \\ \nabla_{x^N x^1}^2 \theta_N(y^N, x^{-N}) & \cdots & \nabla_{x^N x^N}^2 \theta_N(y^N, x^{-N}) \end{pmatrix} \\ &- \begin{pmatrix} \nabla_{x^1 x^1}^2 \theta_1(y^1, x^{-1}) & & 0 \\ & \ddots & \\ 0 & & \nabla_{x^N x^N}^2 \theta_N(y^N, x^{-N}) \end{pmatrix} - \gamma I, \end{aligned}$$

$$\begin{aligned} \nabla_{xy}^2 \Psi_\gamma(x, y) &= - \begin{pmatrix} \nabla_{x^1 x^1}^2 \theta_1(y^1, x^{-1}) & \cdots & \nabla_{x^1 x^N}^2 \theta_N(y^N, x^{-N}) \\ \vdots & \ddots & \vdots \\ \nabla_{x^N x^1}^2 \theta_1(y^1, x^{-1}) & \cdots & \nabla_{x^N x^N}^2 \theta_N(y^N, x^{-N}) \end{pmatrix} \\ &+ \begin{pmatrix} \nabla_{x^1 x^1}^2 \theta_1(y^1, x^{-1}) & & 0 \\ & \ddots & \\ 0 & & \nabla_{x^N x^N}^2 \theta_N(y^N, x^{-N}) \end{pmatrix} + \gamma I, \end{aligned}$$

$$\nabla_{yx}^2 \Psi_\gamma(x, y) = \nabla_{xy}^2 \Psi_\gamma(x, y)^T,$$

$$\nabla_{yy}^2 \Psi_\gamma(x, y) = - \begin{pmatrix} \nabla_{x^1 x^1}^2 \theta_1(y^1, x^{-1}) & & 0 \\ & \ddots & \\ 0 & & \nabla_{x^N x^N}^2 \theta_N(y^N, x^{-N}) \end{pmatrix} - \gamma I.$$

Bemerkung 5.1.3. a) Bei jeder Auswertung der Zielfunktion $V_\gamma(x)$ muss eine Lösung $(y_\gamma(x), \lambda(x), \mu(x))$ des Minimierungsproblems

$$\min_y -\Psi_\gamma(x, y) \quad \text{u.d.N.} \quad g(y) \leq 0, h(y) = 0$$

bestimmt werden. Hierfür verwenden wir die Funktion `fmincon` aus der Optimization Toolbox 3.1.1 von MATLAB.

b) In Schritt (S.1) prüfen wir zusätzlich noch, ob (x^k, λ^k, μ^k) ein KKT-Punkt des Problems

$$\min V_\gamma(x) \quad \text{u.d.N.} \quad g(x) \leq 0, \quad h(x) = 0$$

ist, da es theoretisch möglich ist, dass das SQP-Verfahren KKT-Punkte dieses Problems findet, die keine globalen Minima sind, oder dass das obige Problem keine normalisierten Nash-Gleichgewichte und damit auch keine Nullstellen besitzt.

c) Das quadratische Minimierungsproblem in (S.3) lösen wir mit der Funktion `quadprog` aus der Optimization Toolbox 3.1.1 von MATLAB.

d) Um die Matrix $G_k \in \pi_x \partial \Phi_1(x^k, \lambda^k, \mu^k)$ in (S.2) zu bestimmen, gehen wir folgendermaßen vor:

Wir nehmen an, dass in $(x^*, y_\gamma(x^*), \lambda(x^*), \mu(x^*))$ strikte Komplementarität gilt. Wegen der Stetigkeit der Funktionen $y_\gamma(x)$ und $\lambda(x)$ gilt dann auch für alle Vektoren $(x, y_\gamma(x), \lambda(x), \mu(x))$ mit x aus einer kleinen Umgebung von x^* strikte Komplementarität, d.h die Funktion $\Phi(x, y, \lambda, \mu)$ ist in einer Umgebung des Punktes $(x^*, y_\gamma(x^*), \lambda(x^*), \mu(x^*))$ stetig differenzierbar. Mit Satz 3.3.2 und dem Standard-satz über implizite Funktionen folgt dann, dass die Funktionen $y_\gamma(x)$, $\lambda(x)$ und $\mu(x)$ in einer kleinen Umgebung von $U(x^*)$ nicht nur semismooth, sondern sogar stetig differenzierbar sind. In diesem Fall gilt für alle $x \in U(x^*)$

$$\pi_x \partial \Phi_1(x, \lambda, \mu) = \nabla_{xx}^2 \Psi_\gamma(x, y_\gamma(x)) + \sum_{i=1}^m \lambda_i \nabla^2 g_i(x) + \nabla_{xy}^2 \Psi_\gamma(x, y_\gamma(x)) \cdot y'_\gamma(x).$$

Wie im letzten Abschnitt erhält man für alle $x \in U(x^*)$

$$0 = \begin{pmatrix} D_x \tilde{\Phi}(x, y_\gamma(x), \lambda(x), \mu(x)) & D_{(y, \lambda, \mu)} \tilde{\Phi}(x, y_\gamma(x), \lambda(x), \mu(x)) \end{pmatrix} \cdot \begin{pmatrix} I \\ y'_\gamma(x) \\ \lambda'(x) \\ \mu'(x) \end{pmatrix}.$$

Wählen wir $U(x^*)$ klein genug, so sind alle Matrizen in $\pi_{(y, \lambda, \mu)} \partial \tilde{\Phi}(x, y_\gamma(x), \lambda(x), \mu(x))$ regulär und wir können diese Gleichung wieder umschreiben in

$$y'_\gamma(x) = [-D_{(y, \lambda, \mu)} \tilde{\Phi}(x, y_\gamma(x), \lambda(x), \mu(x))^{-1} D_x \tilde{\Phi}(x, y_\gamma(x), \lambda(x), \mu(x))]_{\{1, \dots, n\} \times \{1, \dots, n\}}.$$

Wir bestimmen daher V_k als

$$V_k = \begin{pmatrix} -\nabla_{yy} \Psi_\gamma(x^k, y_\gamma^k) + \sum_{i=1}^m \lambda_i(x^k) \nabla^2 g_i(y_\gamma^k) & \nabla g_{i \in I_1}(y_\gamma^k) & \nabla g_{i \in I_2}(y_\gamma^k) & \nabla h(y_\gamma^k) \\ -\nabla g_{i \in I_1}(y_\gamma^k)^T & 0 & 0 & 0 \\ 0 & 0 & I_{|I_2| \times |I_2|} & 0 \\ \nabla h(y_\gamma^k)^T & 0 & 0 & 0 \end{pmatrix}$$

mit den folgenden Indextmengen

$$\begin{aligned} I_- &= \{i : g_i(y_\gamma(x^k)) < 0\}, & I_0 &= \{i : g_i(y_\gamma(x^k)) = 0\}, \\ I_{00} &= \{i \in I_0 : \lambda_i(x^k) = 0\}, & I_{0+} &= \{i \in I_0 : \lambda_i(x^k) = 0\}, \\ I_1 &= I_{0+}, & I_2 &= I_- \cup I_{00}. \end{aligned}$$

Zur Abkürzung haben wir hier statt $y_\gamma(x^k)$ nur y_γ^k geschrieben. Genaugenommen berechnen wir hier ein Element $V_k \in \pi_{(y, \lambda, \mu)} \partial_b \tilde{\Phi}(x^k, y_\gamma(x^k), \lambda(x^k), \mu(x^k))$, das aber für alle $x^k \in U(x^*)$ ebenfalls regulär ist. Wir tun dies für den Fall, dass doch keine strikte Komplementariät gilt. Ist jedoch tatsächlich $I_{00} = \emptyset$, dann gilt $V_k = D_{(y, \lambda, \mu)} \tilde{\Phi}(x^k, y_\gamma(x^k), \lambda(x^k), \mu(x^k))$. Dann setzen wir

$$M_k = (V_k^{-1})_{\{1, \dots, n\} \times \{1, \dots, n\}} \nabla_{yx}^2 \Psi_\gamma(x^k, y_\gamma(x^k)),$$

wobei wir

$$D_x \tilde{\Phi}(x, y_\gamma(x), \lambda(x), \mu(x)) = \begin{pmatrix} \nabla_{yx}^2 \Psi_\gamma(x^k, y_\gamma(x^k)) \\ 0 \\ 0 \end{pmatrix}$$

verwendet haben und schließlich

$$G_k = \nabla_{xx}^2 \Psi_\gamma(x^k, y_\gamma(x^k)) + \sum_{i=1}^m \lambda_i^k \nabla^2 g_i(x^k) + \nabla_{xy}^2 \Psi_\gamma(x^k, y_\gamma(x^k)) \cdot M_k.$$

- e) (S.4) globalisiert das lokale SQP-Verfahren mit der exakten l_1 -Penalty-Funktion und einer Armijo-Schrittweiten-Regel. Die hierbei auftretende exakte l_1 -Penalty-Funktion mit Penalty-Parametern $\alpha_i > 0$ ist folgendermaßen definiert

$$P_1(x; \alpha) = V_\gamma(x) + \sum_{i=1}^m \alpha_i \max\{0, g_i(x)\} + \sum_{i=1}^p \alpha_{m+i} |h_i(x)|$$

mit der Richtungsableitung

$$\begin{aligned}
P'_1(x; d; \alpha) &= \nabla V_\gamma(x)^T d + \sum_{i: h_i(x) > 0} \alpha_{m+i} \nabla h_i(x)^T d - \sum_{i: h_i(x) < 0} \alpha_{m+i} \nabla h_i(x)^T d \\
&+ \sum_{i: h_i(x) = 0} \alpha_{m+i} |\nabla h_i(x)^T d| + \sum_{i: g_i(x) > 0} \alpha_i \nabla g_i(x)^T d \\
&+ \sum_{i: g_i(x) = 0} \alpha_i \max\{0, \nabla g_i(x)^T d\}.
\end{aligned}$$

Allerdings sieht man in Beispielen im Abschnitt 5.2, dass dort im Allgemeinen die volle Schrittweite angenommen wird.

f) In Algorithmus 5.1.1 verwenden wir für jede Nebenbedingung einen eigenen Penalty-Parameter α_i . Die Parameter für die Ungleichungen werden nach der folgenden Vorschrift aufdatiert

$$\alpha_i^{k+1} = \begin{cases} \alpha_i^k & \text{falls } \lambda_i^{k+1} + c \leq \alpha_i^k, \\ \lambda_i^{k+1} + 2c & \text{falls } \lambda_i^{k+1} + c > \alpha_i^k, \end{cases}$$

und die Parameter für die Gleichungsnebenbedingungen nach der Vorschrift

$$\alpha_{m+i}^{k+1} = \begin{cases} \alpha_{m+i}^k & \text{falls } |\mu_i^{k+1}| + c \leq \alpha_{m+i}^k, \\ |\mu_i^{k+1}| + 2c & \text{falls } |\mu_i^{k+1}| + c > \alpha_{m+i}^k. \end{cases}$$

Um die komplizierte und nur in einer Umgebung von x^* wohldefinierte Berechnung der Elements $G_k \in \pi_x \partial \Phi(x^k, \lambda^k, \mu^k)$ zu vermeiden, kann man diese Matrix mittels einer modifizierten BFGS-Aufdatierung H_k approximieren. Die folgende Formel von Powell hat den Vorteil, dass alle Matrizen H_k symmetrisch positiv definit sind, falls H_0 symmetrisch positiv definit ist:

$$\begin{aligned}
s^k &= x^{k+1} - x^k, \\
d^k &= \nabla_x L(x^{k+1}, \lambda^k, \mu^k) - \nabla_x L(x^k, \lambda^k, \mu^k) \\
q_k &= \begin{cases} 1 & \text{falls } (s^k)^T d^k \geq 0.2 (s^k)^T H_k s^k, \\ \frac{0.8 (s^k)^T H_k s^k}{(s^k)^T H_k s^k - (s^k)^T d^k} & \text{falls } (s^k)^T d^k < 0.2 (s^k)^T H_k s^k, \end{cases} \\
\eta_k &= q_k d^k + (1 - q_k) H_k s^k, \\
H_{k+1} &= H_k + \frac{\eta^k (\eta^k)^T}{(s^k)^T s^k} - \frac{H_k s^k (s^k)^T H_k}{(s^k)^T H_k s^k}.
\end{aligned}$$

Hierbei ist

$$L(x, \lambda, \mu) = V_\gamma(x) + \lambda^T g(x) + \mu^T h(x)$$

die Lagrange-Funktion des betrachteten Optimierungsproblems und

$$\nabla_x L(x, \lambda, \mu) = \nabla_x \Psi_\gamma(x, y) \Big|_{y=y_\gamma(x)} + \nabla g(x) \lambda + \nabla h(x) \mu$$

der partielle Gradient von L bezüglich der Variablen x . Mit dieser Formel für H_{k+1} und der Aufdatierung für α_k aus Bemerkung 5.1.3 f) ergibt sich dann der folgende Algorithmus:

Algorithmus 5.1.4. (SQP-Verfahren mit BFGS-Aufdatierung)

(S.0) Wähle $\gamma, \varepsilon, c > 0$, $\alpha_0 \in \mathbb{R}_{++}^{m+p}$, $\beta, \sigma \in (0, 1)$ und Startvektoren $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}_+^m$ und $\mu^0 \in \mathbb{R}^p$. Setze $H_0 = I_{n \times n}$ und $k = 0$.

(S.1) Ist $|V_\gamma(x^k)| \leq \varepsilon$: STOPP

(S.2) Finde einen KKT-Punkt $(x^{neu}, \lambda^{k+1}, \mu^{k+1})$ von

$$\begin{aligned} \min \quad & \nabla V_\gamma(x^k)^T(x - x^k) + \frac{1}{2}(x - x^k)^T H_k(x - x^k) \\ \text{u.d.N.} \quad & g(x^k) + \nabla g(x^k)^T(x - x^k) \leq 0 \\ & h(x^k) + \nabla h(x^k)^T(x - x^k) = 0. \end{aligned}$$

Gibt es mehrere KKT-Punkte, so wähle denjenigen, der $\|x^{neu} - x^k\|$ minimiert.

(S.3) Bestimme die Schrittweite $t_k := \max\{\beta^l : l = 0, 1, 2, \dots\}$ mit

$$P_1(x^k + t_k(x^{neu} - x^k); \alpha) \leq P_1(x^k) + t_k \sigma P_1'(x^k; x^{neu} - x^k; \alpha),$$

wobei P_1 die exakte l_1 -Penalty-Funktion bezeichne.

(S.4) Setze $x^{k+1} = x^k + t_k(x^{neu} - x^k)$, datiere H_k und α_k auf. Setze $k \leftarrow k + 1$ und gehe zu (S.1).

5.2 Numerische Illustration

Die folgenden Beispiele sollen das Konvergenzverhalten der obigen Algorithmen illustrieren.

Beispiel 5.2.1. Als erstes Testbeispiel betrachten wir das DSL-Problem aus Abschnitt 1.2. Hier hatte jeder Nutzer $\nu = 1, \dots, N$ das Optimierungsproblem

$$\begin{aligned} \min \quad & - \sum_{f=1}^F \log_2 \left(1 + \frac{x_f^\nu}{\sigma_f^\nu + \sum_{\mu \neq \nu} \alpha_f^{\mu\nu} x_f^\mu} \right) \\ \text{u.d.N.} \quad & 0 \leq x_f^\nu \leq \text{CAP}_f^\nu \quad \forall f = 1, \dots, F, \quad \sum_{f=1}^F x_f^\nu \leq P_{\max}^\nu \end{aligned}$$

zu lösen. Hierbei war F die Anzahl der möglichen Frequenzen, $\sigma_f^\nu > 0$ ein Koeffizient für die Störgeräusche, die den Nutzer ν auf der Frequenz f beeinträchtigen und $\alpha_f^{\mu\nu} \geq 0$ ein Maß für den Crosstalk von Nutzer μ auf Nutzer ν auf der Frequenz f . $P_{\max}^\nu > 0$ und $\text{CAP}_f^\nu > 0$ sind Energierestriktionen, denen der Nutzer ν unterliegt. Wir wollen die folgenden zwei Fällen testen:

Zunächst betrachten wir den Fall $N = 2$, $F = 32$. Die Werte σ_f^ν wählen wir zufällig nach der Gleichverteilung aus dem Intervall $(0, \frac{0.1}{N-1}]$. Weiter setzen wir $\text{CAP}_f^\nu = \infty$ und wählen P_{\max}^ν ebenfalls gleichverteilt aus dem Intervall $[\frac{F}{2}, F]$. Schließlich wählen wir die Koeffizienten $\alpha_f^{\mu\nu}$ gleichverteilt aus $[0, 1]$, was starken Crosstalk-Interferenzen entspricht.

Mit dem Startwert $x^0 = 0$ und den Parametern $\varepsilon = 10^{-6}$, $\gamma = 0.05$ erhalten wir die Iterationsverläufe in den Tabellen 5.1 und 5.2, wobei wir aus Platzgründen nur die ersten vier Komponenten von x^k ausgeben. Hierbei ist t_k die Schrittweite, die in der Iteration k gewählt wurde. Wir wählen als Abbruchparameter $\varepsilon = 10^{-6}$, da auch die in beiden Algorithmen verwendete MATLAB-Funktion `fmincon` mit dieser Genauigkeit arbeitet.

k	x_1^k	x_2^k	x_3^k	x_4^k	$V_\gamma(x^k)$	t_k
0	0.50000000	0.50000000	0.50000000	0.50000000	28.08693137	-
1	-0.00000000	0.00000000	0.24511668	0.69963700	0.46110309	1
2	0.00000013	0.00000037	0.24511696	0.69963706	0.46110230	$0.5^2 0$
3	0.10740760	0.36400188	0.51452824	0.75329606	0.01403636	1
4	0.15226531	0.37960899	0.52349860	0.76435210	0.00062980	1
5	0.15886547	0.38529983	0.52757858	0.76573366	0.00013709	1
6	0.15887771	0.38526195	0.52750973	0.76536332	0.00000389	1
7	0.15967503	0.38591413	0.52807164	0.76560523	0.00000067	1

Tabelle 5.1: Iterationsverlauf im DSL-Beispiel mit starkem Crosstalk und Algorithmus 5.1.1

Im zweiten Fall wählen $\alpha_f^{\mu\nu}$ gleichverteilt aus dem Intervall $[0, \frac{1}{N-1}]$ um schwachen Crosstalk zu simulieren. Die restlichen Werte bestimmen wir wie in dem obigen Fall. Die Tabellen 5.3 und 5.4 geben die zugehörigen Iterationsverläufe mit Startwert $x^0 = 0$ und den Parametern $\varepsilon = 10^{-6}$, $\gamma = 0.05$ wieder.

Beispiel 5.2.2. Als zweites Beispiel betrachten wir das Duopol-Problem von Krawczyk und Uryasev, zu finden in [14], Beispiel 5.2. In diesem Beispiel betrachten wir zwei Firmen $\nu = 1, 2$, die auf dem selben Markt ein identisches Produkt verkaufen. Beide Firmen versuchen ihre Absatzmenge x^ν so zu wählen, dass sie ihren Gewinn maximieren. Hierbei müssen sie berücksichtigen, dass der Marktpreis für ihr Produkt mit steigender Angebotsmenge sinkt:

$$p(x) = 20 - (x_1 + x_2)$$

Wir schreiben hier x_1, x_2 statt x^1, x^2 , um zu vermeiden, dass die Indizes der Spieler mit Potenzen verwechselt werden. Setzt man Herstellungskosten von 4 Geldeinheiten pro Stück an, so ergeben sich daraus die folgenden (negativen) Nutzenfunktionen

$$\theta_\nu(x_1, x_2) = -(p(x)x^\nu - 4x^\nu) = (x_1 + x_2)x^\nu - 16x^\nu$$

für $\nu = 1, 2$. Wir betrachten in diesem Modell nur Nichtnegativitätsrestriktionen, d.h.

$$X = \{(x_1, x_2) \in \mathbb{R}^2 \mid x \geq 0\}.$$

Mit dem Startwert $x^0 = (2, 0)$ und den Parametern $\varepsilon = 10^{-6}$, $\gamma = 0.05$ erhalten wir die in Tabelle A.1 und A.2 im Anhang angegebenen Ergebnisse.

Beispiel 5.2.3. Auch das folgende Beispiel findet man in [14], Beispiel 5.3. Hierbei geht es um drei Firmen $\nu = 1, 2, 3$, die entlang eines Flusses liegen. Jede Firma produziert eine Menge x^ν des Gutes, wobei Schadstoffe in den angrenzenden Fluss gelangen. Deshalb liegen an diesem Fluss auch zwei staatliche Kontrollstationen $l = 1, 2$, in denen der Grad

k	x_1^k	x_2^k	x_3^k	x_4^k	$V_\gamma(x^k)$	t_k
0	0.50000000	0.50000000	0.50000000	0.50000000	28.08693137	-
1	0.41753530	0.63938079	0.72283806	0.76884532	3.13899519	1
2	0.31315147	0.55412376	0.65527514	0.74566472	1.31910907	0.25
3	0.23576882	0.46333694	0.58572440	0.78099845	0.38090685	0.25
4	0.21042233	0.40469671	0.55461460	0.84912862	0.18846315	0.5
5	0.22063675	0.40553987	0.56292715	0.85161056	0.09583239	1
6	0.20627473	0.43685194	0.57413047	0.82402415	0.04543227	1
7	0.18634443	0.42126072	0.55983975	0.83467859	0.02734787	1
8	0.19141316	0.40832761	0.55227302	0.82858048	0.01679765	1
9	0.18760878	0.41424344	0.55361000	0.81364398	0.01195048	1
10	0.17841179	0.41131588	0.54960494	0.80848958	0.00668269	1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
30	0.16158834	0.38761117	0.52893898	0.76473163	0.00005163	1
31	0.16095539	0.38699491	0.52872004	0.76516305	0.00003199	1
32	0.16093930	0.38715071	0.52906012	0.76633609	0.00002332	1
33	0.16102002	0.38714759	0.52923155	0.76674531	0.00001737	1
34	0.16093781	0.38697845	0.52935703	0.76712655	0.00001231	1
35	0.16080081	0.38683873	0.52918245	0.76742077	0.00000883	1
36	0.16047054	0.38660431	0.52892927	0.76748203	0.00000551	1
37	0.16031297	0.38651951	0.52857924	0.76735849	0.00000286	1
38	0.16020931	0.38646730	0.52849292	0.76712891	0.00000207	1
39	0.16027435	0.38646624	0.52851594	0.76683427	0.00000154	1
40	0.16026752	0.38641312	0.52851835	0.76642518	0.00000100	1

Tabelle 5.2: Iterationsverlauf im DSL-Beispiel mit starkem Crosstalk und Algorithmus 5.1.4

der Wasserverschmutzung gemessen wird. Übersteigt er einen gewissen Grenzwert, so müssen die Firmen eine drakonische Strafe bezahlen. Der zulässige Bereich ergibt sich daher als

$$X = \{x \in \mathbb{R}^3 \mid x \geq 0, q_1(x) \leq 100, q_2(x) \leq 100\},$$

wobei die Emmissionsbeschränkung der Stationen $l = 1, 2$ durch

$$q_l(x) = \sum_{\nu=1}^3 u_{\nu l} e_\nu x^\nu \leq 100$$

beschrieben werden. Hierbei repräsentiert $u_{\nu l}$ die Lage der Firma ν im Vergleich zum Kontrollpunkt l und e_ν die Umweltschutzvorrichtungen der Firma ν . Die entsprechenden Konstanten kann man der Tabelle 5.5 entnehmen. Der Marktpreis für ein Stück des Gutes ergibt sich in Abhängigkeit vom Angebot als

$$p(x) = 3 - 0.01(x_1 + x_2 + x_3).$$

Die Herstellungskosten der Firma ν erhält man aus der folgenden Funktion

$$K_\nu(x^\nu) = (c_{1\nu} + c_{2\nu}x^\nu)x^\nu.$$

k	x_1^k	x_2^k	x_3^k	x_4^k	$V_\gamma(x^k)$	t_k
0	0.50000000	0.50000000	0.50000000	0.50000000	25.21735124	-
1	0.14248777	1.13868488	1.23899289	1.10999604	0.33052202	1
2	0.14248786	1.13868486	1.23899288	1.10999599	0.33052138	$0.5^2 0$
3	0.15879459	1.13106311	1.23575621	1.09505035	0.22852724	0.25
4	0.18000042	1.09173610	1.23239566	1.04052550	0.00275594	1
5	0.18994142	1.09440956	1.22962651	1.04483879	0.00010014	1
6	0.18994142	1.09440956	1.22962651	1.04483879	0.00010014	$0.5^2 0$
7	0.19245875	1.09529958	1.22944880	1.04536382	0.00003302	0.5^5
8	0.19207469	1.09380479	1.22829961	1.04465947	0.00000037	1

Tabelle 5.3: Iterationsverlauf im DSL-Beispiel mit schwachem Crosstalk und Algorithmus 5.1.1

Aus diesen beiden Funktionen folgen dann die (negativen) Nutzenfunktionen der Firmen $\nu = 1, 2, 3$

$$\theta_\nu(x_1, x_2, x_3) = -(p(x)x^\nu - K_\nu(x_\nu)) = (0.01(x_1 + x_2 + x_3) + c_{2\nu}x^\nu)x^\nu + (c_{1\nu} - 3)x^\nu.$$

Mit dem Startwert $x^0 = (0, 0, 0)$ und den Parametern $\varepsilon = 10^{-6}$, $\gamma = 0.05$ erhalten wir die Ergebnisse in Tabelle A.3 und A.4 im Anhang.

Beispiel 5.2.4. Als nächstes betrachten wir das „internet-switching“-Beispiel von Keselman et al. aus [13]. Hierbei teilen sich N Internetnutzer die Bandbreite eines Routers. Jeder Nutzer ν versucht hierbei seinen Datendurchsatz x^ν zu maximieren. Allerdings hat der Router nur einen begrenzt großen Puffer mit der Kapazität $B \gg N$. Ist die Länge der Warteschlange

$$W = \sum_{\nu=1}^N x^\nu$$

größer als die Pufferkapazität, so verstopft das System und ankommende Datenpakete werden abgewiesen. Ziel der Nutzer muss es daher sein, den „Congestion“-Level so gering wie möglich zu halten. Daher hat jeder Nutzer $\nu = 1, \dots, N$ die (negative) Nutzenfunktion

$$\theta_\nu(x^\nu, x^{-\nu}) = -\frac{x^\nu}{W} \left(1 - \frac{W}{B}\right).$$

Im eindeutigen Nash-Gleichgewicht hat jeder Nutzer einen Datendurchsatz von $x^\nu = B \frac{(N-1)}{N^2}$.

Wir testen dieses Beispiel mit den Parametern $N = 10$, $B = 1$, $\varepsilon = 10^{-6}$, $\gamma = 0.05$ und dem Startwert $x^0 = (0.01, \dots, 0.01)$. Um Probleme bezüglich der Wohldefiniertheit der Nutzenfunktionen zu vermeiden, schränken wir den zulässigen Bereich auf

$$X = \{x \in \mathbb{R}^N \mid x^\nu \geq 0.01 \text{ für alle } \nu = 1, \dots, 10\}$$

ein. Dies ändert offensichtlich nichts an der Lage des Nash-Gleichgewichts im Punkt $x^* = (0.09, \dots, 0.09)$. Die zugehörigen Iterationsverläufe entnehme man den Tabellen A.5 und A.6 im Anhang, in der aus Platzgründen immer nur die ersten vier Komponenten der aktuellen Iterierten x^k angegeben sind.

k	x_1^k	x_2^k	x_3^k	x_4^k	$V_\gamma(x^k)$	t_k
0	0.50000000	0.50000000	0.50000000	0.50000000	25.21735124	-
1	0.35431548	0.88430395	1.08600051	0.90526760	2.04911508	1
2	0.51717380	0.90246191	1.03510235	0.92665214	0.51641812	0.5
3	0.53243680	1.00865373	1.05906779	1.01031126	0.13390608	0.5
4	0.49394870	1.02740542	1.14363898	1.04525537	0.05719039	1
5	0.46782617	1.03672290	1.13306566	1.02802357	0.02582250	1
6	0.43746066	1.05798083	1.17213817	1.03077331	0.01704263	1
7	0.41238696	1.07138691	1.19233378	1.04230857	0.01145572	1
8	0.37052258	1.09043851	1.21858175	1.04781813	0.00714402	1
9	0.33738793	1.10309478	1.24395015	1.04752678	0.00515427	1
10	0.30603459	1.10852938	1.24206078	1.05026239	0.00394165	1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
28	0.19313042	1.09439678	1.22700967	1.04501508	0.00002333	1
29	0.19378472	1.09415898	1.22705934	1.04474372	0.00002019	1
30	0.19424197	1.09402019	1.22723161	1.04483413	0.00001715	1
31	0.19481475	1.09381084	1.22729516	1.04450054	0.00001449	1
32	0.19520809	1.09389417	1.22751780	1.04436479	0.00001182	1
33	0.19550305	1.09386185	1.22762507	1.04454739	0.00000928	1
34	0.19546230	1.09422517	1.22752601	1.04446326	0.00000697	1
35	0.19556569	1.09413352	1.22743701	1.04462814	0.00000480	1
36	0.19495297	1.09443632	1.22731744	1.04495798	0.00000246	1
37	0.19418802	1.09428513	1.22727422	1.04483519	0.00000139	1
38	0.19355814	1.09432354	1.22734242	1.04482263	0.00000072	1

Tabelle 5.4: Iterationsverlauf im DSL-Beispiel mit schwachem Crosstalk und Algorithmus 5.1.4

Beispiel 5.2.5. Dieses Beispiel findet man in [3], Case Study 1. Man betrachtet hier drei Firmen $\nu = 1, 2, 3$, die Strom herstellen und auf dem selben Markt verkaufen. Hierbei betreibt die erste Firma $n_1 = 1$ Generator, dessen erzeugte Strommenge mit x_1^1 bezeichnet wird, die zweite Firma $n_2 = 2$ Generatoren mit den Outputs x_1^2, x_2^2 und die dritte Firma $n_3 = 3$ Generatoren mit den Outputs x_1^3, x_2^3 und x_3^3 . Ziel jeder Firma ist es natürlich ihren Gewinn zu maximieren. Der Marktpreis für Strom ergibt sich in Abhängigkeit vom Angebot als

$$p(x) = 378.4 - 2 \sum_{\nu=1}^3 \sum_{j=1}^{n_\nu} x_j^\nu.$$

Bei der Herstellung von Strom entstehen den Firmen ν an den jeweiligen Generatoren $j = 1, \dots, n_\nu$ die folgenden Kosten

$$K_j^\nu(x_j^\nu) = \frac{c_j^\nu}{2} (x_j^\nu)^2 + d_j^\nu x_j^\nu + e_j^\nu$$

mit den Konstanten aus Tabelle 5.6. Damit lauten die (negativen) Nutzenfunktionen der

Fabrik ν	$c_{1\nu}$	$c_{2\nu}$	e_ν	$u_{\nu 1}$	$u_{\nu 2}$
1	0.10	0.01	0.50	6.5	4.583
2	0.12	0.05	0.25	5.0	6.250
3	0.15	0.01	0.75	5.5	3.750

Tabelle 5.5: Die Konstanten im Flussverschmutzungsproblem

Firma ν	Generator j	c_j^ν	d_j^ν	e_j^ν	α_j^ν	β_j^ν
1	1	0.04	2	0	0	80
2	1	0.035	1.75	0	0	80
2	2	0.125	1	0	0	50
3	1	0.0166	3.25	0	0	55
3	2	0.05	3	0	0	30
3	3	0.05	3	0	0	40

Tabelle 5.6: Konstanten im Elektrizitätsmarkt-Beispiel

Firmen $\nu = 1, 2, 3$

$$\theta_\nu(x^\nu, x^{-\nu}) = - \left(p(x) \sum_{j=1}^{n_\nu} x_j^\nu - \sum_{j=1}^{n_\nu} K_j^\nu(x_j^\nu) \right).$$

Der zulässige Bereich wird in diesem Beispiel durch Nichtnegativitätsrestriktionen und maximale Produktionskapazitäten beschränkt:

$$X = \{x \in \mathbb{R}^6 \mid \alpha \leq x \leq \beta\}.$$

Mit den Parametern $\varepsilon = 10^{-6}$, $\gamma = 0.05$ und dem Startwert $x^0 = (0, 0, 0, 0, 0, 0)$ erhalten wir die Iterationsverläufe in den Tabellen A.7 und A.8 im Anhang, wobei wir aus Platzgründen nur die ersten vier Komponenten von x^k ausgeben.

Beispiel 5.2.6. Dieses Beispiel stammt aus [17], Seite 233, und beschreibt einen Markt mit einem Anbieter-Oligopol. Fünf Firmen $\nu = 1, \dots, 5$ bieten auf dem selben Markt eine Menge x^ν des gleichen Gutes an. Der Marktpreis ergibt sich in Abhängigkeit vom Angebot aus

$$p(x) = 5000^{\frac{1}{\eta}} \left(\sum_{\nu=1}^5 x^\nu \right)^{\frac{-1}{\eta}},$$

wobei der Parameter $\eta = 1.1$ beschreibt, wie elastisch der Preis auf Angebotsänderungen reagiert. Den Firmen ν entstehen bei der Produktion die folgenden Kosten

$$K_\nu(x^\nu) = c_\nu x^\nu + \frac{\delta_\nu}{\delta_\nu + 1} \cdot K_\nu^{\frac{-1}{\delta_\nu}} \cdot (x^\nu)^{\frac{\delta_\nu + 1}{\delta_\nu}}$$

mit den Konstanten aus Tabelle 5.7. Der zulässige Bereich wird hier durch Nichtnegativitätsrestriktionen und eine gemeinsame obere Absatzgrenze beschränkt:

$$X = \left\{ x \in \mathbb{R}^5 \mid x \geq 0, \sum_{\nu=1}^5 x^\nu \leq P \right\}.$$

Fabrik ν	c_ν	K_ν	δ_ν
1	10	5	1.2
2	8	5	1.1
3	6	5	1.0
4	4	5	0.9
5	2	5	0.8

Tabelle 5.7: Die Konstanten im Oligopolproblem

Mit den Parametern $\varepsilon = 10^{-6}$, $\gamma = 0.05$ und dem Startwert $x^0 = (0, 0, 0, 0, 0)$ erhalten wir für verschiedene Werte von P die Iterationsverläufe aus den Tabellen A.9 und A.10 im Anhang, wobei wir auch hier aus Platzgründen wieder nur die ersten vier Komponenten von x^k ausgeben. Für verschiedene Werte von P erhält man mit den beiden Algorithmen die Iterationszahlen aus Tabelle 5.8. Für $P = 200$ tritt bei Algorithmus 5.1.1 bei der Berechnung von $y_\gamma(x)$ in der Funktion `fmincon` aus der Optimization Toolbox von MATLAB eine Endlosschleife auf.

	$P = 75$	$P = 100$	$P = 150$	$P = 200$
Algorithmus 5.1.1	15	17	26	-
Algorithmus 5.1.4	7	8	10	10

Tabelle 5.8: Iterationszahlen im Oligopol-Beispiel

Beispiel 5.2.7. Zum Abschluss wollen wir noch ein Beispiel von Rosen, aus [21], betrachten, in dem die strikte Komplementarität verletzt ist. In diesem Beispiel gibt es zwei Spieler mit den Nutzenfunktionen

$$\theta_1(x_1, x_2) = 0.5x_1^2 - x_1x_2 \text{ und } \theta_2(x_1, x_2) = x_2^2 + x_1x_2.$$

Der zulässige Bereich ist

$$X = \{x \in \mathbb{R}^2 \mid x \geq 0, x_1 + x_2 \geq 1\}.$$

In diesem GNEP existiert genau ein normalisiertes Nash-Gleichgewicht, nämlich $x^* = (1, 0)$. Der zugehörige Lagrange-Multiplikator zu

$$g(x) = \begin{pmatrix} -x_1 \\ -x_2 \\ -x_1 - x_2 + 1 \end{pmatrix}$$

lautet $\lambda(x^*) = (0, 0, 1)$. In der zweiten Ungleichung gilt also keine strikte Komplementarität. Mit den Parametern $\varepsilon = 10^{-6}$, $\gamma = 0.05$ und dem Startwert $x^0 = (1, 1)$ erhalten wir die Iterationsverläufe aus den Tabellen A.11 und A.12 im Anhang.

Wie erwartet benötigt der Algorithmus 5.1.1 bei den meisten Beispielen weniger Iterationen als der Algorithmus 5.1.4, der ja schließlich nur mit einer Approximation der Ableitung rechnet. Die lokal schnelle Konvergenz von Algorithmus 5.1.1 sieht man besonders schön in den Beispielen 5.2.4 und 5.2.5. Das einzige Beispiel, in dem Algorithmus

5.1.4 deutlich schneller ist, ist das Beispiel 5.2.6. Hier wird der Unterschied umso größer, je größer man die gemeinsame Absatzgrenze P wählt, d.h. je weiter der Startpunkt vom Nash-Gleichgewicht entfernt ist. Dies könnte darauf hindeuten, dass der Algorithmus 5.1.4 stabiler ist.

Anhang

A.1 Tabellen

k	x_1^k	x_2^k	$V_\gamma(x^k)$	t_k
0	2.00000000	0.00000000	82.92682927	-
1	5.33333343	5.33333323	0.00000000	1

Tabelle A.1: Iterationsverlauf im Duopol-Beispiel mit Algorithmus 5.1.1

k	x_1^k	x_2^k	$V_\gamma(x^k)$	t_k
0	2.00000000	0.00000000	82.92682927	-
1	6.63414635	4.87804876	1.16074926	0.25
2	5.78962750	4.87963174	0.10099424	1
3	5.49925965	5.12575011	0.01891793	1
4	5.33217740	5.32551707	0.00009376	1
5	5.33262331	5.33325524	0.00000073	1

Tabelle A.2: Iterationsverlauf im Duopol-Beispiel mit Algorithmus 5.1.4

k	x_1^k	x_2^k	x_3^k	$V_\gamma(x^k)$	t_k
0	0.00000000	0.00000000	0.00000000	77.42990778	-
1	21.78434475	16.97415986	1.93531630	0.04054515	1
2	21.17266304	16.06961160	2.69135288	0.00007705	1
3	21.11820326	16.03111020	2.74592768	0.00000106	1
4	21.19213083	16.02560123	2.68935110	0.00001824	1
5	21.14432223	16.02727781	2.72651043	0.00000000	1

Tabelle A.3: Iterationsverlauf im Flussverschmutzungs-Beispiel mit Algorithmus 5.1.1

k	x_1^k	x_2^k	x_3^k	$V_\gamma(x^k)$	t_k
0	0.00000000	0.00000000	0.00000000	77.42990778	-
1	2.40663168	2.44776941	2.67926174	58.99670141	1
2	11.04297613	10.82175505	12.26257788	2.09961026	1
3	11.10216453	11.23705434	12.09009633	1.91101655	1
4	11.41235047	13.15733342	11.26380466	1.21040348	1
5	12.10184237	16.77887883	9.62313061	0.67130210	1
6	12.21524334	16.82260217	9.52053490	0.65592277	1
7	12.75815352	16.78955548	9.10280163	0.57899180	1
8	15.29687373	16.56318537	7.17119483	0.28159348	1
9	21.14493383	16.02860586	2.72562612	0.00000001	1

Tabelle A.4: Iterationsverlauf im Flussverschmutzungs-Beispiel mit Algorithmus 5.1.4

k	x_1^k	x_2^k	x_3^k	x_4^k	$V_\gamma(x^k)$	t_k
0	0.01000000	0.01000000	0.01000000	0.01000000	3.99007405	-
1	0.08239642	0.08239685	0.08239642	0.08239685	0.01643612	0.5
2	0.08911740	0.08911733	0.08911740	0.08911733	0.00021274	1
3	0.08987471	0.08987471	0.08987471	0.08987471	0.00000427	1
4	0.08998178	0.08998178	0.08998178	0.08998178	0.00000009	1

Tabelle A.5: Iterationsverlauf im Internet-Switching-Beispiel mit Algorithmus 5.1.1

k	x_1^k	x_2^k	x_3^k	x_4^k	$V_\gamma(x^k)$	t_k
0	0.01000000	0.01000000	0.01000000	0.01000000	3.99007405	-
1	0.08831286	0.08831287	0.08831286	0.08831287	0.00078093	0.5^7
2	0.08867861	0.08867860	0.08867861	0.08867860	0.00047802	1
3	0.08997811	0.08997810	0.08997811	0.08997810	0.00000013	1

Tabelle A.6: Iterationsverlauf im Internet-Switching-Beispiel mit Algorithmus 5.1.4

k	x_1^k	x_2^k	x_3^k	x_4^k	$V_\gamma(x^k)$	t_k
0	0.00000000	0.00000000	0.00000000	0.00000000	52011.98982334	-
1	0.00000000	62.73511240	17.61322641	55.00000000	3229.24997503	1
2	46.66162558	32.15384780	15.00331112	22.10729658	0.00000000	1

Tabelle A.7: Iterationsverlauf im Elektrizitätsmarkt-Beispiel mit Algorithmus 5.1.1

k	x_1^k	x_2^k	x_3^k	x_4^k	$V_\gamma(x^k)$	t_k
0	0.00000000	0.00000000	0.00000000	0.00000000	52011.98982334	-
1	40.00000000	40.00000000	25.00000000	27.50000000	2104.65457415	0.5
2	53.84971463	29.14484122	15.31511438	19.75151393	47.18706378	1
3	45.00531748	29.06294341	15.61568516	22.59147444	12.11731280	1
4	45.24127558	30.41208997	16.80657789	21.81524419	2.10791725	0.5
5	47.10947086	30.40659130	16.49006176	21.40772990	0.36852557	1
6	46.68556650	30.57614478	16.55112584	21.46027090	0.21878301	1
7	46.65997618	30.66677349	16.46951787	21.47423101	0.20082759	1
8	46.63353960	31.15796288	15.99499817	21.53035061	0.11602041	1
9	46.64029857	32.40991048	14.76054267	21.66686737	0.02014450	1
10	46.65518004	32.41626550	14.74944568	21.67087519	0.01876334	1
11	46.69624342	32.39764852	14.75410420	21.69647186	0.01410322	1
12	46.71616681	32.34317484	14.80079800	21.73536850	0.00936684	1
13	46.71236313	32.22926245	14.91365941	21.80759964	0.00317036	1
14	46.68447208	32.16201772	14.98831340	21.84639246	0.00089691	1
15	46.66471742	32.15308718	15.00395232	21.85578872	0.00050780	1
16	46.66177637	32.15186811	15.00599955	21.85697504	0.00048032	1
17	46.66041403	32.15172011	15.00642531	21.86036950	0.00045823	1
18	46.65896895	32.15233028	15.00616117	21.87289934	0.00041637	1
19	46.65745623	32.15364631	15.00507252	21.91076889	0.00033067	1
20	46.65539397	32.15651867	15.00228447	22.00180907	0.00019855	1
21	46.65636392	32.15772522	15.00097177	22.09325882	0.00008034	1
22	46.65953494	32.15571965	15.00213652	22.11935831	0.00000520	1
23	46.66016088	32.15509305	15.00261206	22.11423555	0.00000148	0.25
24	46.66044721	32.15465251	15.00279841	22.11161014	0.00000086	0.25

Tabelle A.8: Iterationsverlauf im Elektrizitätsmarkt-Beispiel mit Algorithmus 5.1.4

k	x_1^k	x_2^k	x_3^k	x_4^k	$V_\gamma(x^k)$	t_k
0	10.00000000	10.00000000	10.00000000	10.00000000	1025.53578372	-
1	10.38151906	10.64672094	10.85293026	11.01964351	803.22093821	1
2	12.34156651	12.39124978	12.39266587	12.37753898	452.31190454	1
3	12.50972062	13.09530855	13.56399005	13.94085562	244.17920114	1
4	13.95455863	14.60147647	15.10808328	15.50481291	7.00706926	1
5	11.97983930	13.79090630	15.29736511	16.50900318	1.45671743	1
6	11.07524934	13.41877112	15.39586574	16.97980539	0.30653992	1
7	10.68672825	13.23109583	15.42557767	17.19634814	0.06657473	1
8	10.52344115	13.13430992	15.42792209	17.29597700	0.01496376	1
9	10.45492527	13.08488163	15.42284901	17.34190886	0.00346492	1
10	10.42596425	13.06007717	15.41741439	17.36304978	0.00082192	1
11	10.41354827	13.04773533	15.41342959	17.37290536	0.00019842	1
12	10.40815679	13.04165648	15.41087079	17.37750957	0.00004855	1
13	10.40578616	13.03868461	15.40933795	17.37966226	0.00001201	1
14	10.40472969	13.03723929	15.40845818	17.38066921	0.00000287	1
15	10.40430922	13.03659381	15.40799892	17.38110592	0.00000085	1

Tabelle A.9: Iterationsverlauf im Oligopol-Beispiel für $P = 75$ und Algorithmus 5.1.1

k	x_1^k	x_2^k	x_3^k	x_4^k	$V_\gamma(x^k)$	t_k
0	10.00000000	10.00000000	10.00000000	10.00000000	1025.53578372	-
1	12.32378670	13.62788378	14.96298905	16.33489699	1.55219807	1
2	11.33166817	13.23623899	15.10935136	16.88903594	0.31780406	1
3	10.48858346	12.97716125	15.31816580	17.36458896	0.00646215	1
4	10.43620855	13.00036943	15.36898827	17.38564517	0.00133731	1
5	10.40700335	13.02944597	15.40724027	17.38742128	0.00002310	1
6	10.40529611	13.03296926	15.40797364	17.38406521	0.00000495	1
7	10.40418250	13.03529320	15.40773730	17.38173790	0.00000016	1

Tabelle A.10: Iterationsverlauf im Oligopol-Beispiel für $P = 75$ mit Algorithmus 5.1.4

k	x_1^k	x_2^k	$V_\gamma(x^k)$	t_k
0	1.00000000	1.00000000	1.97500000	-
1	1.00000000	0.00000000	0.00000000	1

Tabelle A.11: Iterationsverlauf im Rosen-Beispiel mit Algorithmus 5.1.1

k	x_1^k	x_2^k	$V_\gamma(x^k)$	t
0	1.00000000	1.00000000	1.97500000	-
1	1.00000000	0.00000000	0.00000000	1

Tabelle A.12: Iterationsverlauf im Rosen-Beispiel mit Algorithmus 5.1.4

A.2 Quellcode

Literaturverzeichnis

- [1] R. Cendrillon und M. Moonen, „Iterative Spectrum Balancing for Digital Subscriber Lines“, IEEE International Conference on Communications Vol. 3, 2005, pp. 1937-1941
- [2] F.H. Clarke, „Optimization and Nonsmooth Analysis“, SIAM, Philadelphia, 1990
- [3] J. Contreras, M. Klusch und J.B. Krawczyk, „Numerical solutions to Nash-Cournot equilibria in coupled constraint electricity markets“, IEEE Transactions on Power Systems Vol. 19, 2004, pp. 195-206
- [4] M.S. Gowda, „Inverse and implicit function theorems for H-differentiable and semismooth functions“, Optimization Methods and Software Vol. 19, 2004, pp. 443-461
- [5] J. Han und D.-F. Sun, „Superlinear convergence of approximate Newton methods for LC^1 optimization problems without strict complementarity“ Recent Advances in Nonsmooth Optimization, World Scientific Publishing Company, Singapur, 1995
- [6] A. von Heusinger und C. Kanzow, „Optimization reformulations of the generalized Nash equilibrium problem using Nikaido-Isoda-type functions“, Computational Optimization and Applications, to appear
- [7] A. von Heusinger und C. Kanzow, „ SC^1 optimization reformulations of the generalized Nash equilibrium problem“, Optimization Methods and Software, to appear
- [8] W.W. Hogan, „Point-to-set maps in mathematical programming“, SIAM Review vol. 15, pp. 591-603, 1973
- [9] H. Jiang und L. Qi, „Local uniqueness and convergence of iterative methods for nonsmooth variational inequalities“, Journal of Mathematical Analysis and Applications Vol. 196, 1995, pp. 314-331
- [10] Z.-Q. Luo und J.-S. Pang, „Analysis of Iterative Waterfilling Algorithm for Multiuser Power Control in Digital Subscriber Lines“, EURASIP Journal on Applied Signal Processing Vol. 2006, 2006, pp. 1-10
- [11] C. Kanzow, Vorlesung „Optimierungsmethoden II“, Universität Würzburg, SS 2007
- [12] C. Geiger und C. Kanzow, „Theorie und Numerik restringierter Optimierungsaufgaben“, Springer-Verlag, Berlin, Heidelberg, 2002

- [13] A. Kesselman, S. Leonardi und V. Bonifaci, „Game-theoretic analysis of internet switching with selfish users“, Lecture Notes in Computer Science Vol. 3828, 2005, pp. 236-245
- [14] J.B. Krawczyk und S. Uryasev, „Relaxation algorithms to find Nash equilibria with economic applications“, Environmental Modeling and Assessment Vol. 5, 2000, pp. 63-73
- [15] K.G. Murty, „Linear Complementarity, Linear and Nonlinear Programming“, Helderman-Verlag, Berlin, 1988
- [16] J.M. Ortega und W.C. Rheinboldt, „Iterative Solution of Nonlinear Equations in Several Variables“, SIAM, Philadelphia, 2000
- [17] J. Outrata, M. Kocvara und J. Zowe, „Nonsmooth Approach to Optimization Problems with Equilibrium Constraints“, Kluwer, Dordrecht, 1998
- [18] L. Qi, „Convergence analysis of some algorithms for solving nonsmooth equations“, Mathematics of Operations Research Vol. 18, 1993, pp. 227-244
- [19] H. Rademacher, „Über partielle und totale Differenzierbarkeit I“, Mathematical Annals 89, 1919, pp. 340-359
- [20] S.M. Robinson, „Strongly regular generalized equations“, Mathematics of Operations Research Vol. 5, 1980, pp. 43-62
- [21] J.B. Rosen, „Existence and uniqueness of equilibrium points for concave N -person games“, Econometrica Vol. 33, 1965, pp. 520-534
- [22] S. Scholtes, „Introduction to piecewise differentiable equations“, Habilitation, Universität Karlsruhe, 1994
- [23] T. Starr, J.M. Cioffi und P.J. Silverman, „Understanding Digital Subscriber Line Technology“, Prentice Hall, Upper Saddle River, 1999
- [24] W. Yu, G. Ginis und J.M. Cioffi, „Distributed Multiuser Power Control for Digital Subscriber Lines“, IEEE Journal on Selected Areas in Communications Vol. 20.5, 2002, pp. 1105-1115