

UNIVERSITÄT DER BUNDESWEHR MÜNCHEN
Fakultät für Elektrotechnik und Informationstechnik

Ein stochastisches Verfahren zur globalen Optimierung bei diskreten und kontinuierlichen Variablen

Ariane Sutor

UNIVERSITÄT DER BUNDESWEHR MÜNCHEN
Fakultät für Elektrotechnik und Informationstechnik

Ein stochastisches Verfahren zur globalen Optimierung bei diskreten und kontinuierlichen Variablen

Ariane Sutor

Vorsitzender des Promotionsausschusses: Prof.Dr.rer.nat. I. Eisele
1. Berichterstatter: Prof.Dr.rer.nat.Dr.-Ing. S. Schäffler
2. Berichterstatter: Prof.Dr.rer.nat. C. Hillermeier

Tag der Prüfung: 16.06.2003

Mit der Promotion erlangter akademischer Grad:
Doktor-Ingenieur
(Dr.-Ing.)

Neubiberg, den 18. Juni 2003

Inhaltsverzeichnis

Einleitung	8
1 Vorbemerkungen zur Theorie stochastischer Prozesse	11
1.1 Stochastische Prozesse	11
1.2 Brownsche Bewegung	12
1.3 Trefferzeiten	13
2 Diskrete stochastische Optimierung mit dem Metropolis-Algorithmus	15
2.1 Das diskrete globale Optimierungsproblem (<i>DGOP</i>)	15
2.2 Markov-Ketten	15
2.3 Der Metropolis-Algorithmus	19
3 Kontinuierliche stochastische Optimierung unter Verwendung stochastischer Integration	24
3.1 Das kontinuierliche globale Optimierungsproblem (<i>KGOP</i>)	24
3.2 Eine spezielle Klasse stochastischer Integralgleichungen	26
3.3 Ein numerisches Verfahren zur Lösung von (<i>KGOP</i>)	28
4 Globale Optimierung mit diskreten und kontinuierlichen Variablen	30
4.1 Problemstellung	30
4.2 Lösungsansatz	31
4.3 Mathematische Modellierung	32

<i>INHALTSVERZEICHNIS</i>	6
4.4 Formulierung des Algorithmus	33
4.4.1 Startdaten	33
4.4.2 Bestimmung des diskreten Parameters	34
4.4.3 Berechnung des kontinuierlichen Parameters	35
4.4.4 Ausgabedaten	37
4.5 Analyse des Verfahrens	37
5 Testergebnisse	44
6 Zusammenfassung und Ausblick: Anwendung in der Vektoroptimierung	51
Literaturverzeichnis	57

Einleitung

Viele Fragen der Wissenschaft, Technik oder Wirtschaft führen zu folgendem globalen Optimierungsproblem (*GOP*): Gesucht wird ein Paar $(i, x) \in S \times \mathbb{R}^n$, für das eine gegebene Zielfunktion

$$f : S \times \mathbb{R}^n \rightarrow \mathbb{R}$$

ihr globales Minimum annimmt. Dabei bezeichnet S eine endliche Menge und \mathbb{R}^n den n -dimensionalen euklidischen Vektorraum. Es liegt also ein globales Optimierungsproblem mit diskreten und kontinuierlichen Variablen vor. In der mathematischen Optimierung werden bisher hauptsächlich die folgenden beiden Teilprobleme von (*GOP*) behandelt:

1. In der diskreten Optimierung wird zu einer gegebenen Zielfunktion $h_1 : S \rightarrow \mathbb{R}$ ein $i \in S$ gesucht, für das h_1 ein Minimum annimmt. Da S endlich ist, kann dieses Problem theoretisch durch den Vergleich aller Funktionswerte gelöst werden. In praktischen Fällen ist die Anzahl der Elemente von S aber immer so groß, dass ein deterministisches Verfahren sehr viel Zeit benötigt. Darum werden stochastische Verfahren eingeführt, bei denen der Zufall helfen soll, schneller das gesuchte Minimum zu finden.
2. In der kontinuierlichen Optimierung wird zu einer im allgemeinen nichtlinearen Zielfunktion $h_2 : \mathbb{R}^n \rightarrow \mathbb{R}$ ein $x \in \mathbb{R}^n$ gesucht, für das h_2 ein globales Minimum annimmt. Von einem gegebenen Startpunkt aus wird in der nichtlinearen Optimierung iterativ ein Minimierer von h_2 approximiert. Hierzu werden in jeder Iteration eine Suchrichtung und eine Schrittweite gewählt. Oft wird h_2 mindestens als stetig differenzierbar vorausgesetzt. Dann wird durch den Gradienten von h_2 die Richtung des lokal steilsten Abstiegs bestimmt. Werden jedoch nur Abstiegsrichtungen, das heißt Richtungen, in denen der Zielfunktionswert abnimmt, zugelassen, so werden abhängig vom Startpunkt im Allgemeinen nur lokale Minima berechnet. Um ein globales Minimum zu finden, ist folgendes Verfahren möglich: Die Verfolgung der Kurve des steilsten Abstiegs, die durch den Gradienten beschrieben wird, wird kombiniert mit einer zufälligen Suche unter Verwendung einer Brownschen Bewegung.

Um (*GOP*) zu lösen, müssen diese beiden Teilprobleme gelöst werden. Es führt jedoch oft nicht zur Lösung des Gesamtproblems, die Teilprobleme separat nacheinander zu behandeln. In dieser Arbeit wird ein Verfahren entwickelt, das beide Teilprobleme simultan löst, indem abwechselnd ein diskreter und ein kontinuierlicher Optimierungsschritt ausgeführt wird.

Als diskretes Optimierungsverfahren wird der Metropolis-Algorithmus gewählt. Der Metropolis-Algorithmus ist ein stochastisches Verfahren der diskreten Optimierung, das nach gewissen festgelegten Regeln Punkte des Lösungsraums generiert und akzeptiert, falls sich der Zielfunktionswert verbessert. Verschlechtert sich der Zielfunktionswert, so wird der generierte Punkt mit einer gewissen positiven Wahrscheinlichkeit, nicht aber sicher akzeptiert. Dabei werden geringe Verschlechterungen wahrscheinlicher akzeptiert als größere.

Als kontinuierliches Optimierungsverfahren wird ein Verfahren von Schäffler gewählt, das die Idee des Metropolis-Algorithmus verbindet mit lokaler Minimierung entlang der Kurve des steilsten Abstiegs. Es wird dabei ausgenutzt, dass der euklidische Vektorraum eine Topologie besitzt und dass in der nichtlinearen Optimierung effiziente Verfahren zur lokalen Minimierung existieren. Die zufällige Suche neuer Punkte wird durch eine Brownsche Bewegung realisiert. Ein Kontrollparameter gewichtet die Zufallssuche und die lokale Minimierung.

Um das Verfahren von Schäffler anwenden zu können, wird vorausgesetzt, dass die Zielfunktion f zweimal stetig partiell nach x differenzierbar ist. Darüberhinaus wird eine Wachstumsbedingung für f "im Unendlichen" vorausgesetzt, um die Existenz eines globalen Minimums zu garantieren. Beide Voraussetzungen müssen durch die Modellierung von (*GOP*) gewährleistet werden, stellen jedoch keine starken Einschränkungen dar.

Der in dieser Arbeit entwickelte Algorithmus zur Lösung von (*GOP*) führt abwechselnd je einen Schritt des Metropolis-Algorithmus sowie einen Schritt des Verfahrens von Schäffler aus. Wir stellen den Algorithmus dar und zeigen wichtige theoretische Eigenschaften. Insbesondere findet der Algorithmus fast sicher eine geeignete Approximation des gesuchten Minimums der Zielfunktion f . Darüberhinaus werden einige einfache numerische Testergebnisse vorgestellt die belegen, dass der Algorithmus mit entsprechend gewählten Kontrollparametern gute Resultate berechnet. Zum Abschluß der Arbeit verallgemeinern wir (*GOP*) noch auf vektorwertige Zielfunktionen und wenden den Algorithmus auf dieses neue Problem an.

Das erste Kapitel stellt einige Grundlagen der Theorie stochastischer Prozesse bereit, die im Rahmen dieser Arbeit insbesondere für die kontinuierliche globale Optimierung benötigt werden. Im zweiten Kapitel wird diskrete stochastische Optimierung mit dem Metropolis-Algorithmus beschrieben. Das dritte Kapitel beschäftigt sich mit kontinuierlicher globaler Optimierung unter Verwendung des

Schäffler-Verfahrens.

Aufbauend auf der bis dahin dargestellten Theorie wird im vierten Kapitel unser Algorithmus zur globalen Optimierung mit diskreten und kontinuierlichen Variablen formuliert. Anschließend werden seine theoretischen Eigenschaften untersucht. Das fünfte Kapitel enthält numerische Ergebnisse. Im sechsten Kapitel werden die Ergebnisse der Arbeit noch einmal zusammengefasst.

Kapitel 1

Vorbemerkungen zur Theorie stochastischer Prozesse

Wir fassen in diesem Kapitel die im weiteren Verlauf der Arbeit benötigte Theorie stochastischer Prozesse zusammen. Die Darstellung ist [2] entnommen.

1.1 Stochastische Prozesse

Definition 1.1.1 Sei $(X_t)_{t \in I}$ eine Familie von Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit Werten in einem gemeinsamen Messraum (E, \mathcal{A}') . Dann heißt $(X_t)_{t \in I}$ oder auch $(\Omega, \mathcal{A}, P, (X_t)_{t \in I})$ ein **stochastischer Prozess**. Man nennt E den **Zustandsraum** des stochastischen Prozesses und $I \neq \emptyset$ die **Indexmenge**. Für jedes $\omega \in \Omega$ heißt die Abbildung

$$X_\omega : I \rightarrow E, \quad t \mapsto X_t(\omega)$$

ein **Pfad** des stochastischen Prozesses.

Sei $I \neq \emptyset$ eine beliebige Indexmenge. Mit $\mathcal{H}(I)$ bezeichnen wir die Menge aller endlichen nichtleeren Teilmengen von I . Für $J \in \mathcal{H}(I)$ bezeichne E^J die Menge aller Abbildungen von J in E . Sei $J \in \mathcal{H}(I)$. Wir betrachten die Produktabbildung

$$X_J : \Omega \rightarrow E^J, \quad X_J := \bigotimes_{t \in J} X_t.$$

Die gemeinsame Verteilung der Familie $(X_t)_{t \in J}$ ist die Verteilung $X_J(P)$ von X_J , nämlich

$$P_J := X_J(P).$$

Die Familie $(P_J)_{J \in \mathcal{H}(I)}$ heißt die Familie der **endlichdimensionalen Verteilungen** des stochastischen Prozesses $(X_t)_{t \in I}$.

Definition 1.1.2 Ein topologischer Raum E heißt **polnisch**, wenn es eine seine Topologie definierende, vollständige Metrik gibt und wenn diese Topologie eine abzählbare Basis besitzt.

Die Existenz einer abzählbaren Basis bedeutet die Existenz abzählbar vieler offener Mengen derart, dass jede offene Menge die Vereinigung gewisser dieser Mengen ist. Für metrisierbare Räume E ist hierzu die Existenz einer in E dichten, abzählbaren Teilmenge äquivalent.

Seien E ein polnischer Raum und I eine beliebige nichtleere Menge. Seien $J, H \in \mathcal{H}(I)$ mit $J \subseteq H$. Mit

$$p_J^H : E^H \rightarrow E^J$$

wird die stetige und somit $\mathcal{B}(E)^H - \mathcal{B}(E)^J$ -messbare Projektionsabbildung bezeichnet. Dabei bezeichnet $\mathcal{B}(E)^J := \mathcal{B}^J$ für $J \in \mathcal{H}(I)$ die Borelsche σ -Algebra über E^J . p_J^H ordnet jedem Element von E^H , also jeder Abbildung von H in E , deren Restriktion auf J zu.

Auf (E^I, \mathcal{B}^I) sei eine Familie $(P_J)_{J \in \mathcal{H}(I)}$ von Wahrscheinlichkeitsmaßen gegeben, die die Bedingung $P_J = p_J^H(P_H)$ für alle $J, H \in \mathcal{H}(I)$ mit $J \subseteq H$ erfüllt, wobei $p_J^H(P_H)$ das Bildmaß von P_H bei der Abbildung p_J^H bezeichnet. Dann gibt es ein eindeutiges Wahrscheinlichkeitsmaß P_I auf (E^I, \mathcal{B}^I) mit

$$P_J = p_J^I(P_I) \quad \text{für alle } J \in \mathcal{H}(I)$$

Wir betrachten den Wahrscheinlichkeitsraum $(E^I, \mathcal{B}^I, P_I)$. Für jedes $t \in I$ wird durch

$$X_t(\omega) := p_{\{t\}}^I(\omega) = \omega(t)$$

eine Zufallsvariable mit Werten in E definiert. Folglich ist $(X_t)_{t \in I}$ ein stochastischer Prozess mit Zustandsraum E und Parametermenge I . Die Familie seiner endlichdimensionalen Verteilungen ist $(P_J)_{J \in \mathcal{H}(I)}$. $(E^I, \mathcal{B}^I, P_I, (X_t)_{t \in I})$ heißt zu $(P_J)_{J \in \mathcal{H}(I)}$ gehörender **kanonischer Prozess**.

Zwei stochastische Prozesse $(X_t)_{t \in I}$ und $(\bar{X}_t)_{t \in I}$ mit gleichem Zustandsraum und gleicher Parametermenge heißen **äquivalent**, wenn sie zur selben Familie endlichdimensionaler Verteilungen führen.

1.2 Brownsche Bewegung

Definition 1.2.1 Ein stochastischer Prozess $(B_t)_{t \geq 0}$ auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit Zustandsraum \mathbb{R}^n heißt eine **n -dimensionale Brownsche Bewegung**, falls die folgenden Bedingungen erfüllt sind:

1. Der Prozess besitzt unabhängige Zuwächse, das heißt für je endlich viele Zeitpunkte $0 = t_0 < t_1 < \dots < t_m$, mit $m \in \mathbb{N}$ sind die Zufallsvariablen

$$B_0, B_{t_1} - B_{t_0}, \dots, B_{t_m} - B_{t_{m-1}}$$

stochastisch unabhängig.

2. Der Prozess besitzt stationäre Zuwächse. Für alle $s, t \geq 0$ mit $s < t$ ist die Zufallsvariable $B_t - B_s$ nach $\mathcal{N}(0, (t-s)I)$ normalverteilt. Dabei bezeichnet I die n -dimensionale Einheitsmatrix.
3. Fast alle Pfade $t \mapsto B_t(\omega)$ sind stetig
4. $B_0(\omega) = 0$ P -fast sicher

\mathbb{R}^n ist mit der durch die euklidische Metrik induzierten Topologie ein polnischer Raum. Somit ist auch $C(\mathbb{R}_+, \mathbb{R}^n)$ ein polnischer Raum. Ein Beweis dieser Aussage findet sich in [1]. Wir setzen

$$\Omega^n := C(\mathbb{R}_+, \mathbb{R}^n) = \{v : [0, \infty) \rightarrow \mathbb{R}^n; v \text{ stetig}\}.$$

Mit $\mathcal{B}(\Omega^n)$ bezeichnen wir die Borelsche σ -Algebra über Ω^n .

Satz 1.2.2 *Für jede Dimension $n \geq 1$ existieren n -dimensionale Brownsche Bewegungen. Jede n -dimensionale Brownsche Bewegung ist zu einem Ω^n -kanonischen Prozess äquivalent. Sie läßt sich also durch ein geeignetes Wahrscheinlichkeitsmaß W^n auf der Borelschen σ -Algebra $\mathcal{B}(\Omega^n)$ realisieren. Das Wahrscheinlichkeitsmaß W^n heißt n -dimensionales **Wiener-Maß**.*

BEWEIS : [2], Satz 40.3 auf S. 349.

In der Äquivalenzklasse einer jeden Brownschen Bewegung gibt es also einen Repräsentanten mit ausschließlich stetigen Pfaden der Form

$$(\Omega^n, \mathcal{B}(\Omega^n), W^n, (B_t)_{t \geq 0}),$$

wobei $B_t(\omega) = \omega(t)$ für alle $\omega \in \Omega^n$ und alle $t \geq 0$ gilt. Im Rahmen dieser Arbeit wird stets dieser Repräsentant als Brownsche Bewegung bezeichnet.

1.3 Trefferzeiten

Die folgende Definition einer Stoppzeit bezieht sich nur auf die in dieser Arbeit benötigte Situation.

Definition 1.3.1 Sei $(\Omega^n, \mathcal{B}(\Omega^n), W^n)$ der in Abschnitt 1.2 definierte Wahrscheinlichkeitsraum. Eine $\mathcal{B}(\Omega^n) - \mathcal{B}(\overline{\mathbb{R}})$ -meßbare Funktion $T : \Omega^n \rightarrow [0, \infty]$ heißt **Stoppzeit**.

Sei $(\Omega^n, \mathcal{B}(\Omega^n), W^n, (X_t)_{t \geq 0})$ ein stochastischer Prozess mit ausschließlich stetigen Pfaden. Der Zustandsraum von $(X_t)_{t \geq 0}$ sei \mathbb{R}^n . Für $A \in \mathcal{B}(\mathbb{R}^n)$ definieren wir die **Trefferzeit** von A durch

$$T_A : \Omega^n \rightarrow [0, \infty], \quad \omega \mapsto \begin{cases} \inf\{t \geq 0 | X_t(\omega) \in A\}, & \text{falls } \{t \geq 0 | X_t(\omega) \in A\} \neq \emptyset, \\ \infty, & \text{sonst.} \end{cases}$$

Die Trefferzeit gibt die kürzeste Zeit an, nach der der Pfad X_ω von $(X_t)_{t \geq 0}$ die Borelmenge A erreicht.

Satz 1.3.2 *Sei $(\Omega^n, \mathcal{B}(\Omega^n), W^n, (X_t)_{t \geq 0})$ ein stochastischer Prozess mit ausschließlich stetigen Pfaden und Zustandsraum \mathbb{R}^n . Dann ist für jede abgeschlossene Menge $A \in \mathcal{B}(\mathbb{R}^n)$ die Trefferzeit T_A eine Stoppzeit.*

BEWEIS : [2], Satz 49.5 auf S. 447.

Kapitel 2

Diskrete stochastische Optimierung mit dem Metropolis-Algorithmus

In diesem Kapitel betrachten wir den diskreten Teil unseres Optimierungsproblems (*GOP*) mit diskreten und kontinuierlichen Variablen, den wir als diskretes globales Optimierungsproblem (*DGOP*) bezeichnen. (*DGOP*) soll mit dem Metropolis-Algorithmus gelöst werden. In Abschnitt 2.2 stellen wir Grundlagen der Theorie endlicher Markov-Ketten zusammen. Diese werden in Abschnitt 2.3 benötigt, um den Metropolis-Algorithmus mathematisch zu modellieren und zu analysieren.

2.1 Das diskrete globale Optimierungsproblem (*DGOP*)

Sei S eine endliche Menge. Zu einer gegebenen Zielfunktion $h_1 : S \rightarrow \mathbb{R}$ betrachten wir das folgende Optimierungsproblem: Gesucht wird ein $i^* \in S$ mit $h_1(i^*) \leq h_1(i)$ für alle $i \in S$. Abkürzend schreiben wir hierfür

$$\min_{i \in S} h_1(i). \quad (DGOP)$$

2.2 Markov-Ketten

Die Darstellung der Theorie endlicher Markov-Ketten in diesem Abschnitt orientiert sich an [3].

Definition 2.2.1 Sei S eine endliche Menge. Sei $(X_n)_{n \in \mathbb{N}_0}$ ein stochastischer Prozess auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit Zustandsraum S

und $\mathbf{p} = (p_{ik})_{i,k \in S}$ eine $|S| \times |S|$ -Matrix mit nichtnegativen Einträgen, für die gilt:

$$\sum_{k \in S} p_{ik} = 1 \quad \text{für alle } i \in S.$$

Dann heißt $(X_n)_{n \in \mathbb{N}_0}$ eine **zeithomogene Markov-Kette**, falls

$$P(X_{n+1} = k | X_0 = i_0, \dots, X_n = i_n) = P(X_{n+1} = k | X_n = i_n)$$

für jedes $n \in \mathbb{N}_0$ und jede Folge i_0, \dots, i_n in S mit $P(X_0 = i_0, \dots, X_n = i_n) > 0$ gilt. Gilt

$$P(X_{n+1} = k | X_n = i) = p_{ik} \quad \text{für alle } i, k \in S,$$

so heißt \mathbf{p} die Matrix der **Übergangswahrscheinlichkeiten** von $(X_n)_{n \in \mathbb{N}_0}$. Die **Startwahrscheinlichkeiten** von $(X_n)_{n \in \mathbb{N}_0}$ werden definiert als

$$\alpha(i) := P(X_0 = i)$$

mit $\alpha(i) \geq 0$ und $\sum_{k \in S} \alpha(k) = 1$.

Unter den Elementen von S kann man sich die möglichen Zustände eines Systems vorstellen. X_n repräsentiert dabei den Zustand zum Zeitpunkt n . Die Folge X_0, X_1, X_2, \dots beschreibt also die Geschichte des Systems, dass sich gemäß den Übergangswahrscheinlichkeiten entwickelt. Die bedingte Verteilung des nächsten Zustands X_{n+1} unter der Bedingung des gegenwärtigen Zustands X_n darf nicht weiter von der Vergangenheit X_0, \dots, X_{n-1} abhängen.

Satz 2.2.2 *Sei S eine endliche Menge. Zu jedem Paar $(i, k) \in S \times S$ gebe es eine Zahl $p_{ik} \geq 0$ und es gelte*

$$\sum_{k \in S} p_{ik} = 1 \quad \text{für alle } i \in S.$$

Sei α eine Wahrscheinlichkeitsverteilung auf S . Dann gibt es einen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) und einen stochastischen Prozess $(X_n)_{n \in \mathbb{N}_0}$ auf (Ω, \mathcal{A}, P) , so dass $(X_n)_{n \in \mathbb{N}_0}$ eine Markov-Kette mit Zustandsraum S mit Startwahrscheinlichkeiten $\alpha(i)$ für $i \in S$ und Übergangswahrscheinlichkeiten p_{ik} für $(i, k) \in S \times S$ ist.

BEWEIS : [3], Theorem 8.1 auf S. 115.

Wir definieren die **höheren Übergangswahrscheinlichkeiten** einer homogenen Markov-Kette $(X_n)_{n \in \mathbb{N}_0}$ mit Zustandsraum S folgendermaßen: Sei $m \in \mathbb{N}$.

$$p_{ik}^{(m)} := P(X_{m+n} = k | X_n = i) \quad \text{für alle } i, k \in S, n \in \mathbb{N}_0$$

ist die Wahrscheinlichkeit, k von i aus in m Schritten zu erreichen. Es gilt

$$p_{ik}^{(m+n)} = \sum_{l \in S} p_{il}^{(m)} p_{lk}^{(n)} \quad \text{für alle } m, n \in \mathbb{N}.$$

Somit ergeben sich für $m \in \mathbb{N}$ die m -Schritt Übergangswahrscheinlichkeiten als m -te Potenz der Matrix der Übergangswahrscheinlichkeiten \mathbf{p} .

Wir führen einige Bezeichnungen ein. Für $i, k \in S$ und $n \in \mathbb{N}$ sei

$$g_{ik}^{(n)} := P(X_1 \neq k, \dots, X_{n-1} \neq k, X_n = k | X_0 = i)$$

die Wahrscheinlichkeit, von $i \in S$ aus in genau n Schritten $k \in S$ zu erreichen. Weiter sei

$$g_{ik} := \sum_{n=1}^{\infty} g_{ik}^{(n)}$$

die Wahrscheinlichkeit, k jemals von i aus zu erreichen.

Definition 2.2.3 Die endliche Menge S sei der Zustandsraum einer Markov-Kette. Ein Zustand $i \in S$ heißt **transient**, falls $g_{ii} < 1$ gilt.

Ein Zustand $i \in S$ heißt **rekurrent**, falls $g_{ii} = 1$ gilt.

Bemerkung 2.2.4 Es gilt:

$$P(X_n = k \text{ u.o.} | X_0 = i) = \begin{cases} 0 & \text{falls } g_{kk} < 1, \\ g_{ik} & \text{falls } g_{kk} = 1. \end{cases}$$

Dabei ist *u.o.* eine Abkürzung für "unendlich oft". Speziell für $k = i$ ergibt sich also

$$P(X_n = i \text{ u.o.} | X_0 = i) = \begin{cases} 0 & \text{falls } g_{ii} < 1, \\ 1 & \text{falls } g_{ii} = 1, \end{cases}$$

das heißt für einen rekurrenten Zustand i ist die Wahrscheinlichkeit, unendlich oft zurückzukehren, eins, während sie für einen transienten Zustand null ist.

Definition 2.2.5 Sei $(X_n)_{n \in \mathbb{N}}$ eine Markov-Kette mit endlichem Zustandsraum S und Übergangsmatrix \mathbf{p} . $(X_n)_{n \in \mathbb{N}_0}$ heißt **irreduzibel**, wenn folgendes gilt: Es gibt zu jedem Paar $(i, k) \in S \times S$ ein $n \in \mathbb{N}$ mit $p_{ik}^{(n)} > 0$, das heißt für alle $(i, k) \in S \times S$ gilt $g_{ik} > 0$.

Die folgende Definition führen wir ein, um einige technische Schwierigkeiten zu vermeiden.

Definition 2.2.6 Sei $(X_n)_{n \in \mathbb{N}}$ eine Markov-Kette mit endlichem Zustandsraum S und Übergangsmatrix \mathbf{p} . Die Periode d eines Punktes $i \in S$ ist definiert als größter gemeinsamer Teiler der Menge $\{m \geq 1 | p_{ii}^{(m)} > 0\}$. Gilt für $i \in S$, dass $d = 1$ ist, so heißt i **aperiodisch**.

Gilt $p_{ii} > 0$ so ist wegen $1 \in \{m \geq 1 | p_{ii}^{(m)} > 0\}$ der Punkt i offensichtlich aperiodisch.

Definition 2.2.7 Seien $(X_n)_{n \in \mathbb{N}_0}$ eine Markov-Kette mit Zustandsraum S und \mathbf{p} die Matrix ihrer Übergangswahrscheinlichkeiten. Eine Wahrscheinlichkeitsverteilung π auf S heißt **stationäre Verteilung** von $(X_n)_{n \in \mathbb{N}_0}$, falls

$$\sum_{i \in S} \pi(i) p_{ik} = \pi(k) \quad \text{für alle } k \in S$$

gilt.

Satz 2.2.8 Sei $(X_n)_{n \in \mathbb{N}_0}$ eine irreduzible, aperiodische Markov-Kette mit Zustandsraum S , Übergangsmatrix \mathbf{p} und stationärer Verteilung π . Dann ist $(X_n)_{n \in \mathbb{N}_0}$ rekurrent, das heißt jedes $i \in S$ ist rekurrent. Weiter gelten

$$\lim_{n \rightarrow \infty} p_{ik}^{(n)} = \pi(k) \quad \text{für alle } i, k \in S$$

und $\pi(i) > 0$ für alle $i \in S$. Die stationäre Verteilung π ist eindeutig.

BEWEIS : [3], Theorem 8.6 auf S. 125.

g_{ii} beschreibt die Verteilung der Wiederkehrzeiten des Zustandes $i \in S$. Ist i rekurrent, das heißt gilt $g_{ii} = 1$, so sind wir an der mittleren Wiederkehrzeit interessiert. Wir betrachten den Erwartungswert

$$\tilde{\mu}_i := \sum_{n=1}^{\infty} n g_{ii}^{(n)}.$$

Satz 2.2.9 Sei i ein rekurrenter Zustand einer Markov-Kette mit endlichem Zustandsraum und Übergangsmatrix \mathbf{p} . Es gelte $\lim_{n \rightarrow \infty} p_{ii}^{(n)} = u$. Dann ist $u > 0$ genau dann, wenn $\tilde{\mu}_i < \infty$ gilt. In diesem Fall ist $u = \frac{1}{\tilde{\mu}_i}$.

BEWEIS : [3], Lemma 3 auf S. 129.

Lemma 2.2.10 Sei $(X_n)_{n \in \mathbb{N}_0}$ eine Markov-Kette mit Zustandsraum S und Übergangsmatrix \mathbf{p} . Gilt für eine Wahrscheinlichkeitsverteilung π auf S

$$\pi(i) p_{ik} = \pi(k) p_{ki} \quad \text{für alle } i, k \in S,$$

so ist π eine stationäre Verteilung von $(X_n)_{n \in \mathbb{N}_0}$.

BEWEIS Summation über alle $i \in S$ liefert

$$\sum_{i \in S} \pi(i) p_{ik} = \sum_{i \in S} \pi(k) p_{ki} = \pi(k) \sum_{i \in S} p_{ki} = \pi(k).$$

◇

2.3 Der Metropolis-Algorithmus

Die Idee des **Metropolis-Algorithmus** geht zurück auf Metropolis, Rosenbluth, Rosenbluth, Teller und Teller, siehe [9]. Eine Beschreibung des Algorithmus findet sich in [7].

Der Metropolis-Algorithmus ist ein stochastisches Optimierungsverfahren zur Lösung des Minimierungsproblems

$$\min_{i \in S} h_1(i). \quad (DGOP)$$

Dabei bezeichnet S eine endliche Menge und $h_1 : S \rightarrow \mathbb{R}$ eine gegebene Zielfunktion. Ausgehend von einem beliebigen Startpunkt $i_0 \in S$ wird iterativ eine Folge von Punkten $(i_n)_{n \in \mathbb{N}}$ erzeugt. Ein von i_j aus generierter Punkt i_{j+1} wird akzeptiert, falls $h_1(i_{j+1}) \leq h_1(i_j)$ gilt. Gilt $h_1(i_{j+1}) > h_1(i_j)$, so wird i_{j+1} mit einer Wahrscheinlichkeit größer als null aber kleiner als eins akzeptiert.

Um festzulegen, welche Punkte von einem gegebenen Punkt aus generiert werden können, wird eine **Nachbarschaftsstruktur** festgelegt. Diese Nachbarschaftsstruktur muß im Hinblick auf die mathematische Modellierung des Metropolis-Algorithmus gewisse Eigenschaften erfüllen. Ihre konkrete Bestimmung ist jedoch abhängig von der Problemstellung und der Struktur des Suchraums S . Im Rahmen dieser Arbeit wird die Nachbarschaftsstruktur mit allen notwendigen Eigenschaften stets als gegeben vorausgesetzt.

Wir bezeichnen die Nachbarschaft von $i \in S$ mit N_i . Man kann einen Punkt $k \in S$ als lokales Minimum bezeichnen, falls $h_1(k) \leq h_1(i)$ für alle $i \in N_k$ gilt. Die Tatsache, dass Verschlechterungen des Funktionswerts akzeptiert werden können, ermöglicht es dem Algorithmus, lokale Minima wieder zu verlassen.

Die Wahl eines neuen Punktes $i_{j+1} \in S$ von $i_j \in S$ aus im Metropolis-Algorithmus wird in zwei Schritte unterteilt:

1. Generierung eines neuen Punktes,
2. Test auf Akzeptanz.

Ist i_{j+1} akzeptiert worden, so übernimmt i_{j+1} die Rolle von i_j . Für Generierung und Akzeptanz neuer Punkte sind nur der aktuelle Punkt und sein Funktionswert entscheidend, nicht aber, wie dieser Punkt erreicht wurde. In diesem Sinne ist der Algorithmus gedächtnislos.

Wird sichergestellt, dass der Algorithmus mit Wahrscheinlichkeit eins jeden Punkt des Suchraums S in endlicher Zeit erzeugt, das heißt generiert und akzeptiert, so finden wir mit Wahrscheinlichkeit eins in endlicher Zeit das globale Minimum der Zielfunktion h_1 .

Darüberhinaus läßt sich für den Metropolis-Algorithmus eine interessante Konvergenzaussage zeigen. Nach unendlich vielen Übergängen des Algorithmus, das heißt nach der Erzeugung unendlich vieler Punkte, ist die Wahrscheinlichkeit, in einem globalen Optimum der Zielfunktion zu sein größer, als in irgendeinem anderen Punkt des Suchraums zu sein.

Wir interpretieren die Wahl eines neuen Punktes im Metropolis-Algorithmus als Zufallsexperiment, das aus den beiden Zufallsexperimenten Generierung und Akzeptanz besteht. Hierzu geben wir Generierungs- und Akzeptanzwahrscheinlichkeiten für die Elemente in S an. Die Gedächtnislosigkeit des Algorithmus legt es nahe, diese Zufallsexperimente als Markov-Kette $(X_n)_{n \in \mathbb{N}_0}$ mit Zustandsraum S zu modellieren. Die Übergangswahrscheinlichkeiten der Markov-Kette ergeben sich aus den Generierungs- und Akzeptanzwahrscheinlichkeiten. Diese hängen nicht vom Zeitpunkt $n \in \mathbb{N}_0$ ab, das heißt $(X_n)_{n \in \mathbb{N}_0}$ ist eine zeithomogene Markov-Kette.

Definition 2.3.1 Sei $i \in S$. Mit N_i bezeichnen wir die **Nachbarschaft von i** . Die Wahrscheinlichkeit, mit der ein Punkt $k \in S$ als Kandidat für einen neuen Punkt von $i \in S$ aus vorgeschlagen wird, heißt **Generierungswahrscheinlichkeit** G_{ik} . Wir definieren für $i \in S$

$$G_{ik} := \frac{1}{|N_i|} \chi_{N_i}(k) \quad \text{für alle } k \in S,$$

wobei χ_{N_i} die charakteristische Funktion der Menge N_i bezeichnet:

$$\chi_{N_i} : S \rightarrow \{0, 1\}, \quad k \mapsto \begin{cases} 1 & \text{falls } k \in N_i, \\ 0 & \text{falls } k \in S \setminus N_i. \end{cases}$$

Es ist also $G_{ik} > 0$ genau dann, wenn $k \in N_i$ gilt. Da die Nachbarschaftsstruktur, die dem Metropolis-Algorithmus zugrunde liegt, vor dem Start des Algorithmus zu definieren ist, muß G_{ik} für alle $i, k \in S$ a priori festgelegt werden. Wir setzen noch voraus, dass die Generierungswahrscheinlichkeiten symmetrisch sind, das heißt, dass

$$G_{ik} = G_{ki} \quad \text{für alle } i, k \in S$$

gilt. Hieraus ergeben sich

$$k \in N_i \iff i \in N_k \text{ und } |N_i| = |N_k| \text{ für alle } i, k \in S.$$

Dies wird im Beweis von Satz 2.3.5 verwendet.

Bemerkung 2.3.2 Für die Modellierung des Metropolis-Algorithmus genügt es, symmetrische Übergangswahrscheinlichkeiten zu fordern. Die Wahl der Generierungswahrscheinlichkeiten ist jedoch nicht Thema dieser Arbeit. Deshalb wurden sie in Definition 2.3.1 festgelegt.

Definition 2.3.3 Seien $i \in S$ und $k \in N_i$ ein gemäß G_{ik} generierter Punkt. Die Wahrscheinlichkeit, dass k als neuer Punkt von i aus akzeptiert wird heißt **Akzeptanzwahrscheinlichkeit** A_{ik} . Wir definieren für $\zeta > 0$

$$A_{ik} := \exp\left(-\frac{(h_1(k) - h_1(i))^+}{\zeta}\right) \quad \text{für alle } i \in S, k \in N_i,$$

wobei

$$(h_1(k) - h_1(i))^+ := \begin{cases} h_1(k) - h_1(i), & \text{falls } h_1(k) - h_1(i) > 0, \\ 0 & \text{sonst.} \end{cases}$$

Für $k \in S \setminus N_i$ wird $A_{ik} = 0$ festgelegt.

Die Konstante $\zeta > 0$ heißt **Kontrollparameter**.

Der Kontrollparameter ζ ist in Abhängigkeit des gegebenen Problems zu wählen. ζ bestimmt, wie wahrscheinlich Verschlechterungen des Zielfunktionswerts akzeptiert werden. Ist nämlich $h_1(k) > h_1(i)$, so gilt $0 < A_{ik} < 1$. Je kleiner ζ ist, desto kleiner wird A_{ik} , das heißt umso weniger wahrscheinlich wird die Verschlechterung von $h_1(i)$ auf $h_1(k)$ akzeptiert. Hingegen beeinflusst ζ nicht die Akzeptanz von Punkten mit mindestens gleich gutem Funktionswert. Diese werden stets angenommen.

Wird der Kontrollparameter zu klein gewählt, so wird die Chance, lokale Minima zugunsten der Weitersuche nach globalen Minima wieder zu verlassen, zu gering. Ein zu großes ζ dagegen beraubt den Algorithmus seiner Tendenz, bessere Zielfunktionswerte den schlechteren vorzuziehen und damit seiner Steuerung. Wir betrachten in dieser Arbeit ζ stets als gegeben.

Wir definieren nun die Übergangswahrscheinlichkeiten des Algorithmus.

Definition 2.3.4 Sei $i \in S$. Die **Übergangswahrscheinlichkeit** p_{ik} von i zu einem beliebigen Punkt $k \in S$ ist gegeben durch

$$p_{ik} := \begin{cases} G_{ik}A_{ik}, & \text{falls } i \neq k, \\ 1 - \sum_{l \in S, l \neq i} p_{il}, & \text{falls } i = k. \end{cases}$$

Wegen $p_{ii} \geq \frac{1}{N_i} > 0$ sind alle Punkte $i \in S$ aperiodisch. Wir fassen die Übergangswahrscheinlichkeiten in der $|S| \times |S|$ -Matrix \mathbf{p} zusammen. Wegen

$$\sum_{k \in S} p_{ik} = 1 \quad \text{für alle } i \in S$$

existiert gemäß Satz 2.2.2 zu jeder Startverteilung eine homogene Markov-Kette $(X_n^M)_{n \in \mathbb{N}_0}$ mit Zustandsraum S und Übergangsmatrix \mathbf{p} .

Wir betrachten die Wahrscheinlichkeitsverteilung $\frac{1}{\mu}$ auf S mit

$$\frac{1}{\mu_i} := \frac{\exp\left(-\frac{h_1(i)}{\zeta}\right)}{\sum_{k \in S} \exp\left(-\frac{h_1(k)}{\zeta}\right)} \quad \text{für alle } i \in S.$$

Der folgende Satz wird in ähnlicher Form in [15] bewiesen.

Satz 2.3.5 *Sei $\zeta > 0$ fest gewählt. Mit p_{ik} für $i, k \in S$ werden die gemäß Definition 2.3.4 gegebenen Übergangswahrscheinlichkeiten des Metropolis-Algorithmus auf S beziehungsweise der ihn modellierenden Markov-Kette $(X_n^M)_{n \in \mathbb{N}_0}$ bezeichnet. Dann ist durch $\frac{1}{\mu}$ eine stationäre Verteilung von $(X_n^M)_{n \in \mathbb{N}_0}$ gegeben.*

BEWEIS Wir zeigen die Bedingung aus Lemma 2.2.10. Wegen $G_{ik} = G_{ki}$ für alle $i, k \in S$ müssen wir nur

$$\frac{1}{\mu_i} A_{ik} = \frac{1}{\mu_k} A_{ki} \quad \text{für alle } i, k \in S \quad (*)$$

zeigen. Es ergibt sich für alle $k \in N_i$

$$\begin{aligned} \frac{1}{\mu_i} A_{ik} &= \frac{\exp\left(-\frac{h_1(i)}{\zeta}\right)}{\sum_{l \in S} \exp\left(-\frac{h_1(l)}{\zeta}\right)} \exp\left(-\frac{(h_1(k) - h_1(i))^+}{\zeta}\right) = \\ &= \frac{1}{\sum_{l \in S} \exp\left(-\frac{h_1(l)}{\zeta}\right)} \exp\left(-\frac{h_1(k)}{\zeta}\right) \exp\left(-\frac{h_1(i) - h_1(k) + (h_1(k) - h_1(i))^+}{\zeta}\right) = \\ &= \frac{\exp\left(-\frac{h_1(k)}{\zeta}\right)}{\sum_{l \in S} \exp\left(-\frac{h_1(l)}{\zeta}\right)} \exp\left(-\frac{(h_1(i) - h_1(k))^+}{\zeta}\right) = \frac{1}{\mu_k} A_{ki}. \end{aligned}$$

Für alle $k \in S \setminus N_i$ gilt $\frac{1}{\mu_i} A_{ik} = 0 = \frac{1}{\mu_k} A_{ki}$. Damit folgt die Behauptung.

◇

Um Satz 2.2.8 anwenden zu können, müssen wir sicherstellen, dass die Markov-Kette $(X_n^M)_{n \in \mathbb{N}_0}$ irreduzibel ist. Dies ist durch geeignete Wahl der Nachbarschaftsstruktur zu gewährleisten und wird im folgenden angenommen. Zusammen mit Satz 2.2.9 erhalten wir:

Satz 2.3.6 *Sei $(X_n^M)_{n \in \mathbb{N}_0}$ die irreduzible, aperiodische, homogene Markov-Kette mit Zustandsraum S und Übergangsmatrix \mathbf{p} , die den Metropolis-Algorithmus mit Kontrollparameter ζ modelliert. Es gelten*

1. $(X_n^M)_{n \in \mathbb{N}_0}$ ist rekurrent, das heißt jedes $i \in S$ ist ein rekurrenter Zustand.

2. $\lim_{n \rightarrow \infty} p_{ik}^{(n)} = \frac{1}{\mu_k}$ für alle $i, k \in S$.
3. $\frac{1}{\mu}$ ist die eindeutige stationäre Verteilung von $(X_n^M)_{n \in \mathbb{N}_0}$.
4. Die mittlere Wiederkehrzeit von i ist μ_i . Es gilt $\mu_i < \infty$.

Die stationäre Verteilung $\frac{1}{\mu}$ hat die wichtige Eigenschaft, dass sie maximal wird für die globalen Minimierer von h_1 . Es ist also nach unendlich vielen Übergängen des Metropolis-Algorithmus wahrscheinlicher, in einem globalen Minimum von h_1 zu sein als in irgendeinem anderen Punkt des Lösungsraums.

Die Existenz der stationären Verteilung liefert mit Satz 2.2.8 die Rekurrenz aller Zustände, das heißt für jedes $i \in S$ ist die Wahrscheinlichkeit, unendlich oft zu i zurückzukehren gemäß Bemerkung 2.2.4, eins. Aus der Positivität der stationären Verteilung folgt gemäß Satz 2.2.9 die Endlichkeit der mittleren Wiederkehrzeit.

Kapitel 3

Kontinuierliche stochastische Optimierung unter Verwendung stochastischer Integration

Wir betrachten in diesem Kapitel den kontinuierlichen Teil des Optimierungsproblems (GOP), den wir als kontinuierliches globales Optimierungsproblem ($KGOP$) bezeichnen. ($KGOP$) wird mit einem in [11] entwickelten stochastischen Verfahren zur globalen Optimierung unter Verwendung stochastischer Integration gelöst. In Abschnitt 3.1 wird ($KGOP$) eingeführt. Abschnitt 3.2 befasst sich mit einer speziellen Klasse stochastischer Integralgleichungen, deren Lösung zur Lösung von ($KGOP$) führt. In Abschnitt 3.3 wird ein numerisches Verfahren zur Lösung von ($KGOP$) beschrieben.

3.1 Das kontinuierliche globale Optimierungsproblem ($KGOP$)

Gegeben sei die zweimal stetig differenzierbare Zielfunktion $h_2 : \mathbb{R}^n \rightarrow \mathbb{R}$. Wir betrachten das folgende Optimierungsproblem: Gesucht wird ein $x^* \in \mathbb{R}^n$ mit $h_2(x^*) \leq h_2(x)$ für alle $x \in \mathbb{R}^n$. Abkürzend schreiben wir hierfür

$$\min_{x \in \mathbb{R}^n} h_2(x). \quad (KGOP)$$

Die Lösung x^* von ($KGOP$) heißt **globales Minimum** von h_2 . Existiert zu einem Punkt $\bar{x} \in \mathbb{R}^n$ eine Umgebung $U(\bar{x}) \subseteq \mathbb{R}^n$ mit $h_2(\bar{x}) \leq h_2(x)$ für alle $x \in U(\bar{x})$, so heißt \bar{x} ein **lokales Minimum**. Notwendig dafür, dass h_2 im Punkt $\bar{x} \in \mathbb{R}^n$ ein lokales Minimum annimmt sind die folgenden Bedingungen:

$$\nabla h_2(\bar{x}) = 0, \quad \nabla^2 h_2(\bar{x}) \text{ ist positiv semidefinit.}$$

Numerische Verfahren zur lokalen Minimierung von h_2 erzeugen ausgehend von x_0 eine Folge $(x_j)_{j \in \mathbb{N}}$ von Punkten mit $h_2(x_{j+1}) < h_2(x_j)$, indem sie eine Folge von Suchrichtungen $(s_j)_{j \in \mathbb{N}_0}$ und von Schrittweiten $(\sigma_j)_{j \in \mathbb{N}_0}$ bestimmen und für $j \in \mathbb{N}_0$ den neuen Punkt gemäß $x_{j+1} = x_j + \sigma_j s_j$ festsetzen. Ein Vektor s heißt **Abstiegsrichtung** im Punkt x_j , falls ein $\bar{\sigma} > 0$ existiert mit $h_2(x_j + \sigma s) < h_2(x_j)$ für alle $\sigma \in (0, \bar{\sigma})$. Für eine Abstiegsrichtung $s \in \mathbb{R}^n$ gilt (siehe [5]):

$$s^T \nabla h_2(x_j) < 0.$$

Entsprechend definieren wir, dass s eine **Aufstiegsrichtung** im Punkt x_j ist, falls es ein $\bar{\sigma} > 0$ gibt mit $h_2(x_j + \sigma s) > h_2(x_j)$ für alle $\sigma \in (0, \bar{\sigma})$. Ist s eine Aufstiegsrichtung im Punkt x_j , so gilt $s^T \nabla h_2(x_j) > 0$.

Die folgende Annahme über die Zielfunktion h_2 ist entscheidend für das weitere Vorgehen:

Annahme A Es gibt ein $\varepsilon > 0$ so dass

$$x^T \nabla h_2(x) \geq \frac{1 + n\varepsilon^2}{2} \max\{1, \|\nabla h_2(x)\|_2\}$$

für alle $x \in \mathbb{R}^n \setminus \{x \in \mathbb{R}^n \mid \|x\|_2 \leq r\}$ mit einem $r > 0$ gilt.

Annahme A beschreibt ein spezielles Verhalten von h_2 außerhalb einer Kugel vom Radius r . Es wird lediglich die Existenz einer solchen Kugel gefordert. Annahme A legt das Verhalten von h_2 "im Unendlichen" fest.

Bemerkung 3.1.1 Jede Funktion, für die Annahme A erfüllt ist, besitzt ein globales Minimum. Außerhalb der Kugel $\{x \in \mathbb{R}^n \mid \|x\|_2 \leq r\}$ ist in jedem Punkt \bar{x} die Richtung $s = \bar{x}$ eine Aufstiegsrichtung. Auf dem Kompaktum $\{x \in \mathbb{R}^n \mid \|x\|_2 \leq r\}$ aber nimmt jede stetige Funktion ihr Minimum an.

Eine große Klasse numerischer Methoden zur Lösung von Minimierungsproblemen mit zweimal stetig differenzierbarer Zielfunktion $h_2 : \mathbb{R}^n \rightarrow \mathbb{R}$ kann interpretiert werden als numerische Lösung des folgenden Anfangswertproblems:

$$\dot{x}(t) = -\nabla h_2(x(t)), \quad t \in [0, \infty), \quad x(0) = x_0, \quad (AWP)$$

wobei $x_0 \in \mathbb{R}^n$ einen beliebigen Startpunkt bezeichnet. Da h_2 Lipschitz-stetig ist, besitzt das Anfangswertproblem (AWP) eine eindeutige Lösung $x : [0, \infty) \rightarrow \mathbb{R}^n$. Ist $\nabla h_2(x(t)) \neq 0$ so ist $-\nabla h_2(x(t))$ eine Abstiegsrichtung im Punkt $x(t)$ und es gilt

$$\dot{h}_2(x(t)) = \nabla h_2(x(t))^T \dot{x}(t) = -\nabla h_2(x(t))^T \nabla h_2(x(t)) < 0.$$

Somit ist $h_2(x(\cdot)) : [0, \infty) \rightarrow \mathbb{R}$ streng monoton fallend, das heißt es ist

$$h_2(x(s)) > h_2(x(t)) \quad \text{für alle } 0 \leq s < t < \infty.$$

Ausgehend von x_0 wird in der numerischen Optimierung iterativ eine Folge von Punkten $(x_j)_{j \in \mathbb{N}}$ bestimmt mit $h_2(x_{j+1}) \leq h_2(x_j)$ für alle $j \in \mathbb{N}_0$. Ob ein globales Minimum von h_2 gefunden wird, hängt von x_0 ab. Um möglichst alle lokalen Minima und damit das globale Minimum der Zielfunktion zu finden, untersuchen wir ein zufällig gestörtes Anfangswertproblem.

3.2 Eine spezielle Klasse stochastischer Integralgleichungen

Wir benötigen im Rahmen dieser Arbeit lediglich eine spezielle Klasse stochastischer Integralgleichungen. Eine allgemeine Einführung in die Theorie stochastischer Differentialgleichungen bieten [13] und [10]. Wir betrachten die folgende Klasse stochastischer Integralgleichungen:

$$X_t(\omega) = x_0 + \varepsilon(B_t(\omega) - B_0(\omega)) - \int_0^t \nabla h_2(X_\tau(\omega)) d\tau. \quad (SIE)$$

Dabei bezeichne $h_2 : \mathbb{R}^n \rightarrow \mathbb{R}$ die gegebene Zielfunktion unseres globalen Optimierungsproblems (*KGOP*), $(B_t)_{t \geq 0}$ sei eine n -dimensionale Brownsche Bewegung und es sei $\omega \in \Omega^n$.

Den ersten Teil der Gleichung, $x_0 + \varepsilon(B_t(\omega) - B_0(\omega))$, kann man als Zufallssuche eines Startpunkts interpretieren. Der zweite Teil, $-\int_0^t \nabla h_2(X_\tau(\omega)) d\tau$, ist gerade der Teil, der bei der nichtlinearen Optimierung betrachtet wird.

Der nächste Satz beschreibt Existenz, Eindeutigkeit und Regularität der Lösungen von (*SIE*).

Satz 3.2.1 *Gegeben sei die Integralgleichung (*SIE*). Für alle $x_0 \in \mathbb{R}^n$ und alle $\varepsilon > 0$, für die Annahme A für h_2 erfüllt ist, erhalten wir:*

1. *Es existiert ein eindeutiger stochastischer Prozess $(X_t)_{t \geq 0}$, der (*SIE*) löst.*
2. *Alle Pfade von $(X_t)_{t \geq 0}$ sind stetig.*
3. *$X_0 \equiv x_0$.*

BEWEIS [11], Theorem 2.1.1 auf S. 20.

Wir betrachten die Stoppzeiten $s_{\bar{x}, \rho}(\omega) : \Omega^n \rightarrow \bar{\mathbb{R}}$,

$$\omega \mapsto \begin{cases} \inf\{t \geq 0 \mid \|X_t(\omega) - \bar{x}\|_2 \leq \rho\} & \text{falls } \{t \geq 0 \mid \|X_t(\omega) - \bar{x}\|_2 \leq \rho\} \neq \emptyset, \\ \infty & \text{sonst.} \end{cases}.$$

Der folgende Satz aus [11] ermöglicht es, die Lösung von (*SIE*) zur Lösung des globalen Optimierungsproblems zu verwenden.

Satz 3.2.2 Gegeben sei (SIE) mit $\varepsilon > 0$, so dass Annahme A für h_2 erfüllt ist. Dann gelten für alle $x_0, \bar{x} \in \mathbb{R}^n$ und alle $\rho > 0$ die folgenden Aussagen:

1. $W^n(\{\omega \in \Omega^n | s_{\bar{x}, \rho}(\omega) < \infty\}) = 1$,
2. $E(s_{\bar{x}, \rho}) < \infty$, (E bezeichnet den Erwartungswert),
3. Jede Zufallsvariable X_t mit $t > 0$ hat eine Dichte p_t bezüglich des Lebesgue-Maßes mit

$$\lim_{t \rightarrow \infty} p_t(x) = p(x),$$

wobei p die eindeutig bestimmte Dichtefunktion ist mit

$$\frac{1}{2}\varepsilon^2 \sum_{j=1}^n \frac{\partial^2 p(x)}{\partial x_j^2} + \sum_{j=1}^n \frac{\partial}{\partial x_j} \left(\frac{\partial h_2(x)}{\partial x_j} p(x) \right) = 0.$$

BEWEIS [11], Corollary 2.2.2 auf S. 26.

Sei $x^* \in \mathbb{R}^n$ ein globales Optimum der Zielfunktion h_2 . Wir können die Dichte p aus Satz 3.2.2 explizit angeben, vergleiche [11]. Sie ist gegeben durch

$$p : \mathbb{R}^n \rightarrow \mathbb{R}, \quad x \mapsto \frac{\exp\left(-\frac{2(f(x)-f(x^*))}{\varepsilon^2}\right)}{\int_{\mathbb{R}^n} \exp\left(-\frac{2(f(x)-f(x^*))}{\varepsilon^2}\right) d\lambda^n(x)}.$$

Satz 3.2.2 besagt, dass W^n -fast alle Pfade von $(X_t)_{t \geq 0}$ jede Kugel mit Mittelpunkt \bar{x} und Radius ρ für beliebiges $\rho > 0$ nach endlicher Zeit treffen. Darüber hinaus ist der Erwartungswert endlich.

Die Dichte p in Satz 3.2.2 hat die wichtige Eigenschaft, dass jeder globale Maximierer von p ein globaler Minimierer von h_2 ist.

Nun betrachten wir den Parameter ε in (SIE) . Er ist ein Maß für die Gewichtung der lokalen Minimierung von h_2 durch

$$X_t = x_0 - \int_0^t h_2(X_\tau) d\tau$$

und einer Zufallssuche durch Realisierungen normalverteilter Zufallsvariablen mit wachsender Varianz gemäß

$$X_t(\omega) = x_0 + B_t(\omega) - B_0(\omega).$$

Wird $\varepsilon > 0$ zu klein gewählt, so dominiert die lokale Minimierung und für festes ω bleibt der zugehörige Pfad lange in der Nähe jedes lokalen Minimums. Wird ε dagegen zu groß gewählt, so dass die Zufallssuche dominiert, so spielen die lokalen Minima keine entscheidende Rolle. Die Wahl von ε hängt von der Problemstellung und der Zielfunktion h_2 ab. Wir betrachten ε im folgenden stets als gegeben.

3.3 Ein numerisches Verfahren zur Lösung von (KGOP)

Es ist mit Satz 3.2.2 naheliegend, die Lösung von (KGOP) zu approximieren, indem ein Pfad der Lösung $(X_t)_{t \geq 0}$ berechnet wird. Wird stets der bis zum aktuellen Zeitpunkt kleinste berechnete Zielfunktionswert auf diesem Pfad als Minimum gespeichert, so kann nach einer festgelegten Anzahl von Iterationen vom dann erreichten Minimum aus eine lokale Optimierung erfolgen.

Für die numerische Approximation eines beliebigen aber fest gewählten Pfades $\omega \in \Omega^n$ der Lösung von (SIE) wählen wir ein semi-implizites Eulerverfahren. Eine Beschreibung findet sich beispielsweise in [6]. Wir betrachten die Integralgleichung

$$X_{t+\sigma}(\omega) = \xi + \varepsilon(B_{t+\sigma}(\omega) - B_t(\omega)) - \int_t^{t+\sigma} \nabla h_2(X_\tau(\omega)) d\tau$$

für $\sigma \geq 0$, $\xi \in \mathbb{R}^n$. Ihre Lösung wird mit $X(t + \sigma; \xi, \sigma, \omega)$ bezeichnet. Mit der Taylor-Entwicklung

$$\nabla h_2(x) = \nabla h_2(\xi) + \nabla^2 h_2(\xi)(x - \xi) + \int_0^1 (\nabla^2 h_2(\xi + \vartheta(x - \xi)) - \nabla^2 h_2(\xi)) d\vartheta(x - \xi)$$

erhalten wir die folgende Approximation $\bar{X}(t + \sigma; \xi, \sigma, \omega)$ von $X(t + \sigma; \xi, \sigma, \omega)$:

$$\bar{X}(t + \sigma; \xi, \sigma, \omega) = \xi + \varepsilon(B_{t+\sigma}(\omega) - B_t(\omega)) - \sigma \nabla h_2(\xi) - \sigma \nabla^2 h_2(\xi)(\bar{X}(t + \sigma; \xi, \sigma, \omega) - \xi)$$

und folglich

$$\bar{X}(t + \sigma; \xi, \sigma, \omega) = \xi - (I + \sigma \nabla^2 h_2(\xi))^{-1} (\sigma \nabla h_2(\xi) - \varepsilon(B_{t+\sigma}(\omega) - B_t(\omega)))$$

für alle σ , für die $(I + \sigma \nabla^2 h_2(\xi))$ regulär ist. Wir können stets ein $\sigma > 0$ angeben, für das $(I + \sigma \nabla^2 h_2(\xi))$ regulär ist:

Bemerkung 3.3.1 Bekanntlich ist eine $n \times n$ -Matrix A mit reellen Einträgen invertierbar, falls $\|A - I\| < 1$ gilt (vergleiche etwa [16]). Somit ist $(I + \sigma \nabla^2 h_2(\xi))$ invertierbar, falls

$$\|\sigma \nabla^2 h_2(\xi)\|_2 = \|(I + \sigma \nabla^2 h_2(\xi)) - I\|_2 < 1.$$

Es ist $\|\nabla^2 h_2(\xi)\|_2 = \max_\lambda |\lambda| := |\lambda_{max}|$, wobei das Maximum über alle Eigenwerte von $\nabla^2 h_2(\xi)$ gebildet wird. Für $\sigma < \frac{1}{|\lambda_{max}|}$ ist also $(I + \sigma \nabla^2 h_2(\xi))$ invertierbar.

Der folgende Satz beschreibt den Integrationsfehler des Verfahrens, vergleiche [11].

Satz 3.3.2 *Zu jedem ε , für das Annahme A erfüllt ist, und zu allen $t \geq 0$, $\xi \in \mathbb{R}^n$ existiert ein $c > 0$ mit*

$$\lim_{\sigma \rightarrow 0} \frac{\|X(t + \sigma; \xi, \sigma, \omega) - \bar{X}(t + \sigma; \xi, \sigma, \omega)\|}{\sqrt{\sigma^3 \ln(\ln(\frac{1}{\sigma}))}} \leq c \quad W^n - \text{fast sicher.}$$

BEWEIS [11], Theorem 2.3.1 auf S. 31.

Es werden nur solche Schrittweiten σ verwendet, für die $(I + \sigma \nabla^2 h_2(\xi))$ positiv definit ist. Von einer Startschrittweite σ_0 aus wird $\sigma \leq \sigma_0$ berechnet, so dass $(I + \sigma \nabla^2 h_2(\xi))$ positiv definit ist. Sodann werden die folgenden drei Punkte $\bar{X}(t + \sigma; \xi, \sigma, \omega)$, $\bar{X}(t + \frac{\sigma}{2}; \xi, \frac{\sigma}{2}, \omega)$ und $\bar{X}(t + \sigma; \bar{X}(t + \frac{\sigma}{2}; \xi, \frac{\sigma}{2}, \omega), \frac{\sigma}{2}, \omega)$ berechnet.

Wir führen einen weiteren Kontrollparameter des Algorithmus ein. Es muß ein $\delta > 0$ gewählt werden, das das folgende Akzeptanzkriterium steuert: Falls

$$\|\bar{X}(t + \sigma; \xi, \sigma, \omega) - \bar{X}(t + \sigma; \bar{X}(t + \frac{\sigma}{2}; \xi, \frac{\sigma}{2}, \omega), \frac{\sigma}{2}, \omega)\| < \delta$$

gilt, so wird $\bar{X}(t + \sigma; \bar{X}(t + \frac{\sigma}{2}; \xi, \frac{\sigma}{2}, \omega), \frac{\sigma}{2}, \omega)$ als Approximation von $X(t + \sigma; \xi, \sigma, \omega)$ akzeptiert, andernfalls werden die drei Punkte mit neuer Schrittweite $\frac{\sigma}{2}$ neu berechnet. Das Akzeptanzkriterium soll sicherstellen, dass die berechnete Approximation in der Nähe des von $\omega \in \Omega^n$ definierten Pfades liegt.

Es ist wichtig festzuhalten, dass $\bar{X}(t + \sigma; \xi, \sigma, \omega)$ durch die Realisierung $n_1 + n_2$ von $\varepsilon(B_{t+\sigma} - B_t)$ berechnet werden muß, wobei die Realisierungen $n_1 := \varepsilon(B_{t+\frac{\sigma}{2}}(\omega) - B_t(\omega))$ und $n_2 := \varepsilon(B_{t+\sigma}(\omega) - B_{t+\frac{\sigma}{2}}(\omega))$ für die Berechnung von $\bar{X}(t + \frac{\sigma}{2}; \xi, \frac{\sigma}{2}, \omega)$ und $\bar{X}(t + \sigma; \bar{X}(t + \frac{\sigma}{2}; \xi, \frac{\sigma}{2}, \omega), \frac{\sigma}{2}, \omega)$ verwendet werden, da alle drei Punkte Approximationen von Punkten auf dem selben Pfad sind.

Eine Formulierung des Algorithmus findet sich in [11]. In dieser Arbeit wird er in Abschnitt 4.4 dargestellt als Teil des dort entwickelten Algorithmus mit diskreten und kontinuierlichen Variablen.

Kapitel 4

Globale Optimierung mit diskreten und kontinuierlichen Variablen

In diesem Kapitel betrachten wir unser globales Optimierungsproblem (*GOP*) mit diskreten und kontinuierlichen Variablen und entwickeln ein stochastisches Verfahren zur Lösung von (*GOP*). Dieses Verfahren baut auf den in Kapitel 2 und Kapitel 3 entwickelten Verfahren auf. Wir gehen in Abschnitt 4.1 zunächst auf die Problemstellung ein. In Abschnitt 4.2 wird der in dieser Arbeit verfolgte Lösungsansatz beschrieben. Abschnitt 4.3 enthält die mathematische Modellierung unseres Algorithmus. Dieser wird in Abschnitt 4.4 detailliert beschrieben. In Abschnitt 4.5 wird das entwickelte Verfahren analysiert.

4.1 Problemstellung

Gegeben sei die Zielfunktion

$$f : S \times \mathbb{R}^n \rightarrow \mathbb{R}, \quad (i, x) \mapsto f(i, x).$$

Dabei bezeichne S eine endliche Menge und \mathbb{R}^n mit $n \in \mathbb{N}$ den n -dimensionalen euklidischen Vektorraum. Wir betrachten das folgende globale Optimierungsproblem (*GOP*) mit diskreten und kontinuierlichen Variablen: Wir suchen zu unserer gegebenen Zielfunktion f ein Parameterpaar $(i^*, x^*) \in S \times \mathbb{R}^n$ mit

$$f(i^*, x^*) \leq f(i, x) \quad \text{für alle } (i, x) \in S \times \mathbb{R}^n.$$

Abkürzend schreiben wir hierfür

$$\min_{(i,x) \in S \times \mathbb{R}^n} f(i, x), \quad (\text{GOP})$$

Wir gehen in dieser Arbeit stets davon aus, dass f zweimal stetig partiell nach $x \in \mathbb{R}^n$ differenzierbar ist. Zu jedem $i \in S$ erhalten wir eine zweimal stetig differenzierbare Funktion

$$f^i : \mathbb{R}^n \rightarrow \mathbb{R}, \quad x \mapsto f(i, x).$$

Die folgende Annahme über die Zielfunktion f ist fundamental für die weiteren Untersuchungen:

Annahme B: Es gibt ein $\varepsilon > 0$ so dass für jedes $i \in S$ gilt

$$x^T \nabla f^i(x) \geq \frac{1 + n\varepsilon^2}{2} \max\{1, \|\nabla f^i(x)\|_2\}$$

für alle $x \in \mathbb{R}^n \setminus \{x \in \mathbb{R}^n \mid \|x\|_2 \leq r\}$ mit einem $r > 0$.

Wie in Bemerkung 3.1.1 folgt, dass für jedes $i \in S$ die Funktion f^i ein globales Minimum besitzt, das heißt es existiert zu jedem $i \in S$ ein $x^{i,*} \in \mathbb{R}^n$ mit $f^i(x^{i,*}) \leq f^i(x)$ für alle $x \in \mathbb{R}^n$. Da S endlich ist, können wir i^* so wählen, dass

$$f^{i^*}(x^{i^*,*}) = \min_{i \in S} f^i(x^{i,*})$$

gilt. Somit besitzt (GOP) eine Lösung. Es gilt

$$f(i^*, x^{i^*,*}) = \min_{(i,x) \in S \times \mathbb{R}^n} f(i, x).$$

Bemerkung 4.1.1 Annahme B ist erfüllt, falls Annahme A aus Kapitel 3 für f^i für jedes $i \in S$ mit einem ε_i erfüllt ist. Da S endlich ist, können wir $\varepsilon := \min_{i \in S} \varepsilon_i$ wählen, so dass Annahme B für f erfüllt ist.

4.2 Lösungsansatz

Wir entwickeln eine Methode, bei der iterativ Paare $(i, x) \in S \times \mathbb{R}^n$ berechnet werden. Hierzu wird jeweils eine Komponente von (i, x) festgehalten, während auf die andere Komponente ein stochastischer Minimierungsalgorithmus angewandt wird. Die jeweils festgehaltene Komponente wird als diskreter beziehungsweise kontinuierlicher Parameter bezeichnet.

Sei also zunächst zu jedem $i \in S$ ein $x^i \in \mathbb{R}^n$ fest gewählt. Dann wird auf die Zielfunktion

$$f_x : S \rightarrow \mathbb{R}, \quad i \mapsto f(i, x^i)$$

der diskrete Metropolis-Algorithmus aus Kapitel 2 angewandt. Hierzu wird zunächst eine Nachbarschaftsstruktur in S festgelegt, die sicherstellt, dass die den

Metropolis-Algorithmus modellierende Markov-Kette irreduzibel ist. Die Irreduzibilität hängt nur von den Generierungswahrscheinlichkeiten ab, nicht von den Akzeptanzwahrscheinlichkeiten und damit auch nicht von den Funktionswerten. Deshalb kann die Nachbarschaftsstruktur fest gewählt werden, unabhängig von den kontinuierlichen Parametern.

Im ersten Schritt wird ein beliebiger Punkt aus S als aktueller diskreter Parameter gewählt. In den folgenden Schritten ist der aktuelle diskrete Parameter der vom Algorithmus erzeugte. Vom aktuellen diskreten Parameter $k \in S$ aus wird wie in Abschnitt 2.3 beschrieben ein neuer Parameter i aus der Nachbarschaft N_k von k generiert und auf Akzeptanz geprüft. Wird i akzeptiert, so wird i der neue aktuelle Parameter. Andernfalls wird ein weiterer Punkt aus N_k generiert und überprüft.

Nun wird umgekehrt der aktuelle diskrete Parameter $i \in S$ festgehalten und es wird ein zugehöriges $x^i \in \mathbb{R}^n$ gesucht. Hierzu wird das Verfahren aus Kapitel 3 auf die Funktion f^i angewandt. Wir definieren die folgende stochastische Integralgleichung:

$$X_t(\omega) = x_0 + \varepsilon(B_t(\omega) - B_0(\omega)) - \int_0^t \nabla f^i(X_\tau(\omega)) d\tau \quad (SIE)_i$$

und approximieren numerisch einen Pfad der Lösung von $(SIE)_i$ mit dem in Abschnitt 3.3 beschriebenen Verfahren. Es wird ein Integrationsschritt dieses Verfahrens ausgeführt; das Ergebnis sei der Punkt \bar{x}^i . Verbessert sich der Funktionswert, so wird $x^i := \bar{x}^i$ gesetzt. Andernfalls wird der alte kontinuierliche Parameter beibehalten, jedoch \bar{x}^i als Startpunkt für den nächsten Integrationschritt gesetzt. Somit wird sichergestellt, dass wir zu jedem diskreten Parameter $i \in S$ eine monoton fallende Folge von Zielfunktionswerten erhalten. Das Verfahren aus Kapitel 3 wird dadurch nicht verändert. Somit gelten die Sätze 3.2.1 und 3.2.2 für $(SIE)_i$.

Es ist wichtig festzuhalten, dass wir einen neuen Punkt nur akzeptieren wollen, falls für die Schrittweite σ gilt $\sigma \geq c$ für eine feste Konstante $c \geq 0$. Dies ist lediglich eine Anforderung an das numerische Verfahren.

4.3 Mathematische Modellierung

Die Wahl eines neuen Punktes aus S kann als Ergebnis eines Zufallsexperiments gedeutet werden. Um dieses Zufallsexperiment mathematisch zu modellieren, müssen wir einen geeigneten Wahrscheinlichkeitsraum angeben. Hierzu zitieren wir den folgenden Satz aus [4].

Satz 4.3.1 *Zu jedem $n \in \mathbb{N}_0$ sei $\Omega_n \neq \emptyset$ eine abzählbare Menge. Sei ρ_0 eine Wahrscheinlichkeitsverteilung auf Ω_0 , und für alle $m \geq 1$ und $\omega_n \in \Omega_n$ mit $n < m$ sei $\rho_m|_{\omega_0, \dots, \omega_{m-1}}$ eine Wahrscheinlichkeitsverteilung auf Ω_m . Seien $\Omega = \prod_{n \geq 0} \Omega_n$ sowie $X_n : \Omega \rightarrow \Omega_n$ die Projektion auf die n -te Koordinate, und $\mathcal{A} = \bigotimes_{n \geq 0} \mathcal{P}(\Omega_n)$ die Produkt- σ -Algebra auf Ω . Dann existiert genau ein Wahrscheinlichkeitsmaß P auf (Ω, \mathcal{A}) mit der Eigenschaft*

$$P(X_0 = \omega_0, \dots, X_m = \omega_m) = \rho_0(\omega_0) \rho_1|_{\omega_0}(\omega_1) \cdot \dots \cdot \rho_m|_{\omega_0, \dots, \omega_{m-1}}(\omega_m)$$

für alle $m \in \mathbb{N}_0$ und $\omega_n \in \Omega_n$.

BEWEIS : [4], Satz 3.12 auf S. 60.

Seien nun $\Omega := \bigotimes_{n \geq 0} S_n$ mit $S_n := S$ für alle $n \in \mathbb{N}_0$ und $\mathcal{A} = \bigotimes_{n \geq 0} \mathcal{P}(S_n)$. Auf S ist die Wahrscheinlichkeitsverteilung ρ_0 gegeben mit $\rho_0(i) := \alpha(i)$ für alle $i \in S$, wobei α die Startverteilung des Optimierungs-Algorithmus bezeichnet. Weiter ist für jedes $m \in \mathbb{N}_0$ und alle $(i_0, \dots, i_{m-1}) \in \bigotimes_{n=0}^{m-1} S_n$ eine Wahrscheinlichkeitsverteilung $\rho_m|_{i_0, \dots, i_{m-1}}$ auf S gegeben mit

$$\rho_m|_{i_0, \dots, i_{m-1}}(k) = \begin{cases} \frac{1}{|N_{i_{m-1}}|} \exp\left(-\frac{(f(k, x^k) - f(i_{m-1}, x^{i_{m-1}}))^+}{\zeta}\right) & \text{falls } k \in N_{i_{m-1}} \\ 0 & \text{falls } k \in S \setminus N_{i_{m-1}} \end{cases}$$

Dabei bezeichnet $N_{i_{m-1}}$ die Nachbarschaft von $i_{m-1} \in S$. Die stetigen Parameter $x^{i_j} \in \mathbb{R}^n$ für $j = 0, \dots, m-1$ sowie $x^k \in \mathbb{R}^n$ sind die jeweils aktuellen stetigen Parameter. Bei der Wahl eines neuen Parameters $i \in S$ zum Zeitpunkt $j \in \mathbb{N}$ wird ein zugehöriger stetiger Parameter $x_j^i \in \mathbb{R}^n$ bestimmt mit $f(i, x_j^i) \leq f(i, x_{j-1}^i)$. Deshalb hängt die Wahl eines neuen Punktes $k \in S$ von der Vergangenheit $(i_0, \dots, i_{m-1}) \in \bigotimes_{n=0}^{m-1} S_n$ ab.

Mathematisch wird also der diskrete Teil des Optimierungs-Algorithmus modelliert durch einen zeitdiskreten stochastischen Prozess $(X_n)_{n \in \mathbb{N}_0}$ mit Zustandsraum S , dem Suchraum des Optimierungsproblems. Dieser stochastische Prozess ist definiert auf dem Raum der unendlichen Folgen mit Werten in S . Das Wahrscheinlichkeitsmaß P bestimmt die Familie der endlich-dimensionalen Verteilungen von $(X_n)_{n \in \mathbb{N}_0}$. Der Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) ist so konstruiert, dass die endlich-dimensionalen Verteilungen des Prozesses den Übergangswahrscheinlichkeiten im Optimierungs-Algorithmus entsprechen.

4.4 Formulierung des Algorithmus

4.4.1 Startdaten

Zur Initialisierung wird für jeden diskreten Parameter $i \in S$ ein $x_0^i \in \mathbb{R}^n$ festgelegt. Der zugehörige Funktionswert $f_{x_0}(i)$ ergibt sich zu $f_{x_0}(i) := f(i, x_0^i)$. Es

wird ein erster diskreter Parameter $i_0 \in S$ gewählt. Das globale Minimum \min von f wird mit $\min := f(i_0, x_0^i)$ vorbesetzt.

Außerdem werden die Kontrollparameter ζ des diskreten Metropolis Algorithmus sowie ε und δ des kontinuierlichen Teils des Algorithmus bestimmt. Die Festlegung dieser Parameter ist vom konkreten Problem abhängig und hat wesentlichen Einfluß auf den Ablauf des Algorithmus.

Für eine vorgegebene Anzahl von Iterationen werden die folgende Schritte ausgeführt. Wir setzen $j := 0$; j zählt die Iterationen.

4.4.2 Bestimmung des diskreten Parameters

Die jeweiligen kontinuierlichen Parameter werden festgehalten, das heißt, wir betrachten im j -ten Iterationsschritt die Funktion $f_{x_j} : S \rightarrow \mathbb{R}$, $i \mapsto f(i, x_j^i)$.

Es sei $i_j \in S$ der j -te diskrete Parameter und N_{i_j} die Menge seiner Nachbarn. Zur Bestimmung von i_{j+1} werden die folgenden Schritte ausgeführt.

Generierung

Zur Generierung eines neuen diskreten Parameters wird eine in N_{i_j} gleichverteilte Zufallsvariable realisiert. Um dies auf einem Computer zu simulieren, werden die Elemente von N_{i_j} numeriert: $N_{i_j} = \{k_1, k_2, \dots, k_{|N_{i_j}|}\}$. Das Einheitsintervall $[0, 1]$ wird in $|N_{i_j}|$ gleichlange Teilintervalle $I_1, I_2, \dots, I_{|N_{i_j}|}$ unterteilt. Nun wird mit einem Pseudozufallszahlengenerator eine in $[0, 1]$ stetig gleichverteilte Zufallszahl $r_1 \in [0, 1]$ erzeugt. Ist $r_1 \in I_n$ für ein $n \in \{1, 2, \dots, |N_{i_j}|\}$, so gilt k_n als neu generierter diskreter Parameter. Wir setzen $k := k_n$.

Akzeptanz

Es wird nun überprüft, ob $i_{j+1} := k$ gesetzt wird. Falls mit den jeweiligen stetigen Parametern $f_{x_j}(k) \leq f_{x_j}(i_j)$ gilt, so wird k als neuer diskreter Parameter akzeptiert: $i_{j+1} := k$. Andernfalls wird eine in $[0, 1]$ stetig gleichverteilte Zufallszahl $r_2 \in [0, 1]$ erzeugt. Gilt dann $\exp\left(-\frac{f_{x_j}(k) - f_{x_j}(i_j)}{\zeta}\right) > r_2$ so wird k ebenfalls akzeptiert und $i_{j+1} := k$ gesetzt. Andernfalls wird ein neues Element von N_{i_j} generiert und auf Akzeptanz überprüft.

Das Akzeptanzkriterium begünstigt Verkleinerungen des Zielfunktionswerts, ermöglicht aber auch Verschlechterungen. Dabei sind geringe Verschlechterungen wahrscheinlicher als größere.

4.4.3 Berechnung des kontinuierlichen Parameters

Der diskrete Parameter $i := i_{j+1}$ wird jetzt festgehalten, der zugehörige aktuelle kontinuierliche Parameter sei x_j^i . Es soll die Zielfunktion $f^i : \mathbb{R}^n \rightarrow \mathbb{R}$, $x \mapsto f(i, x)$ minimiert werden. Dazu wird die stochastische Integralgleichung

$$X_{t_j^i + \sigma}(\omega) = x_j^i + \varepsilon(B_{t_j^i + \sigma}(\omega) - B_{t_j^i}(\omega)) - \int_{t_j^i}^{t_j^i + \sigma} \nabla f^i(X_\tau(\omega)) d\tau \quad (*)$$

betrachtet. Dabei legt $\omega \in \Omega$ den Pfad fest, der nun numerisch mit dem semi-impliziten Euler-Verfahren approximiert wird. Die Schrittweite wird mit σ bezeichnet. t_j^i ist die Summe der Integrationsschrittweiten der numerischen Lösung von $(SIE)_i$ in den ersten j Iterationen. Der neue zu i gehörende kontinuierliche Parameter x_{j+1}^i ist das Ergebnis eines Integrationsschrittes.

Eine Approximation \bar{X} der Lösung von $(*)$ mit dem semi-impliziten Euler-Verfahren ergibt sich zu

$$\bar{X}_{t_j^i + \sigma}(\omega) = x_j^i - (I + \sigma \nabla^2 f^i(x_j^i))^{-1} (\sigma \nabla f^i(x_j^i) - \varepsilon(B_{t_j^i + \sigma}(\omega) - B_{t_j^i}(\omega))) \quad (**),$$

wobei I die Einheitsmatrix bezeichnet.

Simulation

Im semi-impliziten Euler-Verfahren in $(**)$ kommt ω nur als Argument der Brownschen Bewegung vor. Da $B_{t_j^i + \sigma} - B_{t_j^i}$ nach $\mathcal{N}(0, \sigma I)$ normalverteilt ist, also $\frac{1}{\sqrt{\sigma}}(B_{t_j^i + \sigma} - B_{t_j^i})$ nach $\mathcal{N}(0, I)$, wird $B_{t_j^i + \sigma}(\omega) - B_{t_j^i}(\omega)$ durch die Simulation von nach $\mathcal{N}(0, I)$ normalverteilten Zufallsvektoren berechnet.

Im nächsten Schritt werden drei Punkte auf dem von ω bestimmten Pfad der Lösung von $(*)$ approximiert, und zwar $x_{j+1}^i(\sigma)$ mit Schrittweite σ und Startpunkt x_j^i , $x_{j+1}^i(\frac{\sigma}{2})$ mit Schrittweite $\frac{\sigma}{2}$ und Startpunkt x_j^i sowie $\bar{x}_{j+1}^i(\sigma)$ mit Schrittweite $\frac{\sigma}{2}$ und Startpunkt $x_{j+1}^i(\frac{\sigma}{2})$.

Wegen der Unabhängigkeit der Zuwächse einer Brownschen Bewegung können $\sqrt{\frac{2}{\sigma}}(B_{t_j^i + \frac{\sigma}{2}}(\omega) - B_{t_j^i}(\omega))$ und $\sqrt{\frac{2}{\sigma}}(B_{t_j^i + \sigma}(\omega) - B_{t_j^i + \frac{\sigma}{2}}(\omega))$ durch zwei unabhängige Realisierungen n_1 und n_2 einer nach $\mathcal{N}(0, I)$ normalverteilten Zufallsvariablen simuliert werden. Dann ist $n_1 + n_2$ die Realisierung von $\sqrt{\frac{2}{\sigma}}(B_{t_j^i + \sigma} - B_{t_j^i})$ mit dem selben ω , das heißt alle drei Punkte liegen auf dem selben Pfad.

Berechnung von drei Punkten

Es werden drei Punkte auf dem von ω bestimmten Pfad mit dem semi-impliziten Eulerverfahren berechnet. Die auftretenden Ableitungen können durch automatisches Differenzieren berechnet werden. Wir setzen $\sigma := 1$.

$x_{j+1}^i(\sigma)$ ergibt sich mit $L_1 := I + \sigma \nabla^2 f^i(x_j^i)$ und der Lösung y_1 des linearen Gleichungssystems $L_1 y_1 = \sigma \nabla f^i(x_j^i) - \varepsilon \sqrt{\frac{\sigma}{2}}(n_1 + n_2)$ zu $x_{j+1}^i(\sigma) := x_j^i - y_1$. Die Schrittweite σ wird dabei solange halbiert, bis L_1 positiv definit ist. Dies ist stets möglich.

$x_{j+1}^i(\frac{\sigma}{2})$ ergibt sich mit $L_2 := I + \frac{\sigma}{2} \nabla^2 f^i(x_j^i)$ und der Lösung y_2 des linearen Gleichungssystems $L_2 y_2 = \frac{\sigma}{2} \nabla f^i(x_j^i) - \varepsilon \sqrt{\frac{\sigma}{2}} n_1$ zu $x_{j+1}^i(\frac{\sigma}{2}) := x_j^i - y_2$. L_2 ist offensichtlich positiv definit, da L_1 positiv definit ist.

Falls $L_3 := I + \frac{\sigma}{2} \nabla^2 f(x_{j+1}^i(\frac{\sigma}{2}))$ positiv definit ist, so wird $\bar{x}_{j+1}^i(\sigma)$ aus der Lösung y_3 des linearen Gleichungssystems $L_3 y_3 = \frac{\sigma}{2} \nabla f(x_{j+1}^i(\frac{\sigma}{2})) - \varepsilon \sqrt{\frac{\sigma}{2}} n_2$ berechnet: $\bar{x}_{j+1}^i(\sigma) := x_{j+1}^i(\frac{\sigma}{2}) - y_3$. Andernfalls wird die Schrittweite halbiert, das heißt $\sigma := \frac{\sigma}{2}$, und mit der neuen Schrittweite werden alle drei Punkte nochmals berechnet.

Die linearen Gleichungssysteme werden mit Hilfe der Cholesky-Zerlegung gelöst, siehe [14]. Dies ist naheliegend, da die auftretenden Matrizen symmetrisch und positiv definit sind.

Akzeptanz-Kriterium

Es wird überprüft, ob $\|x_{j+1}^i(\sigma) - \bar{x}_{j+1}^i(\sigma)\|_2 \leq \delta$ gilt. Ist dies der Fall, so setzen wir $x_{j+1}^i := \bar{x}_{j+1}^i(\sigma)$, andernfalls wird die Schrittweite halbiert und mit der neuen Schrittweite werden $x_{j+1}^i(\sigma)$, $x_{j+1}^i(\frac{\sigma}{2})$ und $\bar{x}_{j+1}^i(\sigma)$ neu berechnet.

Das Akzeptanz-Kriterium soll sicherstellen, dass der neue stetige Parameter x_{j+1}^i in der Nähe des von ω bestimmten Pfades liegt.

Bestimmung des Funktionswertes

Falls $f(i_{j+1}, x_{j+1}^i) \leq f(i_{j+1}, x_j^i)$ gilt, so wird der neue zu i_{j+1} gehörende Funktionswert festgesetzt als $f_{x_{j+1}}(i_{j+1}) := f(i_{j+1}, x_{j+1}^i)$, andernfalls wird der alte Funktionswert beibehalten: $f_{x_{j+1}}(i_{j+1}) := f(i_{j+1}, x_j^i)$. Somit wird sichergestellt, dass wir zu dem diskreten Parameter i_{j+1} eine monoton fallende Folge von Ziel-funktionswerten erhalten.

Startpunkt für einen nächsten Schritt in der zu i_{j+1} gehörenden Differentialgleichung ist jedoch x_{j+1}^i . Damit ist die Folge der stetigen Parameter zu i_{j+1} genau die Folge, die sich im Verfahren von Schäffler angewandt auf die Integralgleichung

$$X_t(\omega) = x_0^i + \varepsilon(B_t(\omega) - B_0(\omega)) - \int_0^t \nabla f^i(X_\tau(\omega)) d\tau$$

ergibt.

4.4.4 Ausgabedaten

Falls für den aktuellen Funktionswert $f(i_{j+1}, x_{j+1}^i)$ gilt $f(i_{j+1}, x_{j+1}^i) \leq \min$, so wird $\min := f(i_{j+1}, x_{j+1}^i)$ gesetzt und das Tripel $(\min, i_{j+1}, x_{j+1}^i)$ ausgegeben.

Neuer Iterationsschritt

Wir setzen $j := j+1$ und gehen zurück zur Bestimmung des diskreten Parameters.

4.5 Analyse des Verfahrens

Wird jedem diskreten Parameter $i \in S$ ein beliebiger, aber fester Parameter $x^i \in \mathbb{R}^n$ und damit ein fester Funktionswert $f(i) := f(i, x^i)$ zugewiesen, so wird der Metropolis-Algorithmus, der ein Minimum der Funktion $f : S \rightarrow \mathbb{R}, i \mapsto f(i)$ berechnet, durch eine homogene Markov-Kette beschrieben. Die i -Komponente der stationären Verteilung dieser Markov-Kette ist gegeben durch

$$\frac{1}{\mu_i} = \frac{\exp\left(-\frac{f(i, x^i)}{\zeta}\right)}{\sum_{k \in S} \exp\left(-\frac{f(k, x^k)}{\zeta}\right)}.$$

Lemma 4.5.1 *Sei $i \in S$ beliebig aber fest gewählt. Zu $i \in S$ sei x_0^i als kontinuierlicher Parameter gewählt, zu $k \in S \setminus \{i\}$ das jeweilige globale Optimum, das heißt x^{k*} mit $f(k, x^{k*}) = \min_{x \in \mathbb{R}^n} f(k, x)$. Die i -Komponente der stationären Verteilung der Markovkette des entsprechenden Metropolis-Algorithmus wird mit $\frac{1}{\mu_i^*}$ bezeichnet. Dann gilt für jede andere Wahl von stetigen Parametern und damit Funktionswerten*

$$\frac{1}{\mu_i} \geq \frac{1}{\mu_i^*}.$$

BEWEIS Es gilt

$$\exp\left(-\frac{f(i, x_j^i)}{\zeta}\right) \sum_{\substack{k \in S \\ k \neq i}} \exp\left(-\frac{f(k, x^{k*})}{\zeta}\right) \geq \exp\left(-\frac{f(i, x_0^i)}{\zeta}\right) \sum_{\substack{k \in S \\ k \neq i}} \exp\left(-\frac{f(k, x_j^k)}{\zeta}\right)$$

für alle $j \in \mathbb{N}_0$. Damit folgt

$$\exp\left(-\frac{f(i, x_j^i)}{\zeta}\right) \sum_{\substack{k \in S \\ k \neq i}} \exp\left(-\frac{f(k, x^{k*})}{\zeta}\right) + \exp\left(-\frac{f(i, x_j^i)}{\zeta}\right) \exp\left(-\frac{f(i, x_0^i)}{\zeta}\right) \geq$$

$$\geq \exp\left(-\frac{f(i, x_0^i)}{\zeta}\right) \sum_{\substack{k \in S \\ k \neq i}} \exp\left(-\frac{f(k, x_j^k)}{\zeta}\right) + \exp\left(-\frac{f(i, x_j^i)}{\zeta}\right) \exp\left(-\frac{f(i, x_0^i)}{\zeta}\right)$$

und somit

$$\frac{\exp\left(-\frac{f(i, x_j^i)}{\zeta}\right)}{\sum_{k \in S} \exp\left(-\frac{f(k, x_j^k)}{\zeta}\right)} \geq \frac{\exp\left(-\frac{f(i, x_0^i)}{\zeta}\right)}{\sum_{\substack{k \in S \\ k \neq i}} \exp\left(-\frac{f(k, x_j^k)}{\zeta}\right) + \exp\left(-\frac{f(i, x_0^i)}{\zeta}\right)}$$

wie behauptet. ◇

In der folgenden Definition fassen wir einige Bezeichnungen zusammen.

Definition 4.5.2 Seien $i \in S$ und $j \in \mathbb{N}_0$. Mit $i^{(j)}$ bezeichnen wir das j -te Auftreten von i in unserem diskreten stochastischen Prozess $(X_m)_{m \in \mathbb{N}_0}$.

Zu $i \in S$ definieren wir $I := \{i^{(j)} | j \in \mathbb{N}_0\}$ sowie

$$p_{kI} := P(X_{m+1} \in I | X_m = k)$$

für $k \in S$, $m \in \mathbb{N}_0$.

Bemerkungen 4.5.3 (i) Da der Funktionswert bei jedem erneuten Auftreten von $i \in S$ überprüft und gegebenenfalls verbessert wird, ist die Angabe von j in $i^{(j)}$ für die Übergangswahrscheinlichkeiten unseres diskreten stochastischen Prozesses entscheidend.

(ii) Soll nur untersucht werden, wie wahrscheinlich $i \in S$ von $k \in S$ aus erreicht wird, so wird p_{kI} betrachtet. Offensichtlich gilt

$$p_{ki^{(j)}} \leq p_{kI} \quad \text{für alle } j \in \mathbb{N}_0.$$

Lemma 4.5.4 *Zu jedem $i \in S$ werden fast sicher nur endlich viele verschiedene Funktionswerte bestimmt.*

BEWEIS Wir verwenden die Bezeichnungen aus Kapitel 3. Seien $i \in S$ und $\rho > 0$ beliebig aber fest. Es sei x^{i*} das globale Optimum der Funktion $f^i : \mathbb{R}^n \rightarrow \mathbb{R}$, $x \mapsto f(i, x)$. Nach Satz 3.2.2 gilt für die Stoppzeiten $s_{x^{i*}, \rho}(\omega) : \Omega^n \rightarrow \bar{\mathbb{R}}$,

$$\omega \mapsto \begin{cases} \inf\{t \geq 0 | \|X_t(\omega) - x^{i*}\|_2 \leq \rho\} & \text{falls } \{t \geq 0 | \|X_t(\omega) - x^{i*}\|_2 \leq \rho\} \neq \emptyset, \\ \infty & \text{sonst} \end{cases}$$

folgendes:

$$W^n(\{\omega \in \Omega^n | s_{x^{i*}, \rho}(\omega) < \infty\}) = 1 \quad \text{und} \quad E(s_{x^{i*}, \rho}(\omega) < \infty) < \infty.$$

Dies bedeutet, dass W^n -fast sicher jeder Pfad der Lösung der stochastischen Differentialgleichung $(SIE)_i$ in endlicher Zeit eine ρ -Umgebung des globalen Optimums von f^i erreicht für beliebig vorgegebenes $\rho > 0$. Stellen wir sicher, dass in unserem Algorithmus für die Integrationsschrittweite σ gilt $\sigma \geq c$ für eine Konstante $c > 0$, so reichen hierfür endlich viele Integrations Schritte und somit werden erst recht nur endlich viele Funktionswerte bestimmt.

◇

Definition 4.5.5 Es seien $i, k \in S$ und $l, m \in \mathbb{N}_0$, $n \in \mathbb{N}$. Mit $g_{i^{(m)}k^{(l)}}^{(n)}$ bezeichnen wir die Wahrscheinlichkeit, dass $k^{(l)}$ in genau n Schritten von $i^{(m)}$ aus erreicht wird. Mit $g_{i^{(m)}i^{(m+1)}}^{(n)}$ wird die Wahrscheinlichkeit bezeichnet, dass $i^{(m+1)}$ in genau n Schritten zum $m+1$ -ten mal erreicht wird, falls i zuvor m -mal erreicht wurde. Wir definieren

$$g_{i^{(m)}k^{(l)}} := \sum_{n=1}^{\infty} g_{i^{(m)}k^{(l)}}^{(n)}.$$

$$g_{i^{(m)}i^{(m+1)}} := \sum_{n=1}^{\infty} g_{i^{(m)}i^{(m+1)}}^{(n)}.$$

i heißt **rekurrent**, falls $g_{i^{(m)}i^{(m+1)}} = 1$ gilt für alle $m \in \mathbb{N}_0$.

Bemerkungen 4.5.6 (i) Offensichtlich ist $g_{i^{(m)}i^{(m+1)}}$ die Wahrscheinlichkeit, dass $i^{(m+1)}$ jemals von $i^{(m)}$ aus erreicht wird. Folglich gilt $g_{i^{(m)}i^{(m+1)}} \leq 1$.

(ii) Definition 4.5.5 ist ein direktes Analogon zur Definition eines rekurrenten Zustands einer homogenen Markov-Kette.

Lemma 4.5.7 Seien $i \in S$ und $m \in \mathbb{N}_0$. $p^{(n)}$ bezeichne die n -Schritt Übergangswahrscheinlichkeit des stochastischen Prozesses mit Zustandsraum S . Dann gilt mit $I := \{i^{(j)} \mid j \in \mathbb{N}_0\}$

$$g_{i^{(m)}i^{(m+1)}} < 1 \quad \implies \quad \sum_{n=1}^{\infty} p_{i^{(m)}I}^{(n)} < \infty.$$

BEWEIS Es gilt für $k \in \mathbb{N}_0$

$$p_{i^{(m)}I}^{(n)} =$$

$$= \sum_{s=0}^{n-1} P(X_{k+1} \notin I, \dots, X_{k+n-s-1} \notin I, X_{k+n-s} = i^{(m+1)}, X_{k+n} \in I \mid X_k = i^{(m)}) =$$

$$= \sum_{s=0}^{n-1} P(X_{k+1} \notin I, \dots, X_{k+n-s-1} \notin I, X_{k+n-s} = i^{(m+1)} \mid X_k = i^{(m)}) \cdot$$

$$\cdot P(X_{k+n} \in I \mid X_{k+n-s} = i^{(m+1)}) =$$

$$= \sum_{s=0}^{n-1} g_{i^{(m)}i^{(m+1)}}^{(n-s)} p_{i^{(m)}I}^{(s)}.$$

Damit folgt

$$\begin{aligned} \sum_{t=1}^n p_{i^{(m)}I}^{(t)} &= \sum_{t=1}^n \sum_{s=0}^{t-1} g_{i^{(m)}i^{(m+1)}}^{(t-s)} p_{i^{(m)}I}^{(s)} = \sum_{s=0}^{n-1} p_{i^{(m)}I}^{(s)} \sum_{t=s+1}^n g_{i^{(m)}i^{(m+1)}}^{(t-s)} \leq \\ &\leq \sum_{s=0}^n p_{i^{(m)}I}^{(s)} g_{i^{(m)}i^{(m+1)}}. \end{aligned}$$

Wir erhalten

$$\sum_{t=1}^n p_{i^{(m)}I}^{(t)} (1 - g_{i^{(m)}i^{(m+1)}}) \leq g_{i^{(m)}i^{(m+1)}}.$$

Falls $g_{i^{(m)}i^{(m+1)}} < 1$ ist, so sind die Partialsummen der Reihe beschränkt.

◇

Lemma 4.5.8 *Es sei $i \in S$. Dann ist i rekurrent.*

BEWEIS Sei $i \in S$. Nach Lemma 4.5.1 gilt für beliebige aber feste kontinuierliche Parameter

$$\lim_{n \rightarrow \infty} p_{i^{(m)}i^{(m)}}^{(n)} = \frac{1}{\mu_i} \geq \frac{1}{\mu_i^*}$$

für alle $m \in \mathbb{N}_0$. Dabei sei wie in Lemma 4.5.1 $\frac{1}{\mu_i^*}$ die i -Komponente der stationären Verteilung des Metropolis-Algorithmus, bei dem als kontinuierlicher Parameter zu i x_0^i gewählt wurde, zu $k \in S \setminus \{i\}$ das jeweilige globale Optimum. Die zugehörigen Übergangswahrscheinlichkeiten werden mit $p_{i^{(0^*)i^{(0^*)}}}^{(n)}$ bezeichnet. Es gibt für jede feste Wahl von kontinuierlichen Parametern ein $n_0 \in \mathbb{N}$ mit $p_{i^{(m)}i^{(m)}}^{(n)} \geq p_{i^{(0^*)i^{(0^*)}}}^{(n)}$ für alle $n \in \mathbb{N}$ mit $n \geq n_0$. Da zu jedem $i \in S$ nur endlich viele verschiedene Funktionswerte angenommen werden, gibt es nur endlich viele verschiedene Werte für $p_{i^{(m)}i^{(m)}}^{(n)}$. Somit kann ein maximales $N_0 \in \mathbb{N}$ bestimmt werden, so dass für jede feste Wahl von Parametern gilt $p_{i^{(m)}i^{(m)}}^{(n)} \geq p_{i^{(0^*)i^{(0^*)}}}^{(n)}$ für alle $n \geq N_0$.

$(p_{i^{(0^*)i^{(0^*)}}}^{(n)})_{n \in \mathbb{N}_0}$ ist keine Nullfolge, also divergiert $\sum_{n=1}^{\infty} p_{i^{(0^*)i^{(0^*)}}}^{(n)}$ bestimmt gegen ∞ . Folglich divergiert auch $\sum_{n=1}^{\infty} p_{i^{(m)}i^{(m)}}^{(n)}$, da $\sum_{n=1}^{\infty} p_{i^{(0^*)i^{(0^*)}}}^{(n)}$ eine divergente Minorante bildet. Nach obigen Überlegungen divergiert $\sum_{n=1}^{\infty} p_{i^{(m)}i^{(m)}}^{(n)}$ auch bei einem Wechsel der stetigen Parameter in jedem Summanden.

Mit $I = \{i^{(k)} | k \in \mathbb{N}_0\}$ gilt gemäß Lemma 4.5.7

$$g_{i^{(m)}i^{(m+1)}} < 1 \quad \implies \quad \sum_{n=1}^{\infty} p_{i^{(m)}I}^{(n)} < \infty.$$

$$\begin{aligned}
&= g_{ki^{(0)}}^{(n_1)} g_{i^{(0)}i^{(1)}}^{(n_2-n_1)} \cdot \dots \cdot g_{i^{(j-2)}i^{(j-1)}}^{(n_j-n_{j-1})} \cdot P(X_{n_{j+1}} \notin i, \dots, X_{n_{j+1}} = i^{(j)} | X_{n_j} = i^{(j-1)}) = \\
&= g_{ki^{(0)}}^{(n_1)} g_{i^{(0)}i^{(1)}}^{(n_2-n_1)} \cdot \dots \cdot g_{i^{(j-1)}i^{(j)}}^{(n_{j+1}-n_j)}.
\end{aligned}$$

Damit ist der Induktionsbeweis abgeschlossen. Addition über alle j -Tupel n_1, \dots, n_j liefert, dass die Wahrscheinlichkeit, dass $X_n \in I$ für mindestens j verschiedene Werte von n gilt, gleich $g_{ki^{(0)}} g_{i^{(0)}i^{(1)}} \cdot \dots \cdot g_{i^{(j-2)}i^{(j-1)}} = g_{ki^{(0)}}$ ist. Für $j \rightarrow \infty$ folgt die Behauptung.

◇

Lemma 4.5.10 Für alle $i, k \in S$ und alle $j, l \in \mathbb{N}_0$ gelten

$$g_{i^{(j)}k^{(l)}} = 1 \quad \text{und} \quad g_{k^{(l)}i^{(j)}} = 1.$$

BEWEIS Aufgrund der Irreduzibilität der Nachbarschaftsstruktur ist $k^{(l)}$ von $i^{(j)}$ aus erreichbar, und dies ist ohne eine vorherige Rückkehr zu $i^{(j)}$ möglich. Die Wahrscheinlichkeit dieses Ereignisses sei $\gamma > 0$. Ist $k^{(l)}$ einmal erreicht, so ist die Wahrscheinlichkeit, $i^{(j+1)}$ nie zu erreichen, größer oder gleich $1 - \max_{n \geq l} g_{k^{(n)}i^{(j+1)}}$. Die Wahrscheinlichkeit, von $i^{(j)}$ aus nie zu $i^{(j+1)}$ zu kommen, ist also mindestens $\gamma(1 - \max_{n \geq l} g_{k^{(n)}i^{(j+1)}})$. Da i rekurrent ist, ist die Wahrscheinlichkeit keiner Rückkehr Null. Folglich gilt

$$g_{k^{(n_{max})}i^{(j+1)}} := \max_{n \geq l} g_{k^{(n)}i^{(j+1)}} = 1.$$

Da k rekurrent ist, wird $k^{(n_{max})}$ von $k^{(l)}$ aus mit Wahrscheinlichkeit 1 erreicht und somit gilt

$$g_{k^{(l)}i^{(m+1)}} \geq g_{k^{(l)}k^{(n_{max})}} g_{k^{(n_{max})}i^{(m+1)}} = 1.$$

Die Rollen von i und k können vertauscht werden und damit folgt die Behauptung.

◇

In Lemma 4.5.1 wurde gezeigt, dass sich die i -Komponente der stationären Verteilung für jeden diskreten Parameter $i \in S$ abschätzen lässt durch einen für i "schlechtesten Fall", bei dem i der schlechteste Funktionswert zugeordnet wird, jedem $k \in S \setminus \{i\}$ aber der optimale Funktionswert. Damit und da für jedes $i \in S$ nur endlich viele verschiedene Funktionswerte angenommen werden (Lemma 4.5.4), konnte in Lemma 4.5.8 gezeigt werden, dass alle $i \in S$ im Sinne von Definition 4.5.5 rekurrent sind. Lemma 4.5.9 und Lemma 4.5.10 schließlich besagen zusammen, dass jeder Punkt $i \in S$ von jedem anderen Punkt $k \in S$ aus mit Wahrscheinlichkeit eins erreicht wird und anschließend unendlich oft angenommen wird.

Bei jedem Erreichen eines diskreten Parameters $i \in S$ wird ein Integrationsschritt zur Approximation eines Pfades der Lösung der zugehörigen Integralgleichung

$(SIE)_i$ ausgeführt. Jeder Pfad der Lösung von $(SIE)_i$ trifft fast sicher in endlicher Zeit eine beliebig vorgegebene Umgebung des globalen Minimierers x^{i*} der zugehörigen Zielfunktion f^i . Somit wird schließlich zu jedem $i \in S$ eine Approximation von x^{i*} gefunden, von der aus ein Algorithmus zur lokalen Minimierung von f^i das globale Optimum findet.

Kapitel 5

Testergebnisse

Der in Kapitel 4 entwickelte Algorithmus wird nun an einem einfachen Beispiel getestet. Wir betrachten für die Zielfunktion

$$f : \{2, 6, 12\} \times \mathbb{R} \rightarrow \mathbb{R}, \quad (i, x) \mapsto \frac{11}{12} + \frac{1}{i} + i \cdot x^2 - \cos(2ix)$$

das globale Optimierungsproblem

$$\min_{i \in \{2, 6, 12\}, x \in \mathbb{R}} f(i, x). \quad (GOP)$$

Für $x \in \mathbb{R} \setminus \{0\}$ gilt

$$f(i, x) \geq \frac{11}{12} + \frac{1}{i} + i \cdot x^2 - 1 = \frac{1}{i} - \frac{1}{12} + i \cdot x^2 > \frac{1}{i} - \frac{1}{12} = f(i, 0),$$

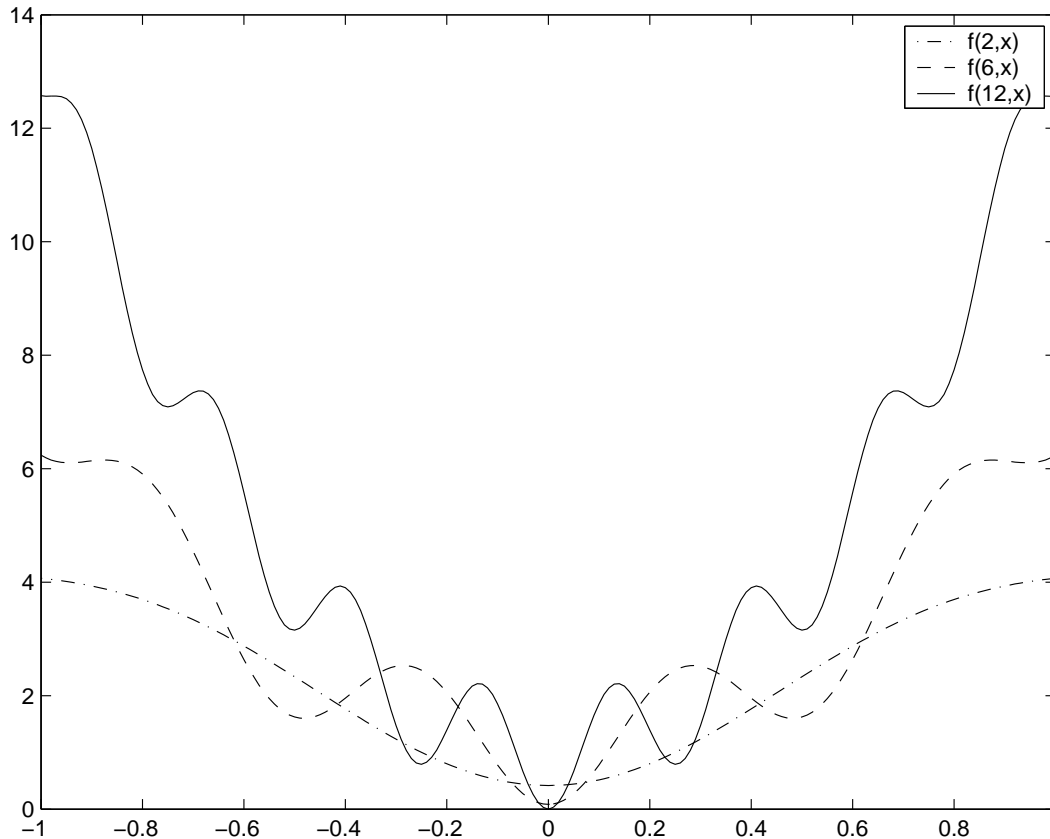
das heißt für $i \in \{2, 6, 12\}$ gilt

$$\min_{x \in \mathbb{R}} f^i(x) = f^i(0) = \frac{1}{i} - \frac{1}{12}.$$

Somit ist die Lösung von (GOP)

$$\min_{i \in \{2, 6, 12\}, x \in \mathbb{R}} f(i, x) = f(12, 0) = 0.$$

Die folgende Abbildung zeigt den Graphen der Zielfunktion f .

Abbildung 5.1: Graph der Zielfunktion f

Der in Abschnitt 4.4 beschriebene Algorithmus wurde mit MATLAB Version 6.1, Release 12.1 [8] programmiert. Es wurde der in MATLAB implementierte Zufallszahlengenerator verwendet. Als Nachbarschaft wurde für 2, 6, 12 jeweils $\{2, 6, 12\}$ gewählt.

Anschließend folgen Abbildungen, die Testläufe des Algorithmus zeigen. Jeder vom Algorithmus berechnete Punkt $(x, f(i, x))$ wird als Kreis in einem Koordinatensystem dargestellt. Auf der Abszisse dieses Koordinatensystems sind die kontinuierlichen Parameter x abgebildet, auf der Ordinate die zugehörigen berechneten Funktionswerte $f(i, x)$. Es wurden jeweils 200 Iterationen gewählt. Startpunkt war stets der Punkt $(2, 2)$ mit dem Funktionswert $f(2, 2) = 9.5622$.

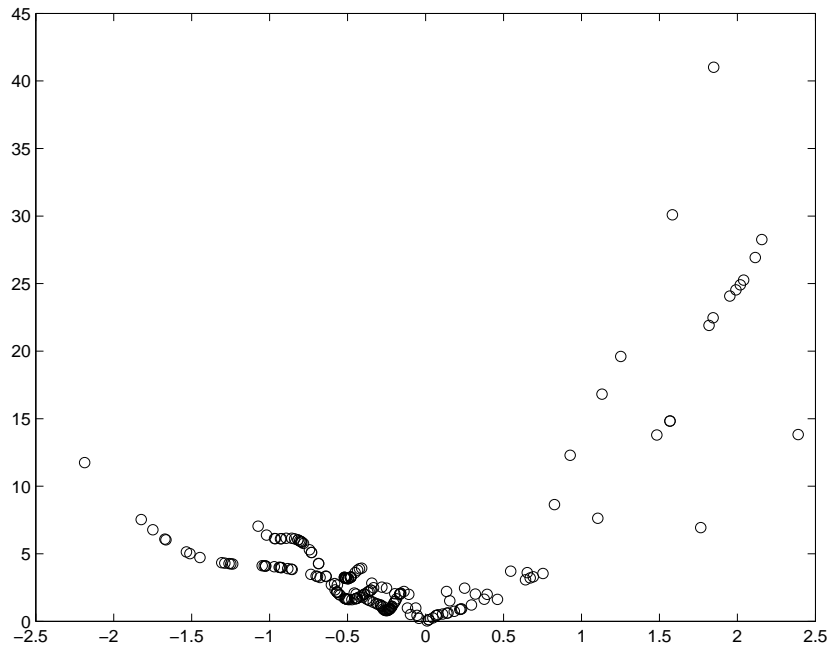


Abbildung 5.2: Testlauf des Algorithmus mit Kontrollparametern $\zeta = 50$, $\varepsilon = 4$, $\delta = 0.1$. Ergebnis: $(i, x) = (12, 0.0096)$ mit $f(i, x) = 0.0274$.

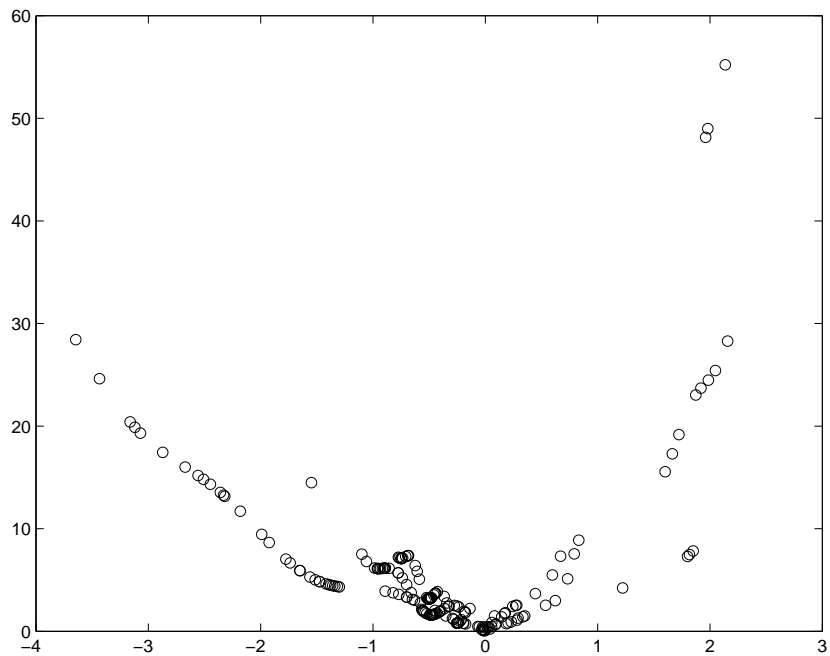


Abbildung 5.3: Testlauf des Algorithmus mit Kontrollparametern $\zeta = 50$, $\varepsilon = 4$, $\delta = 0.1$. Ergebnis: $(i, x) = (12, -0.0108)$ mit $f(i, x) = 0.0349$.

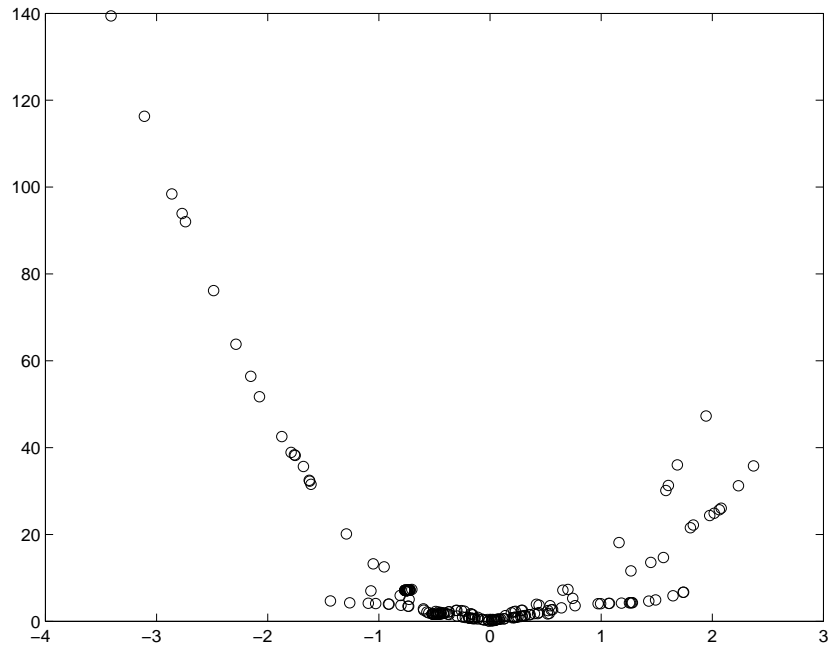


Abbildung 5.4: Testlauf des Algorithmus mit Kontrollparametern $\zeta = 50000$, $\varepsilon = 4$, $\delta = 0.1$. Ergebnis: $(i, x) = (6, -0.0089)$ mit $f(i, x) = 0.0895$.

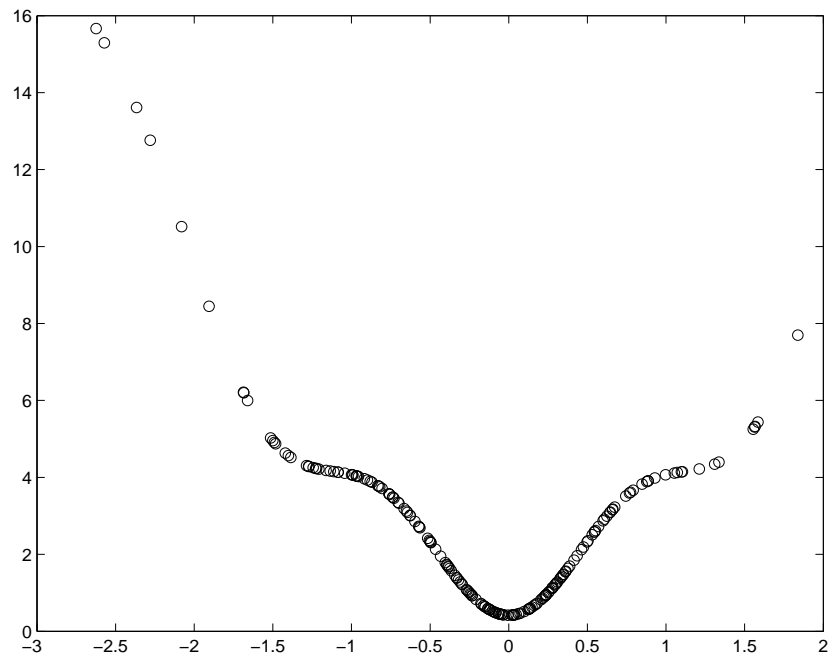


Abbildung 5.5: Testlauf des Algorithmus mit Kontrollparametern $\zeta = 1$, $\varepsilon = 4$, $\delta = 0.1$. Ergebnis: $(i, x) = (2, -0.0050)$ mit $f(i, x) = 0.4169$.

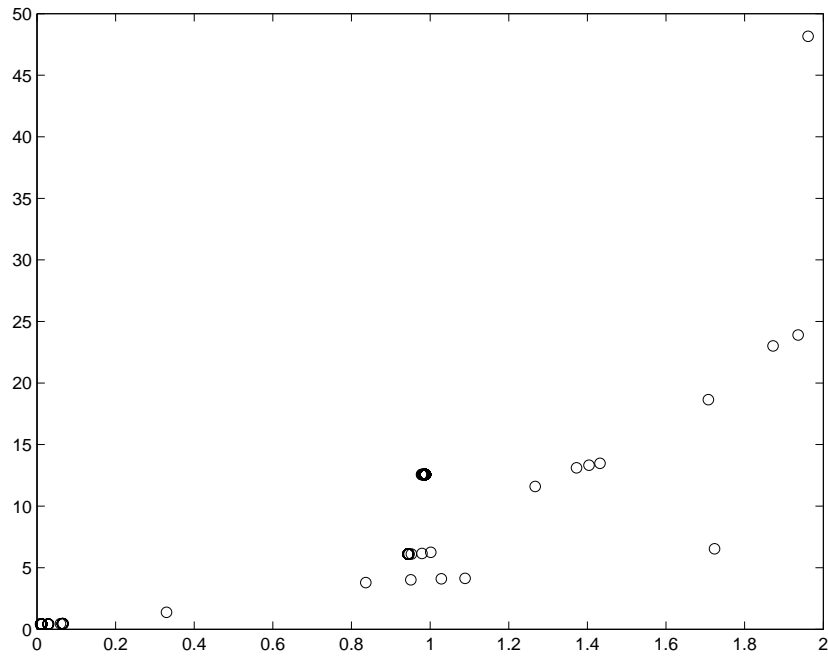


Abbildung 5.6: Testlauf des Algorithmus mit Kontrollparametern $\zeta = 50$, $\varepsilon = 0$, $\delta = 0.1$. Ergebnis: $(i, x) = (2, 0.0093)$ mit $f(i, x) = 0.4175$.

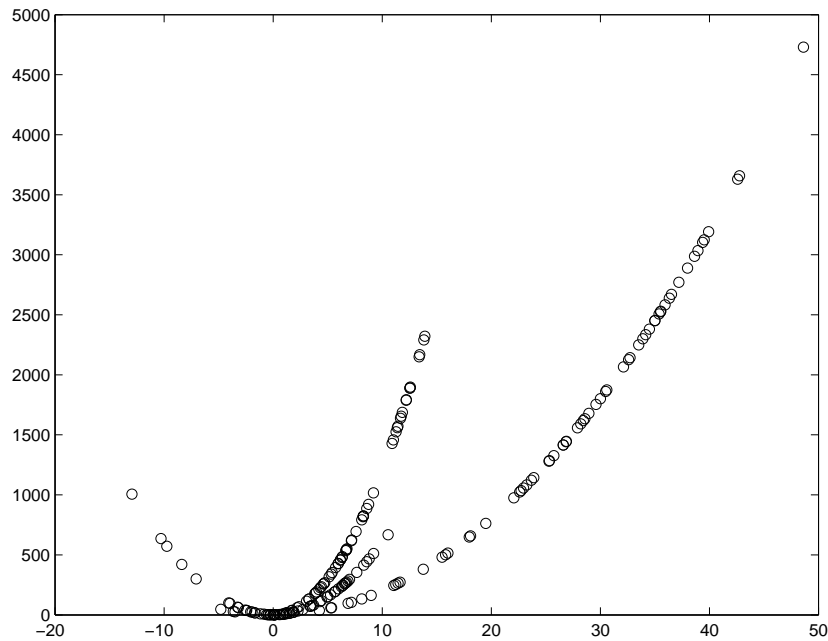


Abbildung 5.7: Testlauf des Algorithmus mit Kontrollparametern $\zeta = 50$, $\varepsilon = 100$, $\delta = 0.1$. Ergebnis: $(i, x) = (6, -0.0248)$ mit $f(i, x) = 0.1311$.

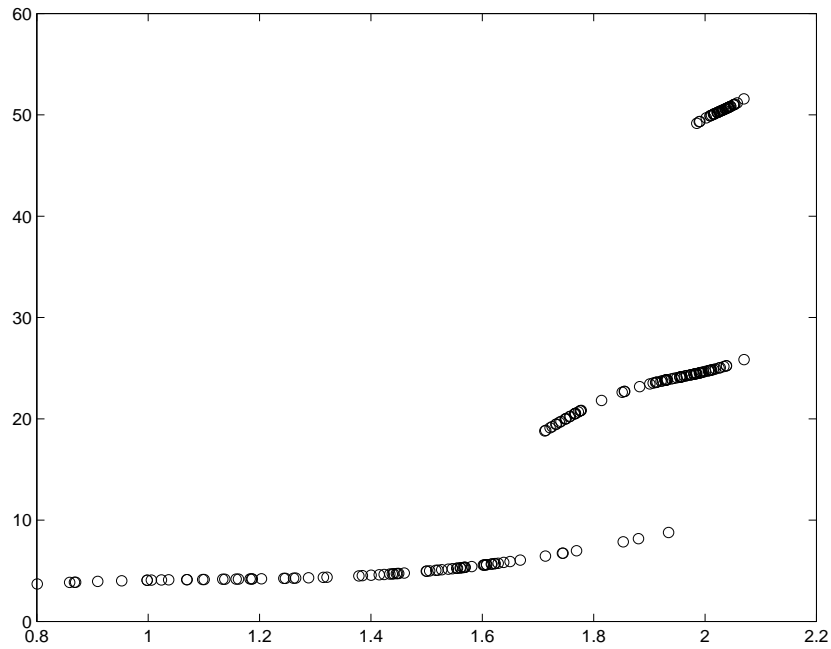


Abbildung 5.8: Testlauf des Algorithmus mit Kontrollparametern $\zeta = 50$, $\varepsilon = 4$, $\delta = 0.0001$. Ergebnis: $(i, x) = (2, 0.8008)$ mit $f(i, x) = 3.6973$.

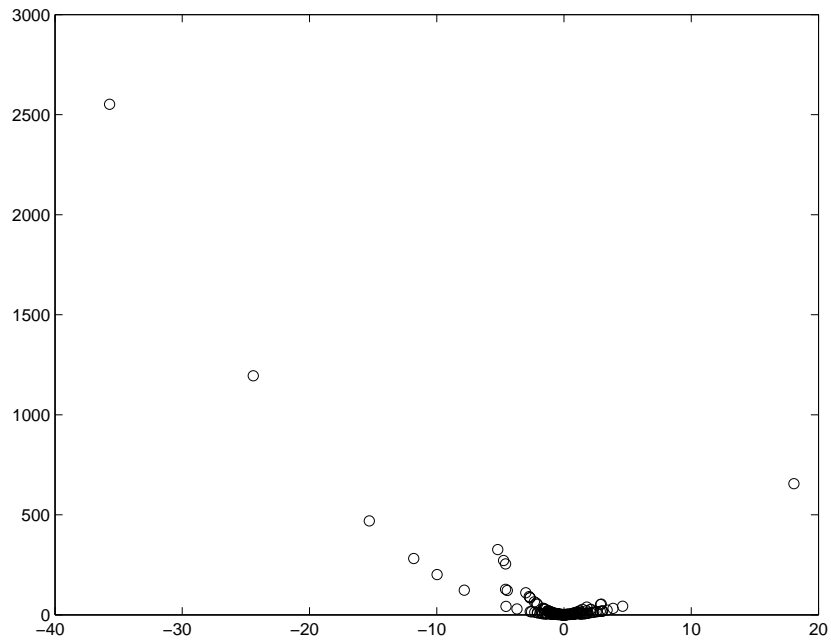


Abbildung 5.9: Testlauf des Algorithmus mit Kontrollparametern $\zeta = 50$, $\varepsilon = 4$, $\delta = 100$. Ergebnis: $(i, x) = (6, -0.0168)$ mit $f(i, x) = 0.1054$.

Abbildung 5.2 und 5.3 zeigen Testläufe des Algorithmus mit identischer Wahl der Kontrollparameter. Diese Parameterdarstellung führt in beiden Fällen zu einem guten Ergebnis, insbesondere wird der richtige diskrete Parameter 12 gefunden. Um den Zielfunktionswert weiter zu verbessern, kann noch eine lokale Minimierung von f^{12} erfolgen.

In Abbildung 5.4 wurde ζ größer gewählt, so dass Verschlechterungen des Zielfunktionswerts im diskreten Teil des Algorithmus mit großer Wahrscheinlichkeit angenommen werden. Dagegen wurde in Abbildung 5.5 ζ sehr klein gesetzt, so dass ein Wechsel des diskreten Parameters unwahrscheinlich wird.

Abbildung 5.6 zeigt einen Testlauf mit $\varepsilon = 0$, das heißt die Berechnung des kontinuierlichen Parameters ist ausschließlich eine lokale Minimierung. In Abbildung 5.7 wurde ε dagegen groß gewählt. Entsprechend verringert sich die Tendenz des Algorithmus zur Minimierung.

In Abbildung 5.8 ist δ sehr klein, das heißt das Akzeptanzkriterium für den kontinuierlichen Parameter ist sehr streng. In Abbildung 5.9 wurde δ dagegen groß gewählt und somit wurden viele berechnete Punkte akzeptiert.

Insgesamt wird deutlich, dass die Wahl der Kontrollparameter wesentlichen Einfluß auf den Ablauf des Algorithmus hat. Bei geeigneter Wahl der Kontrollparameter liefert der Algorithmus gute Ergebnisse.

Kapitel 6

Zusammenfassung und Ausblick: Anwendung in der Vektoroptimierung

In dieser Arbeit wurde ein stochastisches Verfahren zur globalen Optimierung bei diskreten und kontinuierlichen Variablen entwickelt. Aufbauend auf dem Metropolis-Algorithmus zur diskreten Optimierung und einem stochastischen Verfahren zur globalen Optimierung bei kontinuierlichen Variablen von Schäffler wurde ein Algorithmus aufgestellt, der simultan diskrete und kontinuierliche Variable behandelt. Es wurde gezeigt, dass der Algorithmus mit Wahrscheinlichkeit eins bei genügend langer Laufzeit eine beliebig genaue Approximation des gesuchten globalen Optimums berechnet.

Wir wollen nun unser Verfahren aus Kapitel 4 noch verallgemeinern und auf Vektoroptimierungsprobleme mit diskreten und kontinuierlichen Variablen anwenden. Vektoroptimierungsprobleme dienen der Beschreibung von Problemen mit mehreren konkurrierenden Zielen.

Gegeben sei die Zielfunktion

$$f : S \times \mathbb{R}^n \rightarrow \mathbb{R}^m, \quad (i, x) \mapsto f(i, x),$$

wobei S eine endliche Menge sei und $\mathbb{R}^n, \mathbb{R}^m$ den n -dimensionalen beziehungsweise den m -dimensionalen euklidischen Vektorraum bezeichne. f sei zweimal stetig partiell nach x differenzierbar. Zwei weitere Voraussetzungen an f werden wir später einführen. Wir betrachten das folgende Optimierungsproblem:

$$\min_{(i,x) \in S \times \mathbb{R}^n} f(i, x), \quad (VGOP).$$

Zunächst stellen wir ein in [12] entwickeltes Verfahren zur kontinuierlichen Vektoroptimierung vor.

Sei $h_3 : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine zweimal stetig differenzierbare vektorwertige Funktion. Wir betrachten das folgende Vektoroptimierungsproblem:

$$\min_{x \in \mathbb{R}^n} h_3(x). \quad (VKGOP)$$

Wir führen eine Halbordnung in \mathbb{R}^m ein, um einen Optimalitätsbegriff definieren zu können.

Definition 6.0.1 Seien $u, v \in \mathbb{R}^m$. Der Vektor u heißt kleiner oder gleich v , in Zeichen $u \leq_p v$, falls $u_j \leq v_j$ für alle $j \in \{1, \dots, m\}$ gilt. Gleichheit ist erfüllt, falls $u_j = v_j$ gilt für alle $j \in \{1, \dots, m\}$.

Mit dieser Halbordnung definieren wir nun ein Optimalitätskonzept für Vektoroptimierungsprobleme.

Definition 6.0.2 Ein Punkt $x^* \in \mathbb{R}^n$ heißt **Pareto-optimale Lösung** für $(VKGOP)$, falls kein $x \in \mathbb{R}^n$ existiert mit $h_3(x) \neq h_3(x^*)$ und $h_3(x) \leq_p h_3(x^*)$.

Gilt diese Bedingung nur für eine Umgebung $U(x^*) \subseteq \mathbb{R}^n$, so heißt x^* **lokal Pareto-optimale Lösung** für $(VKGOP)$

Ein Punkt $x_2 \in \mathbb{R}^n$ wird **dominiert** von $x_1 \in \mathbb{R}^n$, falls $h_3(x_1) \neq h_3(x_2)$ und $h_3(x_1) \leq_p h_3(x_2)$ gelten.

In Abschnitt 3.2 wurde erwähnt, dass numerische Methoden zur Minimierung einer zweimal stetig differenzierbaren Funktion $h_2 : \mathbb{R}^n \rightarrow \mathbb{R}$ oft als numerische Lösung des Anfangswertproblems

$$\dot{x}(t) = -\nabla h_2(x(t)), \quad x(0) = x_0,$$

interpretiert werden können. Wir wollen diesen Ansatz verallgemeinern auf Optimierungsprobleme $(VKGOP)$ mit Zielfunktion $h_3 : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Wir formulieren ein Anfangswertproblem, dessen eindeutige Lösung $x : [0, \infty) \rightarrow \mathbb{R}^n$ folgendes erfüllt:

$$h_3(x(s)) \geq_p h_3(x(t)) \text{ und } h_3(x(s)) \neq h_3(x(t)) \quad \text{für alle } 0 \leq s < t < \infty.$$

Für die Konstruktion unseres Anfangswertproblems betrachten wir das folgende quadratische Optimierungsproblem für $x \in \mathbb{R}^n$:

$$\min_{\alpha \in \mathbb{R}^m} \left\{ \left\| \sum_{j=1}^m \alpha_j \nabla h_{3,j}(x) \right\|^2 ; \alpha_j \geq 0, j = 1, \dots, m, \sum_{j=1}^m \alpha_j = 1 \right\}, \quad (QOP(x))$$

wobei $\nabla h_{3,j}$ den Gradienten der j -ten Komponente der Zielfunktion bezeichnet:

$$h_3 = (h_{3,1}, \dots, h_{3,m})^T, \quad h_{3,j} : \mathbb{R}^n \rightarrow \mathbb{R}, \quad j = 1, \dots, m.$$

Zu jedem $x \in \mathbb{R}^n$ existiert ein globaler Minimierer $\hat{\alpha}$. Für zwei globale Minimierer $\hat{\alpha}$ und $\bar{\alpha}$ von $(QOP(x))$ für festes x gilt

$$\sum_{j=1}^m \hat{\alpha}_j \nabla h_{3,j}(x) = \sum_{j=1}^m \bar{\alpha}_j \nabla h_{3,j}(x).$$

Damit definieren wir für einen globalen Minimierer $\hat{\alpha}$ von $(QOP(x))$ die Funktion

$$q : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad x \mapsto \sum_{j=1}^m \hat{\alpha}_j \nabla h_{3,j}(x).$$

Satz 6.0.3 *Wir betrachten $(QOP(x))$ und $q : \mathbb{R}^n \rightarrow \mathbb{R}^n$ sei wie eben definiert. Dann gelten:*

1. *Entweder ist $q(x) = 0$ oder $-q(x)$ ist eine Abstiegsrichtung für alle Funktionen $h_{3,1}, \dots, h_{3,m}$ am Punkt x . Jedes x mit $q(x) = 0$ erfüllt die notwendigen Bedingungen erster Ordnung für Pareto-optimale Lösungen.*
2. *Zu jedem $x \in \mathbb{R}^n$ existiert eine Umgebung $U(x)$ und eine Konstante $L_x \geq 0$ so dass für alle $y, z \in U(x)$ gilt*

$$\|q(y) - q(z)\| \leq L_x \|y - z\|.$$

BEWEIS : [12], Theorem 2.1 auf S. 212.

Die Eigenschaften der Funktion q ermöglichen es, das Anfangswertproblem für skalare Optimierungsprobleme zu verallgemeinern auf ein Anfangswertproblem für Vektoroptimierungsprobleme:

$$\dot{x}(t) = -q(x(t)), \quad x(0) = x_0. \quad (AWPV)$$

Satz 6.0.4 *Wir betrachten $(VKGOP)$ und das zugehörige Anfangswertproblem $(AWPV)$ mit $q(x_0) \neq 0$. Sei*

$$R_{\leq p} := \{x \in \mathbb{R}^n | h_3(x) \leq_p h_3(x_0)\}$$

beschränkt. Dann existiert eine eindeutige Lösung $x : [0, \infty) \rightarrow \mathbb{R}^n$ von $(AWPV)$ mit

$$h_3(x(s)) \geq_p h_3(x(t)) \text{ und } h_3(x(s)) \neq h_3(x(t)) \text{ für alle } 0 \leq s < t < \infty.$$

BEWEIS : [12], Theorem 2.2 auf S. 214.

Die numerische Lösung von $(AWPV)$ führt zu einem Punkt $x^* \in \mathbb{R}^n$, der die Bedingungen erster Ordnung für Pareto-optimale Lösungen erfüllt. Da q nicht

stetig differenzierbar ist, müssen explizite numerische Verfahren verwendet werden. Die Anwendung einer stochastischen Störung auf (AWPV) führt zu einer Methode, bei der alle oder eine große Anzahl Pareto-optimaler Lösungen approximiert werden.

Annahme C Es gibt ein $\varepsilon > 0$ so dass

$$x^T q(x) \geq \frac{1 + n\varepsilon^2}{2} \max\{1, \|q(x)\|_2\}$$

für alle $x \in \mathbb{R}^n \setminus \{x \in \mathbb{R}^n \mid \|x\|_2 \leq r\}$ mit einem $r > 0$ gilt.

Seien $(\Omega^n, \mathcal{B}(\Omega^n), W^n)$ der in Abschnitt 1.2 definierte Wahrscheinlichkeitsraum und $(\Omega^n, \mathcal{B}(\Omega^n), W^n, (B_t)_{t \geq 0})$ eine n -dimensionale Brownsche Bewegung.

Für die Lösung von (VKGOP) untersuchen wir für $\varepsilon > 0$ und $x_0 \in \mathbb{R}^n$ die folgende stochastische Integralgleichung:

$$X_t(\omega) = x_0 + \varepsilon(B_t(\omega) - B_0(\omega)) - \int_0^t q(X_\tau(\omega)) d\tau, \quad (SIE)_q$$

wobei q wie oben definiert sei und Annahme C erfülle. Die folgenden beiden Sätze sind direkte Analoga der Sätze 3.2.1 und 3.2.2 aus Abschnitt 3.2.

Satz 6.0.5 *Gegeben sei die Integralgleichung $(SIE)_q$. Für alle $x_0 \in \mathbb{R}^n$ und alle $\varepsilon > 0$, für die Annahme C für q erfüllt ist, erhalten wir*

1. *Es existiert ein eindeutiger stochastischer Prozess $(X_t)_{t \geq 0}$, der $(SIE)_q$ löst.*
2. *Alle Pfade von $(X_t)_{t \geq 0}$ sind stetig.*
3. $X_0 \equiv x_0$

Sei $s_{\bar{x}, \rho}(\omega) : \Omega^n \rightarrow \bar{\mathbb{R}}$,

$$\omega \mapsto \begin{cases} \inf\{t \geq 0 \mid \|X_t(\omega) - \bar{x}\|_2 \leq \rho\} & \text{falls } \{t \geq 0 \mid \|X_t(\omega) - \bar{x}\|_2 \leq \rho\} \neq \emptyset, \\ \infty & \text{sonst.} \end{cases}$$

Satz 6.0.6 *Gegeben sei $(SIE)_q$ mit $\varepsilon > 0$, so dass Annahme C für q erfüllt ist.*

Dann gelten für jedes $x_0 \in \mathbb{R}^n$, jede Pareto-optimale Lösung $\bar{x} \in \mathbb{R}^n$ und alle $\rho > 0$ die folgenden Aussagen:

1. $W^n(\{\omega \in \Omega^n \mid s_{\bar{x}, \rho}(\omega) < \infty\}) = 1$,
2. $E(s_{\bar{x}, \rho}) < \infty$, (E bezeichnet den Erwartungswert),

Satz 6.0.6 besagt, dass für jede Pareto-optimale Lösung \bar{x} W^n -fast jeder Pfad von $(X_t)_{t \geq 0}$ jede Kugel vom Radius ρ um \bar{x} in endlicher Zeit trifft. Darüberhinaus ist der Erwartungswert von $s_{\bar{x}, \rho}$ endlich. Die numerisch Berechnung eines Pfades

von $(X_t)_{t \geq 0}$ erfolgt wie in Abschnitt 3.3 durch die Berechnung dreier Punkte auf diesem Pfad. Da q nicht stetig differenzierbar ist, muß jedoch ein explizites numerisches Verfahren verwendet werden.

Nun betrachten wir $(VGOP)$ mit der Zielfunktion $f : S \times \mathbb{R}^n \rightarrow \mathbb{R}^m$. Der Algorithmus aus Kapitel 4 kann auf $(VGOP)$ angewandt werden. Wir definieren zu jedem $i \in S$ die Funktion

$$f^i : \mathbb{R}^n \rightarrow \mathbb{R}^m, \quad x \mapsto f(i, x).$$

Dabei erfülle jedes f^i Voraussetzung C. Weiter seien die Niveaumengen $\{x \in \mathbb{R}^n \mid f^i(x) \leq_p f^i(x_0)\}$ für alle $i \in S$, $x_0 \in \mathbb{R}^n$ beschränkt. Dies sind die anfangs angekündigten weiteren Voraussetzungen an f .

Seien $i \in S$, $x \in \mathbb{R}^n$ und sei $\hat{\alpha}$ eine Lösung des folgenden quadratischen Optimierungsproblems:

$$\min_{\alpha \in \mathbb{R}^m} \left\{ \left\| \sum_{j=1}^m \alpha_j \nabla f_j^i(x) \right\|^2, \alpha_j \geq 0, j = 1, \dots, m, \sum_{j=1}^m \alpha_j = 1 \right\}, \quad (QOP(x))_i$$

wobei ∇f_j^i den Gradienten der j -ten Komponente der Zielfunktion bezeichnet. Mit diesem $\hat{\alpha}$ definieren wir für $i \in S$ die Funktion

$$q^i : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad x \mapsto \sum_{j=1}^m \hat{\alpha}_j \nabla f_j^i(x)$$

und mit q^i die stochastische Integralgleichung

$$X_t(\omega) = x_0 + \varepsilon(B_t(\omega) - B_0(\omega)) - \int_0^t q^i(X_\tau(\omega)) d\tau. \quad (SIE)_q^i$$

Um $(VGOP)$ zu lösen wird der Algorithmus aus Abschnitt 4.4 geeignet modifiziert. Die Wahl des diskreten Parameters erfolgt genau wie im Fall der skalaren Optimierung. Der kontinuierliche Parameter zu einem fest gewählten diskreten Parameter $i \in S$ wird mit $(SIE)_q^i$ berechnet, indem ein Pfad der Lösung von $(SIE)_q^i$ numerisch approximiert wird. Da q^i nicht stetig nach x differenzierbar ist, muß ein explizites numerisches Verfahren verwendet werden.

Wegen Satz 6.0.6 können wir mit Wahrscheinlichkeit eins mit endlich vielen Integrationsschritten eine Kugel um den globalen Minimierer von $(VGOP)$ vom Radius $\rho > 0$ für beliebiges ρ erreichen. Es reicht sicherzustellen, dass für die Integrationsschrittweite σ gilt $\sigma \geq c$ für eine geeignete Konstante $c > 0$. Somit erhalten wir zu jedem diskreten Parameter $i \in S$ nur endlich viele verschiedene Zielfunktionswerte. Wie in Abschnitt 4.3 können wir einen diskreten stochastischen Prozess definieren, der den eben skizzierten Algorithmus zur diskret-kontinuierlichen Vektoroptimierung modelliert. Dieser Prozess ist im Sinne von

Abschnitt 4.5 rekurrent. Wie dort folgt, dass der Algorithmus bei genügend langer Laufzeit mit Wahrscheinlichkeit eins das globale Optimum von $(VGOP)$ findet.

Literaturverzeichnis

- [1] H. Bauer, *Maß- und Integrationstheorie*, Walter de Gruyter Verlag, Berlin, New York, 1990.
- [2] H. Bauer, *Wahrscheinlichkeitstheorie*, Walter de Gruyter Verlag, Berlin, New York, 2002.
- [3] P. Billingsley, *Probability and Measure*, John Wiley & Sons, New York, Chichester, 1995.
- [4] H. O. Georgii, *Stochastik*, Walter de Gruyter Verlag, Berlin, New York, 2002.
- [5] C. Großmann, J. Terno, *Numerik der Optimierung*, Teubner Verlag, Stuttgart, 1997.
- [6] E. Hairer, S.P. Nørsett, G. Wanner, *Solving Ordinary Differential Equations I*, Springer Verlag, Berlin, Heidelberg, 1987.
- [7] P. J. M. van Laarhoven, E. H. L. Aarts, *Simulated Annealing: Theory and Applications*, D. Reidel Publishing Company, Dordrecht, Boston, 1987.
- [8] Matlab, Version 6.1, Release 12.1. The MathWorks, www.mathworks.com.
- [9] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *Equation of state calculation by fast computing machines*, J. Chem. Phys. 21, 1087-1092, 1953.
- [10] B. Øksendal, *Stochastic Differential Equations*, Springer-Verlag, Berlin, Heidelberg, 2000.
- [11] S. Schäffler, *Global Optimization using stochastic Integration*, S. Roderer Verlag, Regensburg, 1995.
- [12] S. Schäffler, R. Schultz, K. Weinzierl, *Stochastic Method for the Solution of Unconstrained Vector Optimization Problems*, Journal of Optimization Theory and Applications, Vol. 114, Nr. 1, pp.209-222, 2002.
- [13] S. Schäffler, *Stochastische Differentialgleichungen*, Skriptenreihe des IAMS, Band 7, München, 1996.

- [14] H. Schwarz, *Numerische Mathematik*, Teubner Verlag, Stuttgart, 1997.
- [15] R. Sturm, *Numerische Lösung diskret-kontinuierlicher Vektoroptimierungsprobleme*, Dissertation, Universität der Bundeswehr München, 2001.
- [16] D. Werner, *Funktionalanalysis*, Springer, Berlin, Heidelberg, 1995.