

UNIVERSITÄT DER BUNDESWEHR MÜNCHEN
Fakultät für Elektrotechnik und Informationstechnik

Schrittweisensteuerungskonzept zur globalen Optimierung mittels stochastischer Integration

Petra Ehlers

Vorsitzender des Promotionsausschusses: Prof. Dr.-Ing. Marquardt
1. Berichterstatter: Prof. Dr. rer.nat. Dr.-Ing. Schäffler
2. Berichterstatter: Prof. Dr. rer.nat. Gilg

Tag der Prüfung: 02.07.2003

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

Neubiberg, den 2. Juli 2003

UNIVERSITÄT DER BUNDESWEHR MÜNCHEN
Fakultät für Elektrotechnik und Informationstechnik

Schrittweisensteuerungskonzept zur globalen Optimierung mittels stochastischer Integration

Petra Ehlers

Vorsitzender des Promotionsausschusses: Prof. Dr.-Ing. Marquardt
1. Berichterstatter: Prof. Dr. rer.nat. Dr.-Ing. Schäffler
2. Berichterstatter: Prof. Dr. rer.nat. Gilg

Tag der Prüfung: 02.07.2003

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

Neubiberg, den 2. Juli 2003

Inhaltsverzeichnis

Einleitung	3
1 Grundlagen	6
1.1 Notationen	6
1.2 Wahrscheinlichkeitstheorie	8
1.3 Stochastische Integrale	15
1.4 Stochastische Differentialgleichungen	18
1.5 Markovketten	20
2 Diskrete globale Optimierung	31
2.1 Der Metropolis-Algorithmus	31
2.2 Ein Abbruchkriterium zum Metropolis-Algorithmus	35
3 Kontinuierliche globale Optimierung	39
3.1 Voraussetzung A an die Zielfunktion	39
3.2 Eine spezielle Klasse stochastischer Differentialgleichungen	40
4 Schrittweisensteuerungskonzept	44
4.1 Das semi-implizite Eulerverfahren	44
4.2 Allgemeine Vorgehensweise	45
4.3 Voraussetzung B an die Zielfunktion	46
4.4 Dichte der Übergangswahrscheinlichkeiten	48
4.5 Irreduzibilität	50
4.6 Harris-Rekurrenz	51
4.7 Strenge Aperiodizität	54
4.8 Positivität	55
4.9 Konvergenz der Verteilungen	56
5 Numerische Ergebnisse	58
5.1 Anmerkungen zur Implementierung	58
5.2 Formulierung des Algorithmus	59

5.3	Numerische Ergebnisse	61
	Problem 1	63
	Problem 2	64
	Problem 3	65
	Problem 4	66
	Problem 5	67
	Problem 6	68
	Zusammenfassung	69
	Danksagung	72
	Index	74
	Literaturverzeichnis	76
	Symbolverzeichnis	78

Einleitung

Eine große Anzahl von Fragen aus Wissenschaft, Technik und Forschung führen zu dem globalen Optimierungsproblem (GOP): Finde zu einer gegebenen Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^2$, einen Punkt $x^* \in \mathbb{R}^n$, so daß gilt

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

Verfahren zur lokalen Optimierung, die iterativ von einem gewählten Startpunkt aus das Optimum berechnen, können nicht verwendet werden, da als Suchrichtungen für gewöhnlich nur Abstiegsrichtungen gewählt werden. Aus diesem Grund sind die errechneten Minima im allgemeinen nur lokale Minima. Ein Algorithmus zur Auffindung eines globalen Minimierers muß in der Lage sein, ein lokales Minimum wieder zu verlassen.

Eine Möglichkeit, dieses Problem zu lösen, ist eine Reihe zufällig erzeugter Zahlen als Startpunkte zur lokalen Optimierung zu verwenden. In der Praxis erweist sich dieses Verfahren jedoch als zu rechenintensiv.

Kirkpatrick und Černý (vgl. [Kirkpatrick 82],[Černý 85]) stellten für das diskrete globale Optimierungsproblem (DGOP): Finde zu einer gegebenen Funktion $f : R \rightarrow R$, einen Punkt $x^* \in R$, so daß gilt

$$f(x^*) \leq f(x) \quad \text{für alle } x \in R, \quad (\text{DGOP})$$

wobei R ein endlicher Zustandsraum sei, unabhängig voneinander ein Verfahren vor, das auf einem von Metropolis (vgl. [Metropolis 53]) 1953 vorgestellten Verfahren zur effizienten Simulation der Entwicklung eines Festkörpers im Hitzebad zum thermischen Gleichgewicht hin beruhte. Dieser sogenannte Metropolis-Algorithmus setzt sich zusammen aus der zufälligen Wahl eines Punktes aus der „Nachbarschaft“ des Startpunktes und einem nachfolgenden Test auf eine Akzeptanzbedingung. Die Nachbarschaftsstruktur wird dabei separat untersucht und im Verfahren als gegeben vorausgesetzt. Auf diese Weise werden mit positiver Wahrscheinlichkeit auch Punkte akzeptiert, die eine Verschlechterung des Zielfunktionswertes bedeuten. Durch fortgesetzte Durchführung dieser Vorgehensweise erzeugt man nach einer ausreichend großen Anzahl von Iterationen Zufallszahlen, die gemäß einer Verteilung auftreten, deren Maximum dort liegt, wo die

Zielfunktion ihr Minimum hat. Die genaue Vorgehensweise wird in Kapitel 2.1 beschrieben.

Hierbei stellt sich allerdings das Problem zu erkennen, wann eine Verteilung mit diesen Eigenschaften erreicht ist. Ein zu früher Abbruch des Programmdurchlaufs birgt die Gefahr, daß der globale Minimierer noch nicht erreicht wurde; ein zu später Abbruch bedeutet unnötigen numerischen Aufwand. Aus diesem Grund wird in Kapitel 2.2 ein Abbruchkriterium entwickelt, das kaum zusätzlichen numerischen Aufwand erfordert und uns außerdem ein effektives Werkzeug zur Erkennung einer hinreichenden Annäherung an die Grenzverteilung liefert.

Zur Lösung des kontinuierlichen Problems (GOP) wird in [Schäffler 95] ein Verfahren vorgestellt, das den gleichen Grundgedanken verfolgt wie die Metropolis-Vorgehensweise: die Erzeugung von Zufallszahlen gemäß einer Dichte, deren Maximum dort liegt, wo die Funktion ihr globales Minimum besitzt. Die Idee besteht im wesentlichen darin, die stochastische Differentialgleichung

$$X_t^{\epsilon, x_0}(F) = x_0 + \epsilon (B_t(F) - B_0(F)) - \int_0^t \nabla f(X_s^{\epsilon, x_0}(F)) ds \quad (1)$$

zu lösen. Die stochastische Differentialgleichung selbst und inwiefern der Lösungsprozeß zum Ziel führt, wird in Kapitel 3 noch ausführlicher beschrieben.

Leider kann der Lösungsprozeß $\{X_t^{\epsilon, x_0} : t \geq 0\}$ der stochastischen Differentialgleichung (1) in der Praxis nicht genau ermittelt werden. Vielmehr wird in [Schäffler 95] zur numerischen Lösung von (1) das semi-implizite Euler-Verfahren vorgeschlagen, das nur für sehr kleine Schrittweiten ausreichend gut die tatsächliche Lösung approximiert. Die Wahl kleiner Schrittweiten erfordert allerdings sehr lange Laufzeiten des Algorithmus, wird die gewünschte Dichte doch erst für t gegen unendlich erreicht.

Allerdings besteht für die Zwecke der globalen Optimierung keine Notwendigkeit darin, den Lösungsprozeß $\{X_t^{\epsilon, x_0} : t \geq 0\}$ beliebig genau zu berechnen. Vielmehr liegt die Zielsetzung darin, Zufallszahlen mit der oben beschriebenen Verteilung zu erzeugen. Aus diesem Grund wird in Kapitel 4 untersucht, für welche Wahl der Schrittweiten die durch das semi-implizite Euler-Verfahren entstehende Markovkette stabil ist, in dem Sinne, daß eine stationäre Verteilung dieser Markovkette existiert und die Verteilungen im n -ten Schritt für n gegen unendlich gegen diese stationäre Verteilung konvergieren. Zu diesem Zweck wählen wir eine „Schrittweitenfunktion“ auf \mathbb{R}^n , die jedem Ausgangspunkt x eine Schrittweite $\sigma(x)$ zuordnet. Es zeigt sich, daß Anforderungen an die Schrittweitenfunktion zur Sicherstellung der Stabilität nur außerhalb einer Kugel um den Nullpunkt, die den globalen Minimierer enthält, gestellt werden müssen. Innerhalb dieser Kugel

müssen wir lediglich die Stetigkeit der Schrittweitenfunktion sowie die Regularität von $\mathbb{E}_n + \sigma(x)\nabla^2 f(x)$ fordern.

Die Beweisführung in Kapitel 4 ermöglicht es uns, im kontinuierlichen Verfahren das gleiche Abbruchkriterium zu benutzen, wie es zuvor für den Metropolis-Algorithmus entwickelt wurde. In Kapitel 5.1 werden einige Hinweise zur Implementation dieses Schrittweitensteuerungskonzepts angeführt und in Kapitel 5.2 der Algorithmus formuliert. Im Anschluß daran erfolgt die Präsentation numerische Ergebnisse zu einigen Beispielfunktionen.

Kapitel 1

Grundlagen

Bevor die Theorie der Schrittweitensteuerung zur globalen Optimierung untersucht werden kann, werden zunächst einige Grundlagen bereitgestellt. Im ersten Abschnitt dieses Kapitels führen wir lediglich Notationen ein. Im zweiten sowie auch im dritten und vierten Abschnitt untersuchen wir kurz die Grundlagen aus der Stochastik, die für das Verständnis dieser Arbeit erforderlich sind. Im fünften Abschnitt wenden wir uns der Theorie von Markovketten zu, deren Untersuchung einen wesentlichen Bestandteil dieser Arbeit darstellt.

1.1 Notationen

Zu Beginn betrachten wir die für diese Arbeit relevanten Notationen. Unter anderem werden in diesem Abschnitt die in dieser Arbeit verwendeten Normen definiert und für bekannte Operatoren und Funktionen Bezeichnungen eingeführt. Im gesamten Kapitel seien $n, m, k \in \mathbb{N}$.

Notation 1.1 ($\mathbb{R}_0^+, \overline{\mathbb{R}}, \mathbb{N}, \mathbb{N}_0$).

Es bezeichne \mathbb{R}_0^+ den Raum der nichtnegativen reellen Zahlen, also

$$\mathbb{R}_0^+ := \{x \in \mathbb{R} : x \geq 0\}.$$

$\overline{\mathbb{R}}$ sei die Menge der um $\{\pm\infty\}$ erweiterten reellen Zahlen, \mathbb{N} die Menge der natürlichen Zahlen ohne die Null sowie $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.

Notation 1.2 (Potenzmenge).

Für eine Menge A bezeichne $\mathfrak{P}(A)$ die Potenzmenge von A , d.h. die Menge aller Teilmengen von A .

Notation 1.3 (Matrizen, Vektoren).

Sei $A \in \mathbb{R}^{n,m}$ und $i, j \in \mathbb{N}$ mit $1 \leq i \leq n$, $1 \leq j \leq m$. Dann bezeichne A_i die i -te Zeile, A^j die j -te Spalte und A_{ij} den Eintrag an der Stelle (i, j) der Matrix A .

Bemerkung 1.4. Wir werden die Vektoren im \mathbb{R}^n als Spaltenvektoren auffassen, d.h. der \mathbb{R}^n wird identifiziert mit $\mathbb{R}^{n,1}$.

Notation 1.5 (Normen).

i) Mit $\|\cdot\|_2$ bezeichnen wir die euklidische Norm auf dem \mathbb{R}^n , also

$$\|x\|_2 := \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

ii) Mit $\|\cdot\|_F$ bezeichnen wir die Frobeniusnorm auf dem $\mathbb{R}^{n,m}$, also

$$\|A\|_F := \sqrt{\sum_{i=1}^n \sum_{j=1}^m A_{ij}^2}.$$

Bemerkung 1.6. Die Topologien auf dem \mathbb{R}^n bzw. $\mathbb{R}^{n,m}$ seien die von der euklidischen Norm bzw. Frobeniusnorm induzierten Topologien.

Notation 1.7 (δ -Umgebung von $x_0 \in \mathbb{R}^n$).

Für $\delta > 0$ und $x_0 \in \mathbb{R}^n$ definieren wir die δ -Umgebung von x_0 als $U_\delta(x_0) := \{x \in \mathbb{R}^n : \|x - x_0\|_2 < \delta\}$.

Nachdem wir eine Topologie auf dem \mathbb{R}^n definiert haben, sind somit auch stetige Funktionen auf dem \mathbb{R}^n definiert, und wir können Räume stetiger Funktionen auf dem \mathbb{R}^n betrachten.

Notation 1.8 (Funktionsräume).

Sei U eine Teilmenge des \mathbb{R}^k . Dann definieren wir

- a) $C(U, \mathbb{R}^{n,m}) := \{f : U \rightarrow \mathbb{R}^{n,m} : f \text{ ist stetig}\}$,
- b) $C(U) := C(U, \mathbb{R})$,
- c) $C_b(U) := \{f \in C(U) : f \text{ ist beschränkt}\}$,
- d) $C_c(U) := \{f \in C(U) : f \text{ hat einen kompakten Träger in } U\}$.

Ist U offen im \mathbb{R}^k bezüglich der oben genannten Topologie, so setzen wir für $q \in \mathbb{N}$

- e) $C^q(U, \mathbb{R}^{n,m}) := \{f : U \rightarrow \mathbb{R}^{n,m} : f \text{ ist } q\text{-mal stetig differenzierbar}\}$,
- f) $C^q(U) := C^q(U, \mathbb{R})$,
- g) $C_b^q(U) := C^q(U) \cap C_b(U)$,
- h) $C_c^q(U) := C^q(U) \cap C_c(U)$.

Notation 1.9 (Ableitung, Gradient).

- i) Es sei $U \subset \mathbb{R}^k$ offen und $f \in C^1(U, \mathbb{R}^{n,m})$. Dann ist die Ableitung nach der i -ten Komponente $D_i f(x)$ definiert durch

$$D_i f(x) := \begin{pmatrix} \frac{\partial f_{11}(x)}{\partial x_i} & \cdots & \frac{\partial f_{1m}(x)}{\partial x_i} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{n1}(x)}{\partial x_i} & \cdots & \frac{\partial f_{nm}(x)}{\partial x_i} \end{pmatrix}$$

für alle $x \in U$ und $i \in \{1, 2, \dots, k\}$.

- ii) Sei $U \subset \mathbb{R}^n$ offen und $f \in C^1(U, \mathbb{R}^m)$. Dann bezeichnet

$$Jf(x) := (\nabla f_1(x), \nabla f_2(x), \dots, \nabla f_m(x)) := \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1(x)}{\partial x_n} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{pmatrix}$$

den Gradient von f an der Stelle x .

1.2 Wahrscheinlichkeitstheorie

Wesentlicher Bestandteil des Verfahrens der globalen Optimierung mittels stochastischer Methoden ist die Lösung stochastischer Differentialgleichungen. Dazu müssen wir zunächst stochastische Differentialgleichungen definieren. Desweiteren werden wir einführen, was unter einer Lösung einer stochastischen Differentialgleichung zu verstehen ist, und wann sie existiert und eindeutig ist. Stochastische Differentialgleichungen beruhen wiederum auf einer Vielzahl von Definitionen und Sätzen aus der Wahrscheinlichkeitstheorie. Deshalb werden zunächst die Grundlagen aus der Stochastik bereitgestellt (eine ausführlichere Darstellung findet man in [Schäffler Sturm 94], [Schäffler Sturm 95] sowie [Billingsley 85]). Anschließend wird kurz das stochastische Itô-Integral vorgestellt, um dann die für uns wesentlichen Definitionen und Sätze zu stochastischen Differentialgleichungen zu nennen (vgl. [Schäffler 96] sowie [Øksendal 98]).

Notation 1.10.

\mathcal{B}^n bezeichne für $n \in \mathbb{N}$ die Borel- σ -Algebra im \mathbb{R}^n , d.h. die von den offenen Mengen im \mathbb{R}^n erzeugte σ -Algebra, $\mathcal{B}_{[a,b]}$, $a, b \in \mathbb{R}$ mit $a < b$, die Borel- σ -Algebra auf dem Intervall $[a, b]$ und λ^n das n -dimensionale Lebesgue-Maß. Weiterhin bildet die Menge $\overline{\mathcal{B}} := \{A \in \mathfrak{P}(\overline{\mathbb{R}}) : A \cap \mathbb{R} \in \mathcal{B}\}$ eine σ -Algebra über $\overline{\mathbb{R}}$.

Wir werden im folgenden nur einen speziellen Wahrscheinlichkeitsraum – den *Wiener-Raum* – betrachten. Dies ist der Raum Ω^n aller stetigen Funktionen F :

$\mathbb{R}_0^+ \rightarrow \mathbb{R}^n$ mit der σ -Algebra $\mathcal{B}(\Omega^n)$. Dabei bezeichne $\mathcal{B}(\Omega^n)$ die kleinste aller σ -Algebren \mathcal{S} über Ω^n , so daß die Projektionen $p_t : \Omega^n \rightarrow \mathbb{R}^n$, $F \mapsto F(t)$, $t \in \mathbb{R}_0^+$, $\mathcal{S} - \mathcal{B}^n$ -meßbar sind. Um auf diesem Meßraum ein Maß – das Wiener-Maß – definieren zu können, benötigen wir einige Definitionen.

Definition 1.11 (*P-fast überall, P-fast sicher*).

Es sei (Ω, \mathcal{S}, P) ein Maßraum. Eine Eigenschaft gilt P-fast überall, falls es eine Menge A vom Maß 0 gibt, so daß die Eigenschaft außerhalb von A gilt. Ist (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum, so gilt eine Eigenschaft P-fast sicher, falls sie für eine Menge vom Maß 1 gilt.

Definition 1.12 (numerische Funktion).

Eine auf einer nichtleeren Menge $A \subseteq \Omega^n$ definierte Funktion $g : A \rightarrow \overline{\mathbb{R}}$ heißt numerische Funktion.

Definition 1.13 (n-dimensionale Zufallsvariable).

Sei $n \in \mathbb{N}$. Eine Funktion $X : \Omega^n \rightarrow \mathbb{R}^n$ heißt (n-dimensionale) Zufallsvariable, falls sie $\mathcal{B}(\Omega^n)$ - \mathcal{B}^n -meßbar ist.

Wird ein Zufallsexperiment ausgeführt, so zeigt eine Zufallsvariable den Ausgang dieses Experiments als n-dimensionalen reellen Vektor an. Man spricht hier von einer Realisierung der Zufallsvariablen X .

Ein zentraler Begriff dieser Arbeit ist der des stochastischen Prozesses.

Definition 1.14 (stochastischer Prozeß).

Es sei $X_t : \Omega^n \rightarrow \mathbb{R}^n$ für jedes $t \in I$ eine n-dimensionale Zufallsvariable, wobei I eine Indexmenge sei. Dann heißt die Menge $\{X_t : t \in I\}$ stochastischer Prozeß.

Durch Fixierung eines $\omega \in \Omega^n$ erhält man einen Pfad des stochastischen Prozesses.

Definition 1.15 (Pfad eines stochastischen Prozesses).

Es sei $\{X_t : t \geq 0\}$ ein stochastischer Prozeß. Dann heißt die Funktion $X^\omega : \mathbb{R}_0^+ \rightarrow \mathbb{R}^n$, $t \mapsto X_t(\omega)$, $\omega \in \Omega^n$, Pfad des stochastischen Prozesses $\{X_t : t \geq 0\}$.

Die Stetigkeit eines stochastischen Prozesses betrachtet man pfadweise.

Definition 1.16 (Stetigkeit eines stochastischen Prozesses).

Es sei $\{X_t : t \geq 0\}$ ein stochastischer Prozeß. Dann heißt $\{X_t : t \geq 0\}$ stetig, falls jeder Pfad von $\{X_t : t \geq 0\}$ stetig ist.

Sind alle Informationen aus der Vergangenheit in der aktuellen σ -Algebra enthalten, so erhält man eine aufsteigende Folge von σ -Algebren, die man als Filtration bezeichnet.

Definition 1.17 (Filtration).

Es sei (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum, $U \subseteq \mathbb{R}$ und $\{\mathcal{F}_s : s \in U\}$ eine Menge von σ -Algebren auf Ω mit $\mathcal{F}_s \subseteq \mathcal{S}$ für alle $s \in U$. Es gelte $\mathcal{F}_s \subseteq \mathcal{F}_r$ für alle $s, r \in U$ mit $s \leq r$. Dann heißt $\{\mathcal{F}_s : s \in U\}$ eine Filtration in \mathcal{S} .

Definition 1.18 (adaptiert).

Es sei (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum, $U \subseteq \mathbb{R}$ und $\{X_s : s \in U\}$ ein stochastischer Prozeß und $\{\mathcal{F}_s : s \in U\}$ eine Filtration in \mathcal{S} . Dann heißt der stochastische Prozeß $\{X_s : s \in U\}$ der Filtration $\{\mathcal{F}_s : s \in U\}$ adaptiert, falls für jedes $s \in U$ die Zufallsvariable X_s \mathcal{F}_s - \mathcal{B}^n -meßbar ist.

Als *kanonische Filtration* eines stochastischen Prozesses $\{X_s : s \in U\}$ bezeichnen wir die Filtration $\{\mathcal{S}_s : s \in U\}$, gegeben durch $\mathcal{S}_s := \sigma(X_t : t \leq s)$ für alle $s \in U$. Dabei bezeichne $\sigma(X_t : t \leq s)$ die kleinste aller σ -Algebren \mathcal{S} , so daß alle Zufallsvariablen X_t , $t \leq s$, \mathcal{S} - \mathcal{B}^n -meßbar sind.

Satz 1.19 (Integration stetiger stochastischer Prozesse).

Es sei $n \in \mathbb{N}$ und $\{X_t : t \geq 0\}$ mit $X_t : \Omega^n \rightarrow \mathbb{R}^n$, $t \in \mathbb{R}_0^+$, ein stetiger stochastischer Prozeß. Dann ist für jede stetige Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ der stochastische Prozeß $\{Y_t : t \geq 0\}$ mit

$$Y_t : \Omega^n \rightarrow \mathbb{R}^n, \quad F \mapsto \int_0^t g(X_s(F)) ds, \quad t \in \mathbb{R}_0^+$$

stetig.

Wir definieren nun einen wichtigen stochastischen Prozeß - die Brownsche Bewegung. Zunächst betrachten wir sie auf allgemeinen Räumen, dann führen wir einen Satz an, der aussagt, daß es ein Maß auf $(\Omega^n, \mathcal{B}(\Omega^n))$ gibt, so daß eine Brownsche Bewegung existiert.

Definition 1.20 (Brownsche Bewegung auf (Ω, \mathcal{S}, P)).

Es sei (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum und $X_t : \Omega \rightarrow \mathbb{R}^n$ für jedes $t \in \mathbb{R}_0^+$ \mathcal{S} - \mathcal{B}^n -meßbar. Dann heißt der stochastische Prozeß $\{X_t : t \in \mathbb{R}_0^+\}$ n -dimensionale Brownsche Bewegung auf (Ω, \mathcal{S}, P) , falls gilt

- i) $X_0 = 0$ P -fast sicher,
- ii) Für jedes $k \in \mathbb{N}$ und für jedes Tupel $(t_1, t_2, \dots, t_k)^T \in \mathbb{R}^k$ mit $0 \leq t_1 < t_2 < \dots < t_k$ sind die Zufallsvariablen $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_k} - X_{t_{k-1}}$ stochastisch unabhängig,
- iii) für alle $s, t \in \mathbb{R}_0^+$, $s < t$, ist die Zufallsvariable $X_t - X_s$ $N(0, (t-s)\mathbb{E}_n)$ -normal verteilt, wobei \mathbb{E}_n die n -dimensionale Einheitsmatrix bezeichne,
- iv) alle Pfade von $\{X_t : t \in \mathbb{R}_0^+\}$ sind stetig.

Der folgende Satz sagt aus, daß es auf dem Wiener-Raum ein Maß W^n gibt, so daß eine Brownsche Bewegung existiert. Zum Beweis sei auf [Bauer 74] verwiesen.

Satz und Definition 1.21 (Wiener-Maß, Wiener-Raum).

Es sei $(\Omega^n, \mathcal{B}(\Omega^n))$ der Raum aller stetigen Funktionen $F : \mathbb{R}_0^+ \rightarrow \mathbb{R}^n$ mit der σ -Algebra $\mathcal{B}(\Omega^n)$. Dabei bezeichne $\mathcal{B}(\Omega^n)$ die kleinste aller σ -Algebren \mathcal{S} über Ω^n , so daß die Projektionen $p_t : \Omega^n \rightarrow \mathbb{R}^n$, $F \mapsto F(t)$, $t \in \mathbb{R}_0^+$, $\mathcal{S} - \mathcal{B}(\Omega^n)$ -meßbar sind. Es sei $\{B_t : t \geq 0\}$ ein stochastischer Prozeß mit $B_t(\omega) = \omega(t)$ für alle $t \geq 0$ und $\omega \in \Omega^n$. Dann gibt es für jedes $n \in \mathbb{N}$ genau ein Maß W^n auf $(\Omega^n, \mathcal{B}(\Omega^n))$, genannt das Wiener-Maß, mit den folgenden Eigenschaften

- i) $B_0 = 0$ W^n -fast sicher,
- ii) für alle t_0, t_1, \dots, t_m , $m \in \mathbb{N}$ mit $0 = t_0 < t_1 < \dots < t_m$ sind die Zufallsvariablen $B_{t_0}, B_{t_1} - B_{t_0}, \dots, B_{t_m} - B_{t_{m-1}}$ stochastisch unabhängig,
- iii) für alle $s, t \in \mathbb{N}$ mit $0 \leq s < t$ ist die Zufallsvariable $B_t - B_s$ $N(0, (t-s)\mathbb{E}_n)$ -normal verteilt, wobei \mathbb{E}_n die n -dimensionale Einheitsmatrix bezeichne.

Wir bezeichnen den Wahrscheinlichkeitsraum $(\Omega^n, \mathcal{B}(\Omega^n), W^n)$ als Wiener-Raum.

Somit ergibt sich zwingend die folgende Definition.

Definition 1.22 (Brownsche Bewegung auf $(\Omega^n, \mathcal{B}(\Omega^n), W^n)$).

Der stochastische Prozeß $\{B_t : t \geq 0\}$ mit $B_t(\omega) = \omega(t)$ für alle $\omega \in \Omega^n$ und $t \in \mathbb{R}_0^+$, $n \in \mathbb{N}$, heißt die n -dimensionale Brownsche Bewegung auf $(\Omega^n, \mathcal{B}(\Omega^n), W^n)$.

Nachdem wir nun den Maßraum $(\Omega^n, \mathcal{B}(\Omega^n), W^n)$ definiert haben, können wir uns der Verteilung und dem Erwartungswert einer Zufallsvariablen auf diesem Maßraum zuwenden.

Definition 1.23 (Verteilung einer Zufallsvariablen).

Es sei $X : \Omega^n \rightarrow \mathbb{R}^n$ eine Zufallsvariable. Dann heißt das Maß μ auf dem Meßraum $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ mit

$$\mu(A) := W^n(\{\omega \in \Omega^n : X(\omega) \in A\})$$

für alle $A \in \mathcal{B}^n$ die Verteilung von X .

Definition 1.24 (Erwartungswert einer Zufallsvariablen).

Es sei $X : \Omega^n \rightarrow \mathbb{R}^n$ eine Zufallsvariable. Wenn das Integral

$$\int_{\Omega^n} X dW^n$$

existiert, wird es Erwartungswert der Zufallsvariablen X genannt und mit $E(X)$ bezeichnet.

Ein weiterer zentraler Begriff der Wahrscheinlichkeitstheorie ist der der stochastischen Unabhängigkeit. Diese definieren wir zunächst für Ereignisse und auf dieser Grundlage dann auch für Zufallsvariablen.

Definition 1.25 (stochastisch unabhängige Ereignisse).

Es sei (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum und $A_1, A_2, \dots, A_n \in \mathcal{S}$, $n \in \mathbb{N}$. Dann heißen die Ereignisse A_1, A_2, \dots, A_n stochastisch unabhängig, falls für alle $k \in \mathbb{N}$, $k \leq n$, und für alle $i_j \in \mathbb{N}$, $1 \leq j \leq k$, mit $1 \leq i_1 < i_2 < \dots < i_k \leq n$ gilt:

$$P\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k P(A_{i_j}).$$

Aus dieser Definition leitet sich die Definition der stochastischen Unabhängigkeit einer Menge von Ereignissen ab.

Definition 1.26 (stochastische Unabhängigkeit einer Menge von Ereignissen).

Sei (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum und $\{A_i \in \mathcal{S} : i \in I\}$, $I \neq \emptyset$, eine Menge von Ereignissen. Dann heißen diese Ereignisse stochastisch unabhängig, falls die Ereignisse $A_{i_1}, A_{i_2}, \dots, A_{i_n}$ für jedes $n \in \mathbb{N}$ mit $n \leq |I|$ und für jede Menge $\{i_1, i_2, \dots, i_n\} \subseteq I$ stochastisch unabhängig sind.

Um die stochastische Unabhängigkeit von Zufallsvariablen einführen zu können, fehlt nun nur noch der Begriff der stochastischen Unabhängigkeit von Mengensystemen.

Definition 1.27 (stochastische Unabhängigkeit von Mengensystemen).

Sei (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum und $\{\mathcal{F}_i \subseteq \mathcal{S} : i \in I\}$, $I \neq \emptyset$, eine Menge von Mengensystemen über Ω , dann heißen diese Mengensysteme stochastisch unabhängig, falls für jedes $n \in \mathbb{N}$ mit $n \leq |I|$ und für jedes $\{i_1, i_2, \dots, i_n\} \subseteq I$ die Ereignisse $A_{i_1}, A_{i_2}, \dots, A_{i_n}$ für beliebige $A_{i_k} \in \mathcal{F}_{i_k}$, $k = 1, 2, \dots, n$, stochastisch unabhängig sind.

Definition 1.28 (stochastische Unabhängigkeit von Zufallsvariablen).

Seien (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum und (Ω', \mathcal{S}') ein Meßraum. Dann heißen die Zufallsvariablen $X_i : \Omega \rightarrow \Omega'$, $i \in I$, stochastisch unabhängig, falls die Mengensysteme $\{\sigma(X_i); i \in I\}$ stochastisch unabhängig sind. Dabei bezeichne $\sigma(X_i)$ die kleinste unter allen σ -Algebren \mathcal{A} über Ω derart, daß X_i \mathcal{A} - \mathcal{S}' -meßbar ist.

In dieser Arbeit soll ein Verfahren vorgestellt werden, das Zufallszahlen gemäß einer Dichte, die sich für die globale Optimierung als nützlich erweist, erzeugt. Dementsprechend muß hier klargestellt werden, was eine Dichte ist und unter welchen Umständen sie existiert.

Definition 1.29 (Dichte).

Es sei (Ω, \mathcal{S}) ein Meßraum und P und μ zwei Maße auf \mathcal{S} . Existiert eine nicht-negative, \mathcal{S} - $\overline{\mathcal{B}}$ -meßbare Funktion $f : \Omega \rightarrow \overline{\mathbb{R}}$ mit

$$P(A) = \int_A f d\mu \quad \text{für alle } A \in \mathcal{S},$$

so wird f Dichte des Wahrscheinlichkeitsmaßes P bezüglich μ genannt.

Um ein Kriterium für die Existenz einer solchen Dichte aufstellen zu können, benötigen wir die beiden folgenden Definitionen:

Definition 1.30 (absolute Stetigkeit von P bzgl. μ).

Es sei (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum und μ ein Maß auf \mathcal{S} . P heißt absolutstetig bezüglich μ , falls für alle $A \in \mathcal{S}$ mit $\mu(A) = 0$ gilt: $P(A) = 0$.

Gilt die absolute Stetigkeit in beide Richtungen, so spricht man von äquivalenten Maßen:

Definition 1.31 (Äquivalente Maße).

Es sei (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum und μ und ν Maße auf \mathcal{S} . Dann heißen μ und ν äquivalent, falls μ absolutstetig bezüglich ν und ν absolutstetig bezüglich μ ist.

Definition 1.32 (σ -endliches Maß).

Es sei (Ω, \mathcal{S}) ein Meßraum. Ein Maß μ auf \mathcal{S} heißt σ -endlich, falls es eine Folge von Mengen $\{A_i\}_{i \in \mathbb{N}}$, $A_i \in \mathcal{S}$, gibt mit

$$\bigcup_{i=1}^{\infty} A_i = \Omega, \quad \mu(A_i) < \infty \quad \text{für alle } i \in \mathbb{N}.$$

Nun können wir einen wichtigen Satz nennen, der unter bestimmten Bedingungen für ein Maß P die Existenz einer Dichte bezüglich eines Maßes μ sicherstellt.

Satz 1.33 (Radon-Nikodym).

Seien (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum und μ ein σ -endliches Maß auf \mathcal{S} . Dann besitzt P genau dann eine Dichte bezüglich μ , wenn P absolutstetig bezüglich μ ist.

Wir werden in dieser Arbeit nur Dichten bezüglich dem n -dimensionalen Lebesgue-Maß λ^n untersuchen. Hat die Verteilung einer Zufallsvariablen X bezüglich dem Lebesgue-Maß λ^n die Dichte f , so sagt man verkürzt auch X habe die Dichte f .

Sind nun Zufallsvariablen X_1, X_2, \dots, X_n mit Dichten f_1, f_2, \dots, f_n gegeben, so stellt sich die Frage, wie die Dichte der gemeinsamen Verteilung (X_1, X_2, \dots, X_n) aussieht. Für unabhängige Zufallsvariablen gibt darüber der folgende Satz Aufschluß.

Satz 1.34 (Gemeinsame Verteilung unabhängiger Zufallsvariablen).

Es seien X_1, X_2, \dots, X_n unabhängige Zufallsvariablen, und X_i habe die Dichte f_i , $i = 1, \dots, n$, so hat $X = (X_1, X_2, \dots, X_n)$ die Dichte f mit

$$f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2) \dots f(x_n).$$

Um später den Existenz- und Eindeutigkeitsatz für stochastische Differentialgleichungen formulieren zu können, benötigen wir zwei relativ technische Definitionen.

Definition 1.35 (nicht vorgeifende Filtration).

Es sei $(\Omega^n, \mathcal{B}(\Omega^n), W^n)$ der Wiener-Raum und $\{B_t : t \in \mathbb{R}_0^+\}$ die n -dimensionale Brownsche Bewegung mit der kanonischen Filtration $\{\mathcal{F}_t^0 : t \in \mathbb{R}_0^+\}$. Mit \mathcal{A}_s bezeichnen wir die von $\{B_s - B_t : t \leq s < \infty\}$ erzeugte σ -Algebra. Es sei $\{\mathcal{F}_s : s \in [a, b]\}$, $a, b \in \mathbb{R}_0^+$, $a < b$ eine Filtration in $\mathcal{B}(\Omega^n)$ mit

i) $\mathcal{F}_s^0 \subseteq \mathcal{F}_s$ für alle $s \in [a, b]$,

ii) für jedes $s \in [a, b]$ sind die σ -Algebren \mathcal{A}_s und \mathcal{F}_s stochastisch unabhängig.

Dann heißt die Filtration $\{\mathcal{F}_s : s \in [a, b]\}$ nicht vorgeifend bzgl. $\{B_t : t \in \mathbb{R}_0^+\}$.

Der Begriff „nicht vorgeifend“ entsteht aus der Unabhängigkeit der σ -Algebren \mathcal{A}_s und \mathcal{F}_s . Denn dies heißt ja gerade, daß die Filtration $\{\mathcal{F}_s : s \in [a, b]\}$ nicht „in die Zukunft“ sehen kann.

Definition 1.36 (progressiv meßbar).

Es sei $(\Omega^n, \mathcal{B}(\Omega^n), W^n)$ der Wiener-Raum, $a, b \in \mathbb{R}_0^+$, $a < b$. $\{B_t : t \in \mathbb{R}_0^+\}$ sei die n -dimensionale Brownsche Bewegung, und $\{\mathcal{F}_s : s \in [a, b]\}$ sei eine nicht vorgeifende Filtration bezüglich $\{B_t : t \in \mathbb{R}_0^+\}$. Die Funktion $f : [a, b] \times \Omega^n \rightarrow \mathbb{R}^n$ sei $\mathcal{B}_{[a,b]} \times \mathcal{B}(\Omega^n)$ - \mathcal{B}^n -meßbar. Es gelte

$$f_s : \Omega^n \rightarrow \mathbb{R}^n, \omega \mapsto f(s, \omega) \text{ ist } \mathcal{F}_s\text{-}\mathcal{B}^n\text{-meßbar für jedes } s \in [a, b].$$

Dann wird der durch die Funktion f definierte stochastische Prozeß $\{f_s : s \in [a, b]\}$ progressiv meßbar bezüglich $\{\mathcal{F}_s : s \in [a, b]\}$ genannt.

Definition 1.37 (Zufallszeit).

Sei $n \in \mathbb{N}$. Eine numerische Funktion $\tau : \Omega^n \rightarrow \overline{\mathbb{R}}$ heißt Zufallszeit, falls τ $\mathcal{B}(\Omega^n)$ - $\overline{\mathbb{B}}$ -meßbar ist.

Eine wichtige spezielle Zufallszeit gibt der folgende Satz an.

Satz 1.38 (spezielle Zufallszeit).

Sei $n \in \mathbb{N}$ und $\{X_t : t \geq 0\}$ ein n -dimensionaler stetiger stochastischer Prozeß. Dann ist die Funktion $g : \Omega^n \rightarrow \overline{\mathbb{R}}$,

$$\omega \mapsto \begin{cases} \inf(t \geq 0 : X^\omega(t) \in A) & \text{falls } \bigcup_{t \geq 0} (\{X^\omega(t)\} \cap A) \neq \emptyset \\ \infty & \text{sonst} \end{cases}$$

für jede offene oder abgeschlossene Menge $A \in \mathcal{B}^n$ eine Zufallszeit.

Anschaulich zeigt der Funktionswert $g(\omega)$ von g im Punkt ω die kürzeste Zeit, zu welcher der Pfad X^ω von $\{X_t : t \geq 0\}$ auf die Borel-Menge A trifft. Einen Beweis zu Satz 1.38 findet man in [Protter 90].

Satz 1.39.

Seien $n, p \in \mathbb{N}$ und $\{r_k\}_{k \in \mathbb{N}}$ eine Folge von p -dimensionalen Zufallsvariablen auf Ω^n . Ferner existiere für jedes $\omega \in \Omega^n$ der Grenzwert $\lim_{k \rightarrow \infty} r_k(\omega)$ in \mathbb{R}^p . Dann ist die Abbildung $\tau : \Omega^n \rightarrow \mathbb{R}^p, \omega \mapsto \lim_{k \rightarrow \infty} r_k(\omega)$, $\mathcal{B}(\Omega^n)$ - \mathcal{B}^p -meßbar.

Zum Abschluß dieses Abschnitts soll noch eine wichtige Eigenschaft eindimensionaler Brownscher Bewegungen genannt werden.

Satz 1.40 (Gesetze vom iterierten Logarithmus).

Es sei $\{B_t : t \geq 0\}$ eine eindimensionale Brownsche Bewegung. Dann gilt

$$\liminf_{t \rightarrow 0, t > 0} \frac{B_t}{\sqrt{2t \ln(\ln(\frac{1}{t}))}} = -1 \quad W^1\text{-fast sicher,}$$

$$\limsup_{t \rightarrow 0, t > 0} \frac{B_t}{\sqrt{2t \ln(\ln(\frac{1}{t}))}} = 1 \quad W^1\text{-fast sicher.}$$

1.3 Stochastische Integrale

In diesem Abschnitt soll ein kurzer Überblick über das stochastische Itô-Integral gegeben werden. Dabei wird nicht beabsichtigt, die Konstruktion des Integrals in allen Einzelheiten vorzuführen. Vielmehr soll ein allgemeiner Einblick in den Aufbau dieses stochastischen Integrals ermöglicht werden. Eine genauere Darstellung findet man in [Øksendal 98] und [Protter 90].

Die Konstruktion des Itô-Integrals beruht auf der Approximation stochastischer Prozesse durch elementare stochastische Prozesse.

Definition 1.41 (elementarer stochastischer Prozeß).

Es sei (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum, $[a, b]$ ein abgeschlossenes Intervall mit $a, b \in \mathbb{R}_0^+$, $a < b$ und $\{\mathcal{F}_s : s \in [a, b]\}$ eine Filtration in \mathcal{S} . Dann heißt der stochastische Prozeß $\{X_s : s \in [a, b]\}$ elementarer stochastischer Prozeß bezüglich $\{\mathcal{F}_s : s \in [a, b]\}$ auf (Ω, \mathcal{S}, P) , falls die folgenden Bedingungen erfüllt sind:

- i) $\{X_s : s \in [a, b]\}$ ist der Filtration $\{\mathcal{F}_s : s \in [a, b]\}$ adaptiert.
- ii) Es existieren $t_0, \dots, t_k \in [a, b]$ mit $a = t_0 < t_1 < \dots < t_k = b$ derart, daß

$$X_s(\omega) = X_{t_i}(\omega)$$

für alle $s \in [t_i, t_{i+1}[$, $i \in \{0, \dots, k-1\}$ und alle $\omega \in \Omega$.

iii) $E(X_{t_i}^2) < \infty$ für alle $i \in \{0, \dots, k\}$.

Elementare stochastische Prozesse sind also – bildlich gesprochen – solche stochastischen Prozesse, deren Pfade Treppenfunktionen sind.

Für elementare stochastische Prozesse definieren wir das stochastische Integral wie folgt:

Definition 1.42 (stochastisches Integral für elementare Prozesse).

Es sei (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum, $\{B_t : t \geq 0\}$ eine eindimensionale Brownsche Bewegung auf (Ω, \mathcal{S}, P) mit der kanonischen Filtration $\{\mathcal{C}_t : t \geq 0\}$ und $\{X_s : s \in [a, b]\}$ ein elementarer stochastischer Prozeß bezüglich $\{\mathcal{C}_s : s \in [a, b]\}$ mit $a, b \in \mathbb{R}_0^+$, $a < b$. Es sei ferner $\{t_0, t_1, \dots, t_k\}$ eine Zerlegung von $[a, b]$ derart, daß

$$X_s(\omega) = X_{t_i}(\omega)$$

für alle $s \in [t_i, t_{i+1}[$, $i \in \{0, \dots, k-1\}$ und alle $\omega \in \Omega$. Dann definieren wir die \mathcal{C}_b - \mathcal{B} -meßbare Zufallsvariable I_X mit

$$I_X : \Omega \rightarrow \mathbb{R}, \quad \omega \mapsto \sum_{i=1}^k X_{t_{i-1}}(\omega) (B_{t_i}(\omega) - B_{t_{i-1}}(\omega))$$

als stochastisches Integral von $\{X_s : s \in [a, b]\}$. Wir bezeichnen I_X mit

$$\int_a^b X_s dB_s.$$

Das stochastische Integral für beliebige stochastische Prozesse $\{X_s : s \in [a, b]\}$ mit $\sup_{a \leq s \leq b} E(\|X_s\|_2^2) < \infty$ erhält man durch eine Art Grenzübergang der Integrale der elementaren stochastischen Prozesse, die den stochastischen Prozeß $\{X_s : s \in [a, b]\}$ approximieren. Diese Approximation geschieht in drei Schritten: Zunächst wird $\{X_s : s \in [a, b]\}$ durch stochastische Prozesse mit gleichmäßig beschränkten Pfaden approximiert, und diese wiederum durch stetige stochastische Prozesse. Alle stochastischen Integrale verwenden diese Vorgehensweise. Der Unterschied zwischen den verschiedenen stochastischen Integralen liegt nun im dritten Schritt, der Approximation der stetigen stochastischen Prozesse durch elementare stochastische Prozesse.

Bevor wir uns allerdings der Approximation bei der Itô-Integraldefinition genauer zuwenden, nennen wir, für welche stochastischen Prozesse ein Integral überhaupt definiert werden kann.

Definition 1.43 (Menge der approximierbaren stochastischen Prozesse).

Seien (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum, $[a, b]$ ein abgeschlossenes Intervall mit $a, b \in \mathbb{R}_0^+$, $a < b$, und $\{F_s : s \in [a, b]\}$ eine Filtration in \mathcal{S} . Dann betrachten wir die Menge Π aller stochastischen Prozesse $\{Y_s : s \in [a, b]\}$ mit folgenden Eigenschaften:

- i) $Y_s : \Omega \rightarrow \mathbb{R}$ ist für jedes $s \in [a, b]$ \mathcal{F}_s - \mathcal{B} -meßbar.
- ii) Die Abbildung $Y : [a, b] \times \Omega \rightarrow \mathbb{R}$, $(s, \omega) \mapsto Y_s(\omega)$ ist $\mathcal{B}_{[a,b]} \times \mathcal{S}$ - \mathcal{B} meßbar.
- iii)

$$E \left(\int_{[a,b]} Y_s^2 d\lambda(s) \right) < \infty,$$

$$\text{wobei } \int_{[a,b]} Y_s^2 d\lambda(s) : \Omega \rightarrow \overline{\mathbb{R}}_0^+, \quad \omega \mapsto \int_{[a,b]} Y_s^2(\omega) d\lambda(s).$$

Die Menge Π wird als Menge der bezüglich $\{F_s : s \in [a, b]\}$ approximierbaren stochastischen Prozesse bezeichnet.

Um nun die Approximation bei der Itô-Integraldefinition etwas genauer zu betrachten, unterteilen wir das Zeitintervall in n Intervalle gleicher Länge. Wir wählen den elementaren stochastischen Prozeß derart, daß dessen Pfade die Pfade des stetigen stochastischen Prozesses im linken Teilintervallpunkt treffen (vgl. Abbildung 1.1). Wir erhalten eine Folge von elementaren stochastischen Prozessen als Approximation, wenn wir die Intervalllänge gegen 0 gehen lassen.

Bemerkung 1.44. Im Unterschied zum Riemann-Integral, bei dem die Wahl der Position des Aufeinandertreffens zwischen approximierender und zu approximierender Funktion keinerlei Auswirkungen auf das Integral hat, erhält man hier ein völlig anderes Integral, wenn man den elementaren stochastischen Prozeß so wählt, daß er die Pfade des stetigen stochastischen Prozesses an einer anderen Position trifft, z.B. erhält man für die Wahl des Treffens in der Mitte des Intervalls das Stratonovich-Integral.

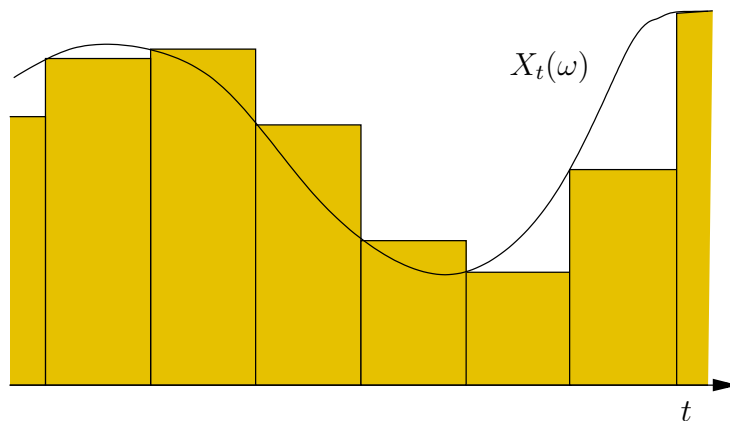


Abbildung 1.1: Konstruktion des Itô-Integrals

Das Integral für matrixwertige elementare stochastische Prozesse kann in ähnlicher Weise wie eben dargestellt definiert werden. Die Approximation matrixwertiger stochastischer Prozesse durch elementare stochastische Prozesse verläuft analog zu den oben vorgestellten Konstruktionen.

Wir bezeichnen das Itô-Integral I_t mit

$$\int_a^t X_s dB_s := I_t.$$

Das Itô-Integral ist seinerseits wieder ein stetiger stochastischer Prozess. Es findet Verwendung in der Definition stochastischer Differentialgleichungen, denen wir uns im nächsten Abschnitt zuwenden.

1.4 Stochastische Differentialgleichungen

Nun haben wir genügend Grundlagen aus der Wahrscheinlichkeitstheorie bereitgestellt, um uns mit stochastischen Differentialgleichungen beschäftigen zu können. Bevor wir uns den Eigenschaften stochastischer Differentialgleichungen, insbesondere der Existenz und Eindeutigkeit einer Lösung, zuwenden können, müssen wir zunächst definieren, was genau unter einer stochastischen Differentialgleichung zu verstehen ist. Wieder seien $n, m \in \mathbb{N}$.

Definition 1.45.

Es seien $a, b \in \mathbb{R}$ mit $a < b$ und $U \subset \mathbb{R}^n$. Weiter sei $\tilde{X} : \Omega^n \rightarrow \mathbb{R}^n$ eine Zufallsvariable mit $E(\|\tilde{X}\|_2^2) \leq \infty$. $\{B_t : t \in \mathbb{R}_0^+\}$ sei eine m -dimensionale Brownsche Bewegung. Die Funktion $h : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ sei $\mathcal{B}_{[a,b]} \times \mathcal{B}^n$ - \mathcal{B}^n -meßbar, die Funktion $G : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^{(n,m)}$ sei $\mathcal{B}_{[a,b]} \times \mathcal{B}^n$ - $\mathcal{B}^{n \cdot m}$ -meßbar. Dann heißt

$$X_t = \tilde{X} + \int_a^t h(s, X_s) ds + \int_a^t G(s, X_s) dB_s, \quad t \in [a, b] \quad (1.1)$$

eine Itô-stochastische Differentialgleichung.

Anstelle von (1.1) verwenden wir häufig die Schreibweise

$$dX_t = h(t, X_t) dt + G(t, X_t) dB_t, \quad t \in [a, b], \quad X_a = \tilde{X}.$$

Wir nennen $h(t, X_t) dt$ den Driftkoeffizienten, $G(t, X_t) dB_t^n$ bezeichnen wir als den Diffusionskoeffizienten.

Nun müssen wir definieren, was unter einer Lösung von (1.1) zu verstehen ist. Darüber gibt die folgende Definition Aufschluß.

Definition 1.46 (Lösung der stochastischen Differentialgleichung).

Ein stochastischer Prozeß $\{X_s : s \in [a, b]\}$ heißt Lösung von (1.1), falls

$$i) \sup_{a \leq s \leq b} E(\|X_s\|_2^2) < \infty.$$

ii) Für alle $a \leq t \leq b$ ist (1.1) erfüllt, d.h. insbesondere die Integrale auf der rechten Seite von (1.1) existieren.

Eine Lösung heißt zulässig, falls $\{X_s : s \in [a, b]\}$ stetig ist.

Für die stochastische Differentialgleichung nach dem Itô-Kalkül können wir eine Bedingung zur Existenz einer Lösung angeben.

Satz 1.47 (Existenz- und Eindeutigkeitsatz).

Es sei (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum, $a, b \in \mathbb{R}_0^+$, $a < b$, $k \in \mathbb{N}$ und $\{B_t : t \in \mathbb{R}_0^+\}$ eine m -dimensionale Brownsche Bewegung. Sei ferner $\{\mathcal{F}_s : s \in [a, b]\}$ eine nicht vorgeifende Filtration in \mathcal{S} bezüglich $\{B_t : t \in \mathbb{R}_0^+\}$, $\tilde{X} : \Omega \rightarrow \mathbb{R}^n$ eine \mathcal{F}_a - \mathcal{B}^n -messbare Zufallsvariable mit $E(\|\tilde{X}\|_F^2) < \infty$ und $h : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ und $G : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^{n,m}$ stetige Funktionen mit folgender Eigenschaft: Es existiert ein $L > 0$ mit

$$\|h(s, x) - h(s, y)\|_F + \|G(s, x) - G(s, y)\|_F \leq L\|x - y\|_F \quad (1.2)$$

für alle $x, y \in \mathbb{R}^n$ und alle $s \in [a, b]$.

Dann gibt es einen stochastischen Prozeß $\{X_s : s \in [a, b]\}$ bestehend aus \mathcal{F}_s - \mathcal{B}^n -meßbaren Zufallsvariablen $X_s : \Omega \rightarrow \mathbb{R}^n$ mit folgenden Eigenschaften:

i) $\{X_s : s \in [a, b]\}$ ist eine Lösung der stochastischen Differentialgleichung

$$dX_s = h(s, X_s) ds + G(s, X_s) dB_s, \quad s \in [a, b], \quad X_a = \tilde{X}. \quad (1.3)$$

ii) $X : [a, b] \times \Omega \rightarrow \mathbb{R}^n$, $(s, \omega) \mapsto X_s(\omega)$ ist progressiv meßbar und $\{X_s : s \in [a, b]\}$ besitzt stetige Pfade.

$$iii) \sup_{a \leq s \leq b} E\|X_s\|_F^2 < \infty.$$

iv) Für jede weitere Lösung $\{Y_s : s \in [a, b]\}$ mit den obigen Eigenschaften gilt:

$$P \left(\left\{ \omega \in \Omega : \sup_{a \leq s \leq b} \|X_s(\omega) - Y_s(\omega)\|_F^2 = 0 \right\} \right) = 1$$

(P -fast sichere Eindeutigkeit).

Zum Abschluß dieses Abschnitts betrachten wir noch eine wichtige Gleichung in der Theorie der stochastischen Differentialgleichungen – die Fokker-Planck-Gleichung, deren Geltung die Vorwärtsgleichung von Kolmogorov sichert.

Satz 1.48 (Vorwärtsgleichung von Kolmogorov).

Seien (Ω, \mathcal{S}, P) ein Wahrscheinlichkeitsraum, $a, b \in \mathbb{R}_0^+$, $a < b$, $\{B_t : t \geq 0\}$ eine m -dimensionale Brownsche Bewegung auf (Ω, \mathcal{S}, P) und $\{F_s : s \in [a, b]\}$ eine nicht vorgreifende Filtration in \mathcal{S} bezüglich $\{B_t : t \geq 0\}$. Seien ferner $\tilde{X} : \Omega \rightarrow \mathbb{R}^n$ eine \mathcal{F}_a - \mathcal{B}^n -meßbare Zufallsvariable mit $E(\|\tilde{X}\|_F^2) < \infty$, deren Verteilung durch die Lebesgue-Dichte $p_a : \mathbb{R}^n \rightarrow \mathbb{R}$ gegeben ist. Weiter seien $h : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ und $G : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^{n,m}$ stetige Funktionen, die der Lipschitzbedingung (1.2) genügen. Man betrachtet die Lösung $\{X_s : s \in [a, b]\}$ der stochastischen Differentialgleichung

$$dX_s = h(s, X_s) ds + G(s, X_s) dB_s, \quad s \in [a, b], \quad X_a = \tilde{X}.$$

Für jede Zufallsvariable X_s , $s \in [a, b]$, sei die Verteilung von X_s durch eine Lebesgue-Dichte $p_s : \mathbb{R}^n \rightarrow \mathbb{R}$ gegeben.

Setzt man die Existenz der Ableitungen $\frac{\partial p}{\partial t}$, $\frac{\partial(h_i, p)}{\partial x_i}$ und $\frac{\partial^2}{\partial x_i \partial x_j}(GG^T)_{ij} \cdot p$ für alle $i, j = 1, \dots, n$ als stetige Funktionen voraus, wobei $p : (a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}$, $(t, x) \mapsto p_t(x)$, so erfüllt die Funktion p die folgende Vorwärtsgleichung von Kolmogorov, die auch als Fokker-Planck-Gleichung bezeichnet wird:

$$\frac{\partial p}{\partial t} + \sum_{i=1}^n \frac{\partial}{\partial x_i}(h_i(t, x)p(t, x)) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} ((G(t, x)G(t, x)^T)_{ij} p(t, x)) = 0 \quad (1.4)$$

für alle $(t, x) \in (a, b] \times \mathbb{R}^n$ mit der Endbedingung

$$\lim_{t \rightarrow a} p(t, x) = p_a(x)$$

für alle $x \in \mathbb{R}^n$.

1.5 Markovketten

Da ein wesentlicher Bestandteil dieser Arbeit aus der Untersuchung eines bestimmten Markovprozesses besteht, sollen in diesem Kapitel die dafür entscheidenden Sätze und Definitionen aus der Theorie der Markovketten genannt werden. Sie sind weitestgehend dem Buch [Meyn Tweedie 96] entnommen.

Im folgenden sei X ein beliebiger topologischer Raum und $\mathcal{B}(X)$ die zugehörige Borelsche σ -Algebra auf X . Zunächst ist die Definition eines Markovprozesses mit den entsprechenden Übergangswahrscheinlichkeiten zu klären.

Definition 1.49 (Übergangswahrscheinlichkeitskern).

Es sei $P = \{P(x, A) : x \in X, A \in \mathcal{B}(X)\}$ gegeben mit

- i) für jedes $A \in \mathcal{B}(X)$ ist $P(\cdot, A)$ eine nicht-negative meßbare Funktion auf X ,

ii) für jedes $x \in X$ ist $P(x, \cdot)$ ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(X)$.

Dann heißt P Übergangswahrscheinlichkeitskern.

Mit Hilfe eines Übergangswahrscheinlichkeitskerns kann nun eine Markovkette definiert werden.

Definition 1.50 (Markovkette).

Ein stochastischer Prozeß $\Phi := \{\Phi_t : t \in \mathbb{N}_0\}$ definiert auf $(X, \mathcal{B}(X))$ heißt zeitlich-homogener Markovprozeß mit Anfangsverteilung μ und Übergangswahrscheinlichkeitskern P , wenn die Verteilung von Φ

$$\begin{aligned} P(\Phi_0 \in A_0, \Phi_1 \in A_1, \dots, \Phi_n \in A_n) \\ = \int_{A_0} \cdots \int_{A_{n-1}} \mu(dy_0) P(y_0, dy_1) \cdots P(y_{n-1}, A_n) \end{aligned}$$

für alle $n \in \mathbb{N}$ erfüllt.

Es bezeichne im folgenden $P_\mu(\Phi_1 \in A_1, \dots, \Phi_n \in A_n)$ die Wahrscheinlichkeit des Ereignisses $\{\Phi_1 \in A_1, \dots, \Phi_n \in A_n\}$, wobei als Verteilung von Φ_0 , also als Anfangsverteilung, μ gewählt wurde. Weiterhin verwenden wir im folgenden die abkürzende Schreibweise $P_x = P_{\delta_x}$, wobei δ_x das Dirac-Maß mit Gewicht auf x darstelle, also

$$\begin{aligned} \delta_x : \mathcal{B}(X) &\rightarrow [0, 1] \\ A &\mapsto \begin{cases} 1 & \text{falls } x \in A \\ 0 & \text{sonst.} \end{cases} \end{aligned}$$

Analog zur Definition von P_x definieren wir E_x als den Erwartungswert bzgl. des Maßes P_x . Wir werden im folgenden häufig die Ausdrucksweise „ P_* -fast sicher“ verwenden. Diese Bezeichnung sagt aus, daß eine Aussage P_μ -fast sicher für jede beliebige Anfangsverteilung μ gilt.

Durch die Übergangswahrscheinlichkeiten läßt sich sehr bequem die Übergangswahrscheinlichkeit nach n Schritten, $P^n(x, A) := P_x(\Phi_n \in A)$, mit den Chapman-Kolmogorov-Gleichungen darstellen.

Satz 1.51 (Gleichungen von Chapman-Kolmogorov).

Es sei Φ eine Markovkette. Dann gilt

$$P^n(x, A) = \int_X P^m(x, dy) P^{n-m}(y, A), \quad x \in X, \quad A \in \mathcal{B}(X)$$

für jedes m mit $0 \leq m \leq n$.

Zu einer gegebenen Markovkette betrachtet man die folgenden wichtigen Stopzeiten.

Definition 1.52 (Besetzungszeit, Rückkehrzeit, Zeitpunkt des ersten Eintreffens in einer Menge A).

Es sei Φ eine Markovkette.

i) Für jede Menge $A \in \mathcal{B}(X)$ bezeichne die Besetzungszeit η_A die Anzahl der Besuche von Φ in A . Sie ist gegeben durch

$$\eta_A := \sum_{n=1}^{\infty} \mathbf{1}_{\{\Phi_n \in A\}}.$$

ii) Für jede Menge $A \in \mathcal{B}(X)$ heißen die Zufallsvariablen

$$\tau_A := \begin{cases} \min\{n \geq 1 : \Phi_n \in A\} & \text{falls } \{n \geq 1 : \Phi_n \in A\} \neq \emptyset \\ \infty & \text{sonst} \end{cases}$$

und

$$\sigma_A := \begin{cases} \min\{n \geq 0 : \Phi_n \in A\} & \text{falls } \{n \geq 0 : \Phi_n \in A\} \neq \emptyset \\ \infty & \text{sonst} \end{cases}$$

die erste Rückkehrzeit bzw. der Zeitpunkt des ersten Eintreffens in der Menge A .

Als eine für die Belange dieser Arbeit sehr wichtige Eigenschaft von Markovketten stellt sich die Irreduzibilität heraus, die im wesentlichen sicherstellt, daß alle Teile des Zustandsraumes unabhängig vom Startpunkt von der Markovkette erreicht werden können. Dabei bezeichnen wir mit $L(x, A)$ die Wahrscheinlichkeit einer endlichen Rückkehrzeit zur Menge A ausgehend vom Punkt x , also

$$L(x, A) := P_x(\tau_A < \infty),$$

und mit $U(x, A)$ den Erwartungswert der Besetzungszeit der Menge A ausgehend vom Punkt x , also

$$\begin{aligned} U(x, A) &:= \sum_{n=1}^{\infty} P^n(x, A) \\ &= E_x[\eta_A], \end{aligned}$$

wobei die Gleichheit aus dem Satz von Beppo-Levi folgt.

Definition 1.53 (φ -Irreduzibilität).

Wir nennen eine Markovkette Φ irreduzibel, wenn ein nicht-triviales Maß φ auf $\mathcal{B}(X)$ existiert, so daß für alle $A \in \mathcal{B}(X)$ mit $\varphi(A) > 0$ folgt $L(x, A) > 0$ für alle $x \in X$. In diesem Fall nennen wir die Markovkette auch φ -irreduzibel.

Satz 1.54.

Äquivalent zur φ -Irreduzibilität ist das folgende Kriterium: Für alle $A \in \mathcal{B}(X)$ mit $\varphi(A) > 0$ folgt $U(x, A) > 0$ für alle $x \in X$.

Aus der Irreduzibilität folgt bereits die Existenz eines maximalen Irreduzibilitätsmaßes ψ (maximal bzgl. der Absolutstetigkeit):

Satz 1.55 (Existenz eines maximalen Irreduzibilitätsmaßes).

Ist Φ φ -irreduzibel für ein Maß φ , so existiert ein Wahrscheinlichkeitsmaß ψ auf $\mathcal{B}(X)$, so daß gilt:

- i) Φ ist ψ -irreduzibel;
- ii) für jedes andere Maß φ' ist die Markovkette Φ φ' -irreduzibel genau dann, wenn φ' absolut stetig ist bzgl. ψ ;
- iii) aus $\psi(A) = 0$ folgt $\psi(\{y \in X : L(y, A) > 0\}) = 0$;
- iv) das Wahrscheinlichkeitsmaß ψ ist äquivalent zu

$$\psi'(A) := \int_X \varphi'(dy) K_{a_{\frac{1}{2}}}(y, A)$$

für jedes endliche Irreduzibilitätsmaß φ' , wobei

$$K_{a_{\frac{1}{2}}}(y, A) := \sum_{n=0}^{\infty} P^n(y, A) 2^{-(n+1)}, \quad y \in X, \quad A \in \mathcal{B}(X). \quad (1.5)$$

Im folgenden bezeichne ψ immer ein maximales Irreduzibilitätsmaß, das die Eigenschaften aus Satz 1.55 erfüllt. Statt ψ -Irreduzibilität schreiben wir auch abkürzend Irreduzibilität. Desweiteren sei

$$\mathcal{B}^+(X) := \{A \in \mathcal{B}(X) : \psi(A) > 0\}.$$

Als sehr nützlich erweisen sich die Eigenschaften *unbedeutend* und *klein* für Mengen $A \in \mathcal{B}(X)$, die nun vorgestellt werden sollen.

Definition 1.56 (unbedeutende Menge).

Eine Menge $C \in \mathcal{B}(X)$ heißt *unbedeutend*, wenn ein $m > 0$ und ein nicht-triviales Maß ν_m auf $\mathcal{B}(X)$ existieren, so daß für alle $x \in C$, $B \in \mathcal{B}(X)$ gilt

$$P^m(x, B) \geq \nu_m(B). \quad (1.6)$$

Ist Ungleichung (1.6) erfüllt, so nennen wir C auch (m, ν_m) -unbedeutend.

Bevor wir uns der Definition von kleinen Mengen widmen, führen wir die folgende Bezeichnung ein (vgl. auch die Definition von $K_{a_{\frac{1}{2}}}$ in (1.5))

$$K_a(x, A) := \sum_{n=0}^{\infty} P^n(x, A) a(n), \quad x \in X, A \in \mathcal{B}(X),$$

wobei $a(n)$ ein Wahrscheinlichkeitsmaß auf \mathbb{N}_0 und P^0 die Anfangsverteilung seien.

Definition 1.57 (kleine Menge).

Wir nennen eine Menge $C \in \mathcal{B}(X)$ klein, falls eine Wahrscheinlichkeitsverteilung a auf \mathbb{N}_0 und ein nicht-triviales Maß ν_a auf $\mathcal{B}(X)$ existieren, so daß

$$K_a(x, B) \geq \nu_a(B) \quad \text{für alle } x \in C, B \in \mathcal{B}(X). \quad (1.7)$$

Ist Ungleichung (1.7) erfüllt, so heißt die Menge C auch (a, ν_a) -klein.

Wählt man für a das Dirac-Maß δ_m mit Gewicht auf m , so sieht man sofort, daß jede unbedeutende Menge auch klein ist.

Für eine irreduzible Markovkette lassen sich immer ein $m \geq 1$ und eine Menge $C \in \mathcal{B}(X)$ finden, so daß C ν_m -unbedeutend ist.

Satz 1.58 (Existenz einer unbedeutenden Menge).

Ist Φ ψ -irreduzibel, so existiert für jede Menge $A \in \mathcal{B}^+(X)$ ein $m \geq 1$, ein nicht-triviales Maß ν_m und eine (m, ν_m) -unbedeutende Menge $C \subseteq A$, so daß $C \in \mathcal{B}^+(X)$ und $\nu_m(C) > 0$.

Lemma 1.59.

- i) Es sei $C \in \mathcal{B}^+(X)$ (n, ν_n) -unbedeutend und $D \in \mathcal{B}(X)$. Für jedes $x \in D$ gelte $P^m(x, C) \geq \delta > 0$. Dann ist D $(m+n, \nu_{m+n})$ -unbedeutend, wobei ν_{m+n} ein Vielfaches von ν_n ist.
- ii) Es sei Φ ψ -irreduzibel. Ist $C \in \mathcal{B}^+(X)$ eine (n, ν_n) -unbedeutende Menge, so können wir ein $M \in \mathbb{N}$ und ein Maß ν_M finden, so daß C (M, ν_M) -unbedeutend ist und $\nu_M(C) > 0$.

Um die „Stabilität“ einer Markovkette bewerten zu können, muß unter anderem das zyklische Verhalten jenes Prozesses untersucht werden.

Dazu betrachten wir für eine beliebige Menge $C \in \mathcal{B}(X)$ und ein beliebiges Maß ν auf $(X, \mathcal{B}(X))$ die Menge

$$E_C^\nu := \{n \geq 1 : \text{Die Menge } C \text{ ist } (n, \nu_n)\text{-unbedeutend mit } \nu_n = c_n \nu \text{ für ein } c_n > 0\}.$$

Wählt man speziell eine unbedeutende Menge $C \in \mathcal{B}^+(X)$, kann nach Lemma 1.59 ein $M \in \mathbb{N}$ gefunden werden, so daß $\nu_M(C) > 0$. Zur Vereinfachung schreiben wir nur ν statt ν_M . Dann ist die Menge E_C^ν nicht leer.

Definition 1.60 (*d*-Zyklizität).

Es sei Φ eine ψ -irreduzible Markovkette auf X . Φ heißt *d*-zyklisch, falls disjunkte Mengen $D_1, \dots, D_d \in \mathcal{B}(X)$ existieren mit den folgenden Eigenschaften:

- i) für $x \in D_i$ gilt $P(x, D_{i+1}) = 1$, $i = 0, \dots, d-1 \pmod{d}$;
- ii) Die Menge $N = [\cup_{i=1}^d D_i]^c$ hat das ψ -Maß 0.

Mit Hilfe der Menge E_C^ν kann das zyklische Verhalten der Markovkette beschrieben werden.

Satz 1.61 (*d*-Zykel).

Es sei Φ eine ψ -irreduzible Markovkette auf X . Desweiteren sei $C \in \mathcal{B}^+(X)$ eine (M, ν) -unbedeutende Menge mit $\nu(C) > 0$ und d der größte gemeinsame Teiler der Menge E_C^ν . Dann existieren disjunkte Mengen $D_1, \dots, D_d \in \mathcal{B}(X)$, so daß Φ *d*-zyklisch bzgl. der Mengen D_1, \dots, D_d ist. Der *d*-Zykel $\{D_i\}$ ist maximal in dem Sinne, daß für jedes andere $\{d', D'_k : k = 1, \dots, d'\}$, das die Eigenschaften (i) und (ii) aus 1.60 erfüllt, d' d teilt, während im Falle $d' = d$ – eventuell durch Umordnung der Indizes – $D'_i = D_i$ ψ -fast sicher für alle $i \in \{1, \dots, d\}$ gilt.

Nun sind wir in der Lage, periodische und aperiodische Markovketten zu definieren.

Definition 1.62 (Periode, aperiodisch).

Es sei Φ eine ψ -irreduzible Markovkette. Das größte d , für das ein *d*-Zykel für Φ auftritt, heißt die Periode von Φ . Desweiteren heißt die Kette Φ

- i) aperiodisch, falls $d = 1$ gilt und
- ii) streng aperiodisch, falls eine $(1, \nu_1)$ -unbedeutende Menge A mit $\nu_1(A) > 0$ existiert.

Aus der Definition der Menge E_C und aus der Maximalität aus Satz 1.61 folgt, daß die Eigenschaft der strengen Aperiodizität bereits die Aperiodizität beinhaltet.

Bevor wir nun die Existenz eines invarianten Wahrscheinlichkeitsmaßes zu der Markovkette Φ untersuchen können, benötigen wir noch die Eigenschaft der Rekurrenz und der Harris-Rekurrenz.

Definition 1.63 (Rekurrenz).

Eine Menge $A \in \mathcal{B}(X)$ heißt rekurrent, falls gilt

$$E_x[\eta_A] = \infty \quad \forall x \in A.$$

Eine Markov-Kette heißt rekurrent, falls jede Menge $A \in \mathcal{B}(X)$ rekurrent ist.

Tatsächlich benötigen wir hier die stärkere Eigenschaft der Harris-Rekurrenz.

Definition 1.64 (Harris-Rekurrenz).

Eine Menge $A \in \mathcal{B}(X)$ heißt Harris-rekurrent, falls

$$Q(x, A) := P_x(\eta_A = \infty) = 1 \quad \forall x \in A.$$

Eine Markov-Kette Φ heißt Harris-rekurrent, falls sie irreduzibel ist und jede Menge $A \in \mathcal{B}^+(X)$ Harris-rekurrent ist.

Der entscheidende Unterschied zwischen Rekurrenz und Harris-Rekurrenz besteht also darin, daß bei Harris-Rekurrenz nicht nur der Erwartungswert der Anzahl des Eintreffens der Markov-Kette in der Menge A unendlich sein soll für jeden Ausgangspunkt $x \in A$, sondern daß tatsächlich das Ereignis, die Menge A unendlich oft zu treffen, P -fast sicher eintritt.

Erinnern wir uns an die Bezeichnungen

$$L(x, A) := P_x(\tau_A < \infty), \quad U(x, A) := \sum_{n=1}^{\infty} P^n(x, A) = E_x(\eta_A),$$

so zeigt der folgende Satz eine vereinfachende Implikationen im Falle eines abzählbaren Zustandsraumes.

Satz 1.65 (spezielle Äquivalenz im diskreten Fall).

Es sei X abzählbar. Dann gilt für jedes $x \in X$, daß $U(x, \{x\}) = \infty$ dann und nur dann gilt, wenn $L(x, \{x\}) = 1$.

Allgemein genügt zur Feststellung der Harris-Rekurrenz einer Menge $A \in \mathcal{B}(X)$, daß $L(x, A) \equiv 1$ für $x \in A$ gilt.

Satz 1.66.

Angenommen, es existiert eine Menge $A \in \mathcal{B}(X)$ mit $L(x, A) \equiv 1$, $x \in A$. Dann gilt $Q(x, A) = L(x, A)$ für alle $x \in X$, und insbesondere ist A Harris-rekurrent.

Rekurrenz und Harris-Rekurrenz sind schwer direkt zu überprüfen, da die Berechnung dieses Erwartungswertes bzw. dieser Wahrscheinlichkeit die Berechnung aller Übergangswahrscheinlichkeiten P^n erfordert. Da dies eine in der Praxis nur schwer lösbare Aufgabe ist, stellen wir Rekurrenz bzw. Harris-Rekurrenz mittels eines „Drift-Kriteriums“ fest. Hierzu definieren wir zu einer meßbaren Funktion $V : X \rightarrow \mathbb{R}^+$ den Drift-Operator Δ durch

$$\Delta V(x) := \int P(x, dy)V(y) - V(x), \quad x \in X.$$

Definition 1.67 (unbeschränkt außerhalb von kleinen Mengen).

Wir nennen eine meßbare Funktion $V : X \rightarrow \mathbb{R}^+$ unbeschränkt außerhalb von kleinen Mengen für Φ , wenn für jedes $n < \infty$ die „Sublevel-Menge“

$$C_V(n) := \{y \in X : V(y) \leq n\}$$

klein ist.

Das Kriterium, das wir nutzen werden, um Harris-Rekurrenz nachzuweisen, ist das folgende:

Drift-Kriterium für Harris-Rekurrenz

(V1) Es existiert eine positive Funktion V und eine Menge $C \in \mathcal{B}(X)$, so daß

$$\Delta V(x) \leq 0, \quad x \in C^c, \quad (1.8)$$

erfüllt ist.

Satz 1.68 impliziert, daß die Gültigkeit von (1.8) für eine unbedeutende Menge C und eine außerhalb von kleinen Mengen unbeschränkte Funktion V auch tatsächlich Harris-Rekurrenz impliziert.

Satz 1.68 (Drift-Kriterium für Harris-Rekurrenz).

Angenommen Φ ist ψ -irreduzibel. Weiter gebe es eine kleine Menge $C \subset X$ und eine Funktion V , welche unbeschränkt außerhalb von kleinen Mengen ist, so daß (1.8) erfüllt ist. Dann ist Φ Harris-rekurrent, also insbesondere auch rekurrent.

Nun können wir uns der Existenz eines invarianten Wahrscheinlichkeitsmaßes zuwenden.

Definition 1.69 (invariantes Maß).

Ein σ -endliches, nicht-triviales Maß π auf $\mathcal{B}(X)$ mit der Eigenschaft

$$\pi(A) = \int_X \pi(dx) P(x, A), \quad A \in \mathcal{B}(X), \quad (1.9)$$

heißt invariant.

Markovketten, die ein invariantes Wahrscheinlichkeitsmaß besitzen, werden als positive Ketten bezeichnet.

Definition 1.70 (positiv, positiv-Harris).

Angenommen Φ ist ψ -irreduzibel und besitzt ein invariantes Wahrscheinlichkeitsmaß π . Dann heißt Φ positive Kette. Ist Φ Harris-rekurrent und positiv, so heißt Φ positive Harris-Kette.

Satz 1.71. *Ist die Markovkette Φ positiv, so ist sie auch rekurrent.*

Definition 1.72 (subinvariantes Maß).

Ist μ ein σ -endliches, nicht-triviales Maß und genügt der Bedingung

$$\mu(A) \geq \int_X \mu(dx)P(x, A), \quad A \in \mathcal{B}(X),$$

so wird μ subinvariant genannt.

Ist ein subinvariantes Maß μ endlich, so ist es auch invariant, denn für $A \in \mathcal{B}(X)$ erhalten wir folgende Abschätzung

$$\begin{aligned} \mu(A) &\geq \int_X \mu(dx)P(x, A) \\ &= \int_X \mu(dx)(1 - P(x, A^c)) \\ &= \mu(X) - \int_X \mu(dx)P(x, A^c) \\ &\geq \mu(X) - \mu(A^c) \\ &= \mu(A) \end{aligned}$$

und daraus die Gleichheit (1.9).

Damit die Markovkette ein invariantes Maß besitzt, genügt die Eigenschaft der Rekurrenz:

Satz 1.73.

Die Markovkette Φ sei rekurrent. Dann besitzt Φ ein eindeutiges (bis auf skalare Multiplikation) subinvariantes Maß, das invariant ist.

Um sicher zu stellen, daß dieses Maß endlich und somit als Wahrscheinlichkeitsmaß gewählt werden kann, benötigen wir noch eine zusätzliche Eigenschaft.

Satz 1.74 (Existenz eines invarianten Wahrscheinlichkeitsmaßes).

Angenommen Φ ist ψ -irreduzibel und μ ein subinvariantes Maß. Dann ist μ endlich und somit Φ positiv und damit rekurrent, wenn für eine kleine Menge $C \in \mathcal{B}^+(X)$

$$\sup_{y \in C} E_y[\tau_C] < \infty \tag{1.10}$$

gilt. Die Markovkette ist positiv-Harris, wenn zusätzlich

$$E_x[\tau_C] < \infty, \quad x \in X,$$

erfüllt ist.

Die Eigenschaft (1.10) heißt auch Regularitätseigenschaft.

Definition 1.75 (regulär).

Eine Menge $C \in \mathcal{B}^+(X)$ heißt regulär, falls Φ ψ -irreduzibel ist und

$$\sup_{x \in C} E_x[\tau_B] < \infty, \quad B \in \mathcal{B}^+(X).$$

Auch für die Überprüfung der Regularitätseigenschaft kann ein Drift-Kriterium herangezogen werden.

Zweites Drift-Kriterium

(V2) Für eine Menge $C \in \mathcal{B}(X)$, eine Konstante $b < \infty$ und eine Funktion $V : X \rightarrow [0, \infty]$ gilt

$$\Delta V(x) \leq -1 + b\mathbf{1}_C(x), \quad x \in X. \quad (1.11)$$

Satz 1.76.

Es sei $C \in \mathcal{B}(X)$, und V genüge Bedingung (1.11). Dann gilt

$$E_x[\tau_C] \leq V(x) + b\mathbf{1}_C(x)$$

für alle $x \in X$. Daher ist, falls C klein und V auf C beschränkt, Φ positiv-Harris.

Ist die Markovkette positiv-Harris, so konvergieren die Verteilungen im n -ten Schritt, wie der folgende Satz aussagt.

Satz 1.77 (Konvergenz der Übergangswahrscheinlichkeiten).

Angenommen, die Markovkette Φ ist aperiodisch und Harris-rekurrent, und es existiere ein invariantes Maß π . Die Markovkette Φ ist positiv-Harris genau dann, wenn ein eindeutiges invariantes Wahrscheinlichkeitsmaß π existiert, so daß für jeden Startpunkt $x \in X$ gilt

$$\sup_{A \in \mathcal{B}(X)} |P^n(x, A) - \pi(A)| \rightarrow 0 \quad \text{für } n \rightarrow \infty.$$

Der für die Zwecke dieser Arbeit bedeutendste Satz - wie sich später herausstellen wird - sagt aus, daß bereits alle Informationen zum invarianten Wahrscheinlichkeitsmaß P_x -fast sicher für jeden Startpunkt $x \in X$ auf einem einzigen Pfad enthalten sind.

Satz 1.78.

Für die Markovkette $\Phi = \{\Phi_t\}$ mit Parameter $t \in \mathbb{N}$ existiere ein invariantes Wahrscheinlichkeitsmaß π . Dann sind die folgenden Aussagen äquivalent:

i) Φ ist positiv-Harris.

ii) Für alle Funktionen $f \in L_1(X, \mathcal{B}(X), \pi)$ gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n f(\Phi_k) = \int f d\pi \quad P_*\text{-fast sicher}$$

für jeden Anfangszustand $x \in X$.

Kapitel 2

Diskrete globale Optimierung

2.1 Der Metropolis-Algorithmus

Wir betrachten nun einen Algorithmus zur Auffindung von globalen Minima auf endlichen Mengen. Die Theorie dieser diskreten globalen Optimierung mittels des Metropolis-Algorithmus beruht auf einem Verfahren, das Metropolis et. al. [Metropolis 53] zur effizienten Simulation der Entwicklung eines Festkörpers im Hitzebad zum thermischen Gleichgewicht hin vorstellten. Ca. 30 Jahre später nutzen Kirkpatrick et. al. [Kirkpatrick 82] und Černý [Černý 85] unabhängig voneinander dieses Verfahren zur Minimierung der Kostenfunktion eines kombinatorischen Optimierungsproblems. Die Idee dieser Vorgehensweise liegt im wesentlichen in der Erweiterung von den reinen Abstiegsverfahren hin zu einem Verfahren, das Abstiegsverfahren und Zufall kombiniert. Dieser Algorithmus trägt in der Literatur den Namen *Metropolis-Algorithmus*.

Der Metropolis-Algorithmus erzeugt von einem gegebenen Startpunkt aus einen zufälligen Punkt aus der Nachbarschaft dieses Startpunkts mittels einer Gleichverteilung und akzeptiert diesen als neuen Punkt, wenn der Zielfunktionswert des neuen Punktes verglichen mit dem des Startpunkts kleiner ist. Sollte er größer als dieser sein, akzeptiert man den neuen Punkt mit einer positiven Wahrscheinlichkeit, die von der Differenz der beiden Zielfunktionswerte und einem Kontrollparameter c abhängt. In dieser Arbeit nehmen wir die Nachbarschaftsstruktur als gegeben an.

Um die Vorgehensweise genauer erläutern zu können, betrachten wir das kombinatorische Optimierungsproblem: Es sei ein endlicher Zustandsraum R mit dazugehöriger Nachbarschaftsstruktur $\{R_i : i \in R\}$, wobei R_i die Menge der Nachbarn des Punktes i für jedes $i \in R$ darstellt, gegeben. Das Optimierungsproblem ist dann das folgende: Finde ein $i^* \in R$ mit der Eigenschaft

$$f(i^*) \leq f(i) \quad \forall i \in R.$$

Zu diesem Optimierungsproblem definieren wir zunächst Generierungs- und Akzeptanzwahrscheinlichkeiten.

Definition 2.1 (Generierungswahrscheinlichkeit).

Es sei ein beliebiger Anfangszustand $i \in R$ gegeben. Dann wird ein beliebiger anderer Punkt $j \in R$ gemäß der Wahrscheinlichkeit

$$G_{ij} := \frac{1}{|R_i|} \mathbf{1}_{R_i}(j) \quad \forall i, j \in S \quad (2.1)$$

erzeugt. Dabei bezeichne $\mathbf{1}_{R_i}$ die Indikatorfunktion auf der Menge R_i , also

$$\begin{aligned} \mathbf{1}_{R_i} : R &\rightarrow \{0, 1\} \\ j &\mapsto \begin{cases} 1 & \text{falls } j \in R_i \\ 0 & \text{sonst.} \end{cases} \end{aligned}$$

Die Wahrscheinlichkeit G_{ij} bezeichnet man als Generierungswahrscheinlichkeit des Metropolis-Algorithmus.

Wir nehmen im folgenden an, die Generierungswahrscheinlichkeiten seien symmetrisch, d.h. für alle $i, j \in R$ gilt

$$G_{ij} = G_{ji}.$$

Für die Nachbarschaftsstruktur $\{R_i\}_{i \in R}$ bedeutet dies

$$i \in R_j \Leftrightarrow j \in R_i \quad \text{und} \quad |R_i| = |R_j| \quad \text{für alle } i, j \in R.$$

Zunächst ermittelt der Metropolis-Algorithmus gemäß der Generierungswahrscheinlichkeit einen neuen Punkt aus der Nachbarschaft des Startpunkts. Dieser wird anschließend auf die Erfüllung einer Akzeptanzbedingung geprüft:

Definition 2.2 (Akzeptanzwahrscheinlichkeit).

Ein mittels der Generierungswahrscheinlichkeit ausgehend vom Punkt $i \in R$ erzeugter Punkt $j \in R$ wird mit der Wahrscheinlichkeit

$$A_{ij}(c) := \begin{cases} 1 & \text{falls } f(j) \leq f(i) \\ \exp\left(-\frac{f(j)-f(i)}{c}\right) & \text{sonst} \end{cases} \quad \forall i \in R, j \in R_i \quad (2.2)$$

als neue Lösung akzeptiert, wobei $c \in \mathbb{R}^+$ konstant gewählt sei. Die Größe $A_{ij}(c)$ wird als Akzeptanzwahrscheinlichkeit des Metropolis-Algorithmus bezeichnet.

Die Konstante c bezeichnen wir als Kontrollparameter des Metropolis-Kriteriums. Durch sie wird festgelegt, in welchem Maße Verschlechterungen im Zielfunktionswert akzeptiert werden. Wählt man c zu groß, wird nahezu jeder erzeugte Punkt

akzeptiert; wählt man c zu klein, führt dies dazu, daß nahezu keine Verschlechterung mehr akzeptiert wird und der Metropolisalgorithmus sich zu einem lokalen Optimierungsverfahren hin wandelt.

Ein Schritt des Metropolis-Algorithmus besteht darin, einen Punkt mittels der Generierungswahrscheinlichkeit zu erzeugen und anschließend mittels der Akzeptanzwahrscheinlichkeit zu akzeptieren oder zu verwerfen. Im Falle des Verwerfens führt man den gesamten Metropolis-Schritt erneut vom gleichen Startpunkt aus durch. Im Falle der Akzeptanz ersetzen wir den Startpunkt durch den neu generierten Punkt, und es erfolgt dann ein Metropolis-Schritt mit dem neu erzeugten Punkt als Startpunkt. Durch sukzessive Fortsetzung dieses Verfahrens erhält man eine Markovkette, die — wie sich im folgenden noch herausstellen wird — für die diskrete globale Optimierung fundamental wichtige Eigenschaften hat.

Durch die Anwendung von Generierungs- und Akzeptanzwahrscheinlichkeit nacheinander erhalten wir die Übergangswahrscheinlichkeit dieser Markovkette:

Definition 2.3 (Übergangswahrscheinlichkeit).

Ausgehend von einem Zustand $i \in R$ wird ein neuer Punkt $j \in R$ als neue Lösung gemäß der folgenden Wahrscheinlichkeit

$$P_{ij}(c) := \begin{cases} G_{ij} A_{ij}(c) & \text{falls } i \neq j \\ 1 - \sum_{\substack{l \in R \\ l \neq i}} G_{il} A_{il}(c) & \text{sonst} \end{cases} \quad (2.3)$$

für alle $i, j \in R$, $c \in \mathbb{R}^+$, ausgegeben. Die Wahrscheinlichkeiten $P_{ij}(c)$ bezeichnen wir als Übergangswahrscheinlichkeiten des Metropolis-Algorithmus.

Die durch die Übergangswahrscheinlichkeiten $P_{ij} := P_{ij}(c)$ erhaltene Markovkette bezeichnen wir mit $\Psi := \{\Psi_k : k \in \mathbb{N}\}$.

Satz 2.4 (Irreduzibilität und Aperiodizität von Ψ).

Die Markovkette Ψ ist irreduzibel und aperiodisch.

Beweis. vgl. [Laar Aarts 87]

Definition 2.5 (stationäre Verteilung).

Es sei Φ eine Markovkette auf einem diskreten Zustandsraum R mit den Übergangswahrscheinlichkeiten $P := (P_{ij})_{(i \in R, j \in R)}$. Für jedes $i \in R$ existiere der Grenzwert

$$q_i := \lim_{k \rightarrow \infty} P(\Phi_k = i \mid \Phi_0 = j)$$

unabhängig vom Startpunkt $j \in R$. Dann heißt die Verteilung $q = (q_i)_{i \in R}$ die stationäre Verteilung von Φ .

Der wesentliche Satz für die Zielsetzung des Metropolis-Algorithmus ist der folgende, denn hier wird ausgesagt, daß der Metropolis-Algorithmus eine Grenzverteilung besitzt und diese für die Belange der diskreten globalen Optimierung wichtige Eigenschaften hat.

Satz 2.6 (stationäre Verteilung von Ψ).

Es bezeichne \tilde{g} die Funktion

$$\begin{aligned} \tilde{g} : R &\rightarrow \mathbb{R} \\ i &\mapsto \exp\left(-\frac{f(i)}{c}\right) \end{aligned}$$

und g eine Verteilung mit den Gewichten

$$g(i) := \frac{\tilde{g}(i)}{\sum_{j \in R} \tilde{g}(j)}.$$

Desweiteren setzen wir für die Übergangswahrscheinlichkeiten G_{ij} aus (2.3) voraus, daß für alle $i, j \in R$ ein $n \in \mathbb{N}$ und Zustände l_0, l_1, \dots, l_n mit $l_0 = i$ und $l_n = j$ existieren, so daß

$$G_{l_k, l_{k+1}} > 0 \quad k = 0, 1, \dots, n-1.$$

Dann besitzt die Markovkette Ψ die stationäre Verteilung g , $c \in \mathbb{R}^+$, und

$$\lim_{n \rightarrow \infty} P_{ij}^n = g(j) \quad \forall i, j \in R, c \in \mathbb{R}^+.$$

Beweis. vgl. [Laar Aarts 87]

Diese stationäre Verteilung g hat für die Belange der diskreten globalen Optimierung zwei wesentliche Eigenschaften:

- Sie ist unabhängig vom Startpunkt.
- Sie ist dort maximal, wo die Funktion minimal ist.

Setzt man also die Erzeugung von Punkten durch den Metropolis-Algorithmus ausreichend lang fort, so werden mit hoher Wahrscheinlichkeit Punkte erzeugt, die im Funktionswert nahe dem des globalen Minimierers liegen.

Zum Abschluß dieses Abschnitts sei noch bemerkt, daß es zur Erzielung dieser Ergebnisse genügt hätte, die Generierungswahrscheinlichkeiten lediglich als symmetrisch anzunehmen. Die Wahl der Generierungswahrscheinlichkeiten als Gleichverteilung wurde nur der Einfachheit halber festgesetzt.

2.2 Ein Abbruchkriterium zum Metropolis-Algorithmus

In der Praxis fällt es oft schwer zu erkennen, wann eine Verteilung, die hinreichend nahe an der Grenzverteilung liegt, erreicht ist. Deshalb wird im folgenden ein Kriterium für den Abbruch des Algorithmus vorgestellt.

Um aus statistischen Gesichtspunkten Informationen über die Verteilung im n -ten Schritt zu erhalten, bietet sich die folgende Vorgehensweise als die Naheliegenste an:

Führe m -mal unabhängig einen Metropolis-Schritt vom Startpunkt $i_s \in R$ aus. Von jedem erzeugten Punkt führe jeweils einen Metropolis-Schritt aus. Nach dem starken Gesetz der großen Zahl für unabhängige Zufallsvariablen geht die relative Häufigkeit des Auftretens des Zustandes $i \in R$ in P_* -fast jedem Pfad der Metropoliskette gegen $g(i)$ für jedes $i \in R$.

Ein Abbruchkriterium könnte also die Beobachtung der Konvergenz der relativen Häufigkeiten der einzelnen Zustände sein. Diese Vorgehensweise führt zwar zu dem Ziel, Aussagen über die Verteilung treffen zu können, ist allerdings numerisch extrem aufwendig und somit unbrauchbar.

Grundidee des hier vorgestellten Abbruchkriteriums ist, daß P_* -fast alle Pfade schon die Information über die Grenzverteilung enthalten und man daher nur einen Pfad beobachten muß.

Um zu zeigen, daß die Markovkette, die durch den Metropolisalgorithmus induziert wird, diese Ergodeneigenschaft hat, soll Satz 1.78 angewandt werden. Zunächst definieren wir den Begriff *ergodisch* auf einem endlichen Zustandsraum.

Definition 2.7 (ergodisch).

Es sei Φ eine irreduzible Markovkette. Ein Zustand $i \in R$ heißt ergodisch, wenn er aperiodisch, rekurrent und der Erwartungswert der Rückkehrzeit endlich ist.

Zur Verifizierung der Voraussetzungen des Satzes 1.78 verwenden wir den folgenden Satz.

Satz 2.8.

In einer irreduziblen Markov-Kette Φ auf R mit nur ergodischen Elementen und Übergangsmatrix P existieren die Grenzwerte

$$q_i = \lim_{k \rightarrow \infty} P(\Phi_k = i \mid \Phi_0 = j) = \lim_{k \rightarrow \infty} P_{ji}^k \quad \forall i, j \in R \quad (2.4)$$

unabhängig vom Anfangszustand j . Darüber hinaus ist $q_i \geq 0 \forall i \in R$, und es gelten die beiden Gleichungen

$$\sum_{i \in R} q_i = 1 \quad (2.5)$$

und

$$q_i = \sum_{j \in R} q_j P_{ji} \quad \forall i \in R, \quad (2.6)$$

welche die Gleichung (1.9) im diskreten Fall repräsentieren.

Sei umgekehrt die Markovkette irreduzibel und aperiodisch, und es existieren Zahlen $q_i \geq 0 \forall i \in R$, die (2.5) und (2.6) erfüllen, dann sind alle Zustände bzw. Punkte in R ergodisch, die q_i sind durch (2.4) für alle $i \in R$ gegeben und

$$q_i = \frac{1}{\mu_i} \quad \forall i \in R,$$

wobei μ_i der Erwartungswert der Rückkehrzeit von i ist.

Beweis. [Feller 70]

Da die vom Metropolis-Algorithmus induzierte Markovkette Ψ irreduzibel und aperiodisch ist und außerdem Zahlen $q_i \geq 0 \forall i \in R$ mit den Eigenschaften (2.5) und (2.6) existieren – nämlich $q(i) = g(i)$ für alle $i \in R$, sind alle Zustände ergodisch, also insbesondere rekurrent, d.h. $E_x(\eta_y) = \infty$.

Um Satz 1.78 anwenden zu können, ist nun noch zu zeigen, daß Ψ sogar Harrisrekurrent ist. Es stellt sich jedoch heraus, daß Harris Rekurrenz und Rekurrenz auf abzählbaren Zustandsräumen übereinstimmen:

Für rekurrente Markovketten folgt

$$\infty = E_x(\eta_y) = \sum_{n=1}^{\infty} P^n(x, \{y\}) =: U(x, \{y\}) \quad (2.7)$$

Satz 1.65 zeigt für abzählbare Zustandsräume die Äquivalenz

$$U(x, \{x\}) = \infty \Leftrightarrow L(x, \{x\}) = 1,$$

d.h. aufgrund der Rekurrenz folgt in unserem Fall

$$L(x, \{x\}) = 1 \quad \forall x \in R.$$

Mit Satz 1.66 schließen wir hieraus

$$Q(x, \{x\}) = 1 \quad \forall x \in R,$$

womit Ψ Harrisrekurrent ist.

Der Satz 1.78 ist also anwendbar, und somit gilt für jede Funktion f aus $L_1(R, \mathcal{B}(R), g)$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n f(\Psi_k) = \int f dg = \sum_{i \in R} f(i)g(i) \quad P_*\text{-fast sicher,} \quad (2.8)$$

wobei Ψ_k das Ergebnis des Metropolis-Algorithmus in jedem Schritt k anzeigt.

Insbesondere ist (2.8) anwendbar für die Funktion

$$f^i(\Psi_k) = \begin{cases} 1 & \text{falls } \Psi_k = i \\ 0 & \text{sonst.} \end{cases}$$

Es folgt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n f^i(\Psi_k) = E_g(f^i) = g(i) \quad \forall i \in R \quad P_*\text{-fast sicher} \quad (2.9)$$

für jeden Startzustand $x \in R$.

Die Gleichung (2.9) zeigt, daß auf einem speziellen Pfad der Markovkette Ψ die relative Häufigkeit des Zustandes i gegen die Grenzverteilung $g(i)$ konvergiert für jeden Zustand $i \in R$. Um also Aussagen über die Verteilung des Metropolis-Algorithmus nach k Schritten treffen zu können, muß *nicht* die numerisch aufwendige, oben aufgeführte Vorgehensweise angewandt werden. Gleichung 2.9 zeigt, daß diese parallele m -malige Durchführung des Algorithmus nicht nötig ist, denn die relative Häufigkeit in *einem* Pfad des Metropolis-Algorithmus hat schon die gewünschte Eigenschaft.

Um zu erkennen, ob die Grenzverteilung annähernd erreicht ist, müßte entweder die Häufigkeit jedes Zustandes aus dem Zustandsraum protokolliert werden, was aufgrund der Mächtigkeit von R numerisch nur schwer bis gar nicht zu bewältigen ist oder die Grenzverteilung bekannt sein, was jedoch nur theoretisch, nicht aber praktisch der Fall ist.

Um dennoch Aussagen über die Abweichung der derzeitigen Verteilung von der Grenzverteilung treffen zu können, wählen wir in Satz 1.78 als Funktion f eine beliebige Funktion h aus $L_1(R, \mathcal{B}(R), g)$ und beobachten als Abbruchkriterium die Konvergenz der Größe

$$\frac{1}{n} \sum_{k=1}^n h(\Psi_k). \quad (2.10)$$

Dabei sollte die Funktion h so gewählt sein, daß nahezu jeder Zustand $i \in R$ – möglichst entsprechend seines Gewichts in der Zielfunktion f – einen Beitrag leisten kann, also insbesondere solche Zustände, deren Funktionswerte von f relativ

klein sind. Es bietet sich an, als Funktion h die Funktion \tilde{g} , also

$$h(i) = \tilde{g}(i) := \exp\left(-\frac{f(i)}{c}\right), \quad i \in R,$$

zu wählen, auch im Hinblick darauf, daß dann die Beobachtung der Größe (2.10) gerade der numerischen Berechnung des Normierungsfaktors von g , also

$$\sum_{i \in R} \exp\left(-\frac{f(i)}{c}\right),$$

entspricht.

Wir brechen also den Lauf des Metropolis-Algorithmusses dann ab, wenn sich die Konvergenz der Größe

$$\frac{1}{n} \sum_{k=1}^n \exp\left(-\frac{f(\Psi_k)}{c}\right)$$

abzeichnet. Diese Vorgehensweise wurde bereits von Catthoor in [Catthoor 88] als Abbruchkriterium vorgeschlagen, allerdings nur im Hinblick auf die damit erfolgte numerische Berechnung des Normierungsfaktors, nicht in der Interpretation als Beobachtung der Konvergenz der Verteilungen.

Kapitel 3

Kontinuierliche globale Optimierung

Da das Ziel dieser Arbeit eine Schrittweitensteuerung zur kontinuierlichen globalen Optimierung ist, wird in diesem Abschnitt die Theorie zur globalen Optimierung mittels stochastischer Integration laut [Schäffler 95] vorgestellt. Sie ist im wesentlichen auch dieser Literatur entnommen; deshalb sei für genaueres und Beweise auch darauf verwiesen.

Wenden wir uns der eingangs erwähnten Problemstellung der globalen Optimierung zu. Gegeben sei eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^2$. Das globale Optimierungsproblem besteht darin, einen Punkt $x^* \in \mathbb{R}^n$ zu finden, derart daß

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

3.1 Voraussetzung A an die Zielfunktion

Wir betrachten im folgenden nur Zielfunktionen, die die Voraussetzung A erfüllen.

Voraussetzung A:

Es existiert ein $\epsilon > 0$ derart, daß es ein $r > 0$ gibt mit

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

für alle $x \in \mathbb{R}^n \setminus \{x \in \mathbb{R}^n : \|x\|_2 \leq r\}$, wobei $\nabla f(x)$ den Gradienten der Zielfunktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ an der Stelle x bezeichne.

Voraussetzung A sichert uns, daß die Zielfunktion außerhalb einer Kugel mit Radius r um den Ursprung entlang allen Richtungen s mit $\|s\|_2 = 1$ hinreichend schnell steigt. Das bedeutet, daß die Funktion f einen globalen Optimierer besitzt. Die Umkehrung gilt im allgemeinen nicht (vgl. die Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto \sin(x)$).

Sollte die Zielfunktion f die Voraussetzung A nicht erfüllen, verwenden wir die Hilfsfunktion

$$\bar{f}(x) := f(x) + P(\|x\|_2^2 - c)^4, \quad c > 0, \quad P : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto \begin{cases} x & \text{falls } x \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

Diese Hilfsfunktion \bar{f} besitzt die folgenden Eigenschaften

- $\bar{f} \in C^2$,
- $\bar{f}(x) = f(x)$ für alle $x \in \{x \in \mathbb{R}^n : \|x\|_2^2 \leq c\}$,
- $\bar{f}(x) > f(x)$ für alle $x \in \mathbb{R}^n$ mit $\|x\|_2^2 > c$,
- \bar{f} erfüllt Voraussetzung A.

3.2 Eine spezielle Klasse stochastischer Differentialgleichungen

Um den globalen Minimierer der Zielfunktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ zu finden, betrachten wir für festes $x_0 \in \mathbb{R}^n$ und $\epsilon > 0$ die Menge der stochastischen Differentialgleichungen (3.1)

$$\Phi_t^{\epsilon, x_0}(F) = x_0 + \epsilon((B_t(F) - B_0(F)) - \int_0^t \nabla f(\Phi_s^{\epsilon, x_0}(F)) ds) \quad (3.1)$$

mit $F \in \Omega^n$.

Unterteilen wir (3.1) in zwei Teile, so fällt eine Interpretation leichter. Zunächst betrachten wir den hinteren Teil:

$$x_0 - \int_0^t \nabla f(\Phi_s^{x_0}(F)) ds.$$

Die Lösung dieses Anfangswertproblems ist die Kurve des steilsten Abstiegs vom Startpunkt x_0 aus und kann somit als Verfahren der lokalen Optimierung angesehen werden.

Der vordere Teil von (3.1)

$$x_0 + \epsilon (B_t(F) - B_0(F))$$

ist lediglich eine von x_0 ausgehende, durch ϵ gewichtete Brownsche Bewegung und kann daher als Zufallssuche interpretiert werden.

Insgesamt gesehen kann (3.1) also als ein Verfahren der lokalen Optimierung mit zufälligem Startpunkt angesehen werden. Durch den Faktor ϵ kann die Gewichtung der einzelnen Faktoren verändert werden.

Damit die stochastische Differentialgleichung (3.1) uns an den globalen Minimierer heranführen kann, muß sie eine Lösung besitzen. Dies wird durch den folgenden Satz sichergestellt.

Satz 3.1 (Existenz und Eindeutigkeit von $\{\Phi_t^{\epsilon, x_0}\}$).

Für jedes $x_0 \in \mathbb{R}^n$ und ein $\epsilon > 0$ derart, daß Voraussetzung A erfüllt ist, gilt:

- i) Es existiert ein eindeutiger stochastischer Prozeß $\{\Phi_t^{\epsilon, x_0}\}$, der (3.1) löst,
- ii) alle Pfade von $\{\Phi_t^{\epsilon, x_0}\}$ sind stetig,
- iii) $\Phi_0^{\epsilon, x_0} = x_0$.

Nachdem die Existenz und Eindeutigkeit der Lösung von (3.1) sichergestellt ist, können wir uns der Frage zuwenden, ob der Lösungsprozeß Eigenschaften aufweist, die für die Auffindung des globalen Minimierers hilfreich sind. Tatsächlich lassen sich einige nützliche Eigenschaften aufzeigen.

Für jedes $\bar{x} \in \mathbb{R}^n$ und $\rho \in \mathbb{R}^+$ definieren wir die Zufallszeit $s_{\bar{x}, \rho} : C(\mathbb{R}_0^+, \mathbb{R}^n) \rightarrow \overline{\mathbb{R}}$,

$$F \mapsto \begin{cases} \inf\{t \geq 0 : \|\Phi_t^{\epsilon, x_0}(F) - \bar{x}\|_2 \leq \rho\} & \text{falls } \{t \geq 0 : \|\Phi_t^{\epsilon, x_0} - \bar{x}\|_2 \leq \rho\} \neq \emptyset \\ \infty & \text{sonst.} \end{cases}$$

Nun können für uns fundamental wichtige Eigenschaften von $\{\Phi_t^{\epsilon, x_0}\}$ gezeigt werden.

Satz 3.2 (Eigenschaften von $\{\Phi_t^{\epsilon, x_0}\}$).

Betrachtet man das zu Beginn des Kapitels aufgeführte globale Optimierungsproblem mit der Zielfunktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^2$, und $\epsilon > 0$ derart, daß Voraussetzung A erfüllt ist, so erhalten wir für jeden globalen Minimierer x^* von f und für jedes $\rho > 0$:

$$W^n(\{F \in C(\mathbb{R}_0^+, \mathbb{R}^n) : s_{x^*, \rho}(F) < \infty\}) = 1$$

und $E(s_{x^*, \rho}) < \infty$ für alle $x_0 \in \mathbb{R}^n$. Dabei bezeichne W^n das n -dimensionale Wiener-Maß.

Ferner besitzt für jedes $x_0 \in \mathbb{R}^n$ und $t > 0$ die Zufallsvariable Φ_t^{ϵ, x_0} eine λ^n -Dichte $p_{\epsilon, x_0, t} : \mathbb{R}^n \rightarrow \mathbb{R}$, und es gilt

$$p_\epsilon(x) := \lim_{t \rightarrow \infty} p_{\epsilon, x_0, t}(x) = \frac{\exp\left(-\frac{2(f(x)-f(x^*))}{\epsilon^2}\right)}{\int_{\mathbb{R}^n} \exp\left(-\frac{2(f(x)-f(x^*))}{\epsilon^2}\right) dx} \quad (3.2)$$

für alle $x \in \mathbb{R}^n$. Die Lebesgue-Dichten $p_{\epsilon, x_0, t}$, $t \in (0, \infty)$, sind für jedes $x_0 \in \mathbb{R}^n$ durch die Fokker-Planck-Gleichung gegeben:

$$\frac{\partial p_{\epsilon, x_0}(s, x)}{\partial s} + \sum_{i=1}^n \frac{\partial}{\partial x_i} (\nabla f(x) p_{\epsilon, x_0}(s, x)) + \frac{\epsilon^2}{2} \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} p_{\epsilon, x_0}(s, x) = 0$$

für alle $(s, x) \in (0, \infty) \times \mathbb{R}^n$ mit der Endbedingung

$$\lim_{s \rightarrow 0} p_{\epsilon, x_0}(s, t) = \delta(x - x_0),$$

wobei $p_{\epsilon, x_0} : (0, \infty) \times \mathbb{R}^n \rightarrow \mathbb{R}$, $(s, x) \mapsto p_{\epsilon, x_0, s}(x)$. Dabei sei δ die Dirac-Distribution auf \mathbb{R}^n .

Überlegen wir kurz, was diese Eigenschaften für uns bedeuten: Betrachten wir ein globales Optimierungsproblem mit einer Zielfunktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^2$. Für einen beliebigen Startpunkt $x_0 \in \mathbb{R}^n$ und ein $\epsilon > 0$ derart, daß Voraussetzung A für f erfüllt ist, kommt bis auf eine W^n -Nullmenge jeder Pfad des Lösungsprozesses $\{\Phi_t^{\epsilon, x_0}\}$ von (3.1) in endlicher Zeit beliebig nahe an jeden globalen Optimierer x^* von f heran.

Die Zufallsvariablen $\Phi_t^{\epsilon, x_0} : C(\mathbb{R}_0^+, \mathbb{R}^n) \rightarrow \mathbb{R}^n$, $t \in \mathbb{R}_0^+$, konvergieren in Verteilung gegen eine Zufallsvariable $\Phi^* : C(\mathbb{R}_0^+, \mathbb{R}) \rightarrow \mathbb{R}^n$, deren λ^n -Dichte p_ϵ durch (3.2) beschrieben wird. Konvergenz in Verteilung heißt dabei, daß die Dichten der Zufallsvariablen Φ_t^{ϵ, x_0} punktweise gegen die Dichte der Zufallsvariablen Φ^* konvergieren.

Diese Grenzdichte p_ϵ hat zwei wichtige Eigenschaften:

- Sie ist unabhängig vom Startpunkt x_0 und
- sie hat ihr globales Maximum dort, wo die Funktion ihr globales Minimum besitzt.

Das bedeutet, daß zu jedem beliebigen Startpunkt die Zufallszahlen, die gemäß p_ϵ erzeugt werden, mit hoher Wahrscheinlichkeit in der Nähe des globalen Minimierers der Funktion f liegen werden.

Die numerische Lösung von (3.1) gibt uns also ein Verfahren, das Pseudozufallszahlen gemäß einer Dichte erzeugt, die gegen die Grenzdichte p_ϵ konvergieren.

Mit Pseudozufallszahlen bezeichnet man dabei Zahlen bzw. Zahlenfolgen, die gewissen statistischen Tests genügen, dennoch aber streng deterministisch sind. Das Dilemma liegt darin, daß der Computer nicht wirklich zufällig Zahlen erzeugen kann. Wir sprechen also immer dann von Pseudozufallszahlen, wenn Algorithmen zur Erzeugung von Zufallszahlen auf den Computer übertragen werden.

Kapitel 4

Schrittweisensteuerungskonzept

4.1 Das semi-implizite Eulerverfahren

Wir gehen von der stochastischen Differentialgleichung (3.1)

$$\begin{aligned}\Phi_{t+\sigma}^{\epsilon, x_0}(\omega) &= x_0 + \epsilon(B_{t+\sigma}(\omega) - B_0(\omega)) - \int_0^{t+\sigma} \nabla f(\Phi_\tau^{\epsilon, x_0}(\omega)) d\tau \\ &= \Phi_t^{\epsilon, x_0}(\omega) + \epsilon(B_{t+\sigma}(\omega) - B_t(\omega)) - \int_t^{t+\sigma} \nabla f(\Phi_\tau^{\epsilon, x_0}(\omega)) d\tau\end{aligned}\quad (4.1)$$

aus. Nehmen wir an, daß bereits eine Realisierung ξ von Φ_t^{ϵ, x_0} ermittelt wurde, gelangen wir zu einer neuen stochastischen Differentialgleichung

$$\Theta_{t+\sigma}^{\epsilon, \xi}(\omega) = \xi + \epsilon(B_{t+\sigma}(\omega) - B_t(\omega)) - \int_t^{t+\sigma} \nabla f(\Theta_\tau^{\epsilon, \xi}(\omega)) d\tau. \quad (4.2)$$

In [Schäffler 95] wird als Lösungsverfahren zur numerischen Lösung der stochastischen Differentialgleichung (4.2) das semi-implizite Euler-Verfahren vorgeschlagen, das zu gegebener Schrittweite σ einen neuen Punkt $\tilde{\Phi}_{t+\sigma}^{\epsilon, \xi}(\omega)$ ermittelt durch

$$\tilde{\Phi}_{t+\sigma}^{\epsilon, \xi}(\omega) = \xi - (\mathbb{E}_n + \sigma \nabla^2 f(\xi))^{-1} (\sigma \nabla f(\xi) - \epsilon(B_{t+\sigma}(\omega) - B_t(\omega))). \quad (4.3)$$

Für kleine σ stellt dies eine Näherung der tatsächlichen Lösung der stochastischen Differentialgleichung (4.2) dar, denn der Integrationsfehler kann wie folgt abgeschätzt werden:

Satz 4.1.

Für jedes ϵ , das die Voraussetzung A erfüllt, und für jedes $t \geq 0$, $\xi \in \mathbb{R}^n$, erhalten wir die Existenz einer Konstanten $c > 0$ mit

$$\lim_{\sigma \rightarrow 0} \frac{\|\Theta_{t+\sigma}^{\epsilon, \xi}(\omega) - \tilde{\Phi}_{t+\sigma}^{\epsilon, \xi}(\omega)\|_2}{\sqrt{\sigma^3 \ln(\ln(1/\sigma))}} \leq c \quad W^n\text{-fast sicher}$$

Beweis. [Schäffler 95], Theorem 2.3.1, Seite 31.

In [Schäffler 95] wird die Schrittweitensteuerung wie folgt realisiert: Man wählt eine Startschrittweite σ und bestimmt anhand dieser Schrittweite ausgehend von ξ einen neuen Punkt $\tilde{\Phi}_\sigma(\xi)$. Anschließend tätigt man von ξ aus erneut einen Eulerschritt, allerdings mit Schrittweite $\sigma/2$. Von diesem neu erzeugten Punkt $\tilde{\Phi}_{\sigma/2}(\xi)$ führt man noch einmal einen Eulerschritt — ebenfalls mit Schrittweite $\sigma/2$ — durch und erhält $\tilde{\Phi}_{\sigma/2}(\tilde{\Phi}_{\sigma/2}(\xi))$. Beträgt die Differenz zwischen $\tilde{\Phi}_\sigma(\xi)$ und $\tilde{\Phi}_{\sigma/2}(\tilde{\Phi}_{\sigma/2}(\xi))$ maximal ein vorgegebenes δ , so akzeptiert man $\tilde{\Phi}_\sigma(\xi)$ und behält die Schrittweite bei. Übersteigt die Differenz δ , so halbiert man die Schrittweite und führt die drei angegebenen Eulerschritte erneut aus.

Diese Schrittweitensteuerung beruht auf dem Ziel, möglichst die genaue Lösung von (3.1) zu approximieren, was allerdings für die globale Optimierung nicht von Nöten ist und somit unnötigen numerischen Aufwand bedeutet. Wir werden daher im folgenden untersuchen, für welche Wahl von σ die Stabilität der Markovkette $\tilde{\Phi}^\sigma$ erhalten bleibt, soll heißen, für welche σ eine stationäre Verteilung existiert und die Übergangswahrscheinlichkeiten im n -ten Schritt für n gegen unendlich gegen diese konvergieren — unabhängig vom Startpunkt.

4.2 Allgemeine Vorgehensweise

Das Problem der Lösung der stochastischen Differentialgleichung (4.2) mit dem semi-impliziten Eulerverfahren besteht darin, daß nur für kleine σ sichergestellt sein kann, daß $\tilde{\Phi}_{t+\sigma}^{\epsilon,\xi}(\omega)$ ausreichend nahe bei $\Theta_{t+\sigma}^{\epsilon,\xi}(\omega)$ liegt, so daß die Konvergenzeigenschaften erhalten bleiben. Allerdings stellt sich die in Kapitel 3 angesprochene Konvergenz der Verteilungen von $\Phi_t^{\epsilon,\xi}$ erst für t gegen unendlich ein. Bei der Lösung der stochastischen Differentialgleichung (4.2) mit dem semi-impliziten Eulerverfahren müssen wir uns also zwischen den beiden folgenden Alternativen entscheiden:

- Wahl einer kleinen Schrittweite σ und somit einer relativ langen Laufzeit des Algorithmus bei einer theoretisch sichergestellten Konvergenz der Verteilungen,
- Wahl einer großen Schrittweite σ und somit einer relativ kurzen Laufzeit des Algorithmus zu dem Preis, daß die Konvergenz der Verteilungen nicht mehr sichergestellt werden kann.

Für die Zwecke der globalen Optimierung ist allerdings die genaue Lösung $\Phi_{t+\sigma}^{\epsilon,\xi}(\omega)$ gar nicht von Interesse, sondern lediglich die Erzeugung von Zufallszahlen gemäß einer Verteilung, bei der ein relativ großes Gewicht auf den globalen Minimierern der Zielfunktion f liegt.

Aus dieser Motivation heraus untersuchen wir in diesem Kapitel, mit welcher Wahl der Schrittweite σ die Stabilitätseigenschaften erhalten bleiben, in dem Sinne, daß die vom semi-implizierten Eulerverfahren induzierte Markovkette eine stationäre Verteilung π besitzt und die Übergangswahrscheinlichkeiten unabhängig vom Startpunkt im n -ten Schritt für n gegen unendlich gegen diese stationäre Verteilung konvergieren.

Hierzu fassen wir nun die Vorgehensweise des semi-impliziten Eulerverfahrens nicht mehr länger als numerische Lösung der stochastischen Differentialgleichung (3.1) auf, sondern als eigenes theoretisches Verfahren. Wir verwenden also eine Schrittweitenfunktion $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}$, $\sigma(\xi) > 0$ für alle $\xi \in \mathbb{R}^n$, und erzeugen ausgehend vom Startpunkt bzw. von der Realisierung ξ im Schritt zuvor eine Markovkette $\Phi^\sigma := \{\Phi_t^\sigma : t \in \mathbb{N}\}$ gemäß dem Bildungsgesetz

$$\Phi_t^\sigma(\omega) = \xi - (\mathbb{E}_n + \sigma(\xi)\nabla^2 f(\xi))^{-1}(\sigma(\xi)\nabla f(\xi) - \epsilon(B_{t+\sigma(\xi)}(\omega) - B_t(\omega))). \quad (4.4)$$

Unser Ziel ist es, die Funktion $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}$, $\sigma(\xi) > 0$ für alle $\xi \in \mathbb{R}^n$, abhängig vom Ausgangspunkt ξ , so zu bestimmen, daß die durch die Vorschrift (4.4) entstehende Markovkette bereits selbst über die Eigenschaft verfügt, ein invariantes Wahrscheinlichkeitsmaß zu besitzen und in Verteilung gegen eine Zufallsvariable zu konvergieren, die gemäß dieses invarianten Wahrscheinlichkeitsmaßes verteilt ist. Kann so eine Funktion σ gefunden werden, können wir eine Schrittweitensteuerung verwenden, die wir schon a priori festlegen und nicht aus dem Verfahren selbst bestimmen.

Um diese Art von Stabilität zu zeigen, verifizieren wir in den folgenden Abschnitten die Voraussetzungen der Sätze 1.78 und 1.77 für die Markovkette Φ^σ , um diese Sätze anwenden zu können. Wir werden im folgenden also die Irreduzibilität, die Harris-Rekurrenz, die strenge Aperiodizität sowie die Positivität von Φ^σ zeigen. Es stellt sich heraus, daß wir hierzu eine weitere Voraussetzung an die Funktion f stellen müssen, allerdings erneut nur außerhalb einer Kugel um den Nullpunkt, die den oder die globalen Minimierer der Zielfunktion f enthält. Diese führen wir in Abschnitt 4.3 ein. Als Vorüberlegung zum Nachweis der genannten Voraussetzungen betrachten wir in 4.4 die Dichte der Übergangswahrscheinlichkeiten der mit (4.4) generierten Markovkette.

4.3 Voraussetzung B an die Zielfunktion

Zunächst stellen wir an die Zielfunktion f neben der Voraussetzung A eine weitere Voraussetzung, die sicherstellt, daß die Zielfunktion außerhalb einer Kugel um den Nullpunkt nicht nur ausreichend stark steigt, sondern daß diese Steigung auch hinreichend gleichmäßig verläuft.

Voraussetzung B:

Es existiert ein $r' \in \mathbb{R}$, $r' > 0$, und eine stetige Funktion $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}$, $\sigma(x) > 0$ für alle $x \in \mathbb{R}^n$, so daß für alle $x \in \mathbb{R}^n$ die Matrix $A_{\sigma,x} := \mathbb{E}_n + \sigma(x)\nabla^2 f(x)$ invertierbar ist und für alle $x \in \mathbb{R}^n$ mit $\|x\|_2 > r'$ gilt

$$\frac{\sigma(x)\|A_{\sigma,x}^{-1}\nabla f(x)\|_2^2 + \epsilon^2\|A_{\sigma,x}^{-1}\|_F^2 + \frac{1}{\sigma(x)}}{2} \leq \nabla f(x)^T A_{\sigma,x}^{-1} x.$$

Parabelfunktionen wie $f(x_1, x_2, \dots, x_n) = x_1^2 + x_2^2 + \dots + x_n^2$ erfüllen diese Voraussetzung:

Durch die Berechnungen

$$\nabla f(x) = 2x \quad \text{und} \quad \nabla^2 f(x) = 2\mathbb{E}_n$$

erhält man

$$A_{\sigma,x} = (1 + 2\sigma(x))\mathbb{E}_n, \quad \text{also} \quad A_{\sigma,x}^{-1} = \frac{1}{1 + 2\sigma(x)}\mathbb{E}_n.$$

Somit ist zur Verifizierung von Voraussetzung B zu zeigen, daß ein $r' \in \mathbb{R}$, $r' > 0$, und eine stetige Funktion $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}$, $\sigma(x) > 0$ für alle $x \in \mathbb{R}^n$ existieren, so daß

$$\frac{\sigma(x)\frac{4\|x\|_2^2}{(1+2\sigma(x))^2} + \epsilon^2\frac{n}{(1+2\sigma(x))^2} + \frac{1}{\sigma(x)}}{2} \leq 2x^T \frac{1}{1 + 2\sigma(x)} x$$

für alle $x \in \mathbb{R}^n$ mit $\|x\|_2 > r'$.

Diese Ungleichung folgt aus der Gültigkeit der Ungleichung

$$\epsilon^2 n + \frac{(1 + 2\sigma(x))^2}{\sigma(x)} \leq 4r'^2(1 + \sigma(x)).$$

Diese Ungleichung läßt sich leicht erfüllen, z.B. indem man die Funktion σ konstant auf $\{x \in \mathbb{R}^n : \|x\|_2 > r'\}$ und r' ausreichend groß wählt.

Für Punkte $x \in \mathbb{R}^n$ mit $\|x\|_2 \leq r$ stellt die Voraussetzung B keine weitere Einschränkung dar, da die Hessematrix $\nabla^2 f(x)$ stetig in x und \mathbb{E}_n invertierbar ist, so daß eine stetige Funktion σ auf $\mathbb{R}^n \setminus \{x \in \mathbb{R}^n : \|x\|_2 > r\}$ gefunden werden kann, für die $\mathbb{E}_n + \sigma(x)\nabla^2 f(x)$ invertierbar ist.

Aus Voraussetzung A folgt, daß der globale Minimierer in der Kugel $\{x \in \mathbb{R}^n : \|x\|_2 \leq r\}$ liegt. Sollte die Zielfunktion f die Voraussetzung B nicht erfüllen, so ändern wir f außerhalb dieser Kugel ab, ohne dabei den globalen Minimierer zu

verändern, indem wir wie folgt vorgehen:

Wir wählen eine Funktion $g \in C^2(\mathbb{R})$ mit $g(x) = 0$ für $x < r^2$ und $g(x) = 1$ für $x > (r + \delta)^2$. Dann ersetzen wir die Funktion f durch die Funktion \bar{f} , definiert durch

$$\begin{aligned} \bar{f} : \mathbb{R}^n &\rightarrow \mathbb{R} \\ x &\mapsto f(x)(1 - g(\|x\|_2^2)) + \|x\|_2^2 g(\|x\|_2^2). \end{aligned}$$

Diese besitzt die notwendigen Eigenschaften:

- i) $\bar{f} \in C^2(\mathbb{R}^n)$,
- ii) $\bar{f}(x) = f(x)$ für alle $x \in \mathbb{R}^n$ mit $\|x\|_2 \leq r$, d.h. die Funktion \bar{f} besitzt die gleichen globalen Minimierer wie die Funktion f ,
- iii) $\bar{f}(x) = \|x\|_2^2$ für alle $x \in \{x \in \mathbb{R}^n : \|x\|_2^2 > r + \delta\}$, d.h. \bar{f} erfüllt Voraussetzung B.

4.4 Dichte der Übergangswahrscheinlichkeiten

Im folgenden seien $r = r'$ und die Schrittweitenfunktion σ so gewählt, daß sowohl Voraussetzung A als auch B erfüllt sind. Wir betrachten im folgenden den stochastischen Prozeß $\Phi^\sigma := \{\Phi_t^\sigma : t \in \mathbb{N}\}$, der durch die Vorgehensweise (4.4) induziert wird. Offensichtlich ist Φ^σ eine Markovkette, denn der Ausgang von Φ_t^σ hängt nur vom Ausgang von Φ^σ zum Zeitpunkt $t - 1$ ab, nicht aber von den Ausgängen zu den vorherigen Zeitpunkten $0, 1, \dots, t - 2$. Unser Ziel ist die Anwendung des Satzes 1.78 auf die Markovkette Φ^σ . Hierzu sind einige Voraussetzungen zu verifizieren.

Wir bezeichnen die Übergangswahrscheinlichkeit der Markovkette Φ^σ , ausgehend von dem Punkt ξ durch das Bildungsgesetz (4.4) mit Schrittweitenfunktion σ in eine Menge B , $B \in \mathcal{B}(\mathbb{R}^n)$, zu gelangen, mit $P(\xi, B)$ und überprüfen zunächst, ob $P(\xi, B)$ eine Dichte besitzt.

Die Zuwächse der Brownschen Bewegung $B_{t+\sigma(\xi)}(\omega) - B_t(\omega)$ sind $N(0, \sigma(\xi)\mathbb{E}_n)$ normal verteilt, besitzen also die Dichte

$$\tilde{N}(y, \sigma(\xi)) = \frac{1}{\sqrt{(2\pi)^n (\sigma(\xi))^n}} \exp \left\{ -\frac{\|y\|_2^2}{2\sigma(\xi)} \right\}.$$

Somit besitzt Φ_t^σ als lineare Transformation einer normalverteilten Zufallsvariable eine Dichte bzgl. dem Lebesgue-Maß.

Wir ermitteln nun die spezielle Form dieser Dichte, die wir im folgenden benötigen werden.

Satz 4.2 (Dichte der Übergangswahrscheinlichkeiten von Φ^σ).

Die Übergangswahrscheinlichkeiten $P(\xi, B)$ der Markovkette Φ^σ ausgehend vom Punkt ξ durch das Bildungsgesetz (4.4) mit Schrittweitenfunktion σ in eine Menge B , $B \in \mathcal{B}^n$, zu gelangen, besitzen die Dichte k_ξ^σ mit

$$\begin{aligned} k_\xi^\sigma : \mathbb{R}^n &\rightarrow \mathbb{R} \\ y &\mapsto \tilde{N}(m_{\xi, \sigma}^{-1}(y), \sigma(\xi)) \left| \det \frac{1}{\epsilon} A_{\sigma, \xi} \right|, \end{aligned}$$

wobei

$$A_{\sigma, \xi} := \mathbb{E}_n + \sigma(\xi) \nabla^2 f(\xi) \quad (4.5)$$

und

$$m_{\xi, \sigma}(y) := \xi - A_{\sigma, \xi}^{-1}(\sigma(\xi) \nabla f(\xi) - \epsilon y). \quad (4.6)$$

Beweis. Die Übergangswahrscheinlichkeiten von Φ_t^σ hängen nur von einer normalverteilten Zufallsvariable ab. Durch die abkürzenden Schreibweisen (4.5) und (4.6) kann das Bildungsgesetz (4.4) formuliert werden zu

$$\Phi_t^\sigma(\omega) = m_{\xi, \sigma}(B_{t+\sigma(\xi)}(\omega) - B_t(\omega)). \quad (4.7)$$

Somit folgt für eine Menge $B \in \mathcal{B}(\mathbb{R}^n)$ mit

$$\tilde{B}_\xi := m_{\xi, \sigma}^{-1}(B)$$

sofort

$$\Phi_t^\sigma \in B \text{ unter der Bedingung } \Phi_{t-1}^\sigma = \xi \Leftrightarrow B_{t+\sigma(\xi)}(\omega) - B_t(\omega) \in \tilde{B}_\xi.$$

Für die Übergangswahrscheinlichkeiten gilt also

$$\begin{aligned} P(\xi, B) &= \int_{\tilde{B}_\xi} \tilde{N}(y, \sigma(\xi)) dy \\ &= \int_{m_{\xi, \sigma}^{-1}(B)} \tilde{N}(y, \sigma(\xi)) dy \\ &= \int_B \underbrace{\tilde{N}(m_{\xi, \sigma}^{-1}(y), \sigma(\xi)) \left| \det \frac{1}{\epsilon} A_{\sigma, \xi} \right|}_{= k_\xi^\sigma(y)} dy, \end{aligned}$$

wobei die letzte Gleichheit aus der mehrdimensionalen Substitutionsformel folgt. Das bedeutet aber, daß die Verteilung von Φ_t^σ bedingt auf $\Phi_{t-1}^\sigma = \xi$, also der Übergangswahrscheinlichkeitskern $P(\xi, B)$, die Dichte k_ξ^σ besitzt, was zu zeigen war. \square

Die Funktion

$$\begin{aligned} k^\sigma &: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, \\ (\xi, x) &\mapsto k_\xi^\sigma(x) \end{aligned} \quad (4.8)$$

ist stetig in beiden Variablen aufgrund der Stetigkeit von σ mit der Eigenschaft $k^\sigma(\xi, x) > 0$ für alle $\xi \in \mathbb{R}^n$ und für alle $x \in \mathbb{R}^n$. Diese Eigenschaft werden wir im folgenden verwenden.

4.5 Irreduzibilität

Nun wenden wir uns der Irreduzibilität der Markovkette Φ^σ zu.

Satz 4.3 (Irreduzibilität von Φ^σ).

Die durch das Bildungsgesetz (4.4) entstehende Markovkette Φ^σ ist irreduzibel.

Beweis. Es sei $C \in \mathcal{B}^n$ eine kompakte Menge mit $\lambda(C) > 0$. Dann existiert aufgrund der Stetigkeit von k_ξ^σ und der Kompaktheit von C für alle $\xi \in \mathbb{R}^n$ ein $k_\xi^* \in \mathbb{R}$, $k_\xi^* > 0$ mit

$$k_\xi^\sigma(x) \geq k_\xi^* \quad \forall x \in C.$$

Somit folgt für alle $\xi \in X$

$$\begin{aligned} P(\xi, B) &= \int_B k_\xi^\sigma(y) dy \\ &\geq \int_{B \cap C} k_\xi^\sigma(y) dy \\ &\geq k_\xi^* \lambda|_C(B), \end{aligned} \quad (4.9)$$

wobei $\lambda|_C$ das n -dimensionale Lebesgue-Maß eingeschränkt auf C bezeichne. Aus (4.9) folgt aber, daß für jeden Startpunkt $\xi \in \mathbb{R}^n$ gilt:

$$\lambda|_C(B) > 0 \quad \Rightarrow \quad U(\xi, B) > 0.$$

Somit ist Φ nach Satz 1.54 $\lambda|_C$ -irreduzibel. \square

Nach Satz 1.55 existiert ein maximales Irreduzibilitätsmaß ψ . Wie bereits in Kapitel 1 definiert, bezeichnen wir mit $\mathcal{B}^+(\mathbb{R}^n)$ die Menge der Borel-meßbaren Mengen mit positivem maximalem Irreduzibilitätsmaß

$$\mathcal{B}^+(\mathbb{R}^n) := \{A \in \mathcal{B}(\mathbb{R}^n) : \psi(A) > 0\}.$$

Wir werden im nächsten Abschnitt benötigen, daß eine Menge $C \in \mathcal{B}^+(\mathbb{R}^n)$ auch positives Lebesgue-Maß besitzt.

Satz 4.4.

Für jede Menge $C \in \mathcal{B}^+(\mathbb{R}^n)$ gilt $\lambda(C) > 0$.

Beweis. Um die Aussage zu zeigen, bemerken wir, daß aufgrund der ψ -Irreduzibilität von Φ nach Satz 1.54 für alle $x \in \mathbb{R}^n$ ein $m_x \in \mathbb{N}$ existiert, so daß

$$P^{m_x}(x, C) > 0. \quad (4.10)$$

Aus der Darstellung der Übergangswahrscheinlichkeiten nach n Schritten $P^n(x, \cdot)$ nach Chapman-Kolmogorov aus Satz 1.51

$$P^n(x, C) = \int_x P(x, dy) P^{n-1}(y, C), \quad n \in \mathbb{N}, \quad n \geq 2$$

und der Existenz einer Dichte bzgl. des Lebesgue-Maßes von $P(x, \cdot)$, ergibt sich induktiv, daß $P^n(x, \cdot)$ für alle $n \in \mathbb{N}$ ebenfalls eine Dichte bezüglich des Lebesgue-Maßes besitzt. Hieraus und aus (4.10) folgt

$$\lambda(C) > 0,$$

was zu zeigen war. □

4.6 Harris-Rekurrenz

Unser nächstes Ziel ist, die Rekurrenz bzw. die Harris-Rekurrenz der Markovkette Φ^σ zu zeigen. Hierzu verifizieren wir das Driftkriterium (1.8). Da wir hierfür eine nach Definition 1.57 kleine Menge C benötigen, klären wir zunächst, daß kompakte Mengen in Bezug auf die Markovkette Φ^σ immer klein sind.

Satz 4.5.

In Bezug auf die Markovkette Φ^σ ist jede kompakte Menge aus \mathcal{B}^n unbedeutend, also insbesondere auch klein.

Beweis. Nach Satz 1.58 existiert in Bezug auf die Markovkette Φ^σ eine unbedeutende Menge $C \in \mathcal{B}^+(\mathbb{R}^n)$. Es sei $D \in \mathcal{B}^n$ eine beliebige kompakte Menge. Diese hat nach Satz 4.4 positives Lebesgue-Maß. Also ist $P(x, C)$ für alle $x \in D$ darstellbar als ein Integral über eine positive Dichte bzgl. des Lebesgue-Maßes über eine Menge mit positivem Lebesgue-Maß, nämlich

$$P(x, C) = \int_C k_x^\sigma(y) dy.$$

Es gilt somit $P(x, C) > 0$ für alle $x \in D$. Außerdem ist $P(x, C)$ stetig in x für alle $x \in \mathbb{R}^n$. Daher nimmt $P(x, C)$ auf D das Minimum $\delta^* > 0$ an. Somit ist D nach Satz 1.59 (i) unbedeutend, also auch klein. □

Satz 4.6 (Harris-Rekurrenz von Φ^σ).

Die Markovkette Φ^σ ist Harris-rekurrent.

Beweis. Um die Harris-Rekurrenz zu zeigen, verifizieren wir die Voraussetzungen des Satzes 1.68, also das Driftkriterium (1.8). Hierzu wählen wir die Funktion V als

$$V(x) = x^T x.$$

Dann sind die „Sublevel-Mengen“

$$C_V(n) := \{y \in \mathbb{R}^n : \|y\|_2^2 \leq n\}$$

kompakt, also insbesondere klein nach Satz 4.5, d.h. V ist unbeschränkt außerhalb von kleinen Mengen.

Es sei $C := \{x \in \mathbb{R}^n : \|x\|_2 \leq r\}$. Wir berechnen nun $\Delta V(\xi)$ für $\xi \in \mathbb{R}^n$.

$$\begin{aligned} \Delta V(\xi) &:= \int y^T y P(\xi, dy) - \xi^T \xi \\ &= \int y^T y \tilde{N}(m_{\xi, \sigma}^{-1}(y), \sigma(\xi)) |\det \frac{1}{\epsilon} A_{\sigma, \xi}| dy - \xi^T \xi \\ &= \int m_{\xi, \sigma}(y)^T m_{\xi, \sigma}(y) \tilde{N}(y, \sigma(\xi)) dy - \xi^T \xi \\ &= -2\sigma(\xi) \xi^T A_{\sigma, \xi}^{-1} \nabla f(\xi) \\ &\quad + \int [2\epsilon \xi^T A_{\sigma, \xi}^{-1} y - 2\sigma(\xi) \epsilon (A_{\sigma, \xi}^{-1} \nabla f(\xi))^T A_{\sigma, \xi}^{-1} y] dy \\ &\quad + \int \epsilon^2 (A_{\sigma, \xi}^{-1} y)^T (A_{\sigma, \xi}^{-1} y) \tilde{N}(y, \sigma(\xi)) dy \\ &\quad + \sigma(\xi)^2 (A_{\sigma, \xi}^{-1} \nabla f(\xi))^T (A_{\sigma, \xi}^{-1} \nabla f(\xi)). \end{aligned} \tag{4.11}$$

Für Terme der Form

$$\int (l_1 y_1 + l_2 y_2 + \dots + l_n y_n) \tilde{N}(y, \sigma(\xi)) dy \quad \text{mit } l_i \in \mathbb{R}, i \in \{1, 2, \dots, n\},$$

ergibt sich aufgrund des Satzes von Fubini und aufgrund der Symmetrie von $\tilde{N}(y, \sigma(\xi))$ in jeder Komponente y_i , $i = 1, 2, \dots, n$,

$$\int l_i y_i \tilde{N}(y, \sigma(\xi)) dy = 0$$

für alle $i \in \{1, 2, \dots, n\}$ und somit aufgrund der Linearität des Integrals

$$\int (l_1 y_1 + l_2 y_2 + \dots + l_n y_n) \tilde{N}(y, \sigma(\xi)) dy = 0. \tag{4.12}$$

Für Terme der Form

$$\int (l_1 y_1^2 + l_2 y_2^2 + \cdots + l_n y_n^2) \tilde{N}(y, \sigma(\xi)) dy \quad \text{mit } l_i \in \mathbb{R}, i \in \{1, 2, \dots, n\},$$

ergibt sich mit dem Satz von Fubini und aufgrund der folgenden Identität

$$\int x^2 \exp \left\{ -\frac{x^2}{2\sigma(\xi)} \right\} dx = \sigma(\xi) \sqrt{2\pi\sigma(\xi)},$$

die Gleichheit

$$\begin{aligned} & \int (l_1 y_1^2 + l_2 y_2^2 + \cdots + l_n y_n^2) \tilde{N}(y, \sigma(\xi)) dy \\ &= \sum_{i=1}^n \int \cdots \int l_i \exp \left\{ -\frac{y_2^2 + \cdots + y_n^2}{2\sigma(\xi)} \right\} s_i^*(y, \sigma(\xi)) dy_2 \cdots dy_n. \end{aligned}$$

Dabei gilt

$$\begin{aligned} s_i^*(y, \sigma(\xi)) &:= \int y_i^2 \frac{1}{\sqrt{(2\pi)^n (\sigma(\xi))^n}} \exp \left\{ -\frac{y_1^2}{2\sigma(\xi)} \right\} dy_1 \\ &= \begin{cases} \frac{\sigma(\xi)}{\sqrt{(2\pi\sigma(\xi))^{n-1}}} & \text{falls } i = 1 \\ \frac{y_i^2}{\sqrt{(2\pi\sigma(\xi))^{n-1}}} & \text{falls } i \neq 1. \end{cases} \end{aligned}$$

Somit ergibt sich iterativ

$$\int (l_1 y_1^2 + l_2 y_2^2 + \cdots + l_n y_n^2) \tilde{N}(y, \sigma(\xi)) dy = \sum_{i=1}^n l_i \sigma(\xi). \quad (4.13)$$

Mit Hilfe von (4.12) schließen wir daher

$$\int [2\epsilon \xi^T A_{\sigma, \xi}^{-1} y - 2\sigma(\xi) \epsilon (A_{\sigma, \xi}^{-1} \nabla f(\xi))^T A_{\sigma, \xi}^{-1} y] dy = 0$$

und mit Hilfe von (4.13) mit $l_{ij} \in \mathbb{R} \forall i \in \{1, \dots, n\}, \forall j \in \{1, \dots, n\}$

$$\begin{aligned} \int \epsilon^2 (A_{\sigma, \xi}^{-1} y)^T (A_{\sigma, \xi}^{-1} y) \tilde{N}(y, \sigma(\xi)) dy &= \epsilon^2 \sum_{i=1}^n \sum_{j=1}^n \underbrace{\int l_{ij} y_i y_j N(y, \sigma(\xi)) dy}_{=0} \\ &+ \underbrace{\epsilon^2 \sum_{i=1}^n \sum_{j=1}^n \int a_{ji}^2 y_i^2 N(y, \sigma(\xi)) dy}_{= \epsilon^2 \sum_{i=1}^n \sum_{j=1}^n a_{ji}^2 \sigma(\xi)} \\ &= \epsilon^2 \sigma(\xi) \|A_{\sigma, \xi}^{-1}\|_F^2, \end{aligned}$$

wobei a_{ij} die Einträge der Matrix $A_{\sigma,\xi}^{-1}$ bezeichnen.

Insgesamt erhalten wir also

$$\Delta V(\xi) = \sigma(\xi)(A_{\sigma,\xi}^{-1}\nabla f(\xi))^T [\sigma(\xi)A_{\sigma,\xi}^{-1}\nabla f(\xi) - 2\xi] + \epsilon^2\sigma(\xi)\|A_{\sigma,\xi}^{-1}\|_F^2. \quad (4.14)$$

Zum Nachweis der Rekurrenz ist also zu zeigen

$$(A_{\sigma,\xi}^{-1}\nabla f(\xi))^T [\sigma(\xi)A_{\sigma,\xi}^{-1}\nabla f(\xi) - 2\xi] + \epsilon^2\|A_{\sigma,\xi}^{-1}\|_F^2 \leq 0$$

für alle $\xi \in C^c$. Diese Aussage ist aufgrund elementarer Umformungen äquivalent zu

$$\frac{\sigma(\xi)\|A_{\sigma,\xi}^{-1}\nabla f(\xi)\|_2^2 + \epsilon^2\|A_{\sigma,\xi}^{-1}\|_F^2}{2} \leq \nabla f(\xi)^T A_{\sigma,\xi}^{-1T} \xi.$$

Diese Ungleichung folgt sofort aus der Voraussetzung B und aus $\sigma(x) > 0$ für alle $x \in \mathbb{R}^n$. Das Driftkriterium (1.8) ist also erfüllt, und Φ^σ ist Harrisrekurrent. \square

Aus Satz 4.6 folgt sofort:

Satz 4.7 (Existenz eines invarianten Maßes zu Φ^σ).

Zu der Markovkette Φ^σ existiert ein subinvariantes Maß, das invariant ist.

Beweis. Die Aussage folgt mit Satz 4.6 direkt aus Satz 1.73.

4.7 Strenge Aperiodizität

Eine ähnliche Betrachtung der Markovkette Φ^σ wie im Beweis zu Satz 4.3 zeigt, daß Φ^σ auch streng aperiodisch ist.

Satz 4.8 (Strenge Aperiodizität von Φ^σ).

Die Markovkette Φ^σ ist streng aperiodisch.

Beweis. Es sei C wieder wie in Satz 4.6 gewählt, also

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

Weiterhin sei $A \in \mathcal{B}^n$ eine beliebige kompakte Menge. Da die Funktion k^σ – definiert in (4.8) – stetig ist und $k^\sigma(\xi, x) > 0$ für alle $(\xi, x) \in \mathbb{R}^n \times \mathbb{R}^n$, existiert ein $\delta^* \in \mathbb{R}$, $\delta^* > 0$ mit $k^\sigma(\xi, x) \geq \delta^*$ für alle $(\xi, x) \in A \times C$. Somit folgt für alle $x \in A$

$$\begin{aligned} P(x, B) &= \int_B k_x^\sigma(y) dy \\ &\geq \delta^* \lambda|_C(B), \end{aligned}$$

d.h. wir können ν_1 aus Definition 1.62 als $\delta^* \lambda|_C$ wählen. Somit genügt jede kompakte und somit nach Satz 4.5 auch unbedeutende Menge $A \in \mathcal{B}(\mathbb{R}^n)$ mit $\lambda|_C(A) > 0$ (also z.B. abgeschlossene Intervalle in C) den Anforderungen der Definition 1.62, so daß Φ^σ streng aperiodisch ist. \square

4.8 Positivität

Nun müssen wir noch zeigen, daß das nach Satz 4.7 existierende invariante Maß endlich ist, also als Wahrscheinlichkeitsmaß dargestellt werden kann. Hierzu verwenden wir das Drift-Kriterium (1.11).

Satz 4.9 (Positivität von Φ^σ).

Die Markovkette Φ^σ ist positiv-Harris.

Beweis. Nach Satz 1.76 genügt es, das Drift-Kriterium (1.11) zu zeigen. Hierzu wählen wir erneut

$$C := \{x \in \mathbb{R}^n : \|x\|_2 \leq r\}$$

und

$$V(x) := x^T x.$$

Wie bereits in (4.14) gezeigt (denn um diese Gleichung zu zeigen, wurde $\xi \in C^c$ noch nicht verwendet), gilt

$$\Delta V(\xi) = \sigma(\xi)(A_{\sigma,\xi}^{-1}\nabla f(\xi))^T [\sigma(\xi)A_{\sigma,\xi}^{-1}\nabla f(\xi) - 2\xi] + \epsilon^2\sigma(\xi)\|A_{\sigma,\xi}^{-1}\|_F^2.$$

Es ist also für alle $\xi \in \mathbb{R}^n$ die folgende Ungleichung zu zeigen

$$\sigma(\xi)(A_{\sigma,\xi}^{-1}\nabla f(\xi))^T [\sigma(\xi)A_{\sigma,\xi}^{-1}\nabla f(\xi) - 2\xi] + \epsilon^2\sigma(\xi)\|A_{\sigma,\xi}^{-1}\|_F^2 \leq -1 + b\mathbf{1}_C(\xi)$$

für eine Konstante $b < \infty$. Für $\xi \in C^c$ folgt dies wiederum aus Voraussetzung B. Da die Funktion

$$\Delta V(\xi) = \sigma(\xi)(A_{\sigma,\xi}^{-1}\nabla f(\xi))^T [\sigma(\xi)A_{\sigma,\xi}^{-1}\nabla f(\xi) - 2\xi] + \epsilon^2\sigma(\xi)\|A_{\sigma,\xi}^{-1}\|_F^2$$

stetig in ξ und C kompakt ist, nimmt sie ein Maximum $< \infty$ auf C an. Somit erhält man für $\xi \in C$

$$\Delta V(\xi) \leq \max_{x \in C} \Delta V(x) < \infty.$$

Somit ist das Drift-Kriterium mit

$$b := \max_{x \in C} \Delta V(x) + 1$$

erfüllt, und Φ^σ ist positiv-Harris. \square

Da die Markovkette Φ^σ positiv-Harris ist, existiert also insbesondere ein zu Φ^σ invariantes Wahrscheinlichkeitsmaß π . Im folgenden Abschnitt werden wir die Konvergenz der Verteilungen von Φ^σ gegen dieses Wahrscheinlichkeitsmaß π untersuchen.

4.9 Konvergenz der Verteilungen

Durch die Verifikation der Voraussetzung der Sätze 1.78 und 1.77 können wir uns nun Aussagen über das Konvergenzverhalten der Verteilungen der Markovkette Φ^σ zuwenden. Eben diese machen die eingangs erwähnte Stabilität des Eulerverfahrens mit beliebiger Schrittweitenfunktion aus.

Da die Markovkette Φ^σ nach Satz 4.8 und 4.9 aperiodisch und positiv-Harris ist, findet Satz 1.77 Anwendung, und es gilt für jeden Startpunkt $x \in \mathbb{R}^n$

$$\sup_{A \in \mathcal{B}(\mathbb{R}^n)} |P^k(x, A) - \pi(A)| \rightarrow 0 \quad \text{für } k \rightarrow \infty.$$

Darüber hinaus kann auch Satz 1.78 angewendet werden, und für alle Funktionen $h \in L_1(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \pi)$ gilt

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k h(\Phi_i^\sigma) = \int h d\pi \quad P_*\text{-fast sicher.} \quad (4.15)$$

Wählt man insbesondere für eine beliebige Menge $A \in \mathcal{B}^n$ die Funktion h als $h(x) = \mathbf{1}_A(x)$, so sagt (4.15) aus, daß die relative Häufigkeit, daß sich der Markovprozeß in der Menge A befindet, gegen die Wahrscheinlichkeit der Menge A bzgl. π konvergiert für P_* -fast jeden Pfad. Das heißt, wir können uns wieder auf die Beobachtung eines einzigen Pfades beschränken und erhalten damit alle relevanten Informationen. Wir fassen diese wichtigen Ergebnisse in dem folgenden Satz zusammen.

Satz 4.10 (Konvergenz der Verteilungen von Φ^σ).

Die Markovkette Φ^σ besitzt ein invariantes Wahrscheinlichkeitsmaß π , das die folgenden Eigenschaften hat:

i) *Für jeden Startpunkt $x \in \mathbb{R}^n$ gilt*

$$\sup_{A \in \mathcal{B}(\mathbb{R}^n)} |P^k(x, A) - \pi(A)| \rightarrow 0 \quad \text{für } k \rightarrow \infty.$$

ii) *Für alle Funktionen $h \in L_1(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \pi)$ gilt*

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k h(\Phi_i^\sigma) = \int h d\pi \quad P_*\text{-fast sicher.}$$

Die Markovkette Φ^σ ist also stabil in dem oben angesprochenen Sinne — ganz egal, welche stetige Schrittweitenfunktion σ gewählt wurde. Für die Belange der globalen Optimierung reicht es daher, das semi-impliziten Eulerverfahren mit einer beliebigen stetigen Schrittweitenfunktion σ , derart daß Voraussetzung

B erfüllt ist, anzuwenden. Es ist also nicht nötig, spezielle Schrittweitenfunktionen σ zu wählen, um Stabilität zu gewährleisten. Einer konkreten Vorgehensweise zur Schrittweitensteuerung wenden wir uns in der Implementation im nächsten Kapitel zu.

Die vorausgegangene Beweisführung erlaubt es uns, bei Bedarf das gleiche Abbruchkriterium zu verwenden, wie wir es schon beim Metropolisalgorithmus entwickelt haben. Erneut liegt der Schlüssel darin, daß bereits alle Informationen zur Verteilung in *einem* Pfad enthalten sind. Aus Satz 4.10 wissen wir, daß für jede Funktion $h \in L_1(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \pi)$ gilt

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k h(\Phi_i^\sigma) = \int h \, d\pi \quad P_*\text{-fast sicher.} \quad (4.16)$$

Wählen wir also als Funktion h wie bereits in 2.2 erneut eine Funktion, die für alle $x \in \mathbb{R}^n$ die Funktionswerte von f widerspiegelt, möglichst in dem Sinne, daß geringe Funktionswerte ein größeres Gewicht erhalten, so kann die Beobachtung der Konvergenz der Größe

$$\frac{1}{k} \sum_{i=1}^k h(\Phi_i^\sigma)$$

als Abbruchkriterium herangezogen werden. In Anlehnung an die Grenzdichte des Schäffler-Verfahrens wählen wir

$$h(x) = \exp\left(-\frac{f(x)}{\epsilon}\right)$$

und beobachten somit die Konvergenz der Größe

$$\frac{1}{k} \sum_{i=1}^k \exp\left(-\frac{f(\Phi_i^\sigma)}{\epsilon}\right).$$

Kapitel 5

Numerische Ergebnisse

5.1 Anmerkungen zur Implementierung

In der Praxis verliert die Einschränkung der Stetigkeit an die Schrittweitensteuerungsfunktion σ an Bedeutung, da mit Wahrscheinlichkeit 1 kein Punkt mehr als einmal erzeugt wird und die erzeugten Punkte diskret im \mathbb{R}^n liegen. Somit kann immer eine stetige Funktion σ gefunden werden, die an den erzeugten Punkten ξ_t , $t \in \mathbb{N}$, mit den gewählten Schrittweiten $\sigma(\xi_t)$, $t \in \mathbb{N}$, übereinstimmt. Wir können also zu jedem beliebigen Zeitpunkt eine beliebige Schrittweite wählen.

In der hier verwendeten Implementierung, die im nächsten Abschnitt genau aufgeführt wird, geben wir eine Startschrittweite σ_0 an. Diese wird bei Bedarf verkleinert, so daß für den Startpunkt x_0 die Matrix $\mathbb{E}_n + \sigma_0 \nabla^2 f(x_0)$ invertierbar ist. Im folgenden werden aufeinanderfolgend Schritte des Euler-Verfahrens realisiert, wobei wiederum bei Bedarf die Schrittweite $\sigma(\xi)$ für jeden erzeugten Punkt ξ verkleinert wird, so daß $\mathbb{E}_n + \sigma(\xi) \nabla^2 f(\xi)$ invertierbar ist. Dann erfolgt die weitere Schrittweitensteuerung durch Beobachtung der Funktionswerte im Verlauf des Algorithmus. Sollten sich nach längerem Verlauf des Euler-Verfahrens nur unwesentliche Veränderungen der Funktionswerte ergeben, so kann davon ausgegangen werden, daß der Algorithmus in einem Minimierer stagniert. Leider kann nicht sichergestellt werden, daß es sich hierbei um ein globales und nicht um ein lokales Minimum handelt. In der Praxis hat sich gezeigt, daß eine Verkleinerung der Schrittweite die Stagnation des Algorithmus in einem lokalen Minimierer schneller aufheben kann.

Desweiteren hat es sich in der Praxis für die Laufzeit des Verfahrens als günstig erwiesen, die Schrittweite zu erhöhen, sollte der Funktionswert des neu ermittelten Punktes kleiner oder gleich dem Funktionswert des derzeitigen Ausgangspunktes sein, und die Schrittweite im umgekehrten Fall zu erniedrigen.

Mit dieser Vorgehensweise wird bei den Beispielpunkten sehr schnell ein Punkt

nahe dem globalen Minimierer erreicht, so daß ein lokales Minimierungsverfahren mit einem solchen erzeugten Punkt zum globalen Minimierer führt. Auch eine „Schrittweitensteuerung“, bei der die Schrittweite nur insoweit angepaßt wird, als daß die Matrix $\mathbb{E}_n + \sigma_0 \nabla^2 f(x_0)$ regulär ist und ansonsten nicht verändert wird, wurde implementiert. Auch diese Vorgehensweise führt nach einer längeren Laufzeit des Algorithmus in die Nähe des globalen Minimierers.

Wir brechen den Algorithmus gemäß des Abbruchkriteriums 4.16 ab. Dabei vergleichen wir die Größen immer nur nach 50 Schritten, um zu vermeiden, daß ein kleiner erzeugter Funktionswert bei relativ großem k fälschlicherweise zum Abbruch führt. Nach Ablauf des Algorithmus sollte von dem Punkt mit dem kleinsten im Durchlauf des Algorithmus erhaltenen Funktionswert ein Verfahren der lokalen Optimierung angewandt werden.

5.2 Formulierung des Algorithmus

Im folgenden gebe ich den Algorithmus zur Bestimmung einer Realisierung der Markovkette Φ^σ an.

Schritt 1: *Initialisierung*

Wähle einen Startpunkt $\Phi_0^\sigma := x_s$, eine Startschrittweite σ_s , eine Blockgröße l zur Untersuchung der Variation der Funktionswerte, eine minimale akzeptierbare Variationsgröße δ sowie einen Gewichtungsfaktor ϵ .

Setze $k := 1$.

Setze $f_{\max} := f(\Phi_0^\sigma)$.

Setze $f_{\min} := f(\Phi_0^\sigma)$.

Setze $\sigma := \sigma_s$.

Gehe zu Schritt 2.

Schritt 2: *Bestimmung der Ableitungen*

Berechne $\nabla f(\Phi_{k-1}^\sigma)$ und $\nabla^2 f(\Phi_{k-1}^\sigma)$.

Gehe zu Schritt 3.

Schritt 3: *Bestimmung einer Ausgangsschrittweite*

Bestimme $\mathbb{E}_n + \sigma \nabla^2 f(\Phi_{k-1}^\sigma)$.

Fall 1: $\mathbb{E}_n + \sigma \nabla^2 f(\Phi_{k-1}^\sigma)$ ist positiv definit

Gehe zu Schritt 4.

Fall 2: $\mathbb{E}_n + \sigma \nabla^2 f(\Phi_{k-1}^\sigma)$ ist nicht positiv definit

Setze $\sigma := \frac{1}{2} \sigma$.
 Gehe zu Schritt 3.

Schritt 4: *Realisierung einer normalverteilten Zufallszahl*

Berechne eine Realisierung N^* eines $N(0, \sigma \mathbb{E}_n)$ Gauß-verteilten Zufallsvektors.
 Gehe zu Schritt 5.

Schritt 5: *Berechnung von Φ_k^σ*

Berechne Φ_k^σ mittels Lösung des linearen Gleichungssystems
 $(\mathbb{E}_n + \sigma \nabla^2 f(\Phi_{k-1}^\sigma)) \Phi_k^\sigma = \sigma \nabla f(\Phi_{k-1}^\sigma) - \epsilon N^*$ und
 $\Phi_k^\sigma := \Phi_{k-1}^\sigma - \Phi_k^\sigma$
 Gehe zu Schritt 6.

Schritt 6: *Bestimmung einer neuen Schrittweite nach einem Schritt*

Fall 1: $f(\Phi_k^\sigma) \leq f(\Phi_{k-1}^\sigma)$

Setze $\sigma := 2 \sigma$.

falls: $f(\Phi_k^\sigma) < f_{\min}$

Setze $f_{\min} := f(\Phi_k^\sigma)$

Gehe zu Schritt 7.

Fall 2: $f(\Phi_k^\sigma) > f(\Phi_{k-1}^\sigma)$

Setze $\sigma := \frac{1}{2} \sigma$.

falls $f(\Phi_k^\sigma) > f_{\max}$

Setze $f_{\max} := f(\Phi_k^\sigma)$.

Gehe zu Schritt 7.

Schritt 7: *Bestimmung einer neuen Schrittweite nach l Schritten*

falls $k \equiv 0 \pmod{l}$

falls: $f_{\max} - f_{\min} \leq \delta$

Setze $\sigma := \frac{1}{5} \sigma$.

Setze $f_{\max} := f(\Phi_k^\sigma)$.

Setze $f_{\min} := f(\Phi_k^\sigma)$.

Setze $k := k + 1$.

Gehe zu Schritt 2.

5.3 Numerische Ergebnisse

Der Algorithmus wurde mit Mathematica 4.0 implementiert. Der Punkt mit dem jeweils niedrigsten Funktionswert wurde als Startpunkt für ein Verfahren der lokalen Optimierung gewählt.

Wir betrachten die numerischen Ergebnisse für die sechs folgenden Beispielfunktionen, jeweils auf dem \mathbb{R}^n definiert:

Problem 1: $n=50$

$$f(x) := 1 + 6x_1^2 - \cos(12x_1) + 590 \sum_{i=2}^{50} (x_i - x_{i-1})^2$$

Dieses Problem besitzt mehrere lokale Minimierer sowie seinen eindeutigen globalen Minimierer bei $x^* = 0$ mit Funktionswert $f(x^*) = 0$.

Problem 2: $n=50$

$$f(x) := 10 + 30x_{49}^2 - 5 \cos(12x_{49}) + 890 \sum_{i=1}^{48} (x_i + 3x_{i+1}^2 - x_{i+2})^2 + 30x_{50}^2 - 5 \cos(12x_{50}).$$

Dieses Problem hat mehrere lokale Minimierer. Der eindeutige globale Minimierer liegt bei $x^* = 0$ mit Funktionswert $f(x^*) = 0$.

Problem 3: $n=50$

$$f(x) := \sqrt{3 + 19x_1^2 - 2 \cos(19x_1) - 19x_1^2 \cos(19x_1) + 90.25x_1^4 - \sin^2(19x_1)} + 1000 \sum_{i=2}^{50} (x_i - \ln(x_{i-1}^2 + 1))^2 - 1$$

Dieses Problem besitzt mehrere lokale Minimierer, wobei der eindeutige globale Minimierer bei $x^* = 0$ mit Funktionswert $f(x^*) = 0$ liegt.

Problem 4: $n=50$

$$f(x) := 2 + 12x_{50}^2 - 2 \cos(12x_{50}) + 720 \sum_{i=1}^{49} (x_i - \sin(\cos(x_{i+1}) - 1))^2.$$

Dieses Problem hat mehrere lokale Minimierer. Die Funktion besitzt ihren eindeutigen globalen Minimierer bei $x^* = 0$ mit Funktionswert $f(x^*) = 0$.

Problem 5: $n=50$

$$f(x) := 9.5x_1^2 - \cos(19x_1) + 792 \sum_{i=3}^{50} (3x_i + \sin^2(x_{i-1}))^2 \\ + \sqrt{3 + 12x_2^2 - 2 \cos(12x_2) - 12x_2^2 \cos(12x_2) + 36x_2^4 - \sin^2(12x_2)}.$$

Die Funktion besitzt mehrere lokale Minimierer mit dem eindeutigen globalen Minimierer bei $x^* = 0$ mit Funktionswert $f(x^*) = 0$.

Problem 6: $n=50$

$$f(x) := 6x_{50}^2 - \cos(12x_{50}) + 980 \sum_{i=1}^{47} (x_i - x_{i+2}^2)^2 + 15x_{49}^2 \\ - 2.5 \cos(12x_{49}) + 18x_{48}^2 - 3 \cos(12x_{48}) + 6.5.$$

Dieses Problem hat mehrere lokale Minimierer mit eindeutigem globalen Minimum bei $x^* = 0$ mit Funktionswert $f(x^*) = 0$.

Als Startpunkt für die globale Optimierung wurde jeweils der Punkt

$$x_s = (x_{s_1}, x_{s_2}, \dots, x_{s_n})^T \quad \text{mit } x_{s_i} = 1 \text{ für alle } i \in \{1, 2, \dots, n\}$$

gewählt. Die Funktionen erfüllen die Voraussetzung B nicht. Aus diesem Grund wurden die Funktionen wie in 4.3 angegeben modifiziert. Als Radius r wurde $\|x_s\|$ verwendet, als Schrittweite $\sigma(x) = 0.1$ für alle x mit $\|x\| > r$. In der Darstellung der Ergebnisse für die einzelnen Funktionen wird der Funktionswert f_{lok} , der durch die lokale Optimierung mit Startpunkt x_s ermittelt wurde, ebenso angegeben, wie auch der Parameter ϵ und die Startschrittweite σ_s sowie der kleinste durch die globale Optimierung ermittelte Funktionswert f_{glob} bei x_{glob} und der Funktionswert f_{glob}^* , der durch die lokale Optimierung mit Startpunkt x_{glob} erreicht wurde. Desweiteren wird die Anzahl an Iterationen, die benötigt wurde, um den Punkt x_{glob} zu erreichen, angeführt.

Problem 1

$$\epsilon = 1, \sigma_s = 1$$

Funktionswert nach lokaler Optimierung von Startpunkt x_s aus:

$$f_{\text{lok}} = 6.02054$$

Funktionswert nach globaler Optimierung:

$$f_{\text{glob}} = 1.2908$$

Funktionswert nach lokaler Optimierung mit Startpunkt x_{glob} :

$$f_{\text{glob}}^* = 3.31826 * 10^{-5}$$

Benötigte Anzahl an Iterationen: 400

Problem 2

$$\epsilon = 1, \sigma_s = 0.1$$

Funktionswert nach lokaler Optimierung von Startpunkt x_s aus:

$$f_{\text{lok}} = 60.205$$

Funktionswert nach globaler Optimierung:

$$f_{\text{glob}} = 0.0976151$$

Funktionswert nach lokaler Optimierung mit Startpunkt x_{glob} :

$$f_{\text{glob}}^* = 5.92687 * 10^{-6}$$

Benötigte Anzahl an Iterationen: 400

Problem 3

$$\epsilon = 7, \quad \sigma_s = 100$$

Funktionswert nach lokaler Optimierung von Startpunkt x_s aus:

$$f_{\text{lok}} = 7.89164$$

Funktionswert nach globaler Optimierung:

$$f_{\text{glob}} = 1.91267$$

Funktionswert nach lokaler Optimierung mit Startpunkt x_{glob} :

$$f_{\text{glob}}^* = 6.10204 * 10^{-6}$$

Benötigte Anzahl an Iterationen: 100

Problem 4

$$\epsilon = 10, \quad \sigma_s = 1$$

Funktionswert nach lokaler Optimierung von Startpunkt x_s aus:

$$f_{\text{lok}} = 3.03193$$

Funktionswert nach globaler Optimierung:

$$f_{\text{glob}} = 0.995625$$

Funktionswert nach lokaler Optimierung mit Startpunkt x_{glob} :

$$f_{\text{glob}}^* = 1.36474 * 10^{-19}$$

Benötigte Anzahl an Iterationen: 450

Problem 5

$$\epsilon = 7, \sigma_s = 1$$

Funktionswert nach lokaler Optimierung von Startpunkt x_s aus:

$$f_{\text{lok}} = 4.75655$$

Funktionswert nach globaler Optimierung:

$$f_{\text{glob}} = 2.30714$$

Funktionswert nach lokaler Optimierung mit Startpunkt x_{glob} :

$$f_{\text{glob}}^* = 3.24613 * 10^{-8}$$

Benötigte Anzahl an Iterationen: 650

Problem 6

$$\epsilon = 7, \quad \sigma_s = 100$$

Funktionswert nach lokaler Optimierung von Startpunkt x_s aus:

$$f_{\text{lok}} = 39.8415$$

Funktionswert nach globaler Optimierung:

$$f_{\text{glob}} = 0.65836$$

Funktionswert nach lokaler Optimierung mit Startpunkt x_{glob} :

$$f_{\text{glob}}^* = 0.$$

Benötigte Anzahl an Iterationen: 250

Zusammenfassung

Die Grundlage dieser Arbeit sind zwei Verfahren zur stochastischen globalen Minimierung: zunächst der Metropolis-Algorithmus zur diskreten globalen Optimierung, dann das von Schäffler in [Schäffler 95] vorgeschlagene Verfahren zur globalen kontinuierlichen Optimierung. Die Zielsetzung dieser Arbeit besteht darin, beide Verfahren in der Praxis effizienter zu gestalten.

- Das Problem des Metropolis-Algorithmus in der praktischen Umsetzung liegt darin, zu erkennen, wann eine Verteilung hinreichend nahe an der Grenzverteilung erreicht ist. In Kapitel 2.2 wird gezeigt, daß als Instrument hierzu die Beobachtung der Konvergenz der Größe

$$\frac{1}{n} \sum_{k=1}^n h(\Psi_k).$$

geeignet ist, wobei Ψ_k die Ausgabe des Metropolis-Algorithmus im k -ten Schritt angibt, sowie h eine beliebige Funktion aus $L_1(R, \mathcal{B}(R), g)$ ist, wobei R der endliche Zustandsraum des Minimierungsproblems und g die Grenzverteilung des Metropolis-Algorithmus darstellt.

- Das Problem des von Schäffler in [Schäffler 95] vorgeschlagenen Verfahrens besteht darin, eine für die Praxis geeignete Schrittweitensteuerung anzugeben, die weniger die genaue numerische Berechnung der Lösung der stochastischen Differentialgleichung,

$$\Phi_t^{\epsilon, x_0}(F) = x_0 + \epsilon ((B_t(F) - B_0(F)) - \int_0^t \nabla f(\Phi_s^{\epsilon, x_0}(F)) ds) \quad (3.1)$$

als vielmehr das Bestehenbleiben der Stabilität der entstehenden Markovkette verfolgt. Das Ziel dieser Vorgehensweise liegt darin, deutlich kürzere Laufzeiten des Algorithmus zu erreichen. Die in [Schäffler 95] vorgeschlagene Schrittweitensteuerung war ausschließlich darauf ausgerichtet, eine numerische Lösung der stochastischen Differentialgleichung (3.1) zu erhalten. Eine Konvergenz des Eulerverfahrens

$$\Phi_t^\sigma(\omega) = \xi - (\mathbb{E}_n + \sigma(\xi) \nabla^2 f(\xi))^{-1} (\sigma(\xi) \nabla f(\xi) - \epsilon (B_{t+\sigma(\xi)}(\omega) - B_t(\omega))) \quad (4.4)$$

gegen die tatsächliche Lösung und somit die Annäherung der Verteilung an eine Dichte, die für die globale Optimierung wichtige Eigenschaften besitzt, kann aber nur für kleine Schrittweiten σ sichergestellt werden, was gleichzeitig eine sehr lange Laufzeit des Algorithmus voraussetzt. Aus diesem Grund wird in Kapitel 4 untersucht, für welche Schrittweiten die Stabilität einer durch das Eulerverfahren (4.4) erzeugten Markovkette erhalten bleibt. Um diese Stabilität zu zeigen, benötigen wir neben der Voraussetzung A aus [Schäffler 95] an die Zielfunktion $f \in C^2$

Voraussetzung A:

Es existiert ein $\epsilon > 0$ derart, daß es ein $r > 0$ gibt mit

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

für alle $x \in \mathbb{R}^n \setminus \{x \in \mathbb{R}^n : \|x\|_2 \leq r\}$, wobei $\nabla f(x)$ den Gradienten der Zielfunktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ an der Stelle x bezeichne,

eine weitere Voraussetzung an die Zielfunktion:

Voraussetzung B:

Es existiert ein $r' \in \mathbb{R}$, $r' > 0$, und eine stetige Funktion $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}$, $\sigma(x) > 0$ für alle $x \in \mathbb{R}^n$, so daß für alle $x \in \mathbb{R}^n$ die Matrix $A_{\sigma,x} := \mathbb{E}_n + \sigma(x)\nabla^2 f(x)$ invertierbar ist und für alle $x \in \mathbb{R}^n$ mit $\|x\|_2 > r'$ gilt

$$\frac{\sigma(x)\|A_{\sigma,x}^{-1}\nabla f(x)\|_2^2 + \epsilon^2\|A_{\sigma,x}^{-1}\|_F^2 + \frac{1}{\sigma(x)}}{2} \leq \nabla f(x)^T A_{\sigma,x}^{-1} x.$$

Wählen wir eine die Voraussetzung B erfüllende Funktion σ und verwenden als Schrittweite für das Eulerverfahren ausgehend von dem Punkt ξ den Funktionswert $\sigma(\xi)$, so kann durch Anwendung einiger Sätze aus der Theorie der Markovketten die Stabilität nachgewiesen werden. Die genannten Voraussetzungen an die Funktion σ lassen bei der Wahl der Schrittweite sehr viel Freiheit.

Die Gültigkeit der Voraussetzung B an die Zielfunktion bezieht sich ausschließlich auf das Verhalten der Zielfunktion außerhalb der Kugel um den

Ursprung mit Radius r' , denn innerhalb dieser Kugel kann aufgrund der Stetigkeit von $\nabla^2 f$ und der Invertierbarkeit von \mathbb{E}_n immer eine stetige Funktion σ gefunden werden, so daß $A_{\sigma, x}$ invertierbar ist. Wir wählen den Radius r' so groß, daß er den Radius r der Voraussetzung A übersteigt. Mit dieser Wahl ist sichergestellt, daß der globale Minimierer in der Kugel um den Ursprung mit Radius r' enthalten ist. Da eine zu minimierende Funktion, die die Voraussetzung B nicht erfüllt, so gegen eine Hilfsfunktion ausgetauscht werden kann, daß diese Hilfsfunktion die Voraussetzung B erfüllt und die gleichen globalen Minimierer besitzt wie die Ausgangsfunktion, bedeutet die Voraussetzung B keine einschneidende Einschränkung an die Zielfunktion.

Die Schrittweitensteuerung wird in Kapitel 5 für einige Beispielfunktionen getestet und liefert gute Ergebnisse, in dem Sinne, daß der globale Minimierer innerhalb relativ kurzer Laufzeit erreicht wird. In diesem Zusammenhang stellt diese Erweiterung des Verfahrens von Schäffler eine Möglichkeit dar, globale Minimierung mittels stochastischer Integration effizienter zu gestalten.

Danksagung

Ich möchte mich ganz herzlich bei meinem Betreuer und Doktorvater, Herrn Prof. Dr. Dr. Schäffler, bedanken. Er stand mir stets mit Rat zur Seite und hat immer ein offenes Ohr für meine mathematischen Probleme gehabt. Desweiteren erwies er mir wertvolle „psychologische Dienste“, indem er mir in Phasen, in denen mich die mir teilweise minimal erscheinenden Fortschritte sehr belasteten, immer wieder Bestärkung gegeben und mich somit aufgebaut hat. Auf seinen Einsatz ist größtenteils mein Stipendium von der Siemens AG — und somit die Finanzierung meiner Promotion — zurückzuführen, und auch dafür möchte ich mich ganz herzlich bei ihm bedanken.

Mein besonderer Dank gilt dem mir von der Siemens AG zur Seite gestellten Betreuer, Herrn Dr. Hoever. Sein unermüdlicher Einsatz, sein großer Zeitaufwand für die Belange meiner Betreuung waren bewundernswert und zugleich vorbildlich. Aus den fachlichen Diskussionen mit ihm entwickelte sich manch neue Idee. Mit seinem mathematischen Wissen ersparte er mir so manchen Irrweg. Mit Freude und großem Engagement übernahm er spontan die Betreuung meiner Arbeit, nachdem mein erster von der Siemens AG gestellter Betreuer, Dr. Meintrup, die Firma verlassen hatte.

Ich danke der Siemens AG — insbesondere der CT PP2 und meinen beiden Vorgesetzten Herrn Prof. Dr. Gilg sowie Herrn Dr. Buchmeier — für das Stipendium, das mir eine Promotion erst ermöglicht hat. Das Stipendium bezog sich nicht nur auf die finanzielle Unterstützung meiner Arbeit sondern auch auf die Bereitstellung von Räumlichkeiten, Rechnerleistung und der entsprechenden Software sowie auf die Unterstützung meiner Arbeit durch die Betreuer vor Ort — zunächst Dr. Meintrup, später dann Dr. Hoever.

Außerordentlicher Dank gilt meinem Mann Mathias, der mich während dieser Promotion immer mit allen Kräften unterstützt hat. Mein Aufenthalt in München unter der Woche und somit die Wochenendbeziehung, die wir führten, bedeutete auch für ihn eine große Anstrengung, über die er sich jedoch nie beklagte, sondern ganz im Gegenteil mich immer wieder in meinem Entschluß bestärkte und mir in allen Phasen der Erstellung dieser Arbeit zur Seite stand. Ich weiß das sehr zu schätzen und möchte ihm daher auch an dieser Stelle meinen Dank aussprechen.

Bis zu seinem Austritt aus der Siemens AG wurde mir Herr Dr. Meintrup als Betreuer zur Seite gestellt. Auch ihm danke ich für seine Zeit, die er für mich aufbrachte, und seinen mathematischen Einsatz.

Außerdem möchte ich auch Frau Stiller aus dem Sekretariat der CT PP2 meinen Dank aussprechen, die mich insbesondere bei meiner Wohnungssuche sehr stark unterstützt hat, aber mir auch bei allen organisatorischen Fragen hilfreich und engagiert zur Seite stand.

Last but not least geht mein Dank an zahlreiche Freunde und Kollegen, die sich angeboten haben, diese Arbeit durchzusehen und sie durch konstruktive Kritik zu bereichern.

Index

- adaptiert, 10
- äquivalente Maße, 13
- Akzeptanzwahrscheinlichkeit, 32
- Algorithmus, 59
- aperiodisch, 25
- Aperiodizität des Metropolis-Algorithmus, 33
- Aperiodizität von Φ^σ , 54

- Besetzungszeit, 22
- Borel- σ -Algebra, 8
- Brownsche Bewegung, 10
- Brownsche Bewegung auf $(\Omega^n, \mathcal{B}(\Omega^n), W^n)$, 11

- Chapman-Kolmogorov-Gleichungen, 21

- δ -Umgebung, 7
- Dichte, 12
- Dichte der Übergangswahrscheinlichkeiten von Φ^σ , 49
- Drift-Kriterium für Rekurrenz, 27

- Eindeutigkeitssatz, 19
- elementarer stochastischer Prozeß, 15
- elementarer stochastischer Prozess
stochastisches Integral für, 16
- ergodisch, 35
- Existenz eines invarianten Maßes für Φ^σ , 54
- Existenz und Eindeutigkeit von $\{\Phi_t^{\epsilon, x_0}\}$, 41
- Existenzsatz, 19

- Filtration, 9
 - kanonische, 10

- Fokker-Planck-Gleichung, 20, 42
- Funktionenräume, 7

- Generierungswahrscheinlichkeit, 32
- Gesetze vom iterierten Logarithmus, 15
- Grenzdichte p_ϵ , 41

- Harris-Rekurrenz, 26
- Harris-Rekurrenz von Φ^σ , 52

- invariantes Maß, 27
- φ -Irreduzibilität, 22
- Irreduzibilität des Metropolis-Algorithmus, 33
- Irreduzibilität von Φ^σ , 50
- Itô-Integral, 17

- kanonische Filtration, 10
- kleine Menge, 24

- Lösung einer stochastischen Differentialgleichung, 19
- Lebesgue-Maß, 8

- Markovkette, 21

- Norm
 - euklidische, 7
 - Frobeniusnorm, 7
- numerische Funktion, 9

- P -fast überall, 9
- P -fast sicher, 9
- Periode, 25
- positiv, 27
- positiv-Harris, 27
- Positivität von Φ^σ , 55

- progressiv meßbar, 14
- Pseudozufallszahlen, 43
- Rückkehrzeit, 22
- Radon-Nikodym, 13
- regulär, 29
- Rekurrenz, 25
- σ -endliches Maß, 13
- Satz
 - Radon-Nikodym, 13
- stationäre Verteilung, 33
- stationäre Verteilung des Metropolis-Algorithmus, 34
- stochastische Differentialgleichung, 18
 - Itô, 18
- stochastische Unabhängigkeit
 - einer Menge von Ereignissen, 12
 - von Mengensystemen, 12
 - von Zufallsvariablen, 12
 - von Ereignissen, 12
- stochastischer Prozeß, 9
 - elementarer, 15
 - Pfad eines stochastischen Prozesses, 9
 - Stetigkeit eines stochastischen Prozesses, 9
- stochastisches Integral für elementare Prozesse, 16
- Strenge Aperiodizität von Φ^σ , 54
- subinvariantes Maß, 28
- Übergangswahrscheinlichkeit des Metropolis-Algorithmus, 33
- Übergangswahrscheinlichkeitskern, 20
- unbedeutende Menge, 23
- unbeschränkt außerhalb von kleinen Mengen, 27
- Verfahren zur globalen Optimierung
 - stochastische Differentialgleichung, 40
- Vorwärtsgleichung von Kolmogorov, 20
- Wiener-Maß, 11
- Wiener-Raum, 11
- Zeitpunkt des ersten Treffens, 22
- Zufallsvariable, 9
 - Verteilung einer Zufallsvariablen, 11
- Zufallszeit, 14
- d -Zykel, 25
- d -Zyklizität, 25

Literaturverzeichnis

- [Bauer 74] H. BAUER: *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*. Walter de Gruyter Verlag, Berlin, New York 1974
- [Bauer 90] H. BAUER: *Maß- und Integrationstheorie*. Walter de Gruyter Verlag, Berlin, New York 1990
- [Billingsley 85] P. BILLINGSLEY: *Probability and Measure*. Wiley & Sons, New York 1985
- [Catthoor 88] F. CATTHOOR, H.DE MAN: *SAMURAI: A general and efficient simulated-annealing schedule with fully adaptive annealing parameters*, *Integration, the VLSI journal* 6 (1988) S.147-178
- [Černý 85] V. ČERNÝ: *Thermodynamical Approach to the traveling salesman problem: An efficient simulation algorithm*, *J. Opt. Theory Appl.*, (1985) S.41-51
- [Ermakov 75] S.M. ERMAKOW: *Die Monte-Carlo-Methode und verwandte Fragen*. R. Oldenburg-Verlag München Wien 1975, S.44-49
- [Feller 68] W. FELLER: *An Introduction to Probability Theory an Its Applications*, Volume I. John Wiley & Sons, New York London Sydney 1968
- [Feller 70] W. FELLER: *An Introduction to Probability Theory an Its Applications*, Volume II. John Wiley & Sons, New York London Sydney 1970
- [Hasminskij 80] R.Z. HASMINSKIJ: *Stochastic Stability of Differential Equations*, Sijthoff & Noordhoff, Amsterdam, 1980
- [Kirkpatrick 82] S. KIRKPATRICK, C.D. GELATT Jr., M.P. VECCHI: *Optimization by Simulated Annealing*, IBM Research Report RC 9355, 1982

- [Laar Aarts 87] P.J.M VAN LAARHOVEN, E.H.L. AARTS: *Simulated Annealing: Theory and Applications*. D.Reidel Publishing Company, Dordrecht 1987
- [Metropolis 53] N. METROPOLIS, A. ROSENBLUTH, A. TELLER, E. TELLER: *Equation of State Calculations by fast computing machines*, J. of Chem. Physics, 21 (1953) 1087-1092
- [Meyn Tweedie 96] S. P. MEYN, R. L. TWEEDIE: *Markov-Chains and Stochastic Stability*. Springer-Verlag Berlin Heidelberg New York 1996
- [Øksendal 98] B. ØKSENDAL: *Stochastic Differential Equations - An Introduction with Applications*. Springer, Berlin Heidelberg 1998
- [Protter 90] P. PROTTER: *Stochastic Integration and Differential Equations*. Berlin, New York 1990
- [Schäffler 95] S. SCHÄFFLER: *Global Optimization Using Stochastic Integration*. S. Roderer Verlag, Regensburg 1995
- [Schäffler 96] S. SCHÄFFLER: *Stochastische Differentialgleichungen - Theorie, Numerik, Anwendungen*. Schriftenreihe des Instituts für Angewandte Mathematik und Statistik der Technischen Universität München, Nr.7 August 1996
- [Schäffler Sturm 94] S. SCHÄFFLER, T. F. STURM: *Wahrscheinlichkeitstheorie und Statistik I*. Schriftenreihe des Instituts für Angewandte Mathematik und Statistik der Technischen Universität München, Nr.5 Oktober 1994
- [Schäffler Sturm 95] S. SCHÄFFLER, T. F. STURM: *Wahrscheinlichkeitstheorie und Statistik II*. Schriftenreihe des Instituts für Angewandte Mathematik und Statistik der Technischen Universität München, Nr.6 März 1995

Symbolverzeichnis

$\mathfrak{P}(A)$	Potenzmenge von A , Seite 6
\mathbb{N}	Menge der natürlichen Zahlen ohne Null, Seite 6
\mathbb{N}_0	$:= \mathbb{N} \cup \{0\}$, Seite 6
\mathbb{R}_0^+	$:= \{x \in \mathbb{R} : x \geq 0\}$, Seite 6
$\overline{\mathbb{R}}$	Menge der um $\{\pm\infty\}$ erweiterten reellen Zahlen, Seite 6
A_i	i -te Zeile der Matrix A , Seite 7
A^j	j -te Spalte der Matrix A , Seite 7
A_{ij}	Eintrag an der Stelle (i, j) der Matrix A , Seite 7
$C(U, \mathbb{R}^{n,m})$	$:= \{f : U \rightarrow \mathbb{R}^{n,m} : f \text{ ist stetig}\}$, Seite 7
$C(U)$	$:= C(U, \mathbb{R})$, Seite 7
$C_b(U)$	$:= \{f \in C(U) : f \text{ ist beschränkt}\}$, Seite 7
$C_c(U)$	$:= \{f \in C(U) : f \text{ hat einen kompakten Träger in } U\}$, Seite 7
$C^q(U, \mathbb{R}^{n,m})$	$:= \{f : U \rightarrow \mathbb{R}^{n,m} : f \text{ ist } q\text{-mal stetig differenzierbar}\}$, Seite 7
$C^q(U)$	$:= C^q(U, \mathbb{R})$, Seite 7
$C_b^q(U)$	$:= C^q(U) \cap C_b(U)$, Seite 7
$C_c^q(U)$	$:= C^q(U) \cap C_c(U)$, Seite 7
$D_i f(x)$	$:= \begin{pmatrix} \frac{\partial f_{11}(x)}{\partial x_i} & \cdots & \frac{\partial f_{1m}(x)}{\partial x_i} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{n1}(x)}{\partial x_i} & \cdots & \frac{\partial f_{nm}(x)}{\partial x_i} \end{pmatrix}$, Seite 8
$Jf(x)$	$:= (\nabla f_1(x), \nabla f_2(x), \dots, \nabla f_m(x)) := \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1(x)}{\partial x_n} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{pmatrix}$, Seite 8

\mathcal{B}^n	Borel- σ -Algebra im \mathbb{R}^n , Seite 8
$\mathcal{B}_{[a,b]}$	Borel- σ -Algebra auf dem Intervall $[a, b]$, Seite 8
$\overline{\mathcal{B}}$	$:= \{A \in \mathfrak{P}(\overline{\mathbb{R}}) : A \cap \mathbb{R} \in \mathcal{B}\}$, Seite 8
$\mathcal{B}^+(X)$	$:= \{A \in \mathcal{B}(X) : \psi(A) > 0\}$, Seite 23
λ^n	n -dimensionales Lebesgue-Maß, Seite 8
$\lambda _C$	n -dimensionales Lebesgue-Maß eingeschränkt auf C , Seite 50
$(\Omega^n, \mathcal{B}(\Omega^n), W^n)$	Wiener-Raum, Seite 11
$\{B_t : t \geq 0\}$	Brownsche Bewegung auf dem Wiener-Raum, Seite 11
$P(x, A)$	Übergangswahrscheinlichkeit vom Punkt x in die Menge A , Seite 21
P_μ	Wahrscheinlichkeit bedingt auf die Anfangsverteilung μ , Seite 21
P_x	Wahrscheinlichkeit bedingt auf $\Phi_0 = x$, Seite 21
$K_{a\frac{1}{2}}(y, A)$	$:= \sum_{n=0}^{\infty} P^n(y, A)2^{-(n+1)}$, Seite 23
$K_a(x, A)$	$:= \sum_{n=0}^{\infty} P^n(x, A)a(n)$, Seite 24
$L(x, A)$	$:= P_x(\tau_A < \infty)$, Seite 22
$Q(x, A)$	$:= P_x(\eta_A = \infty)$, Seite 26
$U(x, A)$	$:= \sum_{n=1}^{\infty} P^n(x, A)$, Seite 22
P_x	$:= P_{\delta_x}$, Seite 21
δ_x	Dirac-Maß mit Gewicht auf x , Seite 21
E_C^ν	$:= \{n \geq 1 : C \text{ ist } (n, \nu_n)\text{-unbedeutend mit } \nu_n = \delta_n \nu \text{ für ein } \delta_n > 0\}$, Seite 24
η_A	$:= \sum_{n=1}^{\infty} \mathbf{1}_{\{\Phi_n \in A\}}$, Seite 22
σ_A	$:= \begin{cases} \min\{n \geq 0 : \Phi_n \in A\} & \text{falls } \{n \geq 0 : \Phi_n \in A\} \neq \emptyset \\ \infty & \text{sonst} \end{cases}$, Seite 22

τ_A	$:= \begin{cases} \min\{n \geq 1 : \Phi_n \in A\} & \text{falls } \{n \geq 1 : \Phi_n \in A\} \neq \emptyset \\ \infty & \text{sonst} \end{cases}$, Seite 22
ψ	maximales Irreduzibilitätsmaß, Seite 23
$C_V(n)$	$:= \{y \in X : V(y) \leq n\}$, Seite 27
$\Delta V(x)$	$:= \int P(x, dy)V(y) - V(x)$, Seite 26
R	Zustandsraum, Seite 31
R_i	Nachbarschaft des Punktes i , Seite 31
$\mathbf{1}_{R_i}(j)$	Indikatorfunktion, Seite 32
G_{ij}	Generierungswahrscheinlichkeit des Metropolis-Algorithmus, Seite 32
$A_{ij}(c)$	Akzeptanzwahrscheinlichkeit des Metropolis-Algorithmus, Seite 32
Ψ_k	Ausgabe des Metropolis-Algorithmus im k -ten Schritt, Seite 33
Ψ	$:= \{\Psi_k : k \in \mathbb{N}\}$, Seite 33
$P_{ij}(c)$	Übergangswahrscheinlichkeit des Metropolis-Algorithmus, Seite 33
$\tilde{g}(i)$	$:= \exp\left(-\frac{f(i)}{c}\right)$, Seite 34
$g(i)$	$:= \frac{\tilde{g}(i)}{\sum_{j \in R} \tilde{g}(j)}$, Seite 34
$A_{\sigma, x}$	$:= \mathbb{E}_n + \sigma(x)\nabla^2 f(x)$, Seite 47
$m_{\xi, \sigma}(y)$	$:= \xi - A_{\sigma, \xi}^{-1}(\sigma(\xi)\nabla f(\xi) - \epsilon y)$, Seite 49
Φ_t^σ	Ausgabe des semiimpliziten Euleralgorithmus zum Zeitpunkt t mit Schrittweitenfunktion σ , Seite 46
Φ^σ	$:= \{\Phi_t^\sigma : t \in \mathbb{N}\}$, Seite 46
$\tilde{N}(y, \sigma(\xi))$	Dichte einer $N(0, \sigma(\xi))$ normal-verteilten Zufallsvariable, Seite 48
$k_\xi^\sigma(y)$	$:= \tilde{N}(m_{\xi, \sigma}^{-1}(y), \sigma(\xi)) \det \frac{1}{\epsilon} A_{\sigma, \xi} $, Seite 49