

# Stochastische Dynamische Optimierung

Lars Grüne  
Mathematisches Institut  
Fakultät für Mathematik und Physik  
Universität Bayreuth  
95440 Bayreuth  
[lars.gruene@uni-bayreuth.de](mailto:lars.gruene@uni-bayreuth.de)  
[www.math.uni-bayreuth.de/~lgruene/](http://www.math.uni-bayreuth.de/~lgruene/)

Vorlesungsskript  
Sommersemester 2007



# Vorwort

Dieses Skript ist im Rahmen einer gleichnamigen Vorlesung entstanden, die ich im Sommersemester 2007 an der Universität Bayreuth gehalten habe.

An dieser Stelle möchte ich mich ganz herzlich bei allen aufmerksamen StudentInnen, die mich auf Fehler und Ungenauigkeiten aufmerksam gemacht haben. Hinweisen möchte ich zudem noch darauf, dass die in Kapitel 4 eingeführten Transversalitätsbedingungen von den in der Literatur üblichen Bedingungen abweichen und mit dem Ziel eingeführt wurden, eine einheitliche Behandlung für verschiedene Typen Stochastischer Dynamischer Optimierungsprobleme zur ermöglichen. Es hat sich aber im weiteren Verlauf der Vorlesung herausgestellt, dass dies nicht in dem geplanten Umfang erreicht wurde (vgl. die Fußnote auf Seite 54) — diese Passagen werden daher in einer neuen Auflage dieses Skriptes überarbeitet werden.

Eine elektronische Version dieses Skripts sowie die zugehörigen Übungsaufgaben finden sich im WWW auf der Seite <http://www.math.uni-bayreuth.de/~lgruene/> unter dem Link “Lehrveranstaltungen”.

Bayreuth, Oktober 2007

LARS GRÜNE



# Inhaltsverzeichnis

<b>Vorwort</b>	<b>i</b>
<b>1 Grundlagen</b>	<b>1</b>
1.1 Grundlegende Konzepte aus der Stochastik . . . . .	1
1.2 Zeitdiskrete stochastische dynamische Systeme . . . . .	6
1.3 Filtrationen . . . . .	9
<b>2 Stochastische Kontrollsysteme</b>	<b>13</b>
2.1 Definition . . . . .	13
<b>3 Stochastische dynamische Optimierungsprobleme</b>	<b>19</b>
3.1 Definition . . . . .	19
3.2 Beispiele . . . . .	20
3.3 Optimale Wertefunktionen, Kontrollprozesse und Lösungen . . . . .	22
<b>4 Das Prinzip der Dynamischen Programmierung</b>	<b>23</b>
4.1 Probleme auf endlichem Horizont . . . . .	23
4.1.1 Formulierung und Beweis des Prinzips . . . . .	23
4.1.2 Charakterisierung Optimaler Kontrollprozesse . . . . .	27
4.2 Probleme auf unendlichem Horizont . . . . .	31
4.2.1 Formulierung und Beweis des Prinzips . . . . .	31
4.2.2 Charakterisierung Optimaler Kontrollprozesse . . . . .	34
<b>5 Numerische Dynamische Programmierung</b>	<b>37</b>
5.1 Ansatz . . . . .	37
5.2 Auswertung von $T(W)$ . . . . .	38
5.3 Berechnung von $\pi T(W)$ . . . . .	40
5.4 Wahl des Berechnungsgebietes . . . . .	42

5.5	Fehleranalyse . . . . .	43
5.6	Adaptive Gitter . . . . .	47
5.7	Unendlicher Zeithorizont . . . . .	48
<b>6</b>	<b>Linear Quadratische Probleme</b>	<b>51</b>
<b>7</b>	<b>Ein stochastisches Maximumprinzip</b>	<b>61</b>
7.1	Formulierung und Beweis . . . . .	61
7.2	Die Perturbationsmethode . . . . .	63
<b>A</b>	<b>Der Gittergenerator GRIDGEN</b>	<b>69</b>
A.1	Routinen für regelmäßige Gitter . . . . .	69
A.1.1	Erzeugen und Löschen von Gittern . . . . .	69
A.1.2	Zugriff auf Knoten und Gitterfunktionen . . . . .	70
A.1.3	Ein- und Ausgabe . . . . .	71
A.2	Routinen für unregelmässige Gitter . . . . .	72
A.2.1	Die Adaptioniteration . . . . .	72
A.2.2	Hängende Knoten . . . . .	75
A.2.3	Entfernen von Gitterpunkten aus dem Startgitter . . . . .	75
A.2.4	Anisotrope Gitter . . . . .	75
A.2.5	Beispielprogramm . . . . .	76
A.3	Grafische Darstellung . . . . .	76
A.4	Technische Hinweise . . . . .	76
A.4.1	Herunterladen der Routinen . . . . .	76
A.4.2	Einbinden der Header-Datei . . . . .	77
A.4.3	Compilieren und Linken . . . . .	77
	<b>Literaturverzeichnis</b>	<b>78</b>
	<b>Index</b>	<b>78</b>

# Kapitel 1

## Grundlagen

In dieser Vorlesung beschäftigen wir uns mit der optimalen Steuerung zeitdiskreter stochastischer Kontrollsysteme. In diesem Kapitel wiederholen wir zunächst einige Tatsachen aus der Stochastik, führen dann stochastische dynamische Systeme ein und betrachten schließlich einige weiterführende stochastische Konzepte für diese Systeme.

### 1.1 Grundlegende Konzepte aus der Stochastik

Wir erinnern zunächst an einige Grundbegriffe und Konzepte aus der Stochastik, wobei wir davon ausgehen, dass diese prinzipiell bekannt sind (diese Einführung ersetzt also keine Einführung in die Stochastik).

Ein *Wahrscheinlichkeitsraum* besteht aus einem Tripel  $(\Omega, \Sigma, P)$  mit:

- $\Omega$  ist die Menge der Elementarereignisse  $\omega \in \Omega$
- $\Sigma$  ist eine  $\sigma$ -Algebra auf  $\Omega$ , d.h. eine Menge von Teilmengen von  $\Omega$  mit den Eigenschaften
  - (1)  $\emptyset, \Omega \in \Sigma$
  - (2) wenn  $A \in \Sigma$  so ist auch  $A^c = \Omega \setminus A \in \Sigma$
  - (3) für  $A_1, A_2, \dots \in \Sigma$  ist auch  $\bigcup_{i=1}^{\infty} A_i \in \Sigma$  und  $\bigcap_{i=1}^{\infty} A_i \in \Sigma$
- $P : \Sigma \rightarrow [0, 1]$  ist ein Wahrscheinlichkeitsmaß auf  $\Sigma$ , d.h. es gilt  $P(\emptyset) = 0$ ,  $P(\Omega) = 1$  und

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \text{ für alle Mengen } A_i \in \Sigma \text{ mit } A_i \cap A_j = \emptyset \text{ für alle } i \neq j.$$

Für eine Menge  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  von Teilmengen von  $\Omega$  ( $\mathcal{P}$  bezeichnet die Potenzmenge) bezeichnen wir mit  $\sigma(\mathcal{A})$  die kleinste  $\sigma$ -Algebra, die  $\mathcal{A}$  enthält.

Um analytisch arbeiten zu können, werden wir  $\Omega$  und  $P$  wie üblich meist nicht direkt sondern durch (i.A. vektorwertige) Zufallsvariablen betrachten, also durch messbare Abbildungen

$$X : \Omega \rightarrow \mathbb{R}^n.$$

Hierbei heißt die Abbildung  $X$  *messbar*, wenn gilt

$$X^{-1}(B) := \{\omega \in \Omega \mid X(\omega) \in B\} \in \Sigma$$

für alle Mengen  $B$  aus der *Borel- $\sigma$ -Algebra*  $\mathcal{B}$ .

$$\mathcal{B} = \sigma(\{(a_1, b_1] \times \dots \times (a_n, b_n] \mid a_i, b_i \in \mathbb{R}, a_i < b_i\})$$

auf  $\mathbb{R}^n$ . Statt  $X^{-1}(B)$  schreiben wir alternativ auch " $X \in B$ ". Ebenso werden wir z.B. für eine zweidimensionale vektorwertige Zufallsvariable  $X = (X_1, X_2)^T$  Schreibweisen wie " $X_1 \leq X_2$ " verwenden, was kurz für  $X \in B = \{(x_1, x_2)^T \in \mathbb{R}^2 \mid x_1 \leq x_2\}$  steht.

Für messbare Zufallsvariablen definiert man ihre *Wahrscheinlichkeitsverteilung*

$$P_X : \mathcal{B} \rightarrow [0, 1], \quad P_X(B) := P(X^{-1}(B)).$$

**Beispiel 1.1** Es sei  $\Omega = (0, 1]$ ,  $\Sigma = \mathcal{B}|_{(0,1]}$ , und  $P$  die Gleichverteilung, also  $P((a, b]) = b - a$ . Betrachte die Zufallsvariable

$$X(\omega) = \begin{cases} -1, & \omega \in (0, 1/2] \\ 1, & \omega \in (1/2, 1] \end{cases}$$

Dann sind die möglichen Urbilder  $X^{-1}(B)$  gerade  $\emptyset$ ,  $(0, 1/2]$ ,  $(1/2, 1]$  und  $\Omega$ , die natürlich allesamt in  $\Sigma = \mathcal{B}|_{(0,1]}$  liegen.  $X$  ist also messbar. Für  $P_X$  ergibt sich

$$P_X(B) = \begin{cases} 0, & -1 \notin B, 1 \notin B \\ 1/2, & -1 \in B, 1 \notin B \\ 1/2, & -1 \notin B, 1 \in B \\ 1, & -1 \in B, 1 \in B \end{cases}.$$

Tatsächlich ist die gewählte Borel'sche  $\sigma$ -Algebra  $\Sigma = \mathcal{B}|_{(0,1]}$  viel größer als nötig, um Messbarkeit von  $X$  zu erhalten. Aus den oben angegebenen Urbilder sieht man leicht, dass die *minimale*  $\sigma$ -Algebra auf  $\Omega$ , bezüglich der  $X$  messbar ist, durch

$$\{\emptyset, (0, 1/2], (1/2, 1], \Omega\}$$

gegeben ist. □

Die in diesem Beispiel angegebene minimale  $\sigma$ -Algebra bezeichnen wir mit  $\sigma(X)$ . Formal ist diese gegeben durch

$$\sigma(X) := \sigma(\{X^{-1}(B) \mid B \in \mathcal{B}\}).$$

Messbarkeit von  $X$  bzgl. einer  $\sigma$ -Algebra  $\Sigma$  ist dann äquivalent zu der Inklusion

$$\sigma(X) \subseteq \Sigma.$$

Wenn wir ein  $\omega \in \Omega$  wählen und  $X(\omega)$  auswerten (was dem Durchführen eines Zufallsexperiments mit nachfolgender Auswertung von  $X$  entspricht), so nennen wir den Wert  $X(\omega)$  eine *Realisierung* von  $X$ .

Falls  $X$  nur endlich viele Werte  $x_1, \dots, x_m$  annehmen kann, ist ihr *Erwartungswert* als

$$E(X) = \sum_{i=1}^m x_i P_X(\{x_i\})$$

gegeben. Falls die Wertemenge von  $X$  kontinuierlich ist, werden wir stets annehmen, dass zu  $P_X$  eine *Dichtefunktion*  $f_X$  existiert, so dass

$$P_X(B) = \int_B f_X(x) dx \quad (1.1)$$

gilt. Der Erwartungswert von  $X$  ist dann als

$$E(X) = \int_{\mathbb{R}^n} x f_X(x) dx$$

definiert. Die *Varianz* ist für eine reellwertige Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  gegeben durch

$$\text{Var}(X) = E((X - E(X))^2) = E(X^2) - 2E(XE(X)) + E(X)^2 = E(X^2) - E(X)^2.$$

Für eine vektorwertige Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}^n$  definiert man die *Kovarianzmatrix*

$$\text{Cov}(X) = E\left((X - E(X))(X - E(X))^T\right) \in \mathbb{R}^{n \times n}.$$

Beachte, dass die Diagonale dieser Matrix gerade die Varianzen  $\text{Var}(X_j)$  der Komponenten von  $X = (X_1, \dots, X_n)^T$  enthält.

Wenn wir eine Zufallsvariable  $X$  betrachten und zusätzlich eine Menge  $A \subset \Omega$  kennen, in denen  $\omega$  liegt, so können wir diese Information verwenden, wenn wir die Wahrscheinlichkeit von  $X \in B$  berechnen wollen. Formal wird dies über die *bedingte Wahrscheinlichkeit*

$$P_X(B|A) = \frac{P(X^{-1}(B) \cap A)}{P(A)}$$

gemessen. In dieser Definition muss  $P(A) \neq 0$  vorausgesetzt werden, allerdings gibt es allgemeinere Definitionen, die auch für  $P(A) = 0$  sinnvoll sind, allerdings weitergehende maßtheoretische Techniken benötigen, die wir hier nicht einführen wollen (vgl. S. Cyganowski, P.E. Kloeden, J. Ombach, *From Elementary Probability to Stochastic Differential Equations with MAPLE*, Springer 2002, Definition 7.1.1). Mittels der charakteristischen Funktion

$$\chi_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

ist der *bedingte Erwartungswert* einer Zufallsvariablen durch

$$E(X|A) = \frac{E(X \cdot \chi_A)}{P(A)}$$

definiert. Auch diese Definition lässt sich mit anspruchsvolleren maßtheoretischen Überlegungen auf den Fall  $P(A) = 0$  erweitern (vgl. S. Cyganowski, P.E. Kloeden, J. Ombach, *From Elementary Probability to Stochastic Differential Equations with MAPLE*, Springer 2002, Definition 7.1.2 oder T. Mikosch, *Elementary Stochastic Calculus with Finance in View*, World Scientific, 1998, Abschnitt 1.4.3).

Ein wichtiger Spezialfall dieser bedingten Größen tritt auf, wenn wir zwei Zufallsvariablen  $X_1$  und  $X_2$  betrachten und die Menge  $A$  als  $X_2^{-1}(B_2)$  gegeben ist, wenn wir also die Größen

$$P_{X_1}(B_1 | X_2 \in B_2) \quad \text{bzw.} \quad E(X_1 | X_2 \in B_2)$$

betrachten.

**Beispiel 1.2** Es sei  $(\Omega, \Sigma, P)$  der Wahrscheinlichkeitsraum, der das (gleichzeitige) Würfeln mit zwei Würfeln beschreibt. Jedes Elementarereignis ist also ein Paar  $\omega = (\omega_1, \omega_2) \in \Omega = \{1, \dots, 6\}^2$  von Würfelwerten, wobei die erste Komponente  $\omega_1$  das Ergebnis des ersten Würfels und  $\omega_2$  das Ergebnis des zweiten Würfels beschreibt (wir nehmen an, dass die zwei Würfel unterscheidbar sind). Als  $\Sigma$  wählen wir alle möglichen Teilmengen von  $\Omega$ . Offenbar ist (für ideale Würfel)  $P(A) = |A|/36$  für jede Teilmenge  $A \subseteq \Omega$  mit  $|A|$  Elementen, da  $\Omega$  gerade  $36 = 6 \times 6$  Ereignisse enthält, die alle gleich wahrscheinlich sind. Wir betrachten nun drei Zufallsvariablen, die als Wert das Ergebnis des ersten Würfels, des zweiten Würfels und die Summe der Würfelwerte ausgeben, also  $X_1(\omega) = \omega_1$ ,  $X_2(\omega) = \omega_2$  und  $X_3(\omega) = \omega_1 + \omega_2$ . Wir betrachten die folgenden Ereignismengen für die Zufallsvariablen  $X_1$ ,  $X_2$  bzw.  $X_3$ :

$$\begin{aligned} X_1 \in B_1 &= \{1\} && \text{(erster Würfel würfelt 1)} \\ X_2 \in B_2 &= \{6\} && \text{(zweiter Würfel würfelt 6)} \\ X_3 \in B_3 &= \{10, 11, 12\} && \text{(Würfelsumme ist } \geq 10) \end{aligned}$$

Hierfür ergibt sich

$$\begin{aligned} X_1^{-1}(B_1) &= \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)\} \\ X_2^{-1}(B_2) &= \{(1, 6), (2, 6), (3, 6), (4, 6), (5, 6), (6, 6)\} \\ X_3^{-1}(B_3) &= \{(4, 6), (5, 5), (5, 6), (6, 4), (6, 5), (6, 6)\} \end{aligned}$$

und wir erhalten die unbedingten Wahrscheinlichkeiten

$$\begin{aligned} P_{X_1}(B_1) &= 1/6 \\ P_{X_2}(B_2) &= 1/6 \\ P_{X_3}(B_3) &= 1/6 \end{aligned}$$

und die bedingten Wahrscheinlichkeiten

$$\begin{aligned} P_{X_2}(B_2 | X_1 \in B_1) &= 1/6 \\ P_{X_3}(B_3 | X_1 \in B_1) &= 0 \\ P_{X_3}(B_3 | X_2 \in B_2) &= 1/2 \end{aligned}$$

□

Der bedingte Erwartungswert  $E(X_1 | X_2 \in B_2)$  liefert eine reelle Zahl. Man kann dieses Konzept so erweitern, dass man eine Zufallsvariable erhält, nämlich indem man die Abbildung

$$E(X_1 | X_2) : \omega \mapsto E(X_1 | X_2 = X_2(\omega))$$

definiert. Dies ist eine Zufallsvariable, die messbar bzgl.  $\sigma(X_2)$  ist, und für deren Erwartungswert gerade

$$E(E(X_1 | X_2)) = E(X_1). \quad (1.2)$$

Gleichung (1.2) lässt sich für den Fall, dass  $X_2$  nur abzählbar viele Werte annimmt, leicht beweisen: In diesem Fall existieren abzählbar viele Mengen der Form  $X_2^{-1}(X_2(\omega))$ , die wir mit  $A_1, A_2, \dots$  bezeichnen. Dann gilt

$$E(E(X_1 | X_2)) = \sum_{i=1}^{\infty} E(X_1 | A_i) P_{X_1}(A_i) = \sum_{i=1}^{\infty} E(X_1 \chi_{A_i}) = E\left(\sum_{i=1}^{\infty} X_1 \chi_{A_i}\right) = E(X_1)$$

Im überabzählbaren Fall gilt (1.2) auch, benötigt aber wiederum tiefere maßtheoretische Techniken zum Beweis (vgl. S. Cyganowski, P.E. Kloeden, J. Ombach, *From Elementary Probability to Stochastic Differential Equations with MAPLE*, Springer 2002, Theorem 7.1.1(3), oder T. Mikosch, *Elementary Stochastic Calculus with Finance in View*, World Scientific, 1998, Section 1.4.4, Rule 2). Statt  $E(X_1 | X_2)$  schreibt man auch

$$E(X_1 | \sigma(X_2)). \quad (1.3)$$

Ein wichtiges Konzept ist die stochastische Unabhängigkeit zweier (oder mehrerer) Zufallsvariablen. Anschaulich bedeutet dies, dass uns die Kenntnis der Werte einer Zufallsvariablen  $X_1$  keinerlei Informationen über die Werte der anderen Zufallsvariablen  $X_2$  liefert. Formal:

**Definition 1.3** Die Zufallsvariablen  $X_1, X_2, \dots, X_k$  heißen *unabhängig*, falls

$$P(X_1^{-1}(B_1) \cap \dots \cap X_k^{-1}(B_k)) = P(X_1^{-1}(B_1)) \cdot \dots \cdot P(X_k^{-1}(B_k)) \quad (1.4)$$

für alle  $B_i \in \mathcal{B}$ ,  $i = 1, \dots, k$ , gilt. □

Für unabhängige Zufallsvariablen  $X_1, X_2, \dots, X_k$  gilt

$$E(X_1 \cdot \dots \cdot X_k) = E(X_1) \cdot \dots \cdot E(X_k).$$

Für eine Abbildung  $h : (\mathbb{R}^n)^k \rightarrow \mathbb{R}^l$  und  $k$  Zufallsvariablen  $X_1, \dots, X_k$  ist  $Y = h(X_1, \dots, X_k)$  wieder eine Zufallsvariable. Wenn dann  $X_1, X_2, \dots, X_k, Z$  unabhängig sind, so sind auch  $Y$  und  $Z$  unabhängig.

**Beispiel 1.4** Betrachte die in Beispiel 1.2 definierten Zufallsvariablen. Für beliebige Mengen  $B_1, B_2 \subseteq \{1, \dots, 6\}$  gilt hier

$$X_1^{-1}(B_1) = \{(\omega_1, \omega_2) \mid \omega_1 \in B_1, \omega_2 \in \{1, \dots, 6\}\},$$

$$X_2^{-1}(B_2) = \{(\omega_1, \omega_2) \mid \omega_2 \in B_2, \omega_1 \in \{1, \dots, 6\}\}$$

und

$$X_1^{-1}(B_1) \cap X_2^{-1}(B_2) = \{(\omega_1, \omega_2) \mid \omega_1 \in B_1, \omega_2 \in B_2\} = B_1 \times B_2.$$

Also gilt

$$|X_1^{-1}(B_1)| = 6|B_1|, \quad |X_2^{-1}(B_2)| = 6|B_2|, \quad |X_1^{-1}(B_1) \cap X_2^{-1}(B_2)| = |B_1| |B_2|$$

und damit

$$\begin{aligned} P(X_1^{-1}(B_1) \cap X_2^{-1}(B_2)) &= |B_1| |B_2| / 36 \\ &= (6|B_1|/36)(6|B_2|/36) = P(X_1^{-1}(B_1))P(X_2^{-1}(B_2)). \end{aligned}$$

Also sind  $X_1$  und  $X_2$  unabhängig. Andererseits gilt für die in Beispiel 1.2 definierten Mengen  $B_1$  und  $B_3$

$$P(X_1^{-1}(B_1) \cap X_3^{-1}(B_3)) = P(\emptyset) = 0$$

aber

$$P(X_1^{-1}(B_1))P(X_3^{-1}(B_3)) = P_{X_1}(B_1)P_{X_3}(B_3) = (1/6)(1/9) = 1/54 \neq 0,$$

weswegen  $X_1$  und  $X_3$  nicht unabhängig sind.  $\square$

## 1.2 Zeitdiskrete stochastische dynamische Systeme

Wir betrachten nun das Basismodell eines — zunächst noch unkontrollierten — stochastischen dynamischen Systems, die sogenannte stochastische Differenzgleichung.

**Definition 1.5** Ein zeitdiskretes stochastisches System ist gegeben durch eine Abbildung

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

Hierbei ist  $x \in \mathbb{R}^n$  der Zustand und  $z \in \mathbb{R}^m$  der stochastische Einfluss. Für einen Anfangswert  $x_0 \in \mathbb{R}^n$  und eine Folge von Zufallsvariablen  $Z_0, Z_1, \dots$  definieren wir die *Lösung*  $X_t = X_t(x_0)$  des Systems für  $t \geq 0$  induktiv mittels

$$X_0 = x_0, \quad X_{t+1} = f(X_t, Z_t). \tag{1.5}$$

$\square$

**Bemerkung 1.6** (i) Wir verwenden im Folgenden oft die Kurzschreibweise  $X_t$  und führen das Argument  $(x_0)$  nur dann explizit auf, wenn ansonsten Uneindeutigkeiten entstünden.

(ii) Die Lösung  $X_t$  ist zu jeder Zeit  $t > t_0$  eine Zufallsvariable, da sie von den Zufallsvariablen  $Z_{t_0}, \dots, Z_{t-1}$  abhängt. Für jeden Anfangswert  $x_0$  ist  $X_t(x_0)$  ein sogenannter *stochastischer Prozess*, eine zeitabhängige Zufallsvariable.

(iii) Wenn wir Realisierungen  $Z_t(\omega)$  der Zufallsvariablen  $Z_t$  betrachten, erhalten wir eine Realisierung  $X_t(\omega) = X_t(x_0, \omega)$ . Diese Realisierung ist nun eine “normale” Funktion  $t \mapsto X_t(x_0, \omega)$  von  $\mathbb{N}_0$  nach  $\mathbb{R}^n$  und wird *Pfad* von  $X_t$  genannt.

(iv) Falls  $Z_0, \dots, Z_{t-1}$  und  $Z_t$  unabhängige Zufallsvariablen sind, sind auch  $X_t$  und  $Z_t$  unabhängig, da  $X_t$  nur von  $Z_0, \dots, Z_{t-1}$  abhängt.  $\square$

**Beispiel 1.7 (Binomialmodell eines Aktienkurses)** Das einfachste Modell des Verlaufes eines Aktienkurses ist das sogenannte *Binomialmodell*. Ausgehend von dem Anfangskurs  $x_0$  kann der Kurs dabei in jedem Zeitschritt entweder mit Wahrscheinlichkeit  $p$  um einen Faktor  $\alpha_u > 1$  steigen (“up”) oder mit Wahrscheinlichkeit  $1 - p$  um einen Faktor  $\alpha_d < 1$  fallen (“down”). Hierbei soll die Bewegung im Zeitschritt  $t + 1$  jeweils stochastisch unabhängig von der Bewegung im vorhergehenden Zeitschritt  $t$  sein.

Zur Modellierung als zeitdiskretes stochastisches System betrachten wir daher unabhängige Zufallsvariablen  $Z_t : \Omega \rightarrow \{\alpha_u, \alpha_d\}$  mit

$$P_{Z_t}(\{\alpha_u\}) = p \quad \text{und} \quad P_{Z_t}(\{\alpha_d\}) = 1 - p$$

und definieren

$$f(x, z) = x \cdot z.$$

Die Lösung ergibt sich damit für  $t \geq 1$  zu

$$X_t = Z_{t-1}Z_{t-2} \dots Z_0x_0.$$

Abbildung 1.1 zeigt einige Beispielpfade dieses Modells für  $\alpha_u = 1.03$ ,  $\alpha_d = 0.98$ ,  $p = 1/2$  und  $x_0 = 1$ .

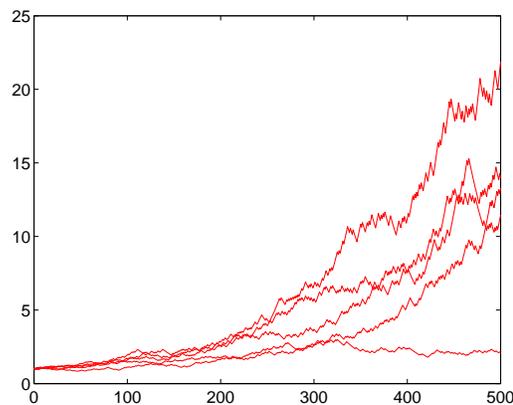


Abbildung 1.1: Beispielpfade des Binomialmodells

□

Die Zufallsvariablen  $Z_t$  in diesem Modell haben zwei Eigenschaften, die wir in Zukunft stets voraussetzen werden: sie sind unabhängig voneinander und sie besitzen alle identische Verteilung. Formal ist dies in der folgenden Definition festgehalten.

**Definition 1.8** Eine Folge von Zufallsvariablen  $Z_0, Z_1, Z_2, \dots$  heißt unabhängig und identisch verteilt (kurz *iid* = independent and identically distributed), wenn die folgenden zwei Bedingungen erfüllt sind:

- (i) die Zufallsvariablen sind alle *auf dem gleichen Wahrscheinlichkeitsraum*  $(\Omega, \Sigma, P)$  *definiert* und *unabhängig*, d.h., für alle  $t \in \mathbb{N}$  erfüllen die Zufallsvariablen  $Z_0, \dots, Z_t$  Definition 1.3 mit  $k = t$
- (ii) die Zufallsvariablen  $Z_i$  sind *identisch verteilt*, d.h. es gilt  $P_{Z_i} = P_{Z_j}$  für alle  $i, j = 0, 1, 2, \dots$

□

**Beispiel 1.9** In diesem Beispiel konstruieren wir explizit iid Zufallsvariablen  $Z_t$ ,  $t = 0, 1, 2, \dots$ . Eigentlich ist dies für unsere Modellierung nicht nötig: hierzu reicht es, wie in Beispiel 1.7 die gewünschten stochastischen Eigenschaften der verwendeten Zufallsvariablen  $Z_t$  zu definieren, ohne sich darum zu kümmern, wie die  $Z_t$  als Abbildungen auf einem geeigneten Wahrscheinlichkeitsraum tatsächlich konstruiert werden können. Zur Illustration von Definition 1.8(i) und insbesondere als Beispiel für die Überlegungen im nachfolgenden Abschnitt wollen wir in diesem Beispiel aber einmal eine mögliche Wahl der  $Z_t$  aus Beispiel 1.7 explizit konstruieren.

Wir betrachten dazu den gleichen Wahrscheinlichkeitsraum wie in Beispiel 1.1, also  $\Omega = (0, 1]$ ,  $\Sigma = \mathcal{B}|_{(0,1]}$  und  $P([a, b]) = b - a$ . Zur Vereinfachung der Konstruktion wählen wir den Parameter  $p$  aus Beispiel 1.7 als  $p = 1/2$ . Nun definieren wir die Zufallsvariablen  $Z_t$  für  $t = 0, 1, 2, \dots$  mittels

$$Z_t(\omega) = \begin{cases} \alpha_u, & \omega \in (q2^{-t}, (q+1/2)2^{-t}], \quad q = 0, \dots, 2^t - 1 \\ \alpha_d, & \text{sonst} \end{cases}$$

Abbildung 1.2 zeigt  $Z_0, Z_1, Z_2$ , und  $Z_3$  für  $\alpha_u = 1.1$  und  $\alpha_d = 0.9$ .

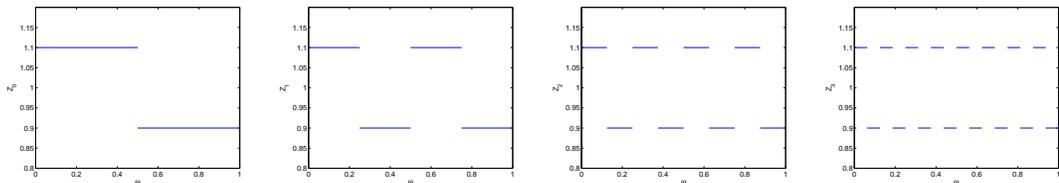


Abbildung 1.2: Zufallsvariablen  $Z_0, Z_1, Z_2$ , und  $Z_3$

Jede der Zufallsvariablen nimmt also genau dann den Wert  $\alpha_u$  an, wenn

$$\omega \in I_t^q := (q2^{-t}, (q+1/2)2^{-t}]$$

gilt und den Wert  $\alpha_d$ , wenn

$$\omega \in J_t^q := ((q+1/2)2^{-t}, (q+1)2^{-t}]$$

erfüllt ist. Jedes der Intervalle hat gerade die Länge  $2^{-t-1}$ , es gilt also  $P(I_t^q) = P(J_t^q) = 2^{-t-1}$ .

Wir zeigen nun zunächst, dass die Zufallsvariablen die gewünschte Verteilung besitzen. Wir definieren die Mengen

$$A_t^u := Z_t^{-1}(\{\alpha_u\}) = \bigcup_{q=0, \dots, 2^t-1} I_t^q$$

und

$$A_t^d := Z_t^{-1}(\{\alpha_d\}) = \bigcup_{q=0, \dots, 2^t-1} J_t^q.$$

Damit folgt

$$P_{Z_t}(\{\alpha_u\}) = P(A_t^u) = \sum_{q=0}^{2^t-1} P(I_t^q) = 2^t 2^{-t-1} = 2^{-1} = 1/2$$

und analog

$$P_{Z_t}(\{\alpha_d\}) = 1/2,$$

also die gewünschte Verteilung.

Zum Beweis der Unabhängigkeit müssen wir Gleichung (1.4) (mit  $k = t$ ) für alle Mengen  $B_t$  überprüfen. Hierzu reicht es, jeweils die Mengen  $B_t = \{\alpha_u\}$  und  $B_t = \{\alpha_d\}$  zu betrachten, da das Urbild  $Z_t^{-1}(B_t)$  jeder anderen Menge  $B_t \in \mathcal{B}$  entweder mit diesem übereinstimmt, gleich  $\emptyset$  ist (womit (1.4) trivialerweise gilt) oder gleich  $\Omega$  ist (womit die entsprechenden Terme in (1.4) wegfallen). Da nach Konstruktion  $P(Z_t^{-1}(\{\alpha_u\})) = P(Z_t^{-1}(\{\alpha_d\})) = 1/2$  ist, ist die rechte Seite von (1.4) stets gleich  $1/2^{t+1}$ . Um zu beweisen, dass die linke Seite den gleichen Wert ergibt, beweisen wir per Induktion über  $t$ , dass eine Menge der Form

$$A_1^{b_1} \cap A_2^{b_2} \cap \dots \cap A_t^{b_t}$$

mit  $b_i \in \{u, d\}$  stets aus genau einem Intervall  $I_t^{q_t}$  oder  $J_t^{q_t}$  besteht. Für  $t = 1$  folgt dies sofort aus der Form der Mengen  $A_1^u$  und  $A_1^d$ . Für  $t \rightarrow t + 1$  folgt die Behauptung, da die Menge  $A_1^{b_1} \cap \dots \cap A_t^{b_t}$  nach Induktionsvoraussetzung aus einem Intervall  $I_t^{q_t}$  oder  $J_t^{q_t}$  besteht, in dem jeweils genau eines der Intervalle  $I_{t+1}^q$  oder  $J_{t+1}^q$  aus  $A_{t+1}^{b_{t+1}}$  liegt, welches in der Schnittmenge gerade übrig bleibt. Damit folgt nun entweder

$$P(A_1^{b_1} \cap A_2^{b_2} \cap \dots \cap A_t^{b_t}) = P(I_t^{q_t}) = 2^{-t-1} = 1/2^{t+1}$$

oder

$$P(A_1^{b_1} \cap A_2^{b_2} \cap \dots \cap A_t^{b_t}) = P(J_t^{q_t}) = 2^{-t-1} = 1/2^{t+1}$$

und damit in beiden Fällen die gleiche Wahrscheinlichkeit wie auf der rechten Seite.  $\square$

## 1.3 Filtrationen

Im nächsten Kapitel werden wir die stochastische Differenzgleichung (1.5) um eine Kontrollfunktion  $\mathbf{u} = (u_t, t \in \mathbb{N}_0)$  erweitern. Diese Kontrollfunktion ist dabei eine Funktion, die wir beeinflussen können, insbesondere um das Verhalten der Lösungen  $X_t$  bezüglich eines Kriteriums zu optimieren.

Hierbei stellt sich eine fundamentale Frage: Welche Information über den stochastischen Einfluss wollen wir verwenden, wenn wir  $\mathbf{u}$  wählen und wie kann man das formalisieren? Dies ist letztendlich eine Modellierungsfrage, bei der wir die folgenden Aspekte berücksichtigen wollen:

- (i) Wenn wir die Kontrollfunktion  $\mathbf{u}$  zum Zeitpunkt  $t = 0$  wählen, wollen wir uns die Möglichkeit offen halten, zu einem späteren Zeitpunkt  $t > 0$  auf die (dann bekannte, jetzt aber noch unsichere) zukünftige Situation zu reagieren: Wenn der Aktienkurs im Intervall  $[0, t]$  gefallen ist, werde ich als Anleger zum Zeitpunkt  $t$  anders reagieren als wenn er gestiegen ist. Die Kontrollfunktion sollte also von der (zum Zeitpunkt  $t = 0$ ) ungewissen Zukunft abhängen und deswegen ebenfalls ein stochastischer Prozess sein, der sogenannte *Kontrollprozess*.
- (ii) Wir haben unterschiedliche Informationen über die Vergangenheit und die Zukunft von  $X_t$ : Wir nehmen an, dass wir zu einem Zeitpunkt  $t$  die realisierten Werte  $X_0(\omega), \dots, X_t(\omega)$  kennen (also z.B. den vergangenen Verlauf des Aktienkurses), für die zukünftigen Werte  $X_{t+1}, X_{t+2}, \dots$  aber lediglich die aus dem Modell (1.5) abgeleiteten stochastischen Informationen verwenden können.
- (iii) Da die Zufallsvariablen  $Z_t$  sind nur ein Hilfskonstrukt in unserer Modellierung. Sie entsprechen i.A. keinen realen "Daten" und sind folglich unbekannt. Deshalb wollen wir die Werte  $Z_s(\omega)$  (egal zu welchem Zeitpunkt) bei der Wahl der Kontrollfunktion nicht verwenden.

Zusammengefasst folgt aus diesen Modellierungsannahmen, dass die Kontrollfunktion  $u_t$  selbst wieder ein stochastischer Prozess ist, in dessen genauer Definition wir nun aber "einbauen" müssen, dass wir zum Zeitpunkt  $t$  nur die Informationen aus  $X_0, \dots, X_t$ , bei der Wahl der zukünftigen Kontrollwerte verwenden dürfen. Wir werden im nächsten Kapitel zwei äquivalente Definitionen solcher Kontrollprozesse geben — eine stochastische in Definition 2.3 und eine kontrolltheoretische in Korollar 2.5.

Um diese zu formulieren benötigen wir ein technisches Hilfsmittel aus der Theorie der stochastischen Prozesse, die sogenannten *Filtrationen*. Die Grundidee der Filtrationen ist die Beobachtung, die wir bereits in Beispiel 1.1 gemacht haben: Dort haben wir auf  $\Omega = (0, 1]$  die Borel'sche  $\sigma$ -Algebra  $\mathcal{B}|_{(0,1]}$  verwendet. Um Messbarkeit der dort definierten Zufallsvariablen

$$X(\omega) = \begin{cases} -1, & \omega \in (0, 1/2] \\ 1, & \omega \in (1/2, 1] \end{cases}$$

zu erhalten, wäre aber die viel kleinere  $\sigma$ -Algebra

$$\sigma(X) = \{\emptyset, (0, 1/2], (1/2, 1], \Omega\}$$

ausreichend gewesen.

Diese Beobachtung ist keineswegs rein technischer Natur, denn tatsächlich gibt uns  $\sigma(X)$  Auskunft über den Informationsgehalt der Zufallsvariablen  $X$ : Wenn wir  $X(\omega)$  messen können, können wir damit lediglich entscheiden, in welcher der in  $\sigma(X)$  enthaltenen nicht-trivialen Mengen  $(0, 1/2]$  und  $(1/2, 1]$  das Ereignis  $\omega$  enthalten ist, eine genauere Bestimmung von  $\omega$  ist prinzipiell nicht möglich.

Diese Beobachtung können wir nun auf stochastische Prozesse erweitern, indem wir die zeitliche Abfolge der Zufallsvariablen berücksichtigen. Wir illustrieren diese Idee an Hand der Lösung  $X_t$  aus Beispiel 1.7:

**Beispiel 1.10** Wir betrachten das System aus Beispiel 1.7 mit den Zufallsvariablen  $Z_t$  aus Beispiel 1.9. Wir wählen den Anfangswert  $x_0 = 1$  und nehmen an, es sei ein  $\omega$  gegeben, welches den Verlauf der Lösung  $X_t$  bestimmt, das uns aber unbekannt ist. Bekannt sind uns zu jedem Zeitpunkt  $t$  lediglich die Realisierungen  $X_0(\omega), \dots, X_t(\omega)$ . Dies entspricht gerade dem (realistischen!) Szenario, dass wir zum Zeitpunkt  $t$  die Aktienkurse für  $s \leq t$  kennen, die Kurse für  $s > t$  aber natürlich nicht. Wir wollen nun untersuchen, wie genau wir  $\omega$  zum Zeitpunkt  $t$  bestimmen können.

$t = 0$ : Zu diesem Zeitpunkt kennen wir nur den Anfangswert  $x_0$ . Das Einzige, was wir über  $\omega$  aussagen können ist daher die "triviale" Aussage  $\omega \in \Omega$ .

$t = 1$ : Nun kennen wir  $x_0$  und  $X_1(\omega)$  und damit auch  $Z_0(\omega)$ . Da  $Z_0$  bis auf Skalierung der Werte gerade mit  $X$  aus Beispiel 1.1 übereinstimmt, können wir nun also entscheiden, in welcher der beiden Mengen

$$(0, 1/2], (1/2, 1] \quad (1.6)$$

$\omega$  liegt.

$t = 2$ : Nun kennen wir zusätzlich  $X_2(\omega)$ , woraus wir  $Z_1(\omega)$  berechnen können. Damit können wir  $\omega$  nun eindeutig einem der Intervalle

$$(0, 1/4], (1/4, 1/2], (1/2, 3/4], (3/4, 1] \quad (1.7)$$

zuordnen.

$t = 3$ : Kennen wir außerdem  $X_3(\omega)$ , so können wir  $Z_2(\omega)$  ausrechnen und damit entscheiden, in welchem der Intervalle

$$(0, 1/8], (1/8, 1/4], (1/4, 3/8], (3/8, 1/2], (1/2, 5/8], (5/8, 3/4], (3/4, 7/8], (7/8, 1] \quad (1.8)$$

$\omega$  liegt.

Dies setzt sich für wachsende  $t$  fort: Je mehr die Zeit voranschreitet, desto mehr Informationen  $X_t(\omega)$  erhalten wir und desto genauer können wir  $\omega$  bestimmen.

Mathematisch formal kann man diese Eigenschaft durch die  $\sigma$ -Algebren  $\sigma(X_0, \dots, X_t)$  machen, die ja gerade die minimalen  $\sigma$ -Algebren sind, für die die  $tn$ -dimensionalen Zufallsvariablen  $(X_0, \dots, X_t)$  messbar sind. Für  $t = 0$  ist dies gerade die triviale  $\sigma$ -Algebra

$$\mathcal{F}_0 = \{\emptyset, \Omega\},$$

für  $t = 1, 2, 3$  erhält man gerade diejenigen  $\sigma$ -Algebren  $\mathcal{F}_t$ , die durch die Mengen in (1.6)–(1.8) erzeugt werden, also

$$\mathcal{F}_1 = \sigma((0, 1/2], (1/2, 1]),$$

$$\mathcal{F}_2 = \sigma((0, 1/4], (1/4, 1/2], (1/2, 3/4], (3/4, 1]),$$

$$\mathcal{F}_3 = \sigma((0, 1/8], (1/8, 1/4], (1/4, 3/8], (3/8, 1/2], (1/2, 5/8], (5/8, 3/4], (3/4, 7/8], (7/8, 1]).$$

□

Je mehr Informationen wir haben, desto feiner wird die Intervallunterteilung in  $\mathcal{F}_t$  und desto größer wird  $\mathcal{F}_t$ . Dies ist genau die Idee der *Filtration* aus der folgenden Definition.

**Definition 1.11** Gegeben sei ein Wahrscheinlichkeitsraum  $(\Omega, \Sigma, P)$ .

(i) Eine *Filtration*  $\mathcal{F} = (\mathcal{F}_t, t = 0, 1, 2, \dots)$  ist eine Folge von  $\sigma$ -Algebren  $\mathcal{F}_t \subseteq \Sigma$ ,  $t = 0, 1, \dots$ , für die die Monotonieeigenschaft

$$\mathcal{F}_s \subseteq \mathcal{F}_t \quad \text{für alle } 0 \leq s \leq t$$

gilt.

(ii) Ein stochastischer Prozess  $X_t$  heißt *adaptiert* zur Filtration  $\mathcal{F}$ , falls  $(X_0, \dots, X_t)$  messbar bzgl.  $\mathcal{F}_t$  ist, also

$$\sigma((X_0, \dots, X_t)) \subseteq \mathcal{F}_t, \quad t = 0, 1, \dots$$

gilt. Die Menge aller stochastischen Prozesse, die adaptiert zu einer gegebenen Filtration  $\mathcal{F}_t$  sind, bezeichnen wir mit  $\mathcal{X}(\mathcal{F})$ .

(iii) Die *natürliche* (oder *induzierte*) *Filtration* eines stochastischen Prozesses  $X_t$  ist gegeben durch

$$\mathcal{F}_t = \sigma((X_0, \dots, X_t)), \quad t = 0, 1, \dots$$

□

**Bemerkung 1.12** Bedingung (ii) ist äquivalent zu der — formal schwächeren — Bedingung

$$\sigma(X_t) \subseteq \mathcal{F}_t, \quad t = 0, 1, \dots,$$

denn daraus folgt wegen der Monotonie der Mengen  $\mathcal{F}_t$  für jedes  $t$  sofort

$$\sigma(X_s) \subseteq \mathcal{F}_t, \quad s = 0, 1, \dots, t-1$$

und damit auch

$$\sigma((X_0, \dots, X_t)) \subseteq \mathcal{F}_t.$$

□

**Bemerkung 1.13** Für Lösungen  $X_t = X_t(x_0)$  der stochastischen Differenzgleichung hängt die natürliche Filtration vom Anfangswert ab:

In Beispiel 1.10 haben wir gerade die ersten Terme der natürlichen Filtration für  $X_t$  für das Modell aus Beispiel 1.7 zum Anfangswert  $x_0 = 1$  konstruiert. Tatsächlich überlegt man sich leicht, dass man hier für alle  $x_0 \neq 0$  die gleiche Filtration erhält.

Für den Anfangswert  $x_0 = 0$  allerdings gilt  $X_t(x_0) = 0$  für alle  $t \geq 0$  unabhängig von den Zufallsvariablen  $Z_t$ , weswegen man als natürliche Filtration für alle  $t \geq 0$  die trivialen  $\sigma$ -Algebren  $\mathcal{F}_t = \{\emptyset, \Omega\}$  erhält. □

Für natürliche Filtrationen und  $s \leq t$  kann man den verallgemeinerten Erwartungswert

$$E_s(X_t) := E(X_t | \mathcal{F}_s) = E(X_t | \sigma(X_0, \dots, X_s)) \quad (1.9)$$

(vgl. (1.3)) definieren. Aus der Definition und  $X_0(\omega) = x_0$  für alle  $\omega \in \Omega$  folgt dann

$$E_0(X_t)(\omega) = E(X_t | \mathcal{F}_0)(\omega) = E(X_t | X_0)(\omega) = E(X_t | X_0 = X_0(\omega)) = E(X_t | \Omega) = E(X_t),$$

d.h.,  $E_0(X_t)$  ist der übliche Erwartungswert. Zudem folgt aus (1.2)

$$E(E_s(X_t)) = E(E(X_t | \mathcal{F}_s)) = E(E(X_t | (X_0, \dots, X_s))) = E(X_t). \quad (1.10)$$

# Kapitel 2

## Stochastische Kontrollsysteme

### 2.1 Definition

Wir kommen nun zur Definition unseres stochastischen zeitdiskreten Kontrollsystems.

**Definition 2.1** Ein zeitdiskretes stochastisches Kontrollsystem ist gegeben durch eine Abbildung

$$f : \mathbb{R}^n \times U \times \mathbb{R}^m \rightarrow \mathbb{R}^n, \quad (x, u, z) \mapsto f(x, u, z).$$

Hierbei ist  $x \in \mathbb{R}^n$  der Zustand,  $u \in U \subseteq \mathbb{R}^l$  der Kontrollwert und  $z \in \mathbb{R}^m$  der stochastische Einfluss. Für einen Anfangswert  $x_0 \in \mathbb{R}^n$ , einen bzgl.  $\Sigma$  messbaren Kontrollprozess  $\mathbf{u} = (u_t, t \in \mathbb{N}_0)$  und eine Folge von Zufallsvariablen  $Z_0, Z_1, \dots$  definieren wir die *Lösung*  $X_t = X_t(x_0, \mathbf{u})$  des Systems für  $t \geq 0$  induktiv mittels

$$X_0 = x_0, \quad X_{t+1} = f(X_t, u_t, Z_t). \quad (2.1)$$

□

**Bemerkung 2.2** (i) Wir verwenden im Folgenden wiederum oft die Kurzschreibweise  $X_t$  und führen das Argument  $(x_0, \mathbf{u})$  nur dann explizit auf, um eventuelle Uneindeutigkeiten zu vermeiden.

(ii) Der Kontrollprozess  $\mathbf{u}$  ist selbst wieder ein stochastischer Prozess, der messbar bzgl.  $\Sigma$  ist. □

Diese Wahl von  $\mathbf{u}$  stellt sicher, dass  $X_t$  wieder ein bzgl.  $\Sigma$  messbarer stochastischer Prozess wird, die so entstehende Klasse von Kontrollprozessen ist aber für die allermeisten Anwendungen viel zu groß. Messbarkeit bezüglich  $\Sigma$  bedeutet nämlich, dass wir die gesamte Zukunft der stochastischen Störungen bei der Wahl von  $u_t$  berücksichtigen können, dass wir also in die Zukunft sehen können: für die meisten Anwendungen sicherlich unrealistisch.

Statt dessen wollen wir nun das formalisieren, was wir bereits am Beginn von Abschnitt 1.3 diskutiert haben: Zum Zeitpunkt  $t$  wollen wir genau diejenige stochastische Information bei der Wahl von  $u_t$  verwenden, die wir aus den bekannten Zustandsinformationen  $X_0, \dots, X_t$

gewinnen können. Wir wollen also keine Informationen aus der Zukunft verwenden, andererseits aber auch keine Information aus der Vergangenheit und Gegenwart “verschenken”. Dies machen wir mit Hilfe der Filtrationen mathematisch präzise.

**Definition 2.3** Gegeben sei ein stochastisches Kontrollsystem, ein Anfangswert  $x_0$  und ein bzgl.  $\Sigma$  messbarer Kontrollprozess  $\mathbf{u}$ . Betrachte die *natürliche Filtration*  $\mathcal{F}$  gegeben durch

$$\mathcal{F}_t = \sigma((X_0, \dots, X_t)), \quad t = 0, 1, \dots$$

des entstehenden stochastischen Prozesses  $X_t = X_t(x_0, \mathbf{u})$ .

Dann nennen wir  $\mathbf{u}$  einen *zulässigen Kontrollprozess* zum Anfangswert  $x_0$ , falls  $\mathbf{u}$  adaptiert zur Filtration  $\mathcal{F}$  ist, d.h. (vgl. Bemerkung 1.12), wenn

$$\sigma(u_t) \subseteq \mathcal{F}_t$$

für alle  $t \in \mathbb{N}_0$  gilt.

Die Menge aller zulässigen Kontrollprozesse zum Anfangswert  $x_0$  bezeichnen wir mit  $\mathcal{U}_{x_0}$ . □

Die Definition mag auf den ersten Blick nicht wohldefiniert erscheinen, da die Zulässigkeit von  $\mathbf{u}$  ja von  $\mathcal{F}$  abhängt, welches wiederum via  $X_t$  von  $\mathbf{u}$  abhängt. Sie ist aber lediglich rekursiv in  $t$ , da  $\mathcal{F}_t$  nur von  $X_0, \dots, X_t$  und damit von  $u_0, \dots, u_{t-1}$  abhängt. Die Wahl von  $u_0, \dots, u_{t-1}$  bestimmt also, welche Zufallsvariablen  $u_t$  als  $t$ -te Komponente von  $\mathbf{u}$  zulässig sind. Beachte, dass  $\mathcal{F}_0$  stets die triviale  $\sigma$ -Algebra  $\{\emptyset, \Omega\}$  ist, weil  $X_0 = x_0$  ja ein deterministischer Anfangswert und keine Zufallsvariable ist. Folglich muss auch  $u_0$  ein deterministischer Wert sein, darf also nicht von  $\omega$  abhängen.

Der folgende Satz gibt eine alternative Darstellung von Definition 2.3.

**Satz 2.4** Ein Kontrollprozess  $\mathbf{u}$  ist genau dann zulässig, wenn er für alle  $\omega_1, \omega_2 \in \Omega$  und alle  $t \in \mathbb{N}_0$  die Implikation

$$X_s(\omega_1) = X_s(\omega_2) \text{ für alle } s = 0, \dots, t \quad \implies \quad u_t(\omega_1) = u_t(\omega_2)$$

erfüllt.

**Beweis:** “ $\implies$ ”: Seien  $\mathbf{u}$  zulässig,  $t \in \mathbb{N}_0$  und  $\omega_1, \omega_2 \in \Omega$  mit  $X_s(\omega_1) = X_s(\omega_2)$  für alle  $s = 0, \dots, t$ . Dies bedeutet, dass für alle Borel-Mengen  $B$  der Dimension  $(t+1)n$  und  $A = (X_0, \dots, X_t)^{-1}(B)$  entweder

$$\omega_1 \in A \text{ und } \omega_2 \in A \tag{2.2}$$

oder

$$\omega_1 \notin A \text{ und } \omega_2 \notin A \tag{2.3}$$

gilt. Da  $\mathcal{F}_t$  gerade von den Mengen  $(X_0, \dots, X_t)^{-1}(B)$  erzeugt ist, gilt entweder (2.2) oder (2.3) für alle  $A \in \mathcal{F}_t$  (da diese Eigenschaften unter der Bildung der  $\sigma$ -Algebra nicht verloren gehen können).

Nehmen wir nun an, dass  $u_t(\omega_1) \neq u_t(\omega_2)$ . Dann gilt für die Borel-Menge  $\tilde{B} = \{u_t(\omega_1)\}$  und  $\tilde{A} = u_t^{-1}(\tilde{B})$

$$\omega_1 \in \tilde{A} \text{ und } \omega_2 \notin \tilde{A}. \quad (2.4)$$

Andererseits ist  $u_t$  aber messbar bezüglich  $\mathcal{F}_t$ , also gilt nach Definition der Messbarkeit

$$\tilde{A} \in \mathcal{F}_t.$$

Da alle Mengen  $A \in \mathcal{F}_t$  nun aber entweder (2.2) oder (2.3), die Menge  $\tilde{A}$  aber wegen (2.4) keine dieser zwei Eigenschaften erfüllt, ergibt sich ein Widerspruch. Folglich muss  $u_t(\omega_1) = u_t(\omega_2)$  gelten.

“ $\Leftarrow$ ”: Wir müssen die Messbarkeit  $\sigma(u_t) \in \mathcal{F}_t$  zeigen, wozu es genügt,

$$u_t^{-1}(\tilde{B}) \in \{(X_0, \dots, X_t)^{-1}(B) \mid B \in \mathcal{B}\} \quad (2.5)$$

für alle Borel-Mengen  $\tilde{B}$  zu zeigen. Sei dazu  $\tilde{B}$  gegeben und setze  $\tilde{A} = u_t^{-1}(\tilde{B})$ ,  $B = (X_0, \dots, X_t)(\tilde{A})$  sowie  $A = (X_1, \dots, X_t)^{-1}(B)$ . Wir zeigen  $\tilde{A} = A$ , woraus (2.5) und damit die Behauptung folgt.

“ $\tilde{A} \subseteq A$ ”: Für  $\omega \in \tilde{A}$  gilt  $(X_1, \dots, X_t)(\omega) \in B$  und damit  $\omega \in (X_0, \dots, X_t)^{-1}(B) = A$ . Daraus folgt  $\tilde{A} \subseteq A$ .

“ $\tilde{A} \supseteq A$ ”: Sei  $\omega_1 \in A$  und damit  $x = (X_1, \dots, X_t)(\omega_1) \in B$ . Nach Definition von  $B$  existiert dann  $\omega_2 \in \tilde{A}$  mit  $x = (X_0, \dots, X_t)(\omega_2) \in B$  und damit schließlich ein  $\tilde{x} \in \tilde{B}$  mit  $u_t(\omega_2) = \tilde{x}$ . Da  $\omega_1$  und  $\omega_2$  nun das gleiche Bild unter  $X_0, \dots, X_t$  haben, muss auch  $u_t(\omega_1) = u_t(\omega_2) = \tilde{x}$  gelten. Wegen  $\tilde{x} \in \tilde{B}$  und der Definition von  $\tilde{A}$  muss daher  $\omega_1 \in \tilde{A}$  liegen, woraus  $A \subseteq \tilde{A}$  folgt.  $\square$

Das folgende Korollar macht nun explizit, dass zulässige Kontrollprozesse tatsächlich genau die gewünschte Eigenschaft haben: zu jedem Zeitpunkt  $t$  kann  $u_t$  auf Basis der Informationen in  $X_0, \dots, X_t$  gewählt werden, hängt aber nicht von weiteren Informationen ab. Es liefert damit eine äquivalente kontrolltheoretische Definition der Zulässigkeit, indem  $u_t$  in geeigneter *verallgemeinerter Feedbackform* dargestellt wird.

**Korollar 2.5** Ein Kontrollprozess  $\mathbf{u}$  ist genau dann zulässig, wenn Abbildungen  $h_t : (\mathbb{R}^n)^{t+1} \rightarrow U$ ,  $t \in \mathbb{N}_0$ , existieren, so dass

$$u_t = h_t(X_0, \dots, X_t)$$

für alle  $t \in \mathbb{N}_0$  gilt.

**Beweis:** “ $\Rightarrow$ ”: Sei  $\mathbf{u}$  zulässig. Dann ist die Abbildung

$$h_t(x_0, \dots, x_t) = \begin{cases} u_t(\omega), & \text{falls } \omega \in \Omega \text{ existiert mit } x_0 = X_0(\omega), \dots, x_t = X_t(\omega) \\ 0, & \text{sonst} \end{cases}$$

wohldefiniert, denn wenn die Bedingung in der Definition für verschiedene  $\omega_1 \neq \omega_2$  erfüllt ist, so muss nach Satz 2.4  $u_t(\omega_1) = u_t(\omega_2)$  gelten. Diese Abbildung erfüllt offenbar  $u_t = h_t(X_0, \dots, X_t)$ .

“ $\Leftarrow$ ”: Wenn  $u_t = h_t(X_0, \dots, X_t)$  gilt, so ist die Bedingung aus Satz 2.4 offensichtlich erfüllt und  $\mathbf{u}$  damit zulässig.  $\square$

Wir beenden dieses Kapitel mit zwei technischen Lemmas über die zulässigen Kontrollprozesse.

**Lemma 2.6** Zu jedem Anfangswert  $x$  sei ein zulässiger Kontrollprozess  $\mathbf{u}^x = (u_t^x)$  gegeben. Für eine Zeit  $s \in \mathbb{N}$ , einem Anfangswert  $x_0$  und einem für den Anfangswert  $x_0$  zulässigen Kontrollprozess  $\mathbf{u}$  konstruieren wir einen neuen Kontrollprozess mittels der *Konkatenation*  $(\tilde{u}_t) = \mathbf{u} \&_s \mathbf{u}^x$  definiert durch

$$\tilde{u}_t = \begin{cases} u_t, & t = 0, 1, \dots, s-1 \\ u_{t-s}^{X_s}, & t = s, s+1, \dots \end{cases}$$

Dann ist  $\mathbf{u} \&_s \mathbf{u}^x$  zulässig für die Lösung  $\tilde{X}_t = X_t(x_0, \mathbf{u} \&_s \mathbf{u}^x)$  und für die Lösungen  $X_t^x = X_t(x, \mathbf{u}^x)$  gilt

$$P((u_0^x, \dots, u_t^x, X_0^x, \dots, X_t^x) \in B) = P((\tilde{u}_s, \dots, \tilde{u}_{s+t}, \tilde{X}_s, \dots, \tilde{X}_{s+t}) \in B \mid \tilde{X}_s = x) \quad (2.6)$$

für alle  $t \in \mathbb{N}_0$  und alle Borel-Mengen  $B$  passender Dimension gilt. Mit anderen Worten: die gemeinsame Verteilung der Prozesse  $u_t^x$  und  $X_t^x$  stimmt mit der mittels  $\tilde{X}_s = x$  bedingten gemeinsamen Verteilung der verschobenen Prozesse  $\tilde{u}_{t+s}$  und  $\tilde{X}_{t+s}$  überein.

**Beweis:** Beachte, dass das Subtile an der Definition in der Tatsache liegt, dass die Auswahl von  $u^{X_s}$  wiederum vom Zufall abhängt. Mit welchem Kontrollprozess ab dem Zeitpunkt  $t = s$  gesteuert wird, hängt also davon ab, in welchem (zufälligen) Zustand  $X_s$  wir uns zu diesem Zeitpunkt befinden.

Zum Beweis verwenden wir die Darstellung aus Korollar 2.5 und bezeichnen die zu  $\mathbf{u}^x$  und  $\mathbf{u}$  gehörigen Abbildungen mit  $h_t^x$  bzw.  $h_t$ . Dann sind die zu  $\mathbf{u} \&_s \mathbf{u}^x$  gehörigen Abbildungen  $\tilde{h}_t$  gegeben durch

$$\tilde{h}_t(x_0, \dots, x_t) = \begin{cases} h_t(x_0, \dots, x_t), & t = 0, 1, \dots, s-1 \\ h_{t-s}^{X_s}(x_s, \dots, x_t), & t = s, s+1, \dots \end{cases}$$

Damit ist  $\mathbf{u} \&_s \mathbf{u}^x$  wieder von der in Korollar 2.5 angegebenen Form und damit zulässig.

Zum Beweis von (2.6) betrachten wir zunächst die deterministischen Differenzgleichungen

$$\tilde{x}_{s+t+1} = f(\tilde{x}_{s+t}, \tilde{u}_{t+s}, z_t), \quad t = 0, 1, \dots$$

und

$$x_{t+1}^x = f(x_t^x, u_t^x, z_t), \quad t = 0, 1, \dots$$

mit Anfangswerten  $\tilde{x}_s = x_0^x = x$ . Diese Gleichungen unterscheiden sich von den Gleichungen für  $\tilde{X}_{s+t}$  bzw.  $X_t^x$  nur dadurch, dass nun Folgen von Zahlen  $z_t$  an Stelle der Zufallsvariablen eingehen. Bezeichnen wir die Lösungen mit

$$\tilde{x}_{s+t}(z_0, \dots, z_{t-1}) \quad \text{bzw.} \quad x_t^x(z_0, \dots, z_{t-1})$$

und die zugehörigen Kontrollfunktionen mit

$$\tilde{u}_{s+t}(z_0, \dots, z_{t-1}) \quad \text{bzw.} \quad u_t^x(z_0, \dots, z_{t-1})$$

so folgt aus der Definition von  $\tilde{u}_t$  leicht per Induktion, dass diese Lösungen übereinstimmen. Zudem überprüft man ebenfalls per leichter Induktion die Gleichungen

$$\tilde{X}_{s+t} = \tilde{x}_{s+t}(Z_s, \dots, Z_{s+t-1}), \quad u_{s+t}^x = u_{s+t}^x(Z_s, \dots, Z_{s+t-1})$$

für alle  $\omega$  mit  $\tilde{X}_s(\omega) = x$  sowie

$$X_t^x = x_t^x(Z_0, \dots, Z_{t-1}), \quad u_t^x = u_t^x(Z_t, \dots, Z_{t-1}).$$

Schreiben wir nun kurz  $y_t^x = ((u_0^x, \dots, u_t^x, x_0^x, \dots, x_t^x)$  und  $\tilde{y}_t = (\tilde{u}_s, \dots, \tilde{u}_{s+t}, \tilde{x}_s, \dots, \tilde{x}_{s+t})$  sowie  $Y_t^x = \tilde{y}_t^x(Z_0, \dots, Z_{t-1})$  und  $\tilde{Y}_t = \tilde{y}_t(Z_s, \dots, Z_{t-1})$  so wird (2.6) zu

$$P(Y_t^x \in B) = P(\tilde{Y}_t \in B \mid \tilde{X}_s = x). \quad (2.7)$$

Zudem gilt  $(y_t^x)^{-1}(B) = \tilde{y}_t^{-1}(B)$ . Setzen wir also  $C := (y_t^x)^{-1}(B)$  so folgt für die Ausdrücke in (2.7)

$$P(Y_t^x \in B) = P(y_t^x(Z_0, \dots, Z_{t-1}) \in B) = P((Z_0, \dots, Z_{t-1}) \in C)$$

und

$$\begin{aligned} P(\tilde{Y}_t \in B \mid \tilde{X}_s = x) &= P(\tilde{y}_t(Z_s, \dots, Z_{t+s-1}) \in B \mid \tilde{X}_s = x) \\ &= P((Z_s, \dots, Z_{t+s-1}) \in C \mid \tilde{X}_s = x). \end{aligned}$$

Um (2.7) und damit (2.6) zu beweisen, genügt es also, die Gleichung

$$P((Z_0, \dots, Z_{t-1}) \in C) = P((Z_s, \dots, Z_{t+s-1}) \in C \mid \tilde{X}_s = x) \quad (2.8)$$

für alle Borel-Mengen  $C$  zu zeigen. Da  $\tilde{X}_s$  nun aber lediglich von den Zufallsvariablen  $Z_0, \dots, Z_{s-1}$  abhängt, welche unabhängig von  $Z_s, \dots, Z_{t+s-1}$  sind, folgt

$$P((Z_s, \dots, Z_{t+s-1}) \in C \mid \tilde{X}_s = x) = P((Z_s, \dots, Z_{t+s-1}) \in C)$$

und damit (2.8), da die  $Z_t$  identisch verteilt sind.  $\square$

Lemma 2.6 zeigt, dass wir die Kontrollprozesse  $\mathbf{u}^x$  an einen zulässigen Kontrollprozess  $\mathbf{u}$  “anhängen” können und so wieder einen zulässigen Kontrollprozess erhalten, dessen Lösungen in Wahrscheinlichkeit mit den um  $s$  nach rechts verschobenen Lösungen zu  $\mathbf{u}^x$  übereinstimmen. Das folgende Lemma zeigt in gewissem Sinne die Umkehrung, nämlich wie wir einen zulässigen Kontrollprozess  $\tilde{\mathbf{u}}$  definieren können, dessen Lösungen in Wahrscheinlichkeit mit den um  $s$  nach links verschobenen Lösungen eines gegebenen zulässigen Kontrollprozesses  $\mathbf{u}$  übereinstimmen.

**Lemma 2.7** Es sei  $\mathbf{u}$  ein zu einem Anfangswert  $x_0$  zulässiger Kontrollprozess und  $s \in \mathbb{N}$ . Es seien  $\tilde{x}_0, \dots, \tilde{x}_s \in \mathbb{R}^n$  und

$$A := \{\omega \in \Omega \mid X_t(\omega) = \tilde{x}_t, t = 0, \dots, s\} \in \mathcal{F}_s.$$

Dann existiert eine Kontrollfunktion  $\tilde{\mathbf{u}}$ , die zulässig für den Anfangswert  $\tilde{x}_s$  ist und für die und die zugehörige Lösung  $\tilde{X}_t = X_t(\tilde{x}_s, \tilde{\mathbf{u}})$  die Gleichung

$$P((\tilde{u}_0, \dots, \tilde{u}_t, \tilde{X}_0, \dots, \tilde{X}_t) \in B) = P((u_s, \dots, u_{s+t}, X_s, \dots, X_{s+t}) \in B \mid A) \quad (2.9)$$

für alle  $t \in \mathbb{N}_0$  und alle Borel-Mengen  $B$  passender Dimension gilt. Mit anderen Worten: die gemeinsame Verteilung der Prozesse  $\tilde{u}_t$  und  $\tilde{X}_t$  stimmt mit der mittels  $A$  bedingten gemeinsamen Verteilung der verschobenen Prozesse  $u_{t+s}$  und  $X_{t+s}$  überein.

**Beweis:** Wir definieren  $\tilde{\mathbf{u}}$  wieder über geeignete Abbildungen  $\tilde{h}_t$  gemäß Korollar 2.5. Seien dazu  $h_t$  die zu  $\mathbf{u}$  gehörigen Abbildungen. Dann definieren wir

$$\tilde{h}_t(x_0, \dots, x_t) = h_{s+t}(\tilde{x}_0, \dots, \tilde{x}_s, x_1, \dots, x_t).$$

Der Beweis von (2.9) verläuft völlig analog zum entsprechenden Beweis von (2.6) im Beweis von Lemma 2.6: Wir betrachten wir zunächst die deterministischen Differenzgleichungen

$$x_{s+t+1} = f(x_{s+t}, u_{t+s}, z_t), \quad t = 0, 1, \dots$$

und

$$\tilde{x}_{t+1} = f(\tilde{x}_t, \tilde{u}_t, z_t), \quad t = 0, 1, \dots$$

mit Anfangswert  $\tilde{x}_0 = x_s$ . Diese Gleichungen unterscheiden sich von den Gleichungen für  $X_{s+t}$  bzw.  $\tilde{X}_t$  nur dadurch, dass nun Folgen von Zahlen  $z_t$  an Stelle der Zufallsvariablen eingehen. Bezeichnen wir die Lösungen mit

$$x_{s+t}(z_0, \dots, z_{t-1}) \quad \text{bzw.} \quad \tilde{x}_t(z_0, \dots, z_{t-1})$$

und die zugehörigen Kontrollfunktionen mit

$$u_{s+t}(z_0, \dots, z_{t-1}) \quad \text{bzw.} \quad \tilde{u}_t(z_0, \dots, z_{t-1})$$

so folgt aus der Definition von  $\tilde{h}_t$  leicht per Induktion, dass diese Lösungen übereinstimmen. Zudem überprüft man ebenfalls per leichter Induktion die Gleichungen

$$X_{s+t} = x_{s+t}(Z_s, \dots, Z_{s+t-1}), \quad u_{s+t} = u_{s+t}(Z_s, \dots, Z_{s+t-1})$$

für alle  $\omega \in A$  sowie

$$\tilde{X}_t = \tilde{x}_t(Z_0, \dots, Z_{t-1}), \quad \tilde{u}_t = \tilde{u}_t(Z_t, \dots, Z_{t-1}).$$

Schreiben wir nun kurz  $\tilde{y}_t = ((\tilde{u}_0, \dots, \tilde{u}_t, \tilde{x}_0, \dots, \tilde{X}_t))$  und  $y_t = (u_s, \dots, u_{s+t}, x_s, \dots, x_{s+t})$  sowie  $\tilde{Y}_t = \tilde{y}_t(Z_0, \dots, Z_{t-1})$  und  $Y_t = y_t(Z_s, \dots, Z_{t-1})$  so wird (2.9) zu

$$P(\tilde{Y}_t \in B) = P(Y_t \in B | A). \quad (2.10)$$

Zudem gilt  $\tilde{y}_t^{-1}(B) = y_t^{-1}(B)$ . Setzen wir also  $C := \tilde{y}_t^{-1}(B)$  so folgt für die Ausdrücke in (2.10)

$$P(\tilde{Y}_t \in B) = P(\tilde{y}_t(Z_0, \dots, Z_{t-1}) \in B) = P((Z_0, \dots, Z_{t-1}) \in C)$$

und

$$P(Y_t \in B | A) = P(y_t(Z_s, \dots, Z_{t+s-1}) \in B | A) = P((Z_s, \dots, Z_{t+s-1}) \in C | A).$$

Um (2.10) und damit (2.9) zu beweisen, genügt es also,

$$P((Z_0, \dots, Z_{t-1}) \in C) = P((Z_s, \dots, Z_{t+s-1}) \in C | A) \quad (2.11)$$

für alle Borel-Mengen  $C$  zu zeigen. Dies geht genau wie der Beweis von Gleichung (2.8) im Beweis von Lemma 2.6.  $\square$

## Kapitel 3

# Stochastische dynamische Optimierungsprobleme

Nachdem wir nun geklärt haben, was wir unter einem stochastischen Kontrollsystem und einem zulässigen Kontrollprozess verstehen, können wir endlich die Klassen dynamischer Optimierungsprobleme definieren, die wir in dieser Vorlesung betrachten wollen.

### 3.1 Definition

**Problem 3.1** Gegeben sei ein stochastisches Kontrollsystem gemäß Definition 2.1 mit Lösung  $X_t = X_t(x_0, \mathbf{u})$ , eine *laufende Ertragsfunktion*  $\ell : \mathbb{R}^n \times U \rightarrow \mathbb{R}$ , ein Diskontfaktor  $\beta \in (0, 1]$  und eine *End-Ertragsfunktion*  $L : \mathbb{R}^n \rightarrow \mathbb{R}$ . Das dynamische Optimierungsproblem auf dem endlichen Zeithorizont  $\{0, 1, \dots, T\}$  lautet dann

$$\text{maximiere } J_T(x_0, \mathbf{u}) := E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t, u_t) + \beta^T L(X_T) \right\}$$

über alle für  $x_0$  zulässigen Kontrollfunktionen  $\mathbf{u} = (u_t, t \in \mathbb{N}_0) \in \mathcal{U}_{x_0}$ .

Für  $L \equiv 0$  definieren wir zudem das Problem auf unendlichen Zeithorizont

$$\text{maximiere } J_\infty(x_0, \mathbf{u}) := \lim_{T \rightarrow \infty} J_T(x_0, \mathbf{u})$$

über alle für  $x_0$  zulässigen Kontrollfunktionen  $\mathbf{u} = (u_t, t \in \mathbb{N}_0) \in \mathcal{U}_{x_0}$ . □

Die analogen Minimierungsprobleme erhält man, indem man über  $-\ell$  und  $-L$  maximiert.

Gegebenenfalls kann das Problem noch durch Beschränkungen ergänzt werden:

**Kontrollbeschränkungen:**  $u_t \in U(X_t)$  für gegebene kompakte Mengen  $U(x)$

**Zustandsbeschränkungen:**  $X_t \in \mathcal{X}$  für eine Menge  $\mathcal{X} \subset \mathbb{R}^n$

Bei der Modellierung realistischer Probleme kommt man ohne solche Beschränkungen meist nicht aus. Für unsere analytischen Untersuchungen stellen sie zum Teil aber erhebliche Hindernisse dar, weswegen wir diese i.A. ohne Beschränkungen formulieren werden und den Effekt von Beschränkungen nur von Fall zu Fall nachträglich diskutieren werden.

### 3.2 Beispiele

**Beispiel 3.2** Wir bauen das einfache Aktienkursmodell aus Beispiel 1.7 zu einem Kontrollsystem aus, das ein Portfoliooptimierungsproblem modelliert. Wir wählen dazu den zweidimensionalen Zustand  $X_t = (X_t^1, X_t^2)^1$ , dessen erste Komponente dem in eine Aktie investierten Vermögensanteil eines Anlegers entspricht. Solange dieser den Anteil nicht ändert, verhält sich dieser Vermögensanteil also genau so wie der Kurs der zu Grunde liegenden Aktie, also gemäß Beispiel (1.7) als

$$X_{t+1}^1 = f_1(X_t^1, Z_t) \quad \text{mit } f_1(x_1, z) = x_1 z.$$

Die Komponente  $X_t^2$  modelliert demgegenüber den Vermögensanteil, der in eine risikolose Anleihe mit Zinssatz  $r$  (angegeben in %/100 pro Zeitschritt, ein Zinssatz von 5% pro Zeitschritt entspricht also  $r = 0.05$ ) investiert wurde. Dieser Anteil verhält sich folglich gemäß der (deterministischen) Dynamik

$$X_{t+1}^2 = f_2(X_t^2) \quad \text{mit } f_2(x_2) = (1 + r)x_2.$$

Die Kontrolle  $u$  erlaubt nun, in jedem Zeitschritt (genauer: jeweils vor der Verzinsung bzw. Änderung des Aktienkurses) zwischen den beiden Anlageformen umzuschichten, indem man  $u$  Geldeinheiten von der einen Anlage abzieht und zu der anderen hinzuaddiert. Dies führt auf das Kontrollsystem

$$X_{t+1} = f(X_t, u_t, Z_t) \quad \text{mit } f((x_1, x_2), u, z) = \begin{pmatrix} z(x_1 + u) \\ (1 + r)(x_2 - u) \end{pmatrix}.$$

Wenn man vermeiden will, dass die  $X_t$ -Variablen negativ werden (was ökonomisch nicht unbedingt nötig ist, weil man dies als Kredit bzw. Leerverkauf interpretieren kann), ergeben sich die zustandsabhängigen Kontrollbeschränkungen  $u_t \in U(X_t) = [-X_t^1, X_t^2]$ .

Ziel der Optimierung ist es nun natürlich, den Wert  $X_T^1 + X_T^2$  des Portfolios zum Laufzeitende zu maximieren. In der einfachsten Version führt dies auf den laufenden Ertrag  $\ell \equiv 0$  (der Diskontfaktor spielt dann keine Rolle und kann als  $\beta = 1$  gewählt werden) und  $L((x_1, x_2)) = x_1 + x_2$ . Wir werden aber bald sehen, dass diese Wahl von  $L$  i.A. wenig sinnvoll ist, weil sie zu trivialen Lösungen des Problems führt.  $\square$

**Beispiel 3.3** Das zweite Modellproblem ist ein Wagen, den wir mittels Motorkraft beschleunigen und bremsen können und den wir mit einer geeigneten Kontrollstrategie möglichst nahe am Nullpunkt halten wollen. Durch externe Störungen (Wind, Stöße...) wird der Wagen allerdings stets wieder beschleunigt und dadurch aus dem Nullpunkt entfernt.

Die Dynamik eines solchen Wagens ist zunächst durch eine Differentialgleichung, also ein zeitkontinuierliches Modell gegeben. Bezeichnet  $x_1$  die Position,  $x_2$  die Geschwindigkeit und  $u(t)$  die durch unseren Motor ausgeübte Beschleunigung des Wagens, so gilt

$$\begin{aligned} \dot{x}_1(t) &= x_2(t) \\ \dot{x}_2(t) &= u(t). \end{aligned}$$

---

<sup>1</sup>Da wir in dieser Vorlesung der Konvention der stochastischen Literatur folgen, die Zeit als unteren Index zu schreiben, werden die Komponenten vektorwertiger stochastischer Prozesse durch einen oberen Index gekennzeichnet.

Diese Differentialgleichung ist nun aber zum Glück so einfach, dass man sie explizit lösen kann. Wenn wir einen Zeitschritt  $h > 0$  betrachten und  $u|_{[t,t+h)} \equiv \text{const}$  setzen gilt für  $\tau \in [t, t+h)$

$$x_2(t+\tau) = x_2(t) + \int_t^{t+\tau} u(s)ds = x_2(t) + \tau u(t)$$

und damit

$$x_1(t+\tau) = x_1(t) + \int_t^{t+\tau} x_2(s)ds = x_1(t) + \int_t^{t+\tau} x_2(t) + su(t)ds = x_1(t) + \tau x_2(t) + \frac{\tau^2}{2}u(t).$$

Nehmen wir nun an, dass die Störung  $z(t)$  genau wie die Kontrolle auf die Beschleunigung wirkt und nehmen wir zusätzlich vereinfachend an, dass  $z(t)$  ebenfalls konstant auf dem Intervall  $[t, t+h)$  ist, erhalten wir

$$\begin{aligned} x_1(t+\tau) &= x_1(t) + hx_2(t) + \frac{h^2}{2}(u(t) + z(t)) \\ x_2(t+h) &= x_2(t) + h(u(t) + z(t)) \end{aligned}$$

Fixieren wir schließlich einen Zeitschritt  $h$  und wählen die diskrete Zeiteinheit gleich diesem  $h$ , so erhalten wir als diskretes Modell

$$X_{t+1} = f(X_t, u_t, Z_t)$$

mit

$$f(x, u, z) = \begin{pmatrix} x_1 + hx_2 + \frac{h^2}{2}(u + z) \\ x_2 + h(u + z) \end{pmatrix}$$

Als Modell für die stochastische Störung nehmen wir an, dass  $Z_t$  normalverteilte Zufallsvariablen mit Mittelwert 0 und Varianz  $\sigma^2$  sind, also  $Z \sim N(0, \sigma^2)$ .

Unser Kontrollziel ist nun, den Wagen möglichst nahe bei Null zu halten. Dies soll aber nicht nur für eine feste Endzeit  $T$  sondern gleichmäßig für alle Zeiten geschehen — eine sogenannte Regelungsaufgabe. Daher verwenden wir ein Minimierungsproblem auf unendlichem Horizont, bei dem wir die mittels  $\ell(x, u) = \|x\|^2$  gemessene Abweichung vom Nullpunkt als laufende Kosten verwenden könnten. Zusätzlich sollte aber auch die Kontrolle “bestraft” werden. Dies ist zum einen mathematisch sinnvoll, wie wir später sehen werden, zum anderen aber auch in der Anwendung, denn  $u$  repräsentiert den Energieaufwand, den wir zum Regeln benötigen. Wir setzen daher

$$\ell(x, u) = \|x\|^2 + \alpha u^2,$$

wobei  $\alpha > 0$  ein Gewichtungsparemeter ist, der angibt, wie stark die Kontrolle in der Optimierung berücksichtigt werden soll. Schließlich müssen wir noch überlegen, wie der Diskontfaktor  $\beta$  gewählt werden soll. Wer die Kontrolltheorie gehört hat, weiß dass in deterministischen Regelungsproblemen  $\beta = 1$  gewählt wird. Dies ist hier nicht sinnvoll, denn wir müssen davon ausgehen, dass der Wagen wegen des stochastischen Einflusses selbst bei optimaler Regelung in einem Intervall  $[-\varepsilon, \varepsilon]$  um den Nullpunkt “herumzappelt”, weswegen  $J_T(x, u)$  für  $T \rightarrow \infty$  divergiert. Daher wählen wir einen Diskontfaktor  $\beta < 1$ , damit sicher gestellt ist, dass  $J_T$  konvergiert. Welchen Effekt verschiedene Wahlen von  $\beta$  haben, werden wir später betrachten.  $\square$

### 3.3 Optimale Wertefunktionen, Kontrollprozesse und Lösungen

Für unsere optimalen Steuerungsprobleme definieren wir die folgenden Größen.

**Definition 3.4** Betrachte das Problem 3.1 mit  $T \in \mathbb{N}_0 \cup \{\infty\}$ .

(i) Wir definieren die *optimale Wertefunktionen*  $V_T : \mathbb{R}^n \rightarrow \mathbb{R}$  als

$$V_T(x_0) := \sup_{\mathbf{u} \in \mathcal{U}_{x_0}} J_T(x_0, \mathbf{u}).$$

Die Funktion  $V_T$  ordnet also jedem Anfangswert den maximal erzielbaren Wert des Funktionals  $J_T(x_0, \mathbf{u})$  zu.

(ii) Ein Kontrollprozess  $\mathbf{u}^T \in \mathcal{U}(x_0)$  heißt *optimaler Kontrollprozess* für  $J_T$ , falls

$$J_T(x_0, \mathbf{u}^T) = V_T(x_0)$$

gilt. Die zugehörige Lösung  $X_t(x_0, \mathbf{u}^T)$  heißt dann *optimale Lösung*. □

Beachte, dass wir das Problem auch für  $T = 0$  definiert haben. In diesem Fall ergibt sich  $J_0(x_0, u) = L(x_0)$  und damit auch  $V_0(x_0) = L(x_0)$ .

Im Folgenden wollen wir die Struktur des Problems genauer untersuchen. Insbesondere wollen wir klären, welche Struktur ein optimaler Kontrollprozess besitzt, wann er überhaupt existiert und wie man ihn (zumindest theoretisch) berechnen kann.

## Kapitel 4

# Das Prinzip der Dynamischen Programmierung

In diesem Kapitel betrachten wir das *Prinzip der Dynamischen Programmierung*, welches grundlegend für fast alle weiteren Betrachtungen — analytisch wie numerisch — sein wird. Das Prinzip wird auch *Bellman'sches Optimalitätsprinzip*<sup>1</sup> oder kurz *Optimalitätsprinzip* genannt. Wir beginnen mit Problemen auf endlichem Zeithorizont  $\{0, 1, \dots, T\}$ .

### 4.1 Probleme auf endlichem Horizont

#### 4.1.1 Formulierung und Beweis des Prinzips

**Satz 4.1** Die optimalen Wertefunktionen  $V_0, \dots, V_T$  erfüllen für alle  $\tau = 1, \dots, T$  und alle  $x_0 \in \mathbb{R}^n$  die Gleichung

$$V_\tau(x_0) = \sup_{u \in U} E \{ \ell(x_0, u) + \beta V_{\tau-1}(f(x_0, u, Z_0)) \}. \quad (4.1)$$

Umgekehrt stimmt jede Familie von Funktionen  $W_0, \dots, W_T$ , die  $W_0(x) = L(x)$  und (4.1) für  $\tau = 1, \dots, T$  erfüllt, mit  $V_0, \dots, V_T$  überein.

**Beweis:** Wir beweisen zunächst (4.1).

“ $\leq$ ”: Sei  $x_0 \in \mathbb{R}^n$  und  $\mathbf{u} \in \mathcal{U}(x_0)$  beliebig. Dann gilt für  $X_t = X_t(x_0, \mathbf{u})$

$$\begin{aligned} J_\tau(x_0, \mathbf{u}) &= E \left\{ \sum_{t=0}^{\tau-1} \beta^t \ell(X_t, u_t) + \beta^\tau L(X_\tau) \right\} \\ &= E \left\{ \ell(x_0, u_0) + \sum_{t=1}^{\tau-1} \beta^t \ell(X_t, u_t) + \beta^\tau L(X_\tau) \right\} \\ &= \ell(x_0, u_0) + E \left\{ E_1 \left( \sum_{t=1}^{\tau-1} \beta^t \ell(X_t, u_t) + \beta^\tau L(X_\tau) \right) \right\} \end{aligned} \quad (4.2)$$

---

<sup>1</sup>Benannt nach dem amerikanischen Mathematiker Richard E. Bellman (1920–1984), einem der Erfinder dieses Prinzips, das tatsächlich öfter unabhängig voneinander erfunden wurde.

wobei wir im letzten Schritt die Eigenschaft (1.10) von  $E_s$  für  $s = 1$  verwendet haben und zudem ausgenutzt haben, dass  $\ell(x_0, u_0)$  keine Zufallsvariable ist. Für jedes  $\omega$  gilt nun nach Definition von  $E_s$

$$E_1 \left( \sum_{t=1}^{\tau-1} \beta^t \ell(X_t, u_t) + \beta^\tau L(X_\tau) \right) (\omega) = E \left( \sum_{t=1}^{\tau-1} \beta^t \ell(X_t, u_t) + \beta^\tau L(X_\tau) \middle| A \right)$$

mit

$$A = \{\tilde{\omega} \in \Omega \mid X_1(\tilde{\omega}) = X_1(\omega)\}.$$

Da  $X_0 = x_0$  keine Zufallsvariable ist, ist dieses  $A$  genau von der Form der Menge  $A$  in Lemma 2.7. Ersetzen wir nun für jedes  $\omega \in A$  die Lösung  $X_t$  und den Kontrollprozess  $u_t$  durch  $\tilde{X}_{t-1}$  und  $\tilde{u}_{t-1}$  aus Lemma 2.7 (beachte: dann gilt  $\tilde{X}_0 = X_1(\omega)$ ), so folgt

$$\begin{aligned} E \left( \sum_{t=1}^{\tau-1} \beta^t \ell(X_t, u_t) + \beta^\tau L(X_\tau) \middle| A \right) &= \beta E \left( \sum_{t=0}^{\tau-2} \beta^t \ell(\tilde{X}_t, \tilde{u}_t) + \beta^{\tau-1} L(\tilde{X}_{\tau-1}) \right) \\ &= \beta J_{\tau-1}(X_1(\omega), \tilde{u}_t) \leq \beta V_{\tau-1}(X_1(\omega)), \end{aligned} \quad (4.3)$$

da aus der in (2.9) formulierten identischen Verteilung auch die Gleichheit der Erwartungswerte folgt. Wegen  $X_1 = f(x_0, u_0, Z_0)$  können wir weiter schreiben

$$\beta V_{\tau-1}(X_1(\omega)) = \beta V_{\tau-1}(f(x_0, u_0, Z_0(\omega)))$$

und erhalten damit schließlich

$$E_1 \left( \sum_{t=1}^{\tau-1} \beta^t \ell(X_t, u_t) + \beta^\tau L(X_\tau) \right) \leq \beta V(f(x_0, u_0, Z_0)).$$

Setzen wir dies in (4.2) ein, erhalten wir insgesamt

$$\begin{aligned} J_\tau(x_0, \mathbf{u}) &\leq \ell(x_0, u_0) + \beta E\{V(f(x_0, u, Z_0))\} \\ &= E\{\ell(x_0, u_0) + \beta V(f(x_0, u, Z_0))\} \\ &\leq \sup_{u \in U} E\{\ell(x_0, u) + \beta V_{\tau-1}(f(x_0, u, Z_0))\}. \end{aligned}$$

Da diese Ungleichung nun für alle zulässigen  $\mathbf{u}$  gilt, gilt sie auch für

$$V_\tau(x) = \sup_{\mathbf{u} \in \mathcal{U}(x)} J_\tau(x, \mathbf{u}),$$

womit “ $\leq$ ” folgt.

“ $\geq$ ”: Sei  $\varepsilon > 0$  und  $x_0 \in \mathbb{R}^n$  gegeben. Zu jedem  $x \in \mathbb{R}^n$  wählen wir einen Kontrollprozess  $\mathbf{u}^x \in \mathcal{U}(x)$ , für den

$$J_{\tau-1}(x, \mathbf{u}^x) \geq V_{\tau-1}(x) - \varepsilon$$

gilt. Weiterhin wählen wir einen Kontrollwert  $u^* \in U$  mit

$$E\{\ell(x_0, u^*) + \beta V_{\tau-1}(f(x_0, u^*, Z_0))\} \geq \sup_{u \in U} E\{\ell(x_0, u) + \beta V_{\tau-1}(f(x_0, u, Z_0))\} - \varepsilon,$$

einen Kontrollprozess  $\mathbf{u} \in \mathcal{U}(x_0)$  mit  $u_0 = u^*$  und setzen  $\tilde{\mathbf{u}} = \mathbf{u} \&_1 \mathbf{u}^x$  gemäß Lemma 2.6. Die zugehörige Lösung sei mit  $\tilde{X}_t = X_t(x_0, \tilde{\mathbf{u}})$  bezeichnet. Dann gilt

$$\begin{aligned}
V_\tau(x_0) &= \sup_{\mathbf{u} \in \mathcal{U}(x_0)} \{J_\tau(x_0, \tilde{\mathbf{u}})\} \\
&= \sup_{\mathbf{u} \in \mathcal{U}(x_0)} E \left\{ \sum_{t=0}^{\tau-1} \beta^t \ell(X_t, u_t) + \beta^\tau L(X_\tau) \right\} \\
&= \sup_{\mathbf{u} \in \mathcal{U}(x_0)} E \left\{ \ell(x_0, u_0) + \sum_{t=1}^{\tau-1} \beta^t \ell(X_t, u_t) + \beta^\tau L(X_\tau) \right\} \\
&\geq E \left\{ \ell(x_0, \tilde{u}_0) + \sum_{t=1}^{\tau-1} \beta^t \ell(\tilde{X}_t, \tilde{u}_t) + \beta^\tau L(\tilde{X}_\tau) \right\} \\
&= \ell(x_0, \tilde{u}_0) + E \left\{ E_1 \left( \sum_{t=1}^{\tau-1} \beta^t \ell(\tilde{X}_t, \tilde{u}_t) + \beta^\tau L(\tilde{X}_\tau) \right) \right\}, \tag{4.4}
\end{aligned}$$

wobei der letzte Schritt wie (4.2) folgt. Genau wie in (4.3) erhalten wir (nun basierend auf (2.6) an Stelle von (2.9)) die Gleichung

$$E_1 \left( \sum_{t=1}^{\tau-1} \beta^t \ell(\tilde{X}_t, \tilde{u}_t) + \beta^\tau L(\tilde{X}_\tau) \right) (\omega) = \beta J_{\tau-1}(x, u_t^x)$$

mit  $x = \tilde{X}_1(\omega)$ . Aus der Wahl von  $u^x$  folgt daher

$$E_1 \left( \sum_{t=1}^{\tau-1} \beta^t \ell(\tilde{X}_t, \tilde{u}_t) + \beta^\tau L(\tilde{X}_\tau) \right) (\omega) \geq V_{\tau-1}(x) - \varepsilon = \beta V_{\tau-1}(f(x_0, \tilde{u}_0, Z_0(\omega))) - \varepsilon.$$

da aus der Definition von  $\mathbf{u} \&_1 \mathbf{u}^x$  die Gleichheit  $\tilde{u}_0 = u_0$  folgt.

Setzen wir dies in (4.4) ein und verwenden die aus der Definition von  $\mathbf{u} \&_1 \mathbf{u}^x$  folgende Gleichheit  $\tilde{u}_0 = u_0$ , so ergibt sich

$$\begin{aligned}
V_\tau(x_0) &\geq E \{ \ell(x_0, u_0) + \beta V_{\tau-1}(f(x_0, u_0, Z_0)) \} - \varepsilon \\
&\geq \sup_{u \in U} E \{ \ell(x_0, u) + \beta V_{\tau-1}(f(x_0, u, Z_0)) \} - 2\varepsilon
\end{aligned}$$

wobei wir im letzten Schritt die Wahl von  $\mathbf{u}$  ausgenutzt haben. Da  $\varepsilon > 0$  beliebig war, folgt

$$V_\tau(x_0) \geq \sup_{u \in U} E \{ \ell(x_0, u) + \beta V_{\tau-1}(f(x_0, u, Z_0)) \},$$

also gerade die gewünschte Ungleichung.

Zum Beweis der Eindeigkeitsaussage beweisen wir zunächst die Hilfsaussage

$$\left| \sup_{u \in U} a(u) - \sup_{u \in U} b(u) \right| \leq \sup_{u \in U} |a(u) - b(u)| \tag{4.5}$$

für beliebige Funktionen  $a, b : U \rightarrow \mathbb{R}$ . Zum Beweis von (4.5) nehmen wir o.B.d.A. an, dass

$$\left| \sup_{u \in U} a(u) - \sup_{u \in U} b(u) \right| = \sup_{u \in U} a(u) - \sup_{u \in U} b(u)$$

gilt, fixieren ein beliebiges  $\varepsilon > 0$  und wählen  $u^*$  so, dass  $\sup_{u \in U} a(u) \leq a(u^*) + \varepsilon$  gilt. Dann folgt

$$\sup_{u \in U} a(u) - \sup_{u \in U} b(u) \leq a(u^*) - b(u^*) + \varepsilon \leq |a(u^*) - b(u^*)| + \varepsilon \leq \sup_{u \in U} |a(u) - b(u)| + \varepsilon.$$

Da  $\varepsilon > 0$  beliebig war, folgt (4.5).

Seien nun  $W_0, \dots, W_T$  gegeben, für die  $W_0 = L$  und (4.1) für  $\tau = 1, \dots, T$  gelten. Wir beweisen  $W_\tau = V_\tau$  induktiv über  $\tau$ . Für  $\tau = 0$  ist die Aussage klar. Für  $\tau \rightarrow \tau + 1$  gilt für alle  $x \in \mathbb{R}^n$

$$\begin{aligned} & |W_{\tau+1}(x) - V_{\tau+1}(x)| \\ &= \left| \sup_{u \in U} E \{ \ell(x, u) + \beta W_{\tau-1}(f(x, u, Z_0)) \} - \sup_{u \in U} E \{ \ell(x, u) + \beta V_{\tau-1}(f(x, u, Z_0)) \} \right| \\ &\leq \sup_{u \in U} |E \{ \ell(x, u) + \beta W_{\tau-1}(f(x, u, Z_0)) \} - E \{ \ell(x, u) + \beta V_{\tau-1}(f(x, u, Z_0)) \}| \\ &= |E \{ \beta W_{\tau-1}(f(x, u, Z_0)) - \beta V_{\tau-1}(f(x, u, Z_0)) \}| \\ &\leq \beta \|W_{\tau-1} - V_{\tau-1}\|_\infty = 0, \end{aligned}$$

wobei wir im letzten Schritt die Induktionsannahme verwendet haben. Also erhalten wir

$$|W_{\tau+1}(x) - V_{\tau+1}(x)| = 0$$

für alle  $x \in \mathbb{R}^n$ , weswegen  $W_{\tau+1} = V_{\tau+1}$  gelten muss.  $\square$

**Bemerkung 4.2** Der Satz gilt auch unter Zustandsbeschränkungen und (zustandsabhängigen) Kontrollbeschränkungen, wenn man das Supremum in (4.1) nur über solche  $u$  nimmt, die die gegebenen Beschränkungen nicht verletzen.  $\square$

Das Prinzip der dynamischen Programmierung gibt eine eindeutige Charakterisierung der optimalen Wertefunktionen. Es kann daher — im Prinzip — sowohl zur Berechnung von  $V_T$  verwendet werden, als auch zur Überprüfung, ob eine gegebene Funktion  $V_T$  tatsächlich die optimale Wertefunktion ist. Wir illustrieren dies an Beispiel 3.2.

**Beispiel 4.3** Betrachte das Modell aus Beispiel 3.2 mit  $\ell \equiv 0$ ,  $\beta = 1$  und  $L(x) = x_1 + x_2$  und der Beschränkung  $u \in U(x) = [-x_1, x_2]$ . Bezeichne  $\mu = E(Z_t)$  den Erwartungswert der Zufallsvariablen  $Z_t$  aus dem Aktienkursmodell. Dann lässt sich aus dem Optimalitätsprinzip induktiv

$$V_T(x) = \max\{\mu, 1 + r\}^T L(x) = \max\{\mu, 1 + r\}^T (x_1 + x_2)$$

errechnen. Hierzu geht man induktiv für  $T = 0, 1, \dots$  vor und nutzt die wesentliche Eigenschaft des Optimalitätsprinzips aus: das Optimierungsproblem auf dem gesamten Zeithorizont wird in eine Folge von “kleinen” Optimierungsproblemen in den einzelnen Zeitpunkten des Horizonts zerlegt und ermöglicht so die Berechnung der optimalen Wertefunktion  $V_{T+1}$  aus  $V_T$  (Details: Übungsaufgabe).  $\square$

### 4.1.2 Charakterisierung Optimaler Kontrollprozesse

Das Optimalitätsprinzip kann aber noch zu einem weiteren wichtigen Zweck verwendet werden, nämlich zur Berechnung der optimalen Kontrollprozesse. Wie diese aussehen und insbesondere welche Struktur sie haben, klärt der folgende Satz.

**Satz 4.4** Angenommen, das Supremum im Optimalitätsprinzip (4.1) ist für alle  $\tau = 1, \dots, T$  und alle  $x \in \mathbb{R}^n$  ein Maximum. Dann existieren Funktionen  $F_\tau : \mathbb{R}^n \rightarrow U$ ,  $\tau = 0, \dots, T-1$ , so dass

$$\sup_{u \in U} E \{ \ell(x, u) + \beta V_\tau(f(x, u, Z_0)) \} = E \{ \ell(x, F_\tau(x)) + \beta V_\tau(f(x, F_\tau(x), Z_0)) \}$$

gilt. Diese definieren für jeden Anfangswert  $x_0$  induktiv eine optimale Lösung  $X_t$

$$X_0 = x_0, X_{t+1} = f(X_t, F_{T-t-1}(X_t), Z_t),$$

d.h. für den zugehörigem optimale Kontrollprozess  $\mathbf{u}^T = (u_t^T, t = 0, \dots, T-1)$ ,  $u_t^T = F_{T-t-1}(X_t)$  gilt

$$J_T(x_0, \mathbf{u}^T) = V_T(x_0).$$

**Beweis:** Die Existenz der Funktionen folgt sofort aus der angenommenen Existenz der Maxima und dass der Kontrollprozess  $\mathbf{u}_t^T$  tatsächlich die angegebene Lösung erzeugt folgt leicht mittels (2.1). Zu zeigen bleibt also die Hauptaussage des Satzes, nämlich dass  $X_t$  und  $\mathbf{u}^T$  optimal sind, dass also

$$J_T(x_0, \mathbf{u}^T) = V_T(x_0) \tag{4.6}$$

gilt.

Wenn wir die Abhängigkeit des Kontrollprozesses vom Anfangswert ausführlicher mit  $\mathbf{u}^{T, x_0}$  notieren, so folgt aus der Konstruktion dieses Prozesses

$$\mathbf{u}^{T, x_0} = \mathbf{u}^{T, x_0} \&_1 \mathbf{u}^{T-1, X_1}.$$

Die ‘‘Endstücke’’ dieses Kontrollprozesses sind also wieder von der gleichen Form, allerdings für  $T-1$ . Dies verwenden wir nun, um die Optimalität per Induktion über  $T$  zu beweisen.

Für  $T=0$  ist nichts zu zeigen, da für das zugehörige Funktional  $J_0(x_0, \mathbf{u}) = L(x_0)$  jeder Kontrollprozess optimal ist. Für  $T-1 \rightarrow T$  gilt

$$\begin{aligned} J_T(x, \mathbf{u}^{T, x_0}) &= E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t, u_t^{T, x_0}) + \beta^T L(X_T) \right\} \\ &= E \left\{ \ell(x, u_0^{T, x_0}) + \sum_{t=1}^{T-1} \beta^t \ell(X_t, u_t^{T, x_0}) + \beta^T L(X_T) \right\} \\ &= \ell(x, F_{T-1}(x_0)) + E \left\{ E_1 \left( \sum_{t=1}^{T-1} \beta^t \ell(X_t, u_t^T) + \beta^T L(X_T) \right) \right\} \end{aligned} \tag{4.7}$$

Genau wie im Beweis von “ $\geq$ ” von Satz 4.1 folgt nun

$$E_1 \left( \sum_{t=1}^{T-1} \beta^t \ell(X_t, u_t^{T,x_0}) + \beta^T L(X_T) \right) (\omega) = \beta J_{T-1}(X_1(\omega), \mathbf{u}^{T-1, X_1(\omega)}),$$

also insgesamt

$$J_T(x_0, \mathbf{u}^{T,x_0}) = E \{ \ell(x_0, F_{T-1}(x_0)) + \beta J_{T-1}(X_1, \mathbf{u}^{T-1, X_1}) \}. \quad (4.8)$$

Mit der Induktionsannahme erhalten wir

$$J_{T-1}(X_1, \mathbf{u}^{T-1, X_1}) = V_{T-1}(X_1) = V_{T-1}(f(x_0, F_{T-1}(x_0), Z_0)).$$

Setzen wir dies in (4.8) ein und verwenden die Maximierungseigenschaft von  $F_T$  sowie das Optimalitätsprinzip, so erhalten wir

$$\begin{aligned} J_T(x_0, \mathbf{u}^T) &= E \{ \ell(x, F_{T-1}(x_0)) + \beta V_{T-1}(f(x_0, F_{T-1}(x_0), Z_0)) \} \\ &= \sup_{u \in U} E \{ \ell(x, u) + \beta V_{T-1}(f(x_0, u, Z_0)) \} = V_T(x_0), \end{aligned}$$

und damit die Behauptung (4.6).  $\square$

**Bemerkung 4.5** (i) Dieser Satz zeigt insbesondere eine wichtige Folgerung aus dem Optimalitätsprinzip: Endstücke optimaler Kontrollen bzw. Lösungen sind selbst wieder optimale Kontrollen bzw. Lösungen.

(ii) Um die optimale Kontrolle zu implementieren bzw. in der Praxis anzuwenden, ist die explizite Konstruktion des optimalen Kontrollprozesses  $\mathbf{u}^T$  nicht nötig, da man die optimale Lösung  $X_t$  ganz einfach dadurch erhält, dass man in jedem Zeitpunkt  $t$  den Kontrollwert für den nächsten Schritt als  $u_t = F_{T-t}(X_t)$  wählt. Da wir ja angenommen haben, dass wir  $X_t$  zum Zeitpunkt  $t$  kennen, ist dies im kontrolltheoretischen Sinne gerade eine klassische (zeitabhängige) Feedback- bzw. Rückkopplungsstrategie.

(iii) Der Satz zeigt auch, dass die optimalen Kontrollprozesse eine echte Teilmenge der zulässigen Kontrollprozesse sind. Während diese nach Korollar 2.5 gerade durch eine Funktion  $u_t = h_t(X_0, \dots, X_t)$  gegeben sind, also von der aktuellen Zeit und der *gesamten Vergangenheit* des gesteuerten Prozesses abhängen dürfen, hängt der optimale Kontrollprozess wegen  $u_t = F_{T-t}(X_t)$  nur von der Zeit und dem *aktuellen* Zustand  $X_t$  ab. Beachte hierbei, dass die Konstruktion der Feedback-Abbildungen  $F_t$  nicht von der Horizontlänge  $T$  abhängt.

(iv) Auch dieser Satz gilt unter Beschränkungen, wenn diese bei der Bestimmung von  $F_t$  berücksichtigt werden.  $\square$

**Beispiel 4.6** Wir können nun die optimalen Kontrollen für das Beispiel 3.2 berechnen. Tatsächlich haben wir  $F_t$  implizit bereits in Beispiel 4.3 ausgerechnet, nämlich indem wir die maximierenden  $u$ -Werte berechnet haben. Es gilt also  $F_t(x) = x_2$ , falls  $\mu > 1 + r$ ,  $F_t(x) = -x_1$  falls  $\mu < 1 + r$  und  $F_t(x) \in [-x_1, x_2]$  beliebig, falls  $\mu = 1 + r$ .

Wir wollen uns im Folgenden den (realistischeren) Fall  $\mu > 1 + r$  etwas genauer ansehen, d.h. der Fall in dem der Aktienkurs im Mittel schneller steigt als die risikolose Anlage. Die optimale Anlagestrategie ist in diesem Fall denkbar einfach: Da die Aktie im Mittel ein höheres Wachstum als die festverzinsliche Anleihe verspricht, wird das gesamte Kapital in Aktien angelegt. Dies ist sicherlich eine ziemlich naive Strategie, denn sie berücksichtigt in keiner Weise das Risiko, mit dem das Wachstum der Aktie behaftet ist. Zwar wird man mit dieser Strategie "im Mittel" immer den besten Ertrag erwirtschaften, im Einzelfall kann man aber auch deutliche Verluste erzielen (und wer hat schon so viel Geld und Zeit, dass er/sie die Anlage so oft wiederholen kann, bis sich der tatsächliche Gewinn gemäß dem Gesetz der großen Zahlen dem Mittelwert annähert?). Dies verdeutlicht auch die numerische Simulation der optimalen Strategie. Die folgende Tabelle zeigt die Ergebnisse der Strategie für  $T = 20$ ,  $r = 0.05$ ,  $p = 1/2$ ,  $\alpha_u = 1.4$ ,  $\alpha_d = 0.8$ . In diesem Fall erhalten wir  $\mu = \frac{1}{2}\alpha_u + \frac{1}{2}\alpha_d = 1.1 > 1.05 = r$ , weswegen die optimale Anlagestrategie  $F_t(x) = x_2$  lautet, wodurch sich die Iterationsvorschrift der optimal gesteuerten Differenzgleichung für  $t = 0$  zu

$$f((x_1, x_2), F_{T-t}, z) = \begin{pmatrix} z(x_1 + x_2) \\ 0 \end{pmatrix}$$

ergibt, woraus insbesondere  $X_t^2 = 0$  für alle  $t \geq 1$  folgt. Simulieren wir dies für  $x_{0,1} + x_{0,2} = 1$  zwanzig mal, so ergeben sich die Werte

9.511660	9.511660	0.331152	29.129459	9.511660
3.105848	0.189230	3.105848	0.579517	3.105848
16.645405	1.014155	3.105848	0.331152	5.435234
3.105848	1.774770	0.579517	5.435234	9.511660

Zum Vergleich: der Mittelwert liegt bei 6.7275 und mit der festverzinslichen Anleihe würde man zur Zeit  $T = 20$  den Betrag 2.653298 erhalten. Wie man sieht, ist die Schwankungsbreite recht groß — eine Tatsache, der in diesem Optimierungsproblem aber nicht Rechnung getragen wird.

In der Praxis wird man nun sicherlich zwischen möglichem Gewinn und Verlustrisiko abwägen wollen. Um dies in das Optimierungsproblem einzubeziehen, bedient man sich in der Ökonomie eines einfachen Tricks: Man ändert  $L$  so ab, dass Gewinne weniger stark belohnt und Verluste stärker bestraft werden. Formal wählt man dazu eine End-Ertragsfunktion  $L$ , die konkav in  $x_1 + x_2$  ist, z.B. die Funktion  $L(x) = \ln(x_1 + x_2)$  und nehmen an, dass  $\alpha_d < (1 + r)$  gilt (ansonsten würde man mit der Aktie *sicher* einen höheren Gewinn als mit der risikolosen Anlage machen und damit die Aktie bevorzugen, egal wie stark die Aktie schwankt). Unter der Kontrollbeschränkung  $U(x) = [-x_1, x_2]$  rechnet man damit nach (Übungsaufgabe), dass die optimale Feedback-Kontrollstrategie für dieses Problem gegeben ist durch

$$F_t(x) = \begin{cases} (x_1 + x_2)v - x_1, & v \in [0, 1] \\ x_2, & v > 1 \end{cases}$$

mit

$$v = \frac{1}{2} \frac{R(\alpha_u + \alpha_d - 2R)}{(\alpha_u - R)(R - \alpha_d)} = \frac{R(\mu - R)}{\sigma^2 - (\mu - R)^2},$$

wobei  $R = 1 + r$  ist und  $\sigma^2 = \text{Var}(Z) = (\alpha_u^2 + \alpha_d^2)/4 - \alpha_u\alpha_d/2$  die Varianz der Zufallsvariablen  $Z_t$  bezeichnet. Diese Kontrollstrategie schichtet das Portfolio in jedem Zeitpunkt gerade so um, dass für das Verhältnis  $x_1 + u = v(x_1 + x_2)$  und  $x_2 - u = (1 - v)(x_1 + x_2)$  gilt, d.h., die Werte  $x_1 + u$  und  $x_2 - u$  stehen immer in einem konstanten Verhältnis zueinander. Beachte, dass aus unseren Bedingungen  $\mu > R$  und  $R > \alpha_d$  sofort  $v \geq 0$  folgt.

Die optimale Kontrollstrategie hängt nun also nicht mehr nur von Erwartungswert  $\nu$  sondern auch von der Varianz  $\sigma$  der Aktie ab, was deutlich realistischer ist. Insbesondere ist  $v$  für die zulässigen Werte von  $\alpha_u$ ,  $\alpha_d$  und  $R$  streng monoton fallend in  $\sigma^2$ , d.h. der Anteil der Aktie im Portfolio wird um so geringer, je stärker diese schwankt.

Die zugehörige optimale Wertefunktion lautet

$$V_T(x) = \ln(x_1 + x_2) + T \begin{cases} \ln\left(\frac{\sigma R}{\sqrt{\sigma^2 - (\mu - R)^2}}\right), & v \in [0, 1] \\ \ln\sqrt{\mu^2 - \sigma^2}, & v > 1 \end{cases}$$

und gibt uns insbesondere den erwarteten Ertrag  $E(\ln(X_T^1 + X_T^2))$  des Portfolios gemessen mittels der Ertragsfunktion  $L$ . Für  $\mu \searrow R$  erhält man gerade  $V_T(x) \rightarrow \ln(x_1 + x_2) + \ln(R)$ , was gerade dem Ertrag der risikolosen Anleihe entspricht.

Zum Verifizieren der optimalen Strategie mittels des Optimalitätsprinzips ist die Kenntnis von  $V_T$  völlig ausreichend. Zur Analyse der optimalen Steuerung ist aber auch der erwartete "echte" Endbetrag  $B_T(X) := E(X_T^1 + X_T^2)$  für einen gegebenen Anfangswert interessant. Wir berechnen man diesen? Die naheliegende (und für das angegebene  $V_T$  leicht zu berechnende) Formel  $E_T(X_T) = \exp(V_T(X_T))$  stimmt leider i.A. *nicht*, da sich die Anwendung der Exponentialfunktion  $\exp$  und des Erwartungswertes  $E$  nicht vertauschen lassen!

Um den echten Endbetrag zu errechnen, können wir ausnutzen, dass sich  $B_T$  gemäß der "Optimalitätsprinzip-ähnlichen" Iteration

$$B_0(x) = x_1 + x_2, \quad B_T(x) = E(B_{t-1}(f(x, F_T(x), Z_0)))$$

berechnen lässt. Wir erhalten damit

$$B_T(x) = (x_1 + x_2)\beta^T$$

mit

$$\beta = \begin{cases} \frac{\sigma^2}{\sigma^2 - (\mu - R)^2}, & v \in [0, 1] \\ \mu, & v > 1 \end{cases}$$

Durch Simulation können wir diese Anlagestrategie mit derjenigen für  $L(x) = x_1 + x_2$  vergleichen. Wir berechnen dazu für Anlagedauer  $T = 20$  mit jeweils 100000 Simulationen für die oben angegebenen Daten den erwarteten Wert  $E(X_T^1 + X_T^2)$  des Portfolios mit Anfangswert  $x_1 + x_2 = 1$ , sowie die Wahrscheinlichkeiten, dass dieser um 10% bzw. 50% unterhalb des Werts  $R^T = 2.653298$  der risikolosen Anlage liegt.

	$L(x) = x_1 + x_2$	$L(x) = \ln(x_1 + x_2)$
$E(X_T^1 + X_T^2)$	6.751541	4.650349
$P(X_T < 0.9R^T)$	0.413100	0.249970
$P(X_T < 0.5R^T)$	0.252890	0.130590

Wie man sieht, ist zwar der erwartete Wert im zweiten Fall um gut 30% niedriger, allerdings halbiert sich dabei in etwa die Gefahr eines großen Verlustes.  $\square$

## 4.2 Probleme auf unendlichem Horizont

### 4.2.1 Formulierung und Beweis des Prinzips

Für die Probleme auf unendlicher Zeit formulieren und beweisen wir das Prinzip in etwas allgemeinerer Form und zunächst ohne Eindeutigkeitsaussage.

**Satz 4.7** Die optimale Wertefunktion  $V_\infty$  erfüllt für alle  $x_0 \in \mathbb{R}^n$  und alle  $T \in \mathbb{N}$  die Gleichung

$$V_\infty(x_0) = \sup_{\mathbf{u} \in \mathcal{U}(x_0)} E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t, u_t) + \beta^T V_\infty(X_T) \right\}, \quad (4.9)$$

mit  $X_t = X_t(x_0, \mathbf{u})$ . Insbesondere gilt für  $T = 1$

$$V_\infty(x_0) = \sup_{u \in U} E \{ \ell(x_0, u) + \beta V_\infty(f(x_0, u, Z_0)) \}. \quad (4.10)$$

**Beweis:** Vorbemerkung: die im Folgenden benutzten Gleichungen der bedingten Erwartungswerte  $E_1$  folgen völlig analog zum Beweis von Satz 4.1 und werden hier deswegen nicht näher ausgeführt.

Wir beweisen zunächst (4.10).

“ $\leq$ ”: Sei  $x_0 \in \mathbb{R}^n$  und  $\mathbf{u} \in \mathcal{U}(x_0)$  beliebig. Dann gilt für  $X_t = X_t(x_0, \mathbf{u})$

$$\begin{aligned} J_\infty(x_0, \mathbf{u}) &= E \left\{ \sum_{t=0}^{\infty} \beta^t \ell(X_t, u_t) \right\} \\ &= E \left\{ \ell(x_0, u_0) + \sum_{t=1}^{\infty} \beta^t \ell(X_t, u_t) \right\} \\ &= \ell(x_0, u_0) + \beta E \left\{ E_1 \left( \sum_{t=0}^{\infty} \beta^t \ell(X_{t+1}, u_{t+1}) \right) \right\} \\ &= \ell(x_0, u_0) + \beta E \{ J_\infty(X_1, \tilde{u}_t) \} \\ &\leq E \{ \ell(x_0, u_0) + \beta V_\infty(X_1) \} \\ &\leq \sup_{u \in U} E \{ \ell(x_0, u) + \beta V_\infty(f(x_0, u, Z_0)) \}. \end{aligned}$$

Da diese Ungleichung nun für alle zulässigen  $\mathbf{u}$  gilt, gilt sie auch für

$$V(x_0) = \sup_{\mathbf{u} \in \mathcal{U}(x)} J_\tau(x_0, \mathbf{u}),$$

womit “ $\leq$ ” folgt.

“ $\geq$ ”: Sei  $\varepsilon > 0$ . Zu jedem  $x_0 \in \mathbb{R}^n$  wählen wir einen Kontrollprozess  $\mathbf{u}^{x_0} \in \mathcal{U}(x_0)$ , für den

$$J_\infty(x_0, \mathbf{u}^{x_0}) \geq V_\infty(x_0) - \varepsilon$$

gilt. Weiterhin wählen wir einen Kontrollwert  $u^* \in U$  mit

$$E\{\ell(x_0, u^*) + \beta V_\infty(f(x_0, u^*, Z_0))\} \geq \sup_{u \in U} E\{\ell(x_0, u) + \beta V_\infty(f(x_0, u, Z_0))\} - \varepsilon,$$

einen Kontrollprozess  $\mathbf{u} \in \mathcal{U}(x_0)$  mit  $u_0 = u^*$  und setzen  $\tilde{\mathbf{u}} = \mathbf{u} \&_1 \mathbf{u}^x$  gemäß Lemma 2.6. Die zugehörige Lösung sei mit  $\tilde{X}_t = X_t(x_0, \tilde{\mathbf{u}})$  bezeichnet. Dann gilt

$$\begin{aligned} V_\infty(x_0) &= \sup_{\mathbf{u} \in \mathcal{U}(x_0)} J_\infty(x_0, \tilde{\mathbf{u}}) \\ &= \sup_{\mathbf{u} \in \mathcal{U}(x_0)} E \left\{ \sum_{t=0}^{\infty} \ell(X_t, u_t) \right\} \\ &= \sup_{\mathbf{u} \in \mathcal{U}(x_0)} E \left\{ \ell(x_0, u_0) + \sum_{t=1}^{\infty} \ell(X_t, u_t) \right\} \\ &\geq E \left\{ \ell(x_0, \tilde{u}_0) + \beta \sum_{t=0}^{\infty} \ell(\tilde{X}_{t+1}, \tilde{u}_{t+1}) \right\} \\ &= \ell(x_0, \tilde{u}_0) + \beta E \left\{ E_1 \left( \sum_{t=0}^{\infty} \ell(\tilde{X}_{t+1}, \tilde{u}_t^{X_1}) \right) \right\} \\ &\geq \sup_{u \in U} E \{ \ell(x_0, u) + \beta V_\infty(f(x_0, u, Z)) \} - 2\varepsilon \end{aligned}$$

wobei wir im letzten Schritt die Wahl von  $\mathbf{u}$  ausgenutzt haben. Da  $\varepsilon > 0$  beliebig war, folgt

$$V_\infty(x_0) \geq \sup_{u \in U} E \{ \ell(x_0, u) + \beta V_\infty(f(x_0, u, Z)) \},$$

also gerade die gewünschte Ungleichung.

Die allgemeinere Gleichung (4.9) folgt nun per Induktion über  $T$  aus (4.10). Für  $T = 1$  stimmen die beiden Gleichungen gerade überein. Für  $T \rightarrow T + 1$  erhalten wir

$$\begin{aligned} V_\infty(x_0) &= \sup_{\mathbf{u} \in \mathcal{U}(x_0)} E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t, u_t) + \beta^T V_\infty(X_T) \right\} \\ &= \sup_{\mathbf{u} \in \mathcal{U}(x_0)} E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t, u_t) + \beta^T \sup_{\tilde{u} \in U} E_T \{ \ell(X_T, \tilde{u}) + \beta V_\infty(f(X_T, \tilde{u}, Z_T)) \} \right\} \\ &= \sup_{\mathbf{u} \in \mathcal{U}(x_0)} E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t, u_t) + \beta^T (\ell(X_T, u_T) + \beta V_\infty(f(X_T, u_T, Z_T))) \right\} \\ &= \sup_{\mathbf{u} \in \mathcal{U}(x_0)} E \left\{ \sum_{t=0}^T \beta^t \ell(X_t, u_t) + \beta^{T+1} V_\infty(X_{T+1}) \right\} \end{aligned}$$

□

Um die Eindeutigkeitsaussage zu ergänzen, benötigen wir eine weitere Bedingung, die ein “schönes” Grenzverhalten der rechten Seite von (4.9) für  $T \rightarrow \infty$  sicher stellt.

**Definition 4.8** Wir sagen, dass eine Funktion  $W : \mathbb{R}^n \rightarrow \mathbb{R}$  die *Transversalitätsbedingung* erfüllt, falls für jedes  $x_0 \in \mathbb{R}^n$  und jede Folge von Kontrollprozessen  $\mathbf{u}^T \in \mathcal{U}(x_0)$  mit Lösungen  $X_t^T = X_t(x_0, \mathbf{u}^T)$  die Implikationen

$$\limsup_{T \rightarrow \infty} E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t^T, u_t^T) + \beta^T W(X_T^T) \right\} < +\infty \Rightarrow \liminf_{T \rightarrow \infty} E \{ \beta^T W(X_T^T) \} \geq 0$$

und

$$\liminf_{T \rightarrow \infty} E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t^T, u_t^T) + \beta^T W(X_T^T) \right\} > -\infty \Rightarrow \limsup_{T \rightarrow \infty} E \{ \beta^T W(X_T^T) \} \leq 0$$

gelten. □

Beachte, dass diese Gleichung — sogar für alle  $\mathbf{u}^T$  — stets erfüllt ist, wenn  $W$  beschränkt und  $\beta < 1$  ist, da dann immer  $\beta^T W(X_T^T) \rightarrow 0$  gilt. Die optimale Wertefunktion  $V_\infty$  wiederum ist für  $\beta < 1$  beschränkt, falls  $\ell$  beschränkt ist.

**Satz 4.9** Die optimale Wertefunktion  $V_\infty$  erfülle die Transversalitätsbedingung. Dann stimmt jede weitere Funktion  $W$ , die das Optimalitätsprinzip (4.9) und die Transversalitätsbedingung erfüllt, mit  $V_\infty$  überein.

**Beweis:** Wir zeigen, dass für je zwei Funktionen  $W_1, W_2 : \mathbb{R}^n \rightarrow \mathbb{R}$ , die (4.9) und die Transversalitätsbedingung erfüllen, die Ungleichung

$$W_1(x) \leq W_2(x)$$

für alle  $x \in \mathbb{R}^n$  gilt. Hieraus folgt die Behauptung, indem wir die Ungleichung sowohl auf  $W_1 = V_\infty, W_2 = W$  als auch auf  $W_1 = W, W_2 = V_\infty$  anwenden.

Es sei  $\varepsilon > 0$  und  $x_0 \in \mathbb{R}^n$  gegeben. Da  $W_1$  das Optimalitätsprinzip (4.9) erfüllt, existieren Kontrollfunktionen  $u^T$  mit zugehörigen Lösungen  $X_t^T$ , so dass die Ungleichung

$$W_1(x_0) \leq E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t^T, u_t^T) + \beta^T W_1(X_T^T) \right\} + \varepsilon$$

gilt. Hieraus folgt insbesondere

$$\liminf_{T \rightarrow \infty} E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t^T, u_t^T) + \beta^T W_1(X_T^T) \right\} \geq W_1(x_0) - \varepsilon > -\infty$$

und daher mit der Transversalitätsbedingung

$$\limsup_{T \rightarrow \infty} E \{ \beta^T W_1(X_T^T) \} \leq 0.$$

Da  $W_2$  ebenfalls das Optimalitätsprinzip erfüllt, folgt

$$\begin{aligned} W_2(x_0) &= \sup_{\mathbf{u} \in \mathcal{U}(x_0)} E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t, u_t) + \beta^T W_2(X_T) \right\} \\ &\geq E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t^T, u_t^T) + \beta^T W_2(X_T^T) \right\}. \end{aligned}$$

Damit folgt

$$\limsup_{T \rightarrow \infty} E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t^T, u_t^T) + \beta^T W_2(X_T^T) \right\} \leq W_2(x_0) < +\infty$$

und folglich mit der Transversalitätsbedingung

$$\liminf_{T \rightarrow \infty} E \{ \beta^T W_2(X_T^T) \} \geq 0.$$

Mit diesen Ungleichungen erhalten wir für alle  $T \in \mathbb{N}$

$$\begin{aligned} &W_1(x_0) - W_2(x_0) \\ &\leq E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t^T, u_t^T) + \beta^T W_1(X_T^T) \right\} - E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t^T, u_t^T) + \beta^T W_2(X_T^T) \right\} + \varepsilon \\ &= E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t^T, u_t^T) + \beta^T W_1(X_T^T) - \sum_{t=0}^{T-1} \beta^t \ell(X_t^T, u_t^T) + \beta^T W_2(X_T^T) \right\} + \varepsilon \\ &= E \{ \beta^T W_1(X_T^T) - \beta^T W_2(X_T^T) \} + \varepsilon = E \{ \beta^T W_1(X_T^T) \} - E \{ \beta^T W_2(X_T^T) \} + \varepsilon \end{aligned}$$

Also folgt

$$\begin{aligned} W_1(x_0) - W_2(x_0) &\leq \limsup_{T \rightarrow \infty} (E \{ \beta^T W_1(X_T^T) \} - E \{ \beta^T W_2(X_T^T) \}) + \varepsilon \\ &\leq \limsup_{T \rightarrow \infty} E \{ \beta^T W_1(X_T^T) \} - \liminf_{T \rightarrow \infty} E \{ \beta^T W_2(X_T^T) \} + \varepsilon \leq \varepsilon. \end{aligned}$$

Da  $\varepsilon > 0$  beliebig war, folgt hieraus die zu zeigende Ungleichung  $W_1(x_0) \leq W_2(x_0)$ .  $\square$

**Bemerkung 4.10** Analog zu Satz 4.1 gelten auch diese Sätze unter Zustandsbeschränkungen und (zustandsabhängigen) Kontrollbeschränkungen, wenn man das Supremum in (4.10) nur über solche  $u$  nimmt, die die gegebenen Beschränkungen nicht verletzen.  $\square$

## 4.2.2 Charakterisierung Optimaler Kontrollprozesse

Genau wie im Fall des endlichen Horizontes kann das Optimalitätsprinzip nun dazu verwendet werden, die optimalen Kontrollprozesse zu charakterisieren. Tatsächlich wird die Aussage hier noch etwas schöner, da das resultierende Feedback hier nicht mehr von  $t$  abhängt. Wiederum benötigen wir die Transversalitätsbedingung, die z.B. für  $\beta < 1$  und beschränktes  $\ell$  immer erfüllt ist.

**Satz 4.11** Das Supremum im Optimalitätsprinzip (4.10) sei für alle  $x \in \mathbb{R}^n$  ein Maximum. Dann existiert eine Funktion  $F_\infty : \mathbb{R}^n \rightarrow U$ , so dass

$$\sup_{u \in U} E \{ \ell(x, u) + \beta V_\infty(f(x, u, Z_0)) \} = E \{ \ell(x, F_\infty(x)) + \beta V_\infty(f(x, F_\infty(x), Z_0)) \}$$

gilt. Diese definiert für jeden Anfangswert  $x_0$  induktiv eine optimale Lösung  $X_t$

$$X_0 = x_0, X_{t+1} = f(X_t, F_\infty(X_t), Z_t)$$

d.h. für den zugehörigem optimale Kontrollprozess  $\mathbf{u}^\infty = (u_t^\infty, t = 0, 1, \dots)$ ,  $u_t^\infty = F_\infty(X_t)$  gilt

$$J_\infty(x_0, \mathbf{u}^\infty) = V_\infty(x_0),$$

vorausgesetzt, dass  $V_\infty$  die zweite Transversalitätsbedingung und  $W(x_0) := J_\infty(x_0, \mathbf{u}^\infty)$  die erste Transversalitätsbedingung erfüllt.

**Beweis:** Die Existenz der Funktionen folgt sofort aus der angenommenen Existenz der Maxima und dass der Kontrollprozess  $\mathbf{u}^\infty$  tatsächlich die angegebenen Lösung erzeugt folgt leicht mittels (2.1). Zu zeigen bleibt also die Hauptaussage des Satzes, nämlich dass  $X_t$  und  $\mathbf{u}^\infty$  optimal sind, dass also

$$J_\infty(x_0, \mathbf{u}^{\infty, x_0}) = V_\infty(x_0) \quad (4.11)$$

gilt, wobei wir analog zum Beweis von Satz 4.4 mit  $\mathbf{u}^{\infty, x_0}$  den optimalen Kontrollprozess zum Anfangswert  $x_0$  bezeichnen.

Wegen der Optimalität von  $V_\infty$  ist die Ungleichung  $J_\infty(x_0, \mathbf{u}^{\infty, x_0}) \leq V_\infty(x_0)$  offensichtlich. Zum Beweis von “ $\geq$ ” verwenden wir, dass wir analog zur Gleichung (4.8) im Beweis von Satz 4.4

$$J_\infty(x_0, \mathbf{u}^{\infty, x_0}) = E \{ \ell(x_0, F_\infty(x_0)) + \beta J_\infty(X_1, \mathbf{u}^{\infty, X_1}) \}.$$

mit  $X_1 = f(x, F_\infty(x), Z_0)$  erhalten. Wie im Beweis von Satz 4.7 erhalten wir daraus per Induktion

$$J_\infty(x_0, \mathbf{u}^{\infty, x_0}) = E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t, F_\infty(X_t)) + \beta^T J_\infty(X_T, \mathbf{u}^{\infty, X_T}) \right\}.$$

Hieraus folgt aus der Definition von  $F_\infty$

$$\begin{aligned} & V_\infty(x_0) - J_\infty(x_0, \mathbf{u}^{\infty, x_0}) \\ &= E \left\{ \sum_{t=0}^{T-1} \beta^t \ell(X_t, F_\infty(X_t)) + \beta^T V_\infty(X_T) - \sum_{t=0}^{T-1} \beta^t \ell(X_t, F_\infty(X_t)) + \beta^T J_\infty(X_T, \mathbf{u}^{\infty, X_T}) \right\} \\ &= E \{ \beta^T V_\infty(X_T) \} - E \{ \beta^T J_\infty(X_T, \mathbf{u}^{\infty, X_T}) \} \end{aligned}$$

Nun können wir wie im Beweis von Satz 4.9 die Transversalitätsbedingung ausnutzen, um

$$V_\infty(x_0) - J_\infty(x_0, \mathbf{u}^{\infty, x_0}) \leq \limsup_{T \rightarrow \infty} E \{ \beta^T V_\infty(X_T) \} - \liminf_{T \rightarrow \infty} E \{ \beta^T J_\infty(X_T, \mathbf{u}^{\infty, X_T}) \} \leq 0$$

und damit die gewünschte Ungleichung zu zeigen.  $\square$

**Bemerkung 4.12** (i) Der Satz zeigt, dass die optimalen Kontrollprozesse eine noch kleinere Teilmenge der zulässigen Kontrollprozesse sind als im Fall des endlichen Horizonts. Während die optimalen Prozesse dort durch  $u_t = F_{T-t}(X_t)$  gegeben sind, hängen sie hier wegen  $u_t = F_\infty(X_t)$  nicht direkt von der aktuellen Zeit ab.

Beachte aber, dass die Zeitabhängigkeit wieder auftaucht, wenn wir  $t$ -Abhängigkeit in der Dynamik  $f$  oder in den laufenden Kosten  $\ell$  zulassen.

(ii) Auch dieser Satz gilt unter Beschränkungen, wenn diese bei der Bestimmung von  $F_\infty$  berücksichtigt werden. □

Im Prinzip haben wir hiermit nun alle Hilfsmittel, um das Wagenproblem aus Beispiel 3.3 analytisch lösen zu können. Wir werden dies aber noch etwas aufschieben, da dieses Problem einige strukturelle Eigenschaften besitzt, welche wir im übernächsten Kapitel zunächst abstrakt betrachten werden. Mit den damit erzielbaren Vereinfachungen wird die Lösung dann deutlich einfacher werden. Zunächst aber widmen wir uns im nächsten Kapitel einer Methode zur numerischen Lösung der Probleme.

## Kapitel 5

# Numerische Dynamische Programmierung

In diesem Kapitel wird erläutert, wie man das Prinzip der Dynamischen Programmierung numerisch nutzen kann. Wir werden hier einen Ansatz erläutern, der einfach umzusetzen und auch auf Probleme mit komplexer nichtlinearer Dynamik anwendbar ist. Der “Preis” für diese Flexibilität ist ein numerischer Aufwand, der mit der Dimension  $n$  des Zustandsraums schnell anwächst, so dass der Algorithmus nur für relativ niedrige Dimension effizient ist. Weitere Methoden, die spezielle Strukturen von  $f$  voraussetzen, dafür aber auch in höheren Raumdimensionen funktionieren, werden wir in den folgenden Kapiteln skizzieren und im an die Vorlesung anschließenden Seminar genauer behandeln.

Wir entwickeln die Methode für Probleme mit endlichem Zeithorizont und besprechen am Ende kurz, wie die Methode für einen unendlichen Horizont modifiziert werden kann.

### 5.1 Ansatz

Die Idee des numerischen Ansatzes ist wie folgt: Wir verwenden einen endlich dimensionalen Funktionenraum  $\mathcal{W}$ , dessen Elemente wir im Rechner speichern können und wenden das Optimalitätsprinzip auf diesem Funktionenraum an, um so Approximationen  $\tilde{V}_\tau \in \mathcal{W}$  der optimalen Wertefunktionen  $V_\tau$  zu berechnen.

Dazu bezeichnen wir mit  $F(A, B) := \{W : A \rightarrow B\}$  die Menge aller Funktionen von einer Menge  $A$  in eine Menge  $B$  und führen einen (zunächst abstrakten) Projektionsoperator  $\pi : F(A, B) \rightarrow \mathcal{W}$  ein, der zu einer beliebigen Funktion  $W$  die numerische Approximation

$$\tilde{W} = \pi W$$

in  $\mathcal{W}$  berechnet. Die algorithmische Umsetzung dieses abstrakten Konzepts behandeln wir in Abschnitt 5.3.

Für die rechte Seite des Optimalitätsprinzips (4.1) führen wir nun Operatoren  $T_u : F(\mathbb{R}^n, \mathbb{R}) \rightarrow$

$F(\mathbb{R}^n, \mathbb{R})$  und  $T : F(\mathbb{R}^n, \mathbb{R}) \rightarrow F(\mathbb{R}^n, \mathbb{R})$  ein, die durch

$$\begin{aligned} T_u(W)(x) &:= E \{ \ell(x, u) + \beta W(f(x, u, Z)) \} \\ T(W)(x) &:= \sup_{u \in U(x)} T_u(W)(x) \end{aligned} \tag{5.1}$$

definiert sind. Die Zufallsvariable  $Z$  ist hierbei ein beliebiger Repräsentant der iid Zufallsvariablen  $Z_t$ . Mit dieser Notation ergibt sich das Optimalitätsprinzip abstrakt zu

$$V_\tau = T(V_{\tau-1}).$$

Mit dem iterativen Algorithmus

$$\tilde{V}_0 := L, \quad \tilde{V}_\tau = \pi T(\tilde{V}_{\tau-1}), \quad \tau = 1, \dots, T \tag{5.2}$$

erhalten wir dann numerische Approximationen der optimalen Wertefunktionen. Hierbei nehmen wir an, dass  $L$  als numerisch auswertbare Funktion vorliegt, die nicht approximiert werden muss, weswegen wir  $V_0 = L$  und nicht  $V_0 = \pi L$  schreiben.

Zur Berechnung approximativer optimaler Feedbacks  $\tilde{F}_\tau$  gibt es zwei Möglichkeiten: zum einen kann man diese wie in der analytischen Berechnung in der Iteration (5.2) mitberechnen und ebenfalls als  $\pi \tilde{F}_\tau$  approximiert speichern. Numerisch besser ist es allerdings, für einen gegebenen Zustand  $x$  den optimalen Kontrollwert  $F_\tau(x)$  jeweils “online” aus der Gleichung

$$\tilde{F}_\tau(x) = \operatorname{argmax}_{u \in U(x)} T_u(\tilde{V}_\tau)(x)$$

zu berechnen. Auf diese Weise vermeidet man den Approximationsfehler bei der Projektion  $\pi \tilde{F}_\tau$  und spart zudem noch Speicherplatz. Da sich das “argmax” meist schnell berechnen lässt, ist der Rechenaufwand für die Online-Optimierung i.A. vertretbar.

Um die Berechnung der  $\tilde{V}_\tau$  gemäß (5.2) numerisch umzusetzen, müssen wir die folgenden Teilprobleme lösen, die wir hier für allgemeine Funktionen  $W$  betrachten:

- (1) Auswertung von  $T(W)$
- (2) Berechnung von  $\pi T(W)$

Zudem werden wir unsere numerische Berechnung nicht auf dem gesamten Zustandsraum  $\mathbb{R}^n$  durchführen können und müssen uns daher überlegen, wie wir das Problem auf geeignete kompakte Teilmengen einschränken können.

## 5.2 Auswertung von $T(W)$

Wir werden im nächsten Abschnitt sehen, dass es zur Berechnung von  $\pi T(W)$  ausreicht, wenn wir für gegebene Punkte  $x \in \mathbb{R}^n$  die Werte  $T(W)(x) \in \mathbb{R}$  berechnen können.

Diese Berechnung führt wiederum auf zwei Teilprobleme, zum einen die Auswertung von  $T_u(W)(x)$  und zum anderen die Maximierung über  $u$  in  $T(W)(x)$ .

Da  $\ell$ ,  $f$  und  $W$  bekannt und numerisch auswertbar sind, besteht die einzige Schwierigkeit in der Auswertung von  $T_u(W)(x)$  in der Berechnung des Erwartungswertes. Falls  $Z$  eine

kontinuierliche Zufallsvariable ist, nehmen wir an, dass wir die Dichtefunktion  $f_Z$  kennen und müssen

$$\int_{Z(\Omega)} (\ell(x, u) + \beta W(f(x, u, z))) f_Z(z) dz$$

berechnen. Dies kann durch numerische Integration geschehen. Wenn wir keine Glattheit von  $W$  voraussetzen wollen, ist die Trapezregel (Newton-Cotes Formel vom Grad  $n = 1$ ) hierfür die geeignetste Methode.

Falls  $Z$  nur endlich viele Werte  $z_1, \dots, z_m$  annimmt, was in den Übungsaufgaben der Fall sein wird, so benötigt man gar keine Integrale, sondern kann den Erwartungswert einfach durch die Summe

$$\sum_{i=1}^m (\ell(x, u) + \beta W(f(x, u, z_i))) P_Z(\{z_i\})$$

berechnen, welche bis auf Rundungsfehler exakt ausgewertet werden kann.

Zur Bestimmung des Maximums betrachten wir nur die einfachste Möglichkeit, die für geringe Genauigkeit aber durchaus brauchbare Ergebnisse liefert. Wir nehmen dazu an, dass der Kontrollwertebereich  $U(x)$ , über den maximiert wird, kompakt ist. Dann betrachten wir eine endliche Menge  $\tilde{U}(x) \subset U(x)$  mit

$$\varepsilon_u := \max_{u \in U(x)} \max_{\tilde{u} \in \tilde{U}(x)} \|\tilde{u} - u\| \quad (5.3)$$

und approximieren  $T$  durch

$$\tilde{T}(W)(x) = \max_{\tilde{u} \in \tilde{U}(x)} T_{\tilde{u}}(W)(x).$$

Hierfür gilt die folgende Fehlerabschätzung.

**Lemma 5.1** Die Abbildung  $u \mapsto T_u(W)(x)$  sei Lipschitz stetig mit Konstante  $L_u$  (dies gilt z.B., wenn  $\ell$  und  $f$  Lipschitz in  $u$  und  $W$  Lipschitz in  $x$  ist). Dann gilt

$$|\tilde{T}(W)(x) - T(W)(x)| \leq L_u \varepsilon_u.$$

**Beweis:** Wegen  $\tilde{U}(x) \subset U(x)$  gilt  $\tilde{T}(W)(x) \leq T(W)(x)$ . Da wir  $U(x)$  als kompakt und  $u \mapsto T(W)(x)$  als stetig annehmen, gilt

$$\sup_{u \in U(x)} T_u(W)(x) = \max_{u \in U(x)} T_u(W)(x),$$

also existiert ein maximierendes  $u^* \in U(x)$ . Aus der Definition von  $\varepsilon_u$  folgt die Existenz von  $\tilde{u}^*$  mit  $\|\tilde{u}^* - u^*\| \leq \varepsilon_u$  und damit

$$\begin{aligned} |\tilde{T}(W)(x) - T(W)(x)| &= T(W)(x) - \tilde{T}(W)(x) = T_{u^*}(W)(x) - \max_{\tilde{u} \in \tilde{U}(x)} T_{\tilde{u}}(W)(x) \\ &\leq T_{u^*}(W)(x) - T_{\tilde{u}^*}(W)(x) \leq L_u \|u^* - \tilde{u}^*\| \leq L_u \varepsilon_u. \end{aligned}$$

□

### 5.3 Berechnung von $\pi T(W)$

Die Berechnung von  $\pi T(W)$  hängt entscheidend von der Wahl der Approximation  $\tilde{V}_\tau$  ab, die wir als erstes beschreiben.

Die Idee der meisten Approximation beruht darauf,  $V_\tau$  durch ein Element  $\tilde{V}_\tau$  eines endlich-dimensionalen Funktionenraums zu approximieren. Dies werden wir hier genau so machen. Um die Darstellung einfach zu halten, werden wir uns hier auf den Fall  $n = 2$  einschränken; die verwendeten Techniken lassen sich aber leicht auf beliebige Dimensionen verallgemeinern und werden auch in allgemeiner  $n$ -dimensionaler Form für die Programmierung in der Übungen zur Verfügung gestellt. Ebenfalls zur Vereinfachung der Darstellung nehmen wir an, dass wir  $V_\tau$  auf einer rechteckigen Menge  $X \subset \mathbb{R}^2$  approximieren wollen.

**Definition 5.2** Sei  $X \subset \mathbb{R}^2$  gegeben durch  $X = [a_1, b_1] \times [a_2, b_2]$  mit Werten  $a_1 < b_1$  und  $a_2 < b_2$ . Ein (*regelmäßiges*) *Rechteckgitter*  $\Gamma$  auf  $X$  ist eine Menge von Rechtecken  $R_i$ ,  $i = 0, \dots, P - 1$ ,  $P = P_1 P_2$ , mit Kantenlängen  $k_1 = (b_1 - a_1)/P_1$  und  $k_2 = (b_2 - a_2)/P_2$ , so dass

$$\bigcup_{i=0}^{P-1} R_i = X \quad \text{und} \quad \text{int } R_i \cap \text{int } R_j = \emptyset \quad \text{für alle } i, j = 0, \dots, P - 1, i \neq j.$$

Mit  $E_i$ ,  $i = 0, \dots, N - 1$ ,  $N = (P_1 + 1)(P_2 + 1)$  bezeichnen wir die Eckpunkte (oder Knotenpunkte) des Gitters. Der Wert  $k = \sqrt{k_1^2 + k_2^2}$  bezeichnet den maximalen Durchmesser eines Rechtecks.  $\square$

Abbildung 5.1 zeigt ein solches Gitter.

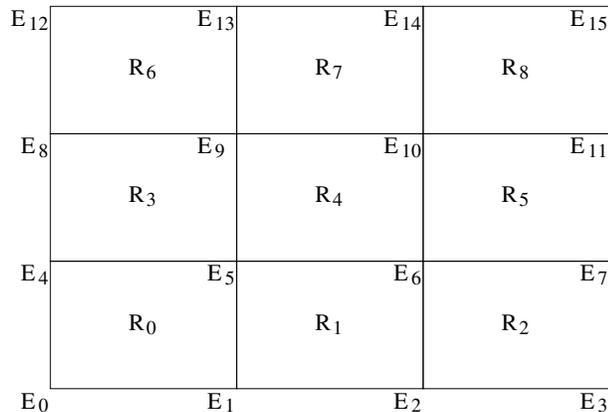


Abbildung 5.1: Beispielgitter

Wir definieren nun den Funktionenraum, den wir zur Approximation von  $V_\tau$  verwenden wollen.

**Definition 5.3** (i) Sei  $A \subset \mathbb{R}^2$ . Eine Funktion  $W : A \rightarrow \mathbb{R}$  heißt *affin bilinear*, falls es Koeffizienten  $\alpha_0, \dots, \alpha_3 \in \mathbb{R}$  gibt, so dass für alle  $x = (x_1, x_2)^T \in A$  die Identität  $W(x) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2$  gilt.

(ii) Betrachte eine rechteckförmige Menge  $X \subset \mathbb{R}^2$  mit Rechteckgitter  $\Gamma$ . Wir definieren den Raum der stetigen und stückweise affin bilinearen Funktionen auf  $X$  bezüglich  $\Gamma$  als

$$\mathcal{W} := \{W : X \rightarrow \mathbb{R} \mid W \text{ ist stetig und } W|_{R_i} \text{ ist affin bilinear für jedes } i = 0, \dots, P-1\}.$$

□

Das folgende Lemma fasst die für uns wichtigen Eigenschaften von  $\mathcal{W}$  zusammen.

**Lemma 5.4** (i) Jede Funktion  $W \in \mathcal{W}$  ist eindeutig durch ihre Werte  $W(E_i)$  in den Eckpunkten des Gitters bestimmt.

(ii) Für jedes Rechteck  $R = [c_1, d_1] \times [c_2, d_2]$  des Gitters mit den Eckpunkten

$$E_{i_0} = (c_1, c_2)^T, \quad E_{i_1} = (d_1, c_2)^T, \quad E_{i_2} = (c_1, d_2)^T, \quad E_{i_3} = (d_1, d_2)^T$$

lässt sich  $W|_{R_i}$  für  $x = (x_1, x_2)^T \in R_i$  schreiben als

$$W(x) = \sum_{j=0}^3 \mu_j(x) W(E_{i_j})$$

mit

$$\begin{aligned} \mu_0(x) &= (1 - y_1(x))(1 - y_2(x)), & \mu_1(x) &= y_1(x)(1 - y_2(x)), \\ \mu_2(x) &= (1 - y_1(x))y_2(x), & \mu_3(x) &= y_1(x)y_2(x) \end{aligned}$$

und

$$y_l(x) = \frac{x_l - c_l}{d_l - c_l} \text{ für } l = 1, 2.$$

Insbesondere gilt hierbei  $\mu_j(x) \geq 0$  für  $j = 0, \dots, 3$  und  $\sum_{j=0}^3 \mu_j(x) = 1$ .

**Beweis:** (i) Auf dem Rechteck  $R = [c_1, d_1] \times [c_2, d_2]$  folgt aus der Definition, dass die Koeffizienten  $\alpha_i$  einer affin bilinearen Funktion  $W$  die Gleichungen

$$\begin{aligned} \alpha_0 + \alpha_1 c_1 + \alpha_2 c_2 + \alpha_3 c_1 c_2 &= W(E_{i_1}) \\ \alpha_0 + \alpha_1 d_1 + \alpha_2 c_2 + \alpha_3 d_1 c_2 &= W(E_{i_2}) \\ \alpha_0 + \alpha_1 c_1 + \alpha_2 d_2 + \alpha_3 c_1 d_2 &= W(E_{i_3}) \\ \alpha_0 + \alpha_1 d_1 + \alpha_2 d_2 + \alpha_3 d_1 d_2 &= W(E_{i_4}) \end{aligned}$$

erfüllen müssen. Berechnet man die Determinante der zugehörigen Matrix (leicht mit MAPLE), so erhält man  $(d_1 - c_1)^2(d_2 - c_2)^2$ . Da die Eckpunkte paarweise verschieden sind, ist diese verschieden von Null und damit besitzt dieses Gleichungssystem also genau eine Lösung. Folglich sind die Koeffizienten  $\alpha_i$  auf jedem Rechteck und damit auch die Funktion  $W$  selbst eindeutig durch die Werte in den Eckpunkten bestimmt.

(ii) Man rechnet leicht nach, dass die angegebene Funktion tatsächlich affin bilinear auf  $R_i$  ist (es tauchen nur Terme der Form  $\alpha_0$ ,  $\alpha_1 x_1$ ,  $\alpha_2 x_2$  oder  $\alpha_3 x_1 x_2$  auf) und in den Eckpunkten des Rechtecks mit  $w$  übereinstimmt. Also folgt die Aussage aus Teil (i). Die behaupteten Eigenschaften für die  $\mu_j(x)$  sind ebenfalls leicht zu sehen, wenn man ausnutzt, dass  $y_l(x) \in [0, 1]$  gilt. □

Wir können also jede Funktion  $W \in \mathcal{W}$  mit ihren Werten  $W(E_i)$  in den Eckpunkten des Gitters identifizieren. Insbesondere ist der Funktionenraum  $\mathcal{W}$  somit ein  $N$ -dimensionaler Vektorraum über  $\mathbb{R}$ .

Damit können wir den Projektionsoperator  $\pi : F(X, \mathbb{R}) \rightarrow \mathcal{W}$  formal definieren als

$$\pi W := \widetilde{W} \in \mathcal{W} \quad \text{mit} \quad \widetilde{W}(E_i) = W(E_i) \quad \text{für alle } i = 0, \dots, N-1.$$

Diese Definition führt sofort zur Berechnung von  $\pi T(W)$ . Um die Funktion  $\widetilde{W} = \pi T(W)$  zu berechnen (und im Rechner zu speichern), genügt es  $\widetilde{W}(E_i) = \pi T(W)(E_i)$  zu berechnen. Der Iterationsschritt in (5.2) ergibt sich also in der numerisch auswertbaren Variante zu

$$\widetilde{V}_\tau(E_i) = \widetilde{T}(\widetilde{V}_{\tau-1})(E_i), \quad i = 0, \dots, N-1. \quad (5.4)$$

Den bei dieser Approximation auftretenden Diskretisierungsfehler werden wir im übernächsten Abschnitt analysieren.

## 5.4 Wahl des Berechnungsgebietes

Die im vorangegangenen Abschnitt beschriebene Approximation kann nur auf kompakten Gebieten  $X$  durchgeführt werden, was in der Praxis zu nicht immer einfach lösbaren Problemen führen kann.

Beachte, dass  $X$  gerade das Definitionsgebiet von  $\widetilde{V}_\tau$  darstellt. Um sicher zu stellen, dass  $\widetilde{V}_\tau(f(x, u, z))$  für alle möglichen Argumente von  $f$  (also  $x \in X$ ,  $u \in U(x)$  und  $z \in Z(\Omega)$ ) definiert ist, muss also

$$f(x, u, z) \in X, \quad \text{für alle } x \in X, u \in U(x), z \in Z(\Omega) \quad (5.5)$$

gelten.

Wenn  $Z(\Omega)$  unbeschränkt ist (z.B. für eine Gauß-verteilte Zufallsvariable, für die  $Z(\Omega) = \mathbb{R}$  gilt), so kann auch  $f(x, u, z)$  Werte in einer unbeschränkten Menge annehmen. In diesem Fall sind wir gezwungen,  $Z(\Omega)$  oder  $\Omega$  einzuschränken, um auf einer kompakten Menge  $X$  rechnen zu können.

Aber auch wenn  $Z$  nur endlich viele Werte annimmt, ist die Wahl einer kompakten Menge  $X$  mit (5.5) oft nicht möglich (versuchen Sie einmal, so ein  $X$  für das Portfoliomodell aus Beispiel 3.2 zu konstruieren...).

Eine oft sinnvolle Abhilfe bietet die Wahl von  $X$  abhängig von  $\tau$ . Zunächst überlegt man sich dazu, auf welchem Gebiet  $X_T$  man die optimale Wertefunktion  $V_T$  berechnen möchte. Dies ist i.W. eine Modellierungs- bzw. Anwendungsfrage. Dann berechnet man sukzessive  $X_{T-1}, X_{T-2}, \dots, X_0$  so, dass die Bedingung

$$f(x, u, z) \in X_{\tau-1}, \quad \text{für alle } x \in X_\tau, u \in U(x), z \in Z(\Omega) \quad (5.6)$$

für alle  $\tau = T, T-1, \dots, 1$  erfüllt ist.

Wollen wir im Portfoliomodell aus Beispiel 3.2 die Funktion  $V_T$  z.B. für  $T = 20$  auf  $X_T = [0, a] \times [0, b]$  berechnen, so erhalten wir unter unserer üblichen Kontrollbeschränkung  $u \in [-x_1, x_2]$

$$0 \leq f_1(x, u, z) \leq (x_1 + x_2)z \leq (a + b)\alpha_u$$

und

$$0 \leq f_2(x, u, z) \leq (x_1 + x_2)R \leq (a + b)R.$$

und erhalten so  $X_{T-1} = [0, (a + b)\alpha_u] \times [0, (a + b)R]$ . Induktiv erhält man so

$$X_{T-t} = [0, (a + b)(\alpha_u + R)^{t-1}\alpha_u] \times [0, (a + b)(\alpha_u + R)^{t-1}R].$$

Hier erkennt man einen typischen Nachteil dieser Methode: Die Menge  $X_\tau$  wächst für kleiner werdendes  $\tau$  enorm an und wir müssen auf riesigen Gebieten rechnen, was numerisch sehr ineffizient ist. Dies liegt daran, dass wir hier Rechteckgebiete verwenden: tatsächlich liegen die Werte  $f(x, u, z)$  in einer recht kleinen Teilmenge  $A_\tau \subset X_\tau$ , die Rechteckmenge  $X_\tau$  ist aber die kleinste rechteckige Menge, die  $A_\tau$  enthält.

Glücklicherweise kann man in einem solchen Fall oft durch eine Koordinatentransformation Abhilfe schaffen. Im Portfoliomodell z.B. können wir die Transformation

$$y_1 = x_1 + x_2, \quad y_2 = \frac{x_1}{x_1 + x_2}$$

mit Umkehrung

$$x_1 = y_1 y_2, \quad x_2 = y_1(1 - y_2)$$

verwenden. Dass dies tatsächlich ein besseres Verhalten liefert, wird in einer Übungsaufgabe gezeigt.

## 5.5 Fehleranalyse

Wir wollen nun berechnen, wie groß der Approximationsfehler

$$\|\tilde{V}_\tau - V_\tau\|_{X_\tau} := \sup_{x \in X_\tau} |\tilde{V}_\tau(x) - V_\tau(x)|_{X_\tau}$$

ist.

Zunächst berechnen wir dazu den Fehler der Projektion  $\pi : F(X, \mathbb{R}) \rightarrow \mathcal{W}$ .

**Lemma 5.5** Sei  $W : X \rightarrow \mathbb{R}$  eine Lipschitz-stetige Funktion mit Lipschitz-Konstante  $L_W$ . Dann gilt

$$\|W - \pi W\|_X \leq L_W k \tag{5.7}$$

mit dem Wert  $k$  aus Definition 5.2. Ist  $W$  darüberhinaus  $C^2$  mit  $\|d^2W(x)/dx^2\| \leq C_W$  für alle  $x \in X$ , so gilt

$$\|W - \pi W\|_X \leq \frac{1}{2} C_W k^2. \tag{5.8}$$

**Beweis:** (5.7): Sei  $x \in X$  ein beliebiger Punkt und  $R$  ein Gitterrechteck, in dem dieser Punkt liegt. Seien  $E_{i_0}, \dots, E_{i_3}$  die Eckpunkte dieses Rechtecks. Dann gilt  $\|x - E_{i_j}\|_2 \leq k$  für  $j = 0, \dots, 3$  und somit  $|W(x) - W(E_{i_j})| \leq L_W k$ . Mit Lemma 5.4 folgt

$$\begin{aligned} |W(x) - \pi W(x)| &= \left| W(x) - \sum_{j=0}^3 \mu_j(x) W(E_{i_j}) \right| \\ &= \left| \sum_{j=0}^3 \mu_j(x) W(x) - \sum_{j=0}^3 \mu_j(x) W(E_{i_j}) \right| \\ &\leq \sum_{j=0}^3 \mu_j(x) |W(x) - W(E_{i_j})| \leq \sum_{j=0}^3 \mu_j(x) L_W k = L_W k \end{aligned}$$

wobei wir im zweiten und im letzten Schritt ausgenutzt haben, dass  $\sum_{j=0}^3 \mu_j(x) = 1$  gilt.

(5.8): Sei  $x \in X$ . Aus der Taylor-Entwicklung folgt für  $x, y \in X$

$$W(y) = W(x) + \frac{dW}{dx}(x)(y-x) + \frac{1}{2}(y-x)^T \frac{d^2W}{dx^2}(\xi)(y-x)$$

für ein  $\xi \in X$ . Seien nun  $R$  und  $E_{i_j}$  wie im ersten Teil des Beweises. Dann folgt

$$W(E_{i_j}) = W(x) + \frac{dW}{dx}(x)(E_{i_j} - x) + \underbrace{\frac{1}{2}(E_{i_j} - x)^T \frac{d^2W}{dx^2}(\xi)(E_{i_j} - x)}_{=: \rho_j}$$

und es folgt

$$\begin{aligned} |W(x) - \pi W(x)| &= \left| W(x) - \sum_{j=0}^3 \mu_j(x) W(E_{i_j}) \right| \\ &= \left| W(x) - \sum_{j=0}^3 \mu_j(x) \left( W(x) + \frac{dW}{dx}(x)(E_{i_j} - x) + \rho_j \right) \right| \\ &= \left| W(x) - \sum_{j=0}^3 \mu_j(x) W(x) - \frac{dW}{dx}(x) \sum_{j=0}^3 \mu_j(x)(E_{i_j} - x) - \sum_{j=0}^3 \mu_j(x) \rho_j \right| \\ &= \left| W(x) - W(x) - \frac{dW}{dx}(x)(x-x) - \sum_{j=0}^3 \mu_j(x) \rho_j \right| \\ &= \left| \sum_{j=0}^3 \mu_j(x) \rho_j \right| \leq \max_{j=0, \dots, 3} |\rho_j|. \end{aligned}$$

Hierbei haben wir  $\sum_{j=0}^3 \mu_j(x) = 1$  sowie  $\sum_{j=0}^3 \mu_j(x) E_{i_j} = x$  ausgenutzt, wobei die letzte Eigenschaft aus der Definition der  $\mu_j$  folgt. Wegen

$$|\rho_j| = \left| \frac{1}{2}(E_{i_j} - x)^T \frac{d^2W}{dx^2}(\xi)(E_{i_j} - x) \right| \leq \frac{1}{2} \left\| \frac{d^2W}{dx^2}(\xi) \right\| \|E_{i_j} - x\|_2^2 \leq \frac{1}{2} C_W k^2$$

folgt (5.8).  $\square$

Als weitere Vorbereitung benötigen wir das folgende Lemma.

**Lemma 5.6** (i) Sei  $\pi$  die zu einem Gitter auf  $X$  gehörige Projektion nach  $\mathcal{W}$  und  $W_1, W_2 \in F(X, \mathbb{R})$ . Dann gilt

$$\|\pi W_1 - \pi W_2\|_X \leq \|W_1 - W_2\|_X.$$

(ii) Seien  $X_1, X_2 \subset \mathbb{R}^n$ , so, dass  $f(x, u, z) \in X_1$  für alle  $x \in X_2$ ,  $u \in \tilde{U}(x)$  und  $z \in Z(\Omega)$  gilt. Dann gelten für alle  $W_1, W_2 \in F(X_1, \mathbb{R})$  die Ungleichungen

$$\|\tilde{T}(W_1) - \tilde{T}(W_2)\|_{X_2} \leq \beta \|W_1 - W_2\|_{X_1}$$

**Beweis:** (i) Für alle  $x \in X$  gilt

$$\begin{aligned} |\pi W_1(x) - \pi W_2(x)| &= \left| \sum_{j=0}^3 \mu_j(x) W_1(E_{i_j}) - \sum_{j=0}^3 \mu_j(x) W_2(E_{i_j}) \right| \\ &\leq \sum_{j=0}^3 \mu_j(x) |W_1(E_{i_j}) - W_2(E_{i_j})| \\ &\leq \sum_{j=0}^3 \mu_j(x) \|W_1 - W_2\|_X = \|W_1 - W_2\|_X \end{aligned}$$

und damit

$$\|\pi W_1 - \pi W_2\|_X \leq \|W_1 - W_2\|_X.$$

(ii) Mit (4.5) gilt für alle  $x \in X_2$

$$\begin{aligned} |\tilde{T}(W_1)(x) - \tilde{T}(W_2)(x)| &= \left| \max_{u \in \tilde{U}(x)} T_u(W_1)(x) - \max_{u \in \tilde{U}(x)} T_u(W_2)(x) \right| \\ &\leq \max_{u \in \tilde{U}(x)} |T_u(W_1)(x) - T_u(W_2)(x)| \\ &\leq \max_{u \in \tilde{U}(x)} E\{|\beta W_1(f(x, u, Z)) - \beta W_2(f(x, u, Z))|\} \\ &\leq \beta \|W_1 - W_2\|_{X_1} \end{aligned}$$

und damit

$$\|\tilde{T}(W_1) - \tilde{T}(W_2)\|_{X_2} \leq \beta \|W_1 - W_2\|_{X_1}.$$

$\square$

**Satz 5.7** Gegeben seien Mengen  $X_0, \dots, X_T$ ,  $T \in \mathbb{N}$ , mit (5.6). Weiterhin seien auf den  $X_\tau$  Gitter  $\Gamma_\tau$  mit Durchmessern  $k_\tau$  gegeben und diskretisierte Kontrollwertemengen  $\tilde{U}(x)$  mit  $\varepsilon_u$  aus (5.3). Es sei  $T_u V_\tau(x)$  Lipschitz mit Konstante  $L_{u,\tau}$ . Weiterhin gelte entweder

(i)  $V_\tau$  ist Lipschitz mit Konstante  $L_\tau$ ,  $\tau = 0, \dots, T$

oder

(ii)  $V_\tau$  ist  $C^2$  mit zweiter Ableitung beschränkt durch  $C_\tau$ ,  $\tau = 0, \dots, T$ .

Dann gilt für  $\tau = 0, \dots, T$  die Fehlerabschätzung

$$(i) \quad \|\tilde{V}_\tau - V_\tau\| \leq \sum_{t=0}^{\tau-1} \beta^{\tau-t-1} (L_{u,t} \varepsilon_u + L_{t+1} k_{t+1})$$

oder

$$(ii) \quad \|\tilde{V}_\tau - V_\tau\| \leq \sum_{t=0}^{\tau-1} \beta^{\tau-t-1} (L_{u,t} \varepsilon_u + C_{t+1} k_{t+1}^2 / 2)$$

Mit  $k = \max_{t=1, \dots, T} k_t$  ist der Fehler also von der Ordnung  $\mathcal{O}(\varepsilon_u + k)$  im Fall (i) und  $\mathcal{O}(\varepsilon_u + k^2)$  im Fall (ii).

**Beweis:** Wir schreiben kurz  $\delta_\tau = L_\tau k_\tau$  in Fall (i) und  $\delta_\tau = C_\tau k_\tau^2 / 2$  in Fall (ii) und zeigen beide Behauptungen per Induktion über  $\tau$ . Für  $\tau = 0$  gilt die Behauptung wegen

$$\tilde{V}_0 = L = V_0.$$

Für den Induktionsschritt  $\tau \rightarrow \tau + 1$  betrachte

$$\begin{aligned} \|\tilde{V}_{\tau+1} - V_{\tau+1}\|_{X_{\tau+1}} &= \|\pi \tilde{T} \tilde{V}_\tau - TV_\tau\|_{X_{\tau+1}} \\ &\leq \|\pi \tilde{T} \tilde{V}_\tau - \pi \tilde{T} V_\tau\|_{X_{\tau+1}} + \|\pi \tilde{T} V_\tau - \pi TV_\tau\|_{X_{\tau+1}} + \underbrace{\|\pi TV_\tau - TV_\tau\|_{X_{\tau+1}}}_{=\pi V_{\tau+1} - V_{\tau+1}} \end{aligned}$$

Den ersten Term können wir nun mit der Induktionsannahme und (beiden Ungleichungen aus) Lemma 5.6 abschätzen, den zweiten mit Lemma 5.1 und der ersten Ungleichung aus Lemma 5.6 und den dritten mit Lemma 5.5. Damit erhalten wir

$$\|\tilde{V}_{\tau+1} - V_{\tau+1}\|_{X_{\tau+1}} \leq \beta \sum_{t=0}^{\tau-1} \beta^{\tau-t-1} (L_{u,t} \varepsilon_u + \delta_{t+1}) + \varepsilon_u L_{u,\tau} + \delta_{\tau+1} = \sum_{t=0}^{\tau} \beta^{\tau-t} (L_{u,t} \varepsilon_u + \delta_{t+1}),$$

also die Behauptung.  $\square$

Aus der numerischen Approximation erhalten wir für jeden Zustand ein numerisch auswertbares Feedback mittels

$$\tilde{F}_\tau(x) := \operatorname{argmax}_{\tilde{u} \in \tilde{U}(x)} T_{\tilde{u}} \tilde{V}_\tau(x). \quad (5.9)$$

Will man optimale Trajektorien simulieren, empfiehlt es sich — wie bereits erwähnt —, dieses Feedback für den jeweils aktuellen Zustand online mit Formel (5.9) zu berechnen und nicht vorab zu berechnen und zu speichern, um Interpolationsfehler in der Approximation von  $\tilde{F}_\tau$  zu vermeiden.

Betrachten wir die zugehörige Lösung

$$\tilde{X}_0 = x_0, \quad \tilde{X}_{t+1} = f(\tilde{X}_t, \tilde{F}_{T-t-1}(\tilde{X}_t), Z_t)$$

und den davon induzierten Kontrollprozess

$$\tilde{\mathbf{u}}^{T, x_0} = (\tilde{F}_{T-t-1}(\tilde{X}_t), t = 0, \dots, T-1)$$

so zeigt Übungsaufgabe 1 von Blatt 7, dass für die approximativen Trajektorien die Abschätzung

$$J_T(x_0, \tilde{\mathbf{u}}^{T,x_0}) \geq V_T(x_0) - \mathcal{O}(\varepsilon_u + k)$$

im Fall (i) bzw.

$$J_T(x_0, \tilde{\mathbf{u}}^{T,x_0}) \geq V_T(x_0) - \mathcal{O}(\varepsilon_u + k^2)$$

im Fall (ii) gilt.

## 5.6 Adaptive Gitter

Adaptive Gitter bilden eine Möglichkeit, die Effizienz der Approximation durch affin bilineare Funktionen deutlich zu erhöhen. Adaptive Gitter sind Gitter, deren Rechtecke nicht gleichmäßig groß sind, sondern — angepasst an die zu approximierende Funktion — in manchen Regionen feiner, in anderen gröber gewählt werden.

Wir stellen das Verfahren hier auf heuristische Weise vor, ohne groß auf die theoretischen Grundlagen einzugehen.

Die Konstruktion der adaptiven Gitter beruht auf der Beobachtung, dass wir in jedem Schritt idealerweise die Gleichung

$$\tilde{V}_{\tau+1}(x) = \tilde{T}\tilde{V}_\tau(x) \tag{5.10}$$

für alle  $x \in X_{\tau+1}$  erfüllen sollten, tatsächlich aber nur

$$\tilde{V}_{\tau+1}(E_i) = \tilde{T}\tilde{V}_\tau(E_i)$$

für  $i = 0, \dots, N-1$  erfüllen, während für alle Punkte  $x \neq E_i$ ,  $i = 0, \dots, N$  i.A.

$$\tilde{V}_{\tau+1}(x) \neq \tilde{T}\tilde{V}_\tau(x)$$

gilt.

Die Idee ist nun, auf einer Menge von Testpunkten  $x_j$ ,  $j = 0, \dots, M$ , im Gitter (die nicht mit den Eckpunkten übereinstimmen sollten) zu testen, wie sehr (5.10) verletzt ist. Dazu werten wir die Größe

$$e_j := |\tilde{V}_{\tau+1}(x_j) - \tilde{T}\tilde{V}_\tau(x_j)|$$

aus. Ist  $e_j$  groß, so wird  $x_j$  als neuer Gitterpunkt (mit dem bereits berechneten Wert  $\tilde{T}\tilde{V}_\tau(x_j)$ ) in das Gitter eingefügt. “Groß” bedeutet hierbei typischerweise  $e_j > tol$  für eine vorgegebene Toleranz  $tol$ , kann aber auch anders definiert werden. Als Testpunkt wählt man hierbei Punkte, die sich leicht als neue Gitterpunkte in die Geometrie einfügen lassen, also in 2d z.B. die Kantenmittelpunkte und die Mittelpunkte der Rechtecke.

Ähnlich kann man testen, ob ein bereits vorhandener Gitterpunkt  $E_i$  tatsächlich notwendig ist. Hierzu entfernt man  $E_i$  aus dem Gitter und betrachtet die Funktion  $V'_{\tau+1}$  auf dem vergrößerten Gitter  $\Gamma'_{\tau+1} = \Gamma_{\tau+1} \setminus \{E_i\}$  mit dem entfernten Knoten  $E_i$ . Ist die Differenz

$$d_i = |V'_{\tau+1}(E_i) - V_{\tau+1}(E_i)|$$

kleiner als die vorgegebene Toleranz (üblicherweise mit einem Sicherheitsfaktor  $\rho < 1$  multipliziert), so kann  $E_i$  aus dem Gitter entfernt werden, ohne dass die Qualität der Approximation — im Rahmen der vorgegebenen Toleranz — beeinträchtigt wird.

Üblicherweise wird man diese Operationen iterativ durchführen, bis die gewünschte Toleranz für alle Testpunkte erfüllt ist. Dies führt auf den folgenden Algorithmus:

- (0) Wähle ein Startgitter  $\Gamma_0$ , setze  $k = 0$
- (1) Berechne  $\tilde{V}_{\tau+1}$  auf  $\Gamma_k$
- (2) Berechne die Werte  $e_j$  für alle Testpunkte  $x_j$
- (3) Berechne die Werte  $d_i$  für alle Gitterpunkte  $E_i$
- (4) Füge alle Testpunkte  $x_j$  mit  $e_j > tol$  in  $\Gamma_k$  ein und entferne alle Gitterpunkte  $E_i$  mit  $d_j < \rho tol$  aus  $\Gamma_k$ ; bezeichne das so entstandene neue Gitter mit  $\Gamma_{k+1}$
- (5) Falls  $e_j \leq tol$  für alle  $j$  beende den Algorithmus, ansonsten setze  $k = k + 1$  und gehe zu (1)

Sicherheitshalber sollte man zusätzlich noch eine Maximalzahl von Elementen definieren, bei deren Erreichen der Algorithmus abbricht, auch wenn die gewünschte Toleranz nicht erreicht wird.

## 5.7 Unendlicher Zeithorizont

Unsere Approximationstechnik lässt sich unter geeigneten Annahmen auf unendlichen Zeithorizont verallgemeinern. Ist  $V_\infty$  z.B. beschränkt und  $\beta < 1$ , so folgt aus Übungsaufgabe 2, Blatt 7, dass  $V_T \rightarrow V_\infty$  gilt, weswegen wir eine numerische Approximation theoretisch als

$$\tilde{V}_\infty = \lim_{T \rightarrow \infty} \tilde{V}_T \quad (5.11)$$

definieren können. Man kann leicht beweisen, dass diese Funktion die eindeutige Lösung der Bellman-Gleichung

$$\tilde{V}_\infty = \tilde{T}\tilde{V}_\infty \quad (5.12)$$

ist. Praktisch kann  $\tilde{V}_\infty$  numerisch approximiert werden, indem man  $\tilde{V}_T$  für  $T = 0, 1, 2, \dots$  berechnet, bis ein Abbruchkriterium der Form

$$\|\tilde{V}_T - \tilde{V}_{T-1}\|_X \leq \varepsilon$$

erfüllt ist, ein Verfahren das als *Werteiteration* bezeichnet wird. Es gibt aber viel effizientere Verfahren. Für stochastische Probleme mit endlichem Zustandsraum werden solche Verfahren im derzeit laufenden Seminar “Reinforcement Learning” behandelt. Für deterministische Probleme mit kontinuierlichem Zustandsraum finden sich eine Reihe von Verfahren in Abschnitt 4.2 des Skripts “Numerische Dynamik von Kontrollsystemen”, Sommersemester 2004<sup>1</sup>. Die dort besprochenen Methoden lassen sich auf stochastische Probleme übertragen.

<sup>1</sup><http://www.math.uni-bayreuth.de/~lgruene/ndks04/>

Für  $\beta = 1$  und unbeschränktes  $V_\infty$  muss das Problem üblicherweise skaliert und reguliert werden, für deterministische Probleme findet sich eine solche Methode in Kapitel 6 des Skripts “Numerische Dynamik von Kontrollsystemen”. Alternativ kann man — falls vorhanden — spezielle Strukturen ausnutzen, was wir im folgenden Kapitel machen werden.



# Kapitel 6

## Linear Quadratische Probleme

In diesem Kapitel betrachten wir optimale Steuerungsprobleme mit der folgenden Struktur:

- (i) Die Dynamik  $X_{t+1} = f(X_t, u_t, Z_t)$  ist gegeben durch eine lineare Abbildung der Form

$$f(x, u, z) = Ax + Bu + Cz$$

mit Matrizen der passenden Dimension, also  $A \in \mathbb{R}^{n \times n}$ ,  $B \in \mathbb{R}^{n \times l}$ ,  $C \in \mathbb{R}^{n \times m}$ .

- (ii) Der Kontrollwertebereich ist  $U = \mathbb{R}^l$  und die Zufallsvariablen  $Z_t$  sind iid mit Erwartungswert  $E(Z_t) = 0$ .

- (iii) Die laufenden Kosten  $\ell(x, u)$  sind von der Form

$$\ell(x, u) = x^T Q x + u^T R u$$

mit symmetrischen und positiv definiten Matrizen  $Q \in \mathbb{R}^{n \times n}$  und  $R \in \mathbb{R}^{l \times l}$ .

- (iv) Die Endkosten  $L(x)$  sind von der Form

$$L(x) = x^T P_0 x$$

mit einer symmetrischen und positiv semidefiniten Matrix  $P_0 \in \mathbb{R}^{n \times n}$ .

Solche Probleme heißen *Linear-quadratische* stochastische optimale Steuerungsprobleme.

In diesem Kapitel betrachten wir — im Gegensatz zu den vorhergehenden Kapiteln — das Minimierungsproblem, da dies die in der Literatur gebräuchlichere Variante des Problems ist. Wir schreiben wieder einfach  $Z$  für einen beliebigen Vertreter der Zufallsvariablen  $Z_t$ .

Wir werden sehen, dass die optimalen Wertefunktionen dieses Problems eine sehr einfache Form besitzen. Es gilt nämlich

$$V_\tau(x) = x^T P_\tau x + c_\tau \tag{6.1}$$

für symmetrische und positiv semidefinite (für  $\tau \geq 1$  sogar positiv definite) Matrizen  $P_\tau$  und Konstanten  $c_\tau$ .

Nehmen wir zunächst einmal an, dass  $V_\tau$  die Form (6.1) besitzt. Dann ergibt sich das Optimalitätsprinzip zu

$$\begin{aligned} V_{\tau+1}(x) &= \inf_{u \in U} E \{ \ell(x, u) + \beta V_\tau(f(x, u, Z)) \} \\ &= \inf_{u \in U} E \{ x^T Q x + u^T R u + \beta (Ax + Bu + CZ)^T P_\tau (Ax + Bu + CZ) + \beta c_\tau \}. \end{aligned}$$

Das folgende Lemma zeigt, wie das Minimum und das minimierende  $u$  für diesen Ausdruck aussieht.

**Lemma 6.1** Für positiv semidefinites  $P$  und

$$g(u) := E \{ x^T Q x + u^T R u + \beta (Ax + Bu + CZ)^T P (Ax + Bu + CZ) \}$$

gelten die Gleichungen

$$\operatorname{argmin}_{u \in \mathbb{R}^l} g(u) = -(R + \beta B^T P B)^{-1} \beta B^T P A x$$

und

$$\min_{u \in \mathbb{R}^l} g(u) = x^T \left( Q + \beta A^T P A - \beta^2 A^T P B (R + \beta B^T P B)^{-1} B^T P A \right) x + \beta E \{ Z^T C^T P C Z \}.$$

**Beweis:** Wir vereinfachen zunächst den dritten Summanden der zu minimierenden Funktion. Es gilt

$$\begin{aligned} & E \{ (Ax + Bu + CZ)^T P (Ax + Bu + CZ) \} \\ &= E \{ x^T A^T P (Ax + Bu + CZ) + u^T B^T P (Ax + Bu + CZ) + Z^T C^T P (Ax + Bu + CZ) \} \\ &= E \{ x^T A^T P A x + x^T A^T P B u + x^T A^T P C Z + u^T B^T P A x + u^T B^T P B u + u^T B^T P C Z \\ &\quad + Z^T C^T P A x + Z^T C^T P B u + Z^T C^T P C Z \} \\ &= E \{ x^T A^T P A x + x^T A^T P B u + u^T B^T P A x + u^T B^T P B u + Z^T C^T P C Z \} \\ &= x^T A^T P A x + x^T A^T P B u + u^T B^T P A x + u^T B^T P B u + E \{ Z^T C^T P C Z \} \end{aligned}$$

Damit folgt

$$g(u) = x^T (Q + \beta A^T P A) x + 2\beta u^T B^T P A x + u^T (R + \beta B^T P B) u + \beta E \{ Z^T C^T P C Z \}$$

Da  $R$  positiv definit und  $P$  positiv semidefinit ist, ist  $u^* = \operatorname{argmin}_{u \in \mathbb{R}^l} g(u)$  die eindeutige Nullstelle von  $dg/du$ , also

$$0 = \frac{dg}{du}(u^*) = 2\beta B^T P A x + 2(R + \beta B^T P B)u^*.$$

Also gilt für die Minimalstelle

$$\operatorname{argmin}_{u \in \mathbb{R}^l} g(u) = -(R + \beta B^T P B)^{-1} \beta B^T P A x.$$

Damit folgt für das Minimum

$$\begin{aligned} g(u^*) &= x^T (Q + \beta A^T P A) x + 2\beta u^{*T} B^T P A x + u^{*T} (R + \beta B^T P B) u^* + \beta E \{ Z^T C^T P C Z \} \\ &= x^T (Q + \beta A^T P A) x - \beta^2 x^T A^T P B (R + \beta B^T P B)^{-1} B^T P A x + \beta E \{ Z^T C^T P C Z \}. \end{aligned}$$

□

**Satz 6.2** Die optimalen Wertefunktionen des linear-quadratischen Problems mit endlichem Zeithorizont sind gegeben durch

$$V_\tau(x) = x^T P_\tau x + c_\tau,$$

mit symmetrischen und positiv semidefiniten Matrizen  $P_\tau$ , die für  $\tau \geq 1$  sogar positiv definit sind, und nichtnegativen reellen Zahlen  $c_\tau$ . Die  $P_\tau$  und  $c_\tau$  sind iterativ gegeben durch

$$P_{\tau+1} = Q + \beta A^T P_\tau A - \beta^2 A^T P_\tau B (R + \beta B^T P_\tau B)^{-1} B^T P_\tau A$$

und

$$c_{\tau+1} = \beta c_\tau + \beta E\{Z^T C^T P_\tau C Z\},$$

für  $\tau \in \mathbb{N}_0$  mit  $P_0$  aus der Definition von  $L$  und  $c_0 = 0$ .

Das zugehörige optimale Feedback ist gegeben durch

$$F_\tau(x) := -(R + \beta B^T P_\tau B)^{-1} \beta B^T P_\tau A x.$$

**Beweis:** Wir beweisen die Formel per Induktion: Für  $\tau = 0$  ist die Aussage wegen

$$V_0(x) = L(x) = x^T P_0 x + c_0$$

klar.

$\tau \rightarrow \tau + 1$ : Mit dem Optimalitätsprinzip gilt

$$V_{\tau+1}(x) = \inf_{u \in U} g_\tau(u) + \beta c_\tau$$

mit

$$g_\tau(u) := E\{x^T Q x + u^T R u + \beta(Ax + Bu + CZ)^T P_\tau (Ax + Bu + CZ)\}.$$

Mit Lemma 6.1 folgt

$$\begin{aligned} V_{\tau+1}(x) &= x^T \underbrace{\left( Q + \beta A^T P_\tau A - \beta^2 A^T P_\tau B (R + \beta B^T P_\tau B)^{-1} B^T P_\tau A \right)}_{=P_{\tau+1}} x \\ &\quad + \underbrace{\beta E\{Z^T C^T P_\tau C Z\} + \beta c_\tau}_{=c_{\tau+1}}. \end{aligned}$$

Die positive Definitheit von  $P_\tau$  für  $\tau \geq 1$  folgt aus der Gleichung

$$x^T P_\tau x + c_\tau = V_\tau(x) \geq x^T R x$$

und der Tatsache, dass  $R$  positiv definit ist.

Die Form des Feedbacks schließlich folgt nun mit (i) aus der allgemeinen nichtlinearen Form des optimalen Feedbacks in Satz 4.4 sowie Lemma 6.1.  $\square$

**Bemerkung 6.3** Tatsächlich funktioniert dieser Beweis auch unter schwächeren Bedingungen an die Matrizen. Z.B. kann die positive Definitheit von  $R$  und  $Q$  wegfallen, wenn man statt dessen annimmt, dass die Matrizen  $R + \beta B^T P_\tau B$  für alle  $\tau \in \mathbb{N}_0$  positiv definit sind, was man dann aber nur iterativ und nicht a priori überprüfen kann.  $\square$

Betrachten wir nun das Problem auf unendlichem Horizont. Leider passen die Transversalitätsbedingungen nicht zu diesem Problem<sup>1</sup>, weswegen wir nicht mit den Sätzen aus Kapitel 4 arbeiten können und die entsprechenden Aussagen direkt beweisen müssen.

Als Grundannahme für das Problem stellen wir die folgende Bedingung:

$$\begin{aligned} &\text{Es existieren Konstanten } C > 0 \text{ und } \sigma \in [0, 1), \text{ so dass} \\ &\text{für alle } x \in \mathbb{R}^n \text{ ein Kontrollprozess } \mathbf{u}^x \in \mathcal{U} \text{ existiert mit} \\ &\beta^t E\{\|X_t\|^2 + \|u_t^x\|^2\} \leq C\sigma^t(\|x_0\|^2 + 1) \text{ für alle } t \in \mathbb{N} \end{aligned} \quad (6.2)$$

**Lemma 6.4** Es gelte die Annahme (6.2). Zudem sei  $V_\infty$  die optimale Wertefunktion für unendlichem Horizont und  $V_\tau$  mit  $V_\tau(x) = x^T P_\tau x + c_\tau$  die optimale Wertefunktion für endlichen Horizont mit Endkosten  $L \equiv 0$ .

Dann gelten für

$$C_\infty = C \frac{\max\{\|Q\|, \|R\|\}}{1 - \sigma}$$

und alle  $\tau \in \mathbb{N}$  die Abschätzungen

$$(i) \ V_\tau(x) \leq V_\infty(x) \leq C_\infty(\|x\|^2 + 1) \quad \text{und} \quad (ii) \ x^T P_\tau x \leq C_\infty \|x\|^2.$$

**Beweis:** (i) Wegen

$$X_t^T Q X_t + u_t^{xT} R u_t^x \leq \|Q\| \|X_t\|^2 + \|R\| \|u_t^x\|^2 \leq \max\{\|Q\|, \|R\|\} (\|X_t\|^2 + \|u_t^x\|^2)$$

folgt aus (6.2) für alle  $\tau \in \mathbb{N}$  die Ungleichung

$$\begin{aligned} V_\infty(x_0) &\leq \sum_{t=0}^{\infty} \beta^t (X_t^T Q X_t + u_t^{xT} R u_t^x) \\ &\leq \sum_{t=0}^{\infty} C\sigma^t \max\{\|Q\|, \|R\|\} (\|x_0\|^2 + 1) \leq C \frac{\max\{\|Q\|, \|R\|\}}{1 - \sigma} (\|x_0\|^2 + 1). \end{aligned}$$

Die Ungleichung  $V_\tau \leq V_\infty$  folgt sofort aus der Definition wegen  $L \equiv 0$ .

(ii) Aus (i) folgt für alle  $x$

$$x^T P_\tau x \leq V_\tau(x) - c_\tau \leq C_\infty(\|x\|^2 + 1) - c_\tau = C_\infty \|x\|^2 + C_\infty - c_\tau.$$

Betrachten wir nun eine positive reelle Folge  $\alpha_n \rightarrow \infty$ , so gilt

$$x^T P_\tau x = \frac{(\alpha_n x)^T P_\tau (\alpha_n x)}{\alpha_n^2} \leq C_\infty \frac{\|\alpha_n x\|^2}{\alpha_n} + \frac{C_\infty - c_\tau}{\alpha_n} = C_\infty \|x\|^2 + \frac{C_\infty - c_\tau}{\alpha_n} \rightarrow C_\infty \|x\|^2$$

für  $n \rightarrow \infty$ . Damit folgt (ii).  $\square$

Zum Beweis unseres Hauptresultates benötigen wir noch das folgende Lemma.

<sup>1</sup>Eigentlich dachte ich beim Aufschreiben von Kapitel 4, dass ich die Transversalitätsbedingungen so formuliert habe, dass sie zum linear-quadratischen Problem passen. Beim Aufschreiben dieses Kapitels musste ich aber feststellen, dass dies doch nicht stimmt — ich hatte einen Vorzeichenfehler in meinen Überlegungen. Dies wird in einer zukünftigen Überarbeitung dieses Skripts repariert werden.

**Lemma 6.5** Betrachte eine Folge symmetrischer und positiv definiter  $n \times n$  Matrizen  $P_t$ ,  $t \in \mathbb{N}$  mit

$$x^T P_{t+1} x \geq x^T P_t x \quad \text{und} \quad x^T P_t x \leq C \|x\|^2$$

für eine Konstante  $C > 0$  und alle  $x \in \mathbb{R}^n$ ,  $t \in \mathbb{N}$ . Dann existiert eine symmetrische und positiv definite Matrix  $P$  mit

$$\lim_{t \rightarrow \infty} P_t = P,$$

wobei der Limes komponentenweise zu verstehen ist.

**Beweis:** Mit  $e_j$  bezeichnen wir den  $j$ -tem Basisvektor. Wir definieren

$$p_{ij} = \lim_{t \rightarrow \infty} (e_i + e_j)^T P_t (e_i + e_j) \quad \text{und} \quad p_j = \lim_{t \rightarrow \infty} e_j^T P_t e_j.$$

Aus unseren Annahmen folgt Monotonie und Beschränktheit der Ausdrücke auf den rechten Seiten, weswegen diese Limes tatsächlich existieren. Wegen

$$[P_t]_{ij} = e_i^T P_t e_j = \frac{1}{2} ((e_i + e_j)^T P_t (e_i + e_j) - e_i^T P_t e_i - e_j^T P_t e_j)$$

gilt nun die Gleichung

$$\lim_{t \rightarrow \infty} [P_t]_{ij} = \frac{1}{2} (p_{ij} - p_i - p_j).$$

Dies zeigt, dass der komponentenweise Limes  $P := \lim_{t \rightarrow \infty} P_t$  existiert. Diese Matrix ist offensichtlich symmetrisch und wegen

$$x^T P x \geq x^T P_t x > 0 \quad \text{für alle } x \neq 0 \text{ und beliebiges } t > 0$$

positiv definit. □

Damit können wir nun den Hauptsatz formulieren.

**Satz 6.6** Betrachte ein linear-quadratisches Problem mit unendlichem Zeithorizont, das Bedingung 6.2 für ein  $\beta < 1$  erfüllt. Dann ist die optimale Wertefunktion gegeben durch

$$V_\infty(x) = x^T P x + c,$$

wobei  $P$  die eindeutige symmetrische und positiv definite Lösung der *Riccati-Gleichung*

$$P = Q + \beta A^T P A - \beta^2 A^T P B (R + \beta B^T P B)^{-1} B^T P A$$

ist und

$$c = \frac{\beta E \{ Z^T C^T P C Z \}}{1 - \beta}.$$

Das zugehörige optimale Feedback ist gegeben durch

$$F_\infty(x) = -(R + \beta B^T P B)^{-1} \beta B^T P A x.$$

**Beweis:** Sei  $P$  eine beliebige positiv definite Lösung der Riccati-Gleichung und setze  $W(x) = x^T P x + c$  mit  $c$  aus dem Satz. Sei

$$G(x) := -(R + \beta B^T P B)^{-1} \beta B^T P A x$$

das zugehörige Feedback. Dann gilt

$$\begin{aligned} W(x) &= x^T P x + c \\ &= x^T (Q + \beta A^T P A - \beta^2 A^T P B (R + \beta B^T P B)^{-1} B^T P A) x + \frac{\beta E\{Z^T C^T P C Z\}}{1 - \beta} \\ &= x^T Q x + G(x)^T R G(x) + \beta (A x + B G(x) + C Z)^T P (A x + B G(x) + C Z) \\ &\quad - \underbrace{\beta E\{Z^T C^T P C Z\} + \frac{\beta E\{Z^T C^T P C Z\}}{1 - \beta}}_{= \frac{\beta^2 E\{Z^T C^T P C Z\}}{1 - \beta} = \beta c} \\ &= E\{\ell(x, G(x)) + \beta W(f(x, G(x), Z))\}. \end{aligned}$$

Zu  $x_0$  betrachte die zugehörige Lösung  $X_{t+1} = f(X_t, G(X_t), Z_t) = A X_t + B G(X_t) + C Z_t$  und bezeichne mit  $\mathbf{u}^{x_0} = (G(X_t), t = 0, 1, \dots)$  den zugehörigen Kontrollprozess.

Per Induktion über  $\tau$  zeigen wir nun

$$J_\tau(x_0, \mathbf{u}^{x_0}) \leq W(x).$$

Für  $\tau = 0$  folgt die Behauptung wegen  $J_0(x, \mathbf{u}^{x_0}) = L(x_0) = 0 \leq W(x)$ . Für  $\tau \rightarrow \tau + 1$  folgt wie im Beweis von Satz 4.4 die Gleichung

$$J_\tau(x_0, \mathbf{u}^{x_0}) = E\{\ell(x_0, G(x_0)) + \beta J_{\tau-1}(X_1, \mathbf{u}^{X_1})\}.$$

Damit folgt

$$\begin{aligned} &J_\tau(x_0, \mathbf{u}^{x_0}) - W(x) \\ &= E\{\ell(x, u_0^x) + J_{\tau-1}(X_1, \mathbf{u}^{X_1})\} - E\{\ell(x_0, G(x_0)) + W(X_1)\} \\ &= E\{\underbrace{J_{\tau-1}(X_1, \mathbf{u}^{X_1}) - W(X_1)}_{\leq 0 \text{ nach Induktionsannahme}}\} \leq 0. \end{aligned}$$

Damit erhalten wir das erste Zwischenergebnis

$$V_\infty(x) \leq J_\infty(x_0, \mathbf{u}^x) = \lim_{\tau \rightarrow \infty} J_\tau(x_0, \mathbf{u}^x) \leq W(x). \quad (6.3)$$

Für  $V_\infty$  gilt (4.9):

$$V_\infty(x_0) = \inf_{\mathbf{u} \in \mathcal{U}} E \left\{ \sum_{t=0}^{\tau-1} \beta^t \ell(X_t, u_t) + \beta^\tau V_\infty(X_\tau) \right\}.$$

Für jedes  $\tau > 0$  sei nun  $\mathbf{u}^\tau$  ein Kontrollprozess mit Lösung  $X_t^\tau$  und

$$V_\infty(x_0) \geq E \left\{ \sum_{t=0}^{\tau-1} \beta^t \ell(X_t^\tau, u_t^\tau) + \beta^\tau V_\infty(X_\tau^\tau) \right\} - \varepsilon.$$

Da  $Q$  positiv definit ist, existiert  $\alpha > 0$  mit  $\ell(x, u) \geq \alpha \|x\|^2$  und wir erhalten

$$\sum_{t=0}^{\tau-1} \beta^t \alpha E \{ \|X_t^\tau\|^2 \} \leq V_\infty(x_0) + \varepsilon$$

für alle  $\tau > 0$ . Daraus folgt die Existenz von Zeiten  $s(\tau) \leq \tau - 1$  mit  $s(\tau) \rightarrow \infty$  und

$$\beta^{s(\tau)} E \{ \|X_{s(\tau)}^\tau\|^2 \} \rightarrow 0$$

für  $\tau \rightarrow \infty$ .

Unter Ausnutzung von Lemma 6.1 erhalten wir

$$W(x) = E \{ \ell(x, G(x)) + \beta W(f(x, G(x), Z)) \} = \min_{u \in \mathbb{R}^l} E \{ \ell(x, u) + \beta W(f(x, u, Z)) \}$$

und damit induktiv

$$\begin{aligned} W(x) &= \min_{\mathbf{u} \in \mathcal{U}} E \left\{ \sum_{t=0}^{s(\tau)-1} \ell(X_t, u_t) + \beta^{s(\tau)} W(X_{s(\tau)}) \right\} \\ &\leq E \left\{ \sum_{t=0}^{s(\tau)-1} \ell(X_t^\tau, u_t^\tau) + \beta^{s(\tau)} W(X_{s(\tau)}^\tau) \right\} \\ &\leq E \left\{ \sum_{t=0}^{s(\tau)-1} \ell(X_t^\tau, u_t^\tau) + \beta^{s(\tau)} (\|P\| \|X_{s(\tau)}^\tau\|^2 + c) \right\} \end{aligned}$$

Damit folgt

$$\begin{aligned} W(x) - V_\infty(x) &\leq E \left\{ \sum_{t=0}^{s(\tau)-1} \ell(X_t^\tau, u_t^\tau) + \beta^{s(\tau)} (\|P\| \|X_{s(\tau)}^\tau\|^2 + c) \right\} \\ &\quad - E \left\{ \sum_{t=0}^{\tau-1} \beta^t \ell(X_t^\tau, u_t^\tau) + \beta^\tau V_\infty(X_\tau^\tau) \right\} + \varepsilon \\ &\leq E \left\{ \sum_{t=0}^{s(\tau)-1} \ell(X_t^\tau, u_t^\tau) + \beta^{s(\tau)} (\|P\| \|X_{s(\tau)}^\tau\|^2 + c) \right\} \\ &\quad - E \left\{ \sum_{t=0}^{s(\tau)-1} \beta^t \ell(X_t^\tau, u_t^\tau) \right\} + \varepsilon \\ &= E \left\{ \beta^{s(\tau)} (\|P\| \|X_{s(\tau)}^\tau\|^2 + c) \right\} + \varepsilon \rightarrow \varepsilon \end{aligned}$$

für  $\tau \rightarrow \infty$ . Da  $\varepsilon > 0$  beliebig war, erhalten wir das zweite Zwischenergebnis

$$W(x) \leq V_\infty(x). \quad (6.4)$$

Die Zwischenergebnisse (6.3) und (6.4) zeigen zusammengefasst:

Wenn eine symmetrische und positiv definite Lösung  $P$  der Riccati-Gleichung existiert, dann gilt

$$V_\infty(x) = x^T P x + c$$

und das angegebene Feedback erzeugt Kontrollprozesse  $\mathbf{u}^x$  mit

$$J_\infty(x, \mathbf{u}^x) \leq V_\infty(x),$$

ist also tatsächlich optimal.

Es bleibt zu zeigen, dass eine solche Lösung  $P$  der Riccati-Gleichung existiert. Betrachte dazu die optimalen Wertefunktionen  $V_\tau(x) = x^T P_\tau x + c_\tau$  des Problems mit Horizont  $\tau$  und Endkosten  $L \equiv 0$ . Aus den Formeln für  $P_\tau$  und  $c_\tau$  folgt

$$x^T P_{\tau+1} x \geq x^T P_\tau x.$$

Aus Lemma 6.4 folgt daher, dass Lemma 6.5 auf  $P_\tau$  anwendbar ist. Also existiert eine symmetrische und positiv definite Matrix  $P$  mit

$$\lim_{\tau \rightarrow \infty} P_\tau = P.$$

Wegen

$$P_{\tau+1} = Q + \beta A^T P_\tau A - \beta^2 A^T P_\tau B (R + \beta B^T P_\tau B)^{-1} B^T P_\tau A$$

und der Konvergenz  $P_\tau \rightarrow P$  und damit natürlich auch  $P_{\tau+1} \rightarrow P$  folgt, dass  $P$  die angegebene Riccati-Gleichung erfüllt und damit eine symmetrische und positiv definite Lösung existiert.  $\square$

**Beispiel 6.7** Wir betrachten unser Wagenmodell aus Beispiel 3.3. Hier gilt:

$$A = \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} \quad \text{und} \quad B = C = \begin{pmatrix} \frac{h^2}{2} \\ h \end{pmatrix}$$

und wir haben gewählt

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{und} \quad R = (\alpha).$$

Wir lösen die Gleichungen mit MATLAB, vgl. das M-File `wagen.m` auf der Vorlesungshomepage. Für  $h = 0.1$ ,  $\alpha = 0.1$ ,  $\beta = 0.9$ ,  $\sigma = 0.1$  erhalten wir z.B.

$$P = \begin{pmatrix} 7.4591 & 1.6909 \\ 1.6909 & 3.78831 \end{pmatrix}, \quad F = (-1.3680 \quad -2.7038), \quad c = 0.0036.$$

Abbildung 6.1 zeigt die Wertefunktion und eine simulierte optimale Trajektorie. Beachte: Je größer die Beträge der Einträge von  $F$  sind, desto schneller wird der Wagen in den Ursprung gelenkt.

Betrachtet man die Lösungen in Abhängigkeit der Parameter, so stellt man fest: Je kleiner  $\alpha$  und je größer  $\beta$  wird, desto schneller wird der Wagen in den Ursprung gelenkt.  $\square$

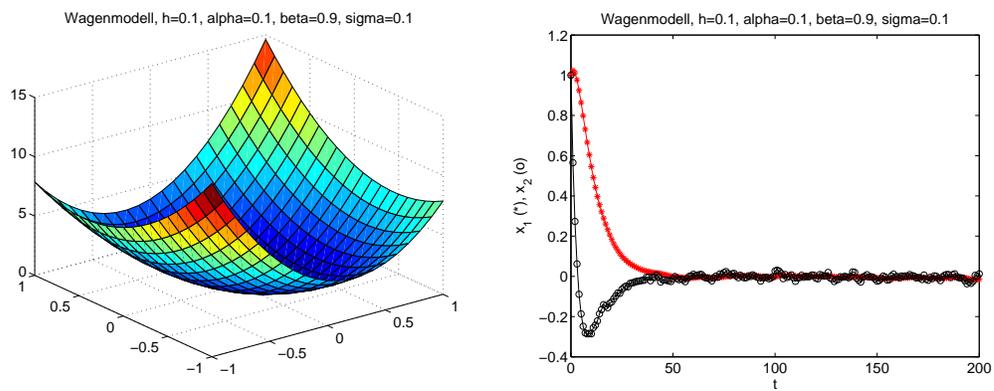


Abbildung 6.1: Wertefunktion und Trajektorie des Wagenmodells



# Kapitel 7

## Ein stochastisches Maximumprinzip

Alle bisherigen Methoden beruhen darauf, dass zuerst die optimale Wertefunktion und darauf aufbauend das optimale Feedback berechnet wird. In diesem Kapitel gehen wir einen anderen Weg, indem wir optimale Trajektorien und Kontrollprozesse direkt ohne Umweg über die Wertefunktion charakterisieren. An deren Stelle treten hier die Kozustands- (oder adjungierten) Variablen, die aber nicht global für alle  $x$  sondern lokal entlang der optimalen Trajektorien definiert sind.

Dieses im deterministischen sehr erfolgreiche Verfahren ist in unserem stochastischen Kontext leider schwierig direkt anzuwenden, weil auch die Kozustandsvariablen stochastische Prozesse werden. Eine interessante indirekte numerische Anwendung ist allerdings die sogenannte Perturbationsmethode, deren Funktionsweise im zweiten Abschnitt skizziert wird.

### 7.1 Formulierung und Beweis

Zur Formulierung des Prinzips benötigen wir die sogenannte Hamilton-Funktion, welche für  $x, \lambda \in \mathbb{R}^n$  und  $u \in \mathbb{R}^l$  gegeben ist durch

$$H^t(x, \lambda, u) = E_t\{\beta^t \ell(x, u) + \lambda^T f(x, u, Z_t)\},$$

wobei  $f, \ell$  und  $Z$  aus der Definition unseres optimalen Steuerungsproblems stammen. Hierbei betrachten wir in diesem Kapitel wieder das Maximierungsproblem und setzen unbeschränkten Kontrollwertebereich  $U = \mathbb{R}^l$  voraus. Darüberhinaus nehmen wir an, dass wir Ableitung und Erwartungswertbildung vertauschen können.

**Satz 7.1** Die optimalen Wertefunktionen  $V_t$  bzw.  $V_\infty$  und die Hamiltonfunktionen  $H^t$  seien stetig differenzierbar und  $(X_t, u_t)$  sei eine optimale Lösung. Dann existiert ein  $\mathbb{R}^n$ -wertiger

stochastischer Prozess  $\lambda_t$ , so dass für alle  $t \in \{0, \dots, T-1\}$  bzw.  $t \in \mathbb{N}_0$  die Gleichungen

$$E_t\{X_{t+1}\} = \left(\frac{\partial H^t}{\partial \lambda}\right)^T (X_t, \lambda_{t+1}, u_t) \quad (7.1)$$

$$E_t\{\lambda_t\} = \left(\frac{\partial H^t}{\partial x}\right)^T (X_t, \lambda_{t+1}, u_t) \quad (7.2)$$

$$0 = \left(\frac{\partial H^t}{\partial u}\right)^T (X_t, \lambda_{t+1}, u_t) \quad (7.3)$$

erfüllt sind. Für das Problem auf endlichem Horizont  $\{0, \dots, T\}$  erfüllt  $\lambda_T$  zusätzlich die Transversalitätsbedingung<sup>1</sup>

$$\lambda_T = \beta^T \frac{dL}{dx}(X_T).$$

**Beweis:** Wir beweisen das Prinzip für endlichen Horizont  $T$ , der Beweis für unendlichen Horizont verläuft analog.

Gleichung (7.1) folgt aus

$$\left(\frac{\partial H^t}{\partial \lambda}\right)^T (X_t, \lambda_{t+1}, u_t) = E_t\{f(X_t, u_t, Z_t)\}.$$

Zum Beweis der Gleichung (7.2) und der Transversalitätsbedingung setzen wir

$$\lambda_t = \beta^t \frac{dV_{T-t}}{dx}(X_t)^T$$

für alle  $t \in \mathbb{N}_0$ . Damit ist die Transversalitätsbedingung wegen  $V_0 = L$  erfüllt, zudem gilt

$$\begin{aligned} \left(\frac{\partial H^t}{\partial x}\right)^T (X_t, \lambda_{t+1}, u_t) &= E_t \left\{ \beta^t \frac{\partial \ell}{\partial x}(X_t, u_t) + \lambda_{t+1}^T \frac{\partial f}{\partial x}(X_t, u_t, Z_t) \right\} \\ &= E_t \left\{ \beta^t \frac{\partial \ell}{\partial x}(X_t, u_t) + \beta^{t+1} \frac{\partial V_{T-t-1}}{\partial x}(X_{t+1}) \frac{\partial f}{\partial x}(X_t, u_t, Z_t) \right\} \\ &= \frac{\partial}{\partial x} E_t \{ \beta^t \ell(X_t, u_t) + \beta^{t+1} V_{T-t-1}(f(X_t, u_t, Z_t)) \} \\ &= \frac{\partial}{\partial x} \max_{u \in U} E_t \{ \beta^t \ell(X_t, u) + \beta^{t+1} V_{T-t-1}(f(X_t, u, Z_t)) \} \\ &= E_t \left\{ \frac{\partial}{\partial x} \beta^t V_{T-t}(X_t) \right\} = E_t \{ \lambda_t^T \}. \end{aligned}$$

<sup>1</sup>Unter geeigneten Annahmen an das Problem kann man auch für unendlichen Horizont Transversalitätsbedingungen erhalten, z.B. die Bedingung  $\lim_{t \rightarrow \infty} E\{\lambda_t\} = 0$ .

Zum Beweis der letzten Gleichung (7.3) benutzen wir

$$\begin{aligned}
\left(\frac{\partial H^t}{\partial u}\right)(X_t, \lambda_{t+1}, u_t) &= E_t \left\{ \beta^t \frac{\partial \ell}{\partial u}(X_t, u_t) + \lambda_{t+1}^T \frac{\partial f}{\partial u}(X_t, u_t, Z_t) \right\} \\
&= E_t \left\{ \beta^t \frac{\partial \ell}{\partial u}(X_t, u_t) + \beta^{t+1} \frac{\partial V_{T-t-1}}{\partial x}(X_{t+1}) \frac{\partial f}{\partial u}(X_t, u_t, Z_t) \right\} \\
&= \frac{\partial}{\partial u} E_t \{ \beta^t \ell(X_t, u_t) + \beta^{t+1} V_{T-t-1}(f(X_t, u_t, Z_t)) \} = 0,
\end{aligned}$$

da  $u_t$  diesen Ausdruck ja gerade maximiert.  $\square$

**Bemerkung 7.2** (i) Beachte, dass das Prinzip in der hier vorgestellten einfachen Form für das Minimierungsproblem exakt die gleiche Form annimmt — letztlich charakterisiert es also nur extremale Trajektorien. Um sicher zu gehen, dass man tatsächlich das gewünschte Maximum (bzw. Minimum) erhält, muss man weitere Annahmen treffen, z.B. Konkavität (bzw. Konvexität) der rechten Seite des Optimalitätsprinzips.

(ii) Die Annahme der stetigen Differenzierbarkeit der optimalen Wertefunktion dient nur der Vereinfachung des Beweises und kann abgeschwächt werden.  $\square$

## 7.2 Die Perturbationsmethode

Im Gegensatz zum Prinzip der dynamischen Programmierung charakterisiert das Maximumprinzip kein Feedback  $F_t(x)$  bzw.  $F_\infty(x)$  sondern eine zeitabhängige Kontrolle  $u_t$ . Dies ist vom Prinzip her deutlich weniger aufwändig, was die Popularität dieses Prinzips für deterministische Probleme begründet. In der stochastischen Situation relativiert sich dieser Vorteil allerdings: Grund dafür ist, dass die durch das Prinzip charakterisierte optimale Kontrolle natürlich auf den “vergangenen Zufall” reagieren muss. Deswegen ist  $u_t$  ein stochastischer Prozess, also nicht nur eine Funktion in  $t$  sondern auch in  $\omega$  oder — äquivalent, vgl. die Bemerkungen 4.5(iii) und 4.12(i) — in  $X_t$ , und damit wieder ein Feedback. Während also im Deterministischen die Berechnung einer rein zeitabhängigen optimalen Kontrollfolge  $u_t$  durchaus sinnvoll ist, landet man im stochastischen Kontext unweigerlich wieder beim Feedback.

Trotzdem gibt es eine — insbesondere in der Ökonomie und dort besonders in der Analyse volkswirtschaftlicher Modelle — populäre Methode, die das Maximumprinzip nutzt, allerdings wieder um Feedbacks auszurechnen.

Wir beschreiben die Methode im Folgenden für Probleme auf unendlichem Horizont und illustrieren dabei jeden Schritt an Hand des sehr einfachen Problems gegeben durch

$$X_{t+1} = X_t + u_t + Z_t, \quad \text{also} \quad f(x, t, z) = x + u + z$$

mit (zu minimierenden) laufenden Kosten

$$\ell(x, u) = x^2 + u^2$$

und normalverteilten Zufallsvariablen  $Z_t \sim N(0, \sigma^2)$ . Dieses Problem kann als einfache ein-dimensionale Variante unseres Wagenproblems interpretiert werden, bei der die Kontrolle direkt auf die Position und nicht auf die Geschwindigkeit wirkt.

Für die Perturbationsmethode ändert man zunächst die Zufallsvariablen so ab, dass die Varianz  $\sigma^2$  als zusätzlicher Parameter explizit in  $f$  auftritt. In unserem Beispiel erreichen wir dass, indem wir  $Z_t \sim N(0, 1)$  wählen und die Dynamik zu

$$f(x, u, \sigma^2, z) = x + u + \sigma^2 z$$

ändern. Ebenso wird auch das optimale Feedback nun stets in Abhängigkeit von  $\sigma$  geschrieben, also  $F(x, \sigma^2)$  (den Index “ $\infty$ ” lassen wir hier der einfacheren Notation wegen wegfallen).

Wir berechnen nun zuerst einmal die Hamilton-Funktion für unser Beispiel mit der geänderten Dynamik. Diese lautet

$$H^t(x, \lambda, u, \sigma^2, z) = E_t\{\beta^t(x^2 + u^2) + \lambda(x + u + \sigma^2 z)\}$$

und Gleichungen (7.1)–(7.3) lauten

$$E_t\{X_{t+1}\} = E_t\{X_t + u_t + \sigma^2 Z_t\} = E_t\{X_t + u_t\} \quad (7.4)$$

$$E_t\{\lambda_t\} = E_t\{2\beta^t X_t + \lambda_{t+1}\} \quad (7.5)$$

$$0 = E_t\{2\beta^t u_t + \lambda_{t+1}\}. \quad (7.6)$$

Die Idee der Perturbationsmethode ist nun, das optimale Feedback

$$F(x, \sigma^2)$$

durch eine Taylor-Entwicklung

$$F(x, \sigma^2) \approx F(x^*, 0) + F_x(x^*, 0)(x - x^*) + F_{\sigma^2}(x^*, 0)\sigma^2 + (x - x^*)^T F_{xx}(x^*, 0)(x - x^*) + \dots$$

in einem Punkt  $x = x^*$  und in  $\sigma = 0$  zu approximieren, wobei  $F$  die partiellen Ableitungen nach den tiefgestellten Argumenten bezeichnen. Zur Berechnung dieser Approximation benötigen wir auch eine Taylor-Approximation der optimalen Dynamik

$$g(x, \sigma^2, z) = f(x, F(x, \sigma^2), \sigma^2, z),$$

die in der Methode gleich mitberechnet wird. Ebenso muss gegebenenfalls noch  $\lambda_t = \lambda(X_t, \sigma^2)$  mit approximiert werden (wie wir sehen werden, ist dies in unserem Beispiel nicht nötig).

Neben den Annahmen in Satz 7.1 müssen wir daher zusätzlich voraussetzen, dass alle auftretenden Funktionen so oft stetig differenzierbar sind, wie wir in der approximierenden Taylor-Entwicklung gehen wollen.

Der Punkt  $x^*$  wird nun so gewählt, dass er ein optimales Gleichgewicht für  $\sigma^2 = 0$  ist, d.h. ein Gleichgewicht für das optimal gesteuerte deterministische System. Ein Punkt  $x^*$  ist genau dann ein optimales Gleichgewicht, wenn für die optimale Lösung  $X_t \equiv x^*$  und

$u_t \equiv u^*$  gilt für alle  $t$ . Solche optimalen Gleichgewichte können aus dem Maximumprinzip errechnet werden: In unserem Beispiel folgt aus Gleichung (7.4)

$$x^* = x^* + u_t$$

weswegen die optimale Kontrolle im optimalen Gleichgewicht gleich Null sein muss. Daraus folgt mit Gleichung (7.6)

$$0 = 2\beta^t u_t + \lambda_{t+1} = \lambda_{t+1},$$

weswegen  $\lambda_t \equiv 0$  gelten muss. Eingesetzt in (7.5) ergibt sich damit schließlich

$$0 = 2\beta^t x^* + 0.$$

Einzig möglicher Kandidat für ein optimales Gleichgewicht ist demnach  $x^* = 0$  mit zugehöriger optimaler Kontrolle  $F(0, 0) = u^* = 0$ .

Die Perturbationsmethode funktioniert nun wie folgt:

- (1) Finde ein optimales Gleichgewicht  $x^*$  für  $\sigma^2 = 0$ .
- (2) Schreibe die Gleichungen (7.1)–(7.3) kompakt in der Form

$$E_t\{G(X_t, X_{t+1}, u_t, u_{t+1}, \lambda_t, \lambda_{t+1})\} = 0$$

und vereinfache sie dabei so weit wie möglich.

- (3) Setze in diese Gleichung für  $t = 0$  und  $X_0 = x$  die zu berechnenden Funktionen  $g$  und  $F$  ein. Damit erhalten wir die notwendige Optimalitätsbedingung

$$\begin{aligned} E\{G(x, g(x, \sigma^2, Z), F(x, \sigma^2), F(g(x, \sigma^2, Z), \sigma^2), \\ \lambda(x, \sigma^2), \lambda(g(x, \sigma^2, Z)))\} = 0, \end{aligned} \quad (7.7)$$

also kurz  $E\{G\} \equiv 0$ .

- (4) Aus (7.7) folgt nun, dass für alle Ableitungen ebenfalls

$$\frac{\partial}{\partial x} E\{G\} \equiv 0, \quad \frac{\partial}{\partial(\sigma^2)} E\{G\} \equiv 0, \quad \frac{\partial^2}{\partial x^2} E\{G\} \equiv 0, \quad \frac{\partial^2}{\partial x \partial(\sigma^2)} E\{G\} \equiv 0, \dots \quad (7.8)$$

gelten muss. Diese Ausdrücke enthalten wegen der Kettenregel die Ableitungen von  $F$ ,  $g$  und  $\lambda$ , bilden also Gleichungssysteme für diese Ableitungen. Diese werden in  $x = x^*$  und  $\sigma^2 = 0$  gelöst und liefern — Lösbarkeit vorausgesetzt — die gesuchten Ableitungen für die Taylor-Approximation.

Wir führen diesen Algorithmus nun für unser Beispiel und die ersten zwei Ableitungsgleichungen  $E\{G_x\} \equiv 0$  und  $E\{G_{\sigma^2}\} \equiv 0$  durch.

Schritt (1) haben wir oben bereits durchgeführt. Für Schritt (2) vereinfachen wir (7.4)–(7.6) zunächst: Dazu ziehen wir (7.6) von (7.5) ab und erhalten

$$E_t\{\lambda_t\} = E_t\{2\beta^t(X_t - u_t)\}.$$

Daraus folgt mit den Rechenregeln für  $E_t$  auch

$$E_{t-1}\{\lambda_t\} = E_{t-1}\{2\beta^t(X_t - u_t)\}.$$

Andererseits folgt aus (7.6) mit  $t - 1$  an Stelle von  $t$  die Gleichung

$$E_{t-1}\{\lambda_t\} = E_{t-1}\{-2\beta^{t-1}u_{t-1}\},$$

was zusammen

$$E_{t-1}\{-2\beta^{t-1}u_{t-1}\} = E_{t-1}\{2\beta^t(X_t - u_t)\}$$

oder — nun wieder für  $t$  statt  $t - 1$  geschrieben —

$$E_t\{u_t\} = E_t\{\beta(u_{t+1} - X_{t+1})\} \quad (7.9)$$

ergibt. Aus Gleichungen (7.5) und (7.6) folgt also (7.9), weswegen wir als vereinfachte notwendige Optimalitätsbedingungen

$$0 = E_t \left( \underbrace{\begin{pmatrix} X_t + u_t - X_{t+1} \\ u_t - \beta(u_{t+1} - X_{t+1}) \end{pmatrix}}_{=: G(X_t, X_{t+1}, u_t, u_{t+1})} \right)$$

erhalten.

Gleichung (7.7) aus Schritt (3) lautet für unser Beispiel also

$$E \left( \begin{pmatrix} x + F(x, \sigma^2) - g(x, \sigma^2, Z) \\ F(x, \sigma^2) - \beta(F(g(x, \sigma^2, Z), \sigma^2) - g(x, \sigma^2, Z)) \end{pmatrix} \right) = 0 \quad (7.10)$$

Schritt (4): Damit ergeben sich die ersten beiden Gleichungen (Ableitung nach  $x$  und  $\sigma^2$ ) unter Berücksichtigung der Kettenregel zu

$$0 = E \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ \beta \end{pmatrix} g_x(x, \sigma^2, Z) + \begin{pmatrix} 1 \\ 1 \end{pmatrix} F_x(x, \sigma^2) \right. \\ \left. + \begin{pmatrix} 0 \\ -\beta \end{pmatrix} F_x(g(x, \sigma^2, Z), \sigma^2) g_x(x, \sigma^2, Z) \right\}$$

und

$$0 = E \left\{ \begin{pmatrix} -1 \\ \beta \end{pmatrix} g_{\sigma^2}(x, \sigma^2, Z) + \begin{pmatrix} 1 \\ 1 \end{pmatrix} F_{\sigma^2}(x, \sigma^2) \right. \\ \left. + \begin{pmatrix} 0 \\ -\beta \end{pmatrix} \left( F_x(g(x, \sigma^2, Z), \sigma^2) g_{\sigma^2}(x, \sigma^2, Z) + F_{\sigma^2}(x, \sigma^2) \right) \right\}.$$

Um den Erwartungswert loszuwerden, müssen wir die Funktion  $g$  geeignet zerlegen: es gilt

$$g(x, \sigma^2, Z) = \underbrace{x + F(x, \sigma^2)}_{=: h(x, \sigma^2)} + \sigma^2 Z,$$

weswegen

$$g_x(x, \sigma^2, Z) = h_x(x, \sigma^2) \quad \text{und} \quad g_{\sigma^2}(x, \sigma^2, Z) = h_{\sigma^2}(x, \sigma^2) + Z$$

gilt. Wegen  $E\{Z\} = 0$  und  $g(0, 0, Z) = 0$  (denn  $x^* = 0$  ist ja gerade das optimale Gleichgewicht für  $\sigma^2 = 0$ ) vereinfachen sich die obigen Gleichungen in  $(x, \sigma^2) = (0, 0)$  daher zu

$$0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ \beta \end{pmatrix} h_x(0, 0) + \begin{pmatrix} 1 \\ 1 \end{pmatrix} F_x(0, 0) + \begin{pmatrix} 0 \\ -\beta \end{pmatrix} F_x(0, 0) h_x(0, 0)$$

und

$$0 = \begin{pmatrix} -1 \\ \beta \end{pmatrix} h_{\sigma^2}(0, 0) + \begin{pmatrix} 1 \\ 1 \end{pmatrix} F_{\sigma^2}(0, 0) + \begin{pmatrix} 0 \\ -\beta \end{pmatrix} (F_x(0, 0) h_{\sigma^2}(0, 0) + F_{\sigma^2}(0, 0)).$$

Diese Gleichungen mit den Unbekannten  $F_x(0, 0)$ ,  $F_{\sigma^2}(0, 0)$ ,  $g_x(0, 0)$  und  $g_{\sigma^2}(0, 0)$  müssen wir nun lösen. Dazu kann man sie in ein verallgemeinertes Eigenwertproblem umwandeln, das sich mit Methoden der numerischen Linearen Algebra lösen lässt (z.B. mit MATLAB, vgl. Stephanie Becker, Numerische Lösung dynamischer Gleichgewichtsmodelle, Diplomarbeit 2005, Universität Bayreuth, [www.math.uni-bayreuth.de/~lgruene/diplom/](http://www.math.uni-bayreuth.de/~lgruene/diplom/)).

In unserem Fall können wir die Gleichungen aber auch einfach direkt auflösen. Die zweite Gleichung besitzt offensichtlich die (für  $F_x(0, 0) \neq 1/\beta$  auch eindeutige) Lösung

$$h_{\sigma^2}(0, 0) = F_{\sigma^2}(0, 0) = 0.$$

Aus der ersten Zeile der ersten Gleichung erhalten wir

$$h_x(0, 0) = F_x(0, 0) + 1,$$

was eingesetzt in die zweite Zeile

$$\beta + F_x(0, 0) - \beta F_x(0, 0)^2 = 0$$

ergibt. Diese quadratische Gleichung besitzt zwei Lösungen, nämlich

$$\frac{1 \pm \sqrt{1 + 4\beta^2}}{2\beta}.$$

Warum gibt es hier zwei Lösungen für den linearen Koeffizienten  $F_x$  der Taylor-Entwicklung des optimalen Feedbacks? Dies kommt daher, dass wir ein linear quadratisches Problem betrachten, dessen Optimalitätsprinzip (also gerade die Riccati-Gleichung) immer mehrere Lösungen besitzt. Zwar gibt es nur eine eindeutige positiv definite Lösung der Riccati-Gleichung, aber i.A. viele weitere. Daher existieren auch mehrere  $\lambda_t$ , für die die Gleichungen des Maximumprinzips erfüllt sind und damit mehrere Feedbacks.

Im Gegensatz zum Ansatz über die Riccati-Gleichung ist es bei der Perturbationsmethode im Falle einer nicht-eindeutigen Lösung im Allgemeinen nicht so einfach, den "richtigen" Feedback-Koeffizienten  $F_x$  zu finden — in unserem Beispiel ist es natürlich derjenige, der die Lösung in Richtung Null steuert, also

$$F_x = \frac{1 - \sqrt{1 + 4\beta^2}}{2\beta},$$

wodurch wir das approximative Feedback

$$F(x, \sigma^2) = F(x^*, 0) + F_x(x^*, 0)(x - x^*) + F_{\sigma^2}(x^*, 0)\sigma^2 = \frac{1 - \sqrt{1 + 4\beta^2}}{2\beta}x$$

erhalten. Für linear-quadratische Probleme ist die Perturbationsmethode also nicht besonders geeignet — allerdings ist sie dafür auch nicht entwickelt worden; wir haben sie hier nur der Einfachheit halber auf ein solches Problem angewendet. Für andere Arten optimaler Steuerungsprobleme kann man durchaus eindeutige  $F_x$  erhalten.

Zudem ist es für allgemeine nichtlineare Probleme natürlich auch sinnvoll, durch Betrachtung der höheren Ableitungen von  $G$  auch Koeffizienten höherer Ordnung in der Taylor-Entwicklung zu erhalten. Interessanterweise führen z.B. die aus den zweiten Ableitungen von  $G$  erhaltenen Gleichungen dabei auf ein lineares Gleichungssystem und nicht auf ein Eigenwertproblem, also auf eine numerisch leichter lösbare Aufgabe, siehe z.B. die oben erwähnte Diplomarbeit.

# Anhang A

## Der Gittergenerator GRIDGEN

In diesem Anhang findet sich eine kurze Beschreibung der für die Übungen zur Verfügung gestellten *C*-Routinen zur Implementierung der Gittererzeugung und der affin multilinearen Funktionen, im Folgenden kurz *Gitterfunktionen* genannt.

Im Gegensatz zu anderen Implementierungen kann der Gittergenerator beliebige unregelmäßige Gitter mit Rechteckelementen (bzw. deren höherdimensionale Analoga) realisieren. Der “Preis” für diese Flexibilität ist, dass man auf die Gitterknoten nicht wie in einem Array direkt über Indizes zugreifen kann, sondern diese über Interfacefunktionen ansprechen muss.

### A.1 Routinen für regelmäßige Gitter

Ein Gitter inklusive der zugehörigen Gitterfunktion wird in der Datenstruktur `qgrid` gespeichert. Jedes in einem Programm verwendete Gitter wird daher durch einen Zeiger vom Typ `qgrid*` repräsentiert.

#### A.1.1 Erzeugen und Löschen von Gittern

Ein regelmäßiges Gitter auf einem Rechteckgebiet wird durch die Routine

```
qgrid *create_grid(double *lo, double *hi, double *Delta,  
                  int dim, int level);
```

erzeugt. Hierbei ist `dim` die Dimension des Gebietes. Die Variablen `lo` und `hi` sind `dim`-dimensionale Arrays, die die untere und obere Ecke des Gebietes definieren. Im Array `Delta` ist die gewünschte Elementgröße für jede Koordinatenrichtung angegeben. Die Variable `level` ist für die Erzeugung unregelmäßiger Gitter wichtig und wird in Abschnitt A.2 besprochen. Für regelmäßige Gitter kann sie immer auf 0 gesetzt werden. Die Gitterfunktion auf einem neu erzeugten Gitter ist immer konstant gleich 0.

Die Routine gibt einen Zeiger auf eine `qgrid`-Struktur zurück, der einer Variablen vom Typ `qgrid*` zugewiesen werden muss, da dieser Zeiger für alle zukünftigen Zugriffe auf Gitter und Gitterfunktion benötigt wird.

**Beispiel:** Erzeugen eines 2d Gitters auf  $X = [0, 1]^2$ :

```
double lo[2], hi[2], Delta[2];
qgrid *g;

lo[0] = 0; lo[1] = 0;
hi[0] = 1; hi[1] = 1;
Delta[0] = 0.2; Delta[1] = 0.2;

g = create_grid(lo, hi, Delta, 2, 0);
```

Zum Löschen eines so erzeugten Gitters dient die Anweisung

```
void delete_grid(qgrid *g);
```

### A.1.2 Zugriff auf Knoten und Gitterfunktionen

Die Gitterknoten (oder Eckpunkte) werden nicht direkt wie in einem Array sondern indirekt über Interface-Funktionen angesprochen. Hierzu dienen die Funktionen

```
int first_node(qgrid *g, double *x);
int next_node(qgrid *g, double *x);
```

Beim Aufruf dieser Funktionen geschieht das Folgende:

- ein interner (nach außen unsichtbarer) Zeiger wird auf den ersten bzw. nächsten Knoten in der Knotenliste des Gitters  $g$  gesetzt. Der Knoten, auf den dieser Zeiger zeigt, wird im Folgenden als *aktueller Knoten* (engl. *current node*) bezeichnet.
- Falls der übergebene Zeiger  $x$  nicht den Wert Null hat, wird davon ausgegangen, dass dieser als Array  $x[0], \dots, x[\text{dim}-1]$  definiert ist, in dem die Koordinaten des aktuellen Knotens zurück gegeben werden
- Als Funktionswert wird der Index (immer  $\geq 0$ ) des aktuellen Knotens zurück gegeben und -1, falls alle Knoten der Liste bereits abgearbeitet wurden. Dies dient als Abbruchkriterium, wenn die Knotenliste in einer Schleife durchlaufen wird.

**Beispiel:** Durchlaufen der Knotenliste in einer `do...while` Schleife:

```
index = first_node(g, x);
do
{
    ...      /* Operationen auf dem aktuellen Knoten */

    index = next_node(g, x);
}
while (index!=-1);
```

Die Basisoperationen auf dem aktuellen Knoten lauten:

```
double current_nodevalue(qgrid *g);
int set_current_nodevalue(qgrid *g, double v);
```

Mit `current_nodevalue` kann der Knotenwert der Gitterfunktion abgefragt werden, mit `set_current_nodevalue` wird er auf den Wert `v` gesetzt. Die Funktion gibt im Erfolgsfalle 0 und ansonsten 1 zurück. Dies passiert, wenn `set_current_nodevalue` ohne vorherigen Aufruf von `first_node` aufgerufen wurde oder wenn alle Knoten der Liste bereits abgearbeitet wurde. In diesen Fällen ist das Ergebnis von `current_nodevalue` undefiniert.

Mit `current_nodevalue` kann nur auf die Werte der Gitterfunktion in den Knoten zugegriffen werden. Für beliebige Punkte `x` liefert die Routine

```
double value(qgrid *g, double *x, int *flag);
```

den Wert in `x`. Wird dieser Routine ein nicht-Null-Zeiger `flag` übergeben, so wird in dieser Variablen der Status der Operation zurück gegeben: 0 bedeutet kein Fehler, 1 zeigt an, dass `x` außerhalb des Gitters liegt und 2 zeigt andere Fehler an, z.B. die Übergabe eines falschen Zeigers `g`.

**Beispiel:** Abfrage des Wertes für  $x = (0.5, 0.2)^T$ :

```
int flag;
qgrid *g;
int flag;

...

x[0] = 0.5; x[1] = 0.2;

wert = value(g, x, &flag);
```

**Beachte:** Wenn Knotenwerte abgefragt werden sollen, ist `current_nodevalue` deutlich schneller als `value`.

### A.1.3 Ein- und Ausgabe

Zum Speichern und Laden einer Gitterfunktion inklusive des Gitters dienen die Anweisungen

```
int save_val(qgrid *g, char *filename);
qgrid *load_val(char *filename);
```

Die Routine `save_val` gibt im Erfolgsfall 0 und sonst 1 zurück, `load_val` gibt im Falle eines Fehlers einen Null-Zeiger zurück und eine Fehlermeldung auf dem Bildschirm aus.

**Beispiel:** Speichern eines Gitters `g` und Laden des Gitters in `g_neu`:

```
save_val(g, "meingitter.val");

g_neu = load_val("meingitter.val");
```

Zum Exportieren nach MATLAB steht die Anweisung

```
int export_gridandval(qgrid *g, char *filename);
```

zur Verfügung. Die Routine gibt im Erfolgsfall 0 und sonst 1 zurück,

**Beispiel:** `export_gridandval(g, "gitter.asc");`

Die so gespeicherte Datei kann mit den unten beschriebenen MATLAB Routinen grafisch dargestellt werden.

## A.2 Routinen für unregelmäßige Gitter

GRIDGEN erlaubt das Einfügen weiterer Gitterpunkte in ein bestehendes Gitter. Ebenso können zuvor eingefügte Gitterpunkte jederzeit wieder entfernt werden.

### A.2.1 Die Adaptionssiteration

Das Grundschema eines Gitteradaptionsschritts ist wie folgt:

- (1) Erzeugung von Testpunkten
- (2) Schleife über alle Testpunkte:
  - (3) Markierung der Testpunkte, die eingefügt werden sollen
- (4) Schleife über alle Gitterpunkte:
  - (5) Markierung der Gitterpunkte, die entfernt werden sollen
- (6) Einfügen bzw. Entfernen der markierten Punkte

Im Folgenden werden die Routinen für die einzelnen Schritte beschrieben.

#### (1) Erzeugung von Testpunkten

Zur Erzeugung der Testpunkte dient die Routine

```
void prepare_adaptation(qgrid *g);
```

Jeder Testpunkt hat — genau wie die Gitterpunkte — einen Wert. Bei der Erzeugung der Testpunkte wird dieser mit dem interpolierten Wert der Gitterfunktion an der entsprechenden Stelle initialisiert.

Die erzeugten Testpunkte entsprechen gerade den Gitterpunkten, die durch eine einmalige Verfeinerung jedes Elementes in alle Koordinatenrichtungen erzeugt werden, vgl. Abbildung A.1 in 2d.

In Abbildung A.2 ist beispielhaft illustriert, zu welchen Verfeinerungen das Einfügen der dargestellten Testpunkte führt.

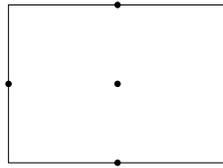


Abbildung A.1: Testpunkte eines Elements in 2d

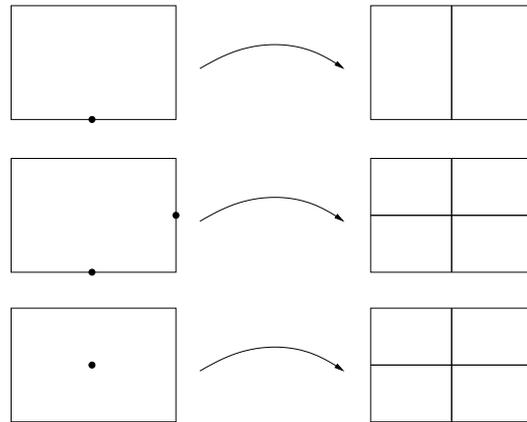


Abbildung A.2: Einzufügende Testpunkte und resultierende Verfeinerung

### (2) Schleife über die Testpunkte

Die Schleife über die Testpunkte wird völlig analog zur Schleife über die Gitterpunkte mit den Funktionen

```
int first_testpoint(qgrid *g, double *x);
```

```
int next_testpoint(qgrid *g, double *x);
```

in einer `do...while`-Schleife programmiert. Intern wird hier der gleiche Zeiger wie für die Gitterpunkte benutzt, weswegen die Routinen

```
double current_nodevalue(qgrid *g);
```

```
int set_current_nodevalue(qgrid *g, double v);
```

aus Abschnitt A.1.2 auch in der Testpunktschleife verwendet werden können.

### (3) Markierung der Testpunkte

Der aktuelle Testpunkt wird mit der Anweisung

```
void insert_current_testpoint(qgrid *g);
```

als "einzufügend" markiert. Um den Fehler in einem Testpunkt zu messen, muss sein interpolierter Wert mit dem gewünschten Wert verglichen werden. Da der interpolierte Wert ja gerade der initialisierte Wert des Testpunkts ist, kann er mit

```
real current_nodevalue(qgrid *g);
```

ausgelesen werden.

Um zu vermeiden, dass die Knotenwerte nach Beenden der Verfeinerung noch einmal aktualisiert werden müssen, kann der Wert des Testpunktes bereits in der Fehlerberechnungsschleife (aber nach der Abfrage des initialisierten interpolierten Wertes, da dieser dadurch überschrieben wird!) mittels der bereits bekannten Anweisung

```
int set_current_nodevalue(qgrid *g, double v);
```

mit dem gewünschten Wert  $v$  belegt werden. Da in der eigentlichen Verfeinerung des Gitters auch Testpunkte eingefügt werden, die nicht markiert wurden (vgl. Punkt (6)), sollte dies stets für alle Testpunkte gemacht werden.

#### (4) Schleife über die Gitterpunkte

siehe Abschnitt A.1.2

#### (5) Markierung der Gitterpunkte

Mit der Anweisung

```
void remove_current_node(qgrid *g);
```

wird der aktuelle Gitterpunkt als “zu entfernen” markiert.

Um festzustellen, ob ein Gitterpunkt entfernt werden soll, muss sein Wert verglichen werden mit dem Wert der Gitterfunktion auf dem um diesen Punkt vergrößerten (“coarsened”) Gitter.

Den Wert des Gitterpunktes erhält man dazu wie üblich mit

```
double current_nodevalue(qgrid *g);
```

und den Wert auf dem vergrößerten Gitter mit

```
double current_coarsevalue(qgrid *g);
```

#### (6) Einfügen bzw. Entfernen der markierten Punkte

Dies geschieht schließlich mit der Anweisung

```
void complete_adaptation(qgrid *g);
```

Die Anweisung fügt alle entsprechend markierten Testpunkte mit ihren Werten in das Gitter ein und entfernt die markierten Gitterpunkte. Im Konfliktfall hat “Einfügen” Priorität über “Entfernen”. Zudem wird sicher gestellt, dass sich zwei benachbarte Elemente an den berührenden Hyperebenen nur in maximal einer Koordinatenrichtung und nur um maximal eine Verfeinerungsstufe unterscheiden. Dies vermeidet zum einen degenerierte Gitter und bietet zum anderen eine gewisse Sicherheit gegen Ungenauigkeiten der Fehlerschätzung, die aus der relativ groben Verteilung der Testpunkte entsteht. Durch diesen Mechanismus werden auch Testpunkte in das Gitter eingefügt, die nicht explizit als “einzufügen” markiert wurden.

Die Routine `complete_adaptation(g)` gibt am Bildschirm eine Statistik über die Veränderungen des Gitters im abgeschlossenen Adaptionsschritt aus.

### A.2.2 Hängende Knoten

Wenn zwei unterschiedlich feine Elemente zusammenstoßen, entstehen sogenannte *hängende Knoten*. Dies sind Gitterpunkte, die nicht für alle angrenzenden Elemente Eckpunkte sind. Um die Stetigkeit der Gitterfunktion zu garantieren, müssen die Werte in diesen Knoten durch Interpolation bestimmt werden. Diese Interpolation wird beim Abschließen der Adaption durch `complete_adaptation(g)` automatisch durchgeführt.

Werden aber auf einem unregelmäßigen Gitter Knotenwerte in der üblichen Gitterpunkt-Schleife zugewiesen, ohne dass danach eine Adaption durchgeführt wird, muss diese Interpolation manuell ausgelöst werden. Dazu dient die Anweisung

```
void update_values(qgrid *g);
```

### A.2.3 Entfernen von Gitterpunkten aus dem Startgitter

Mit `remove_current_node(g)` können nur Knoten entfernt werden, die in einer vorherigen Adaption mit `insert_current_testpoint(g)` eingefügt wurden. Um trotzdem das Entfernen von Knoten aus dem Startgitter zu erlauben, kann dieses mit `create_grid` mit `level ≥ 1` erzeugt werden. Dies bewirkt, dass das Startgitter aus bereits `level` mal verfeinerten Elementen besteht. Dies schränkt den Spielraum der Elementgröße ein, weil die Anzahl der Elemente in jeder Koordinatenrichtung ein Vielfaches von  $2^{\text{level}}$  sein muss, erhöht dafür aber die Flexibilität in der Adaptionroutine.

### A.2.4 Anisotrope Gitter

Unter einer anisotropen Verfeinerung wird die mehrfache Verfeinerung von Elementen in der gleichen Koordinatenrichtung verstanden, vgl. Abb. A.3.

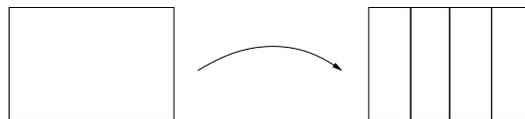


Abbildung A.3: Anisotrope Verfeinerung in  $x_1$ -Richtung

Diese Art von Verfeinerung kann bei gleicher Genauigkeit eine Menge Gitterpunkte (und damit Speicher und Rechenzeit) sparen. Sie kann andererseits aber auch zu numerischen Problemen führen, da Gitterfunktionen auf anisotropen Gittern große Lipschitz-Konstanten aufweisen können. Erfahrungsgemäß überwiegt aber meist der positive Aspekt.

Um anisotrope Gitter zu erzeugen, dürfen nur Testpunkte auf den eindimensionalen Kanten der Elemente mit `insert_current_testpoint(g)` eingefügt werden, denn ansonsten erfolgt automatisch eine Verfeinerung in mehrere Richtungen, vgl. das dritte Beispiel in Abbildung A.2. Um diese zu identifizieren, liefert

```
int current_testpointtype(qgrid *g);
```

die Dimension der Elementkomponente, auf der sich der aktuelle Testpunkt befindet. Also 0 = Ecke (= Gitterpunkt  $\neq$  Testpunkt), 1 = Kante, 2 = Fläche, ...

Die einfachste Implementierung einer anisotropen Verfeinerung besteht darin, den Fehler nur in Testpunkten mit `current_testpointtype(g) = 1` zu messen. Man kann aber auch den Fehler in allen Testpunkten messen und dann abhängig vom Verhältnis der Fehler in den verschiedenen Koordinatenrichtungen entscheiden, welche Testpunkte man in das Gitter aufnehmen möchte.

### A.2.5 Beispielprogramm

Auf der Vorlesungshomepage steht das Beispielprogramm `adaptivbeispiel.c` zur Verfügung, in dem eine komplette Adaptionsiteration realisiert und ausführlich kommentiert ist. Das Compilieren erfolgt analog zur Routine `beispiel.c`, vgl. Abschnitt A.4.3.

## A.3 Grafische Darstellung

Zur grafischen Darstellung stehen die MATLAB-Routinen

```
gridplot1d(filename)
```

```
gridplot2d(filename)
```

```
gridplot3d(filename, vec, off)
```

zur Verfügung. Die 1d und 2d Variante stellen die Funktion inklusive des Gitters in “üblicher” Weise dar. Der 3d Version wird zusätzlich ein dreidimensionaler Vektor `vec` und ein Wert `off` übergeben, durch die eine 2d Hyperebene im  $\mathbb{R}^3$  definiert wird. Grafisch dargestellt wird dann der Schnitt durch das Gitter entlang dieser Hyperebene, in dem die Funktionswerte der Gitterfunktion farbig dargestellt werden. Wird der Nullvektor übergeben, so werden alle Elemente dargestellt (was allerdings bereits bei relativ groben Gittern ziemlich lange dauert).

## A.4 Technische Hinweise

Zur Verwendung der oben beschriebenen C-Routinen müssen drei Schritte durchgeführt werden.

### A.4.1 Herunterladen der Routinen

Zur Verwendung der C-Routinen müssen Sie dazu die Files `gridgen.c` und `gridgen.h` von der Vorlesungs-Webpage<sup>1</sup> herunterladen. Hier finden sich auch die MATLAB-Routinen zur grafischen Darstellung und die Beispielprogramme.

<sup>1</sup>[www.math.uni-bayreuth.de/~lgruene/stochdynopt07/](http://www.math.uni-bayreuth.de/~lgruene/stochdynopt07/)

### A.4.2 Einbinden der Header-Datei

Um die Routinen in Ihrem Programm verwenden zu können, müssen Sie im Kopf die Zeile `#include "gridgen.h"` einfügen.

### A.4.3 Compilieren und Linken

Nehmen wir an, Ihr Programm ist im File `meinprogramm.c` gespeichert, und Sie wollen daraus ein ausführbares Programm mit Namen `meintest` erzeugen. Dazu müssen Sie die folgenden Befehle ausführen

(a) Übersetzung der Routinen:

```
gcc -c gridgen.c
```

(Dies muss nach dem Herunterladen nur einmal durchgeführt werden.)

(b) Übersetzen Ihres Programms:

```
gcc -c meinprogramm.c
```

(c) Zusammenbinden (Linken) der Programme:

```
gcc meinprogramm.o gridgen.o -lm -o meintest
```

Wenn alles fehlerfrei geklappt hat, können Sie Ihr Programm nun mit `meintest` bzw. — je nach Einstellung des Pfades — mit `./meintest` starten.

Auf der Vorlesungshomepage finden Sie unter `beispiel.c` ein Programm, mit dem Sie diese Schritte testen können.

# Index

- affin bilineare Funktionen, 40
- Approximationsfehler, 43
  
- bedingte Wahrscheinlichkeit, 3
- bedingter Erwartungswert, 3
- Binomialmodell, 7
  
- Dichtefunktion, 3
- Dynamische Programmierung, 23, 31
  - Operator, 37
  
- Elementarereignis, 1
- Ertragsfunktion, 19
- Erwartungswert, 3
  - bedingt, 3
  - verallgemeinert, 12
  
- Feedback, 15, 28
- Fehlerschätzer, 47
- Filtration, 10, 12
  - induziert, 12
  - natürlich, 12
  
- Gitter, 40
  - adaptiv, 48
- Gitterfunktionen, 40
- Gittergenerator, 69
- Gleichverteilung, 2
  
- Hamilton-Funktion, 61
  
- Konkatenation, 16
- Kontrollbeschränkung, 19
- Kontrollprozess, 13
  - in verallgemeinerter Feedbackform, 15
  - optimal, 22
    - Charakterisierung, 27, 34
    - zulässig, 14
- Kovarianzmatrix, 3
  
- Lösung
  - optimal, 22
- Linear-quadratisches Problem, 51
  
- Maximumprinzip, 61
- messbar, 2
  
- optimale Lösung, 22
- optimale Wertefunktion, 22
- Optimalitätsprinzip, 23, 31
  
- Perturbationsmethode, 63
- Pfad, 6
- Portfoliooptimierung, 20, 26, 28, 42
- Prinzip der Dynamischen Programmierung, 23, 31
- Projektionsoperator, 37
  
- Realisierung, 3
- Rechteckgitter, 40
- Riccati-Gleichung, 55
- Rückkopplung, 28
  
- $\sigma$ -Algebra, 1
  - minimale, 2
- stochastischer Prozess, 6
  - adaptiert, 12
- stochastisches dynamisches Optimierungsproblem, 19
  - linear-quadratisch, 51
- stochastisches dynamisches System, 6
- stochastisches Kontrollsystem, 13
  
- Transversalitätsbedingung, i, 32
  
- Varianz, 3
  
- Wagenstabilisierung, 20, 58
- Wahrscheinlichkeit
  - bedingt, 3
- Wahrscheinlichkeitsmaß, 1
- Wahrscheinlichkeitsraum, 1

Wahrscheinlichkeitsverteilung, 2

Wertefunktion  
optimal, 22

Zufallsvariable, 2

iid, 7

unabhängig, 5

Zustandsbeschränkung, 19