§10 Das kontinuierliche Entscheidungsprozess

10.1 Einführung

■ Wir behandeln nun das Problem des Werkmeisters.

Nehmen wir an, unser Werkmeister müsse sich für eine Unterhalts- und Reparaturpolitik für die Maschine entscheiden.

Wenn das System im Zustand 1, d.h. in Betrieb ist, muss der Werkmeister entscheiden, welche Art von Unterhalt er anwenden will. Wir nehmen an, dass, wenn er ein gewöhnliches Unterhaltsverfahren anwendet, die Anlage \$6 pro Zeiteinheit einbringt und eine Wahrscheinlichkeit von $5\,dt$ für einen Ausfall in der kurzen Zeit dt aufweist. Das ist dasselbe, wie wenn wir sagen würden, dass die Dauer der Arbeitsintervalle der Maschine exponential mit einem Mittelwert von $\frac{1}{5}$ verteilt ist.

Der Werkmeister hat aber auch die Wahl, sich für ein kostspieligeres Unterhaltsverfahren zu entschliessen, das zwar den Erlös pro Zeiteinheit auf \$4, jedoch auch die Wahrscheinlichkeit eines Ausfalls in der Zeit dt auf $2\,dt$ herabsetzt.

Bei diesen beiden Unterhaltsverfahren entstehen keine Kosten in Verbindung mit dem Ausfall als solchem. Wenn wir die beiden Strategien im Zustand 1 mit 1 und 2 numerieren, dann ergibt sich für die erste Strategie

$$a_{12}^1 = 5, \quad c_{11}^1 = 6, \quad c_{12}^1 = 0$$

und für die zweite Strategie

$$a_{12}^2 = 2, \quad c_{11}^2 = 4, \quad c_{12}^2 = 0.$$

Schliesslich ergibt sich aus Gleichung (9.3)

$$q_i = c_{i\,i} + \sum_{j \neq i} a_{i\,j}\, c_{i\,j}$$

$$q_1^1 = 6 \text{ und } \quad q_1^2 = 4.$$

Nun müssen wir uns vor Auge halten, was geschehen kann, wenn die Maschine ausser Betrieb ist und das System sich im Zustand 2 befindet. Wir nehmen an, dass der Werkmeister auch in diesem Zustand zwei Strategien hat. Erstens kann er die Reparaturarbeiten durch seine eigenen Leute ausführen lassen. Bei dieser Strategie kostet die Reparatur \$1 pro Zeiteinheit an Arbeitszeit zuzüglich \$0.50 fixe Kosten pro Ausfall. Ferner besteht eine Wahrscheinlichkeit von $4\,dt$, dass die Maschine in einer kurzen Zeit dt repariert wird (die Reparaturzeit ist exponential verteilt mit einem mittel von $\frac{1}{4}$). Die Parameter dieser Strategie lauten somit

$$a_{21}^1 = 4, \quad c_{22}^1 = -1, \quad c_{21}^1 = -0.5,$$

und unter Anwendung der Gleichung (9.3)

$$q_i = c_{i\,i} + \sum_{j \neq i} a_{i\,j}\, c_{i\,j}$$

haben wir

$$q_2^1 = -1 + 4 \cdot (-0.5) = -3.$$

Wenn die Maschine still steht, besteht die zweite Strategie für den Werkmeister darin, eine aussenstehende Reparaturfirma zu benützen. Für diese Strategie betragen die festen Kosten pro Ausfall ebenfalls \$0.50. Doch

diese arbeiter kosten pro Zeiteinheit \$1.50, aber die Wahrscheinlichkeit einer Reparatur in der Zeit $\mathrm{d}t$ steigt auf $7\,\mathrm{d}t$. Für diese Alternative gilt daher

$$a_{21}^2 = 7, \quad c_{22}^2 = -1.5, \quad c_{21}^2 = -0.5$$

und

$$q_2^2 = -1.5 + 7\cdot(-0.5) = -5.$$

Nun muss der Werkmeister sich für eine Strategie in jedem Zustand entscheiden, die seinen Verdienst auf lange Sicht maximiert. Die Daten für dieses Problem sind in der Tabelle zusammengefasst:

Zustand i	Strategie k	a_{i1}^k	a_{i2}^k	Erlösrate q_i^k
1 Laufende Anlage	1 (gewöhnliche Unterhalt)	−5	5	6
1	2 (teurer Unterhalt)	−2	2	4
2 Anlage ausser Betrieb	1 (Reparatur in eigener Werkstatt)	4	−4	−3
2	2 (Reparatur ausserhalb)	7	−7	−5

Die Strategie-, Entscheidungs- und Politikbegriffe werden vom diskreten Fall übernommen. Da jede der vier möglichen Politiken, die in dieser Tabelle enthalten sind, einen Prozess mit einem ergodischen Klasse darstellt, hat jede einen eindeutigen Gewinn, der vom Anfangszustand des Systems unabhängig ist. Der Werkmeister möchte die Politik mit dem höchsten Gewinn feststellen. Das ist die optimale Politik.

Ein Weg, die optimale Politik festzustellen, besteht darin, den Gewinn für jede der vier Politiken zu suchen und nachzusehen, welcher Gewinn der höchste ist. Obschon das für kleinere Probleme möglich ist, ist das nicht durchführbar für Probleme mit vielen Zuständen und vielen Strategien in jedem Zustand.

10.1.1 Bemerkung: Man beachte auch, dass die Wertiterationsmethode, die für diskrete Prozesse anwendbar war, sich im kontinuierlichen Fall nicht mehr als brauchbar erweist. Es ist nicht möglich, einfache rekursive Relationen anzuwenden, welche letzlich zur optimalen Politik führen, weil wir jetzt Differentialgleichungen statt Differenzengleichungen behandeln.

Eine Politik-Iterationsmethode wurde für die Lösung des kontinuierlichen Entscheidungsproblems mit langer Dauer entwickelt. Sie ist in allen Hauptpunkten völlig analog zum Verfahren, welches auf diskrete Prozesse angewendet wird. Wie zuvor ist das “Herz” des Verfahrens ein Iterationszyklus, der sich zusammensetzt aus einer Wertbestimmung und einer Politik-Verbesserung. Wir werden nun beide Teile genau diskutieren.

10.2 Die Werbestimmung

Für eine gegebene Politik unterliegt der total zu erwartende Erlös des Systems für die Zeit t den Gleichungen (9.1)

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{v}_i(t) = q_i + \sum_{j=1}^{N} a_{ij}\,\mathrm{v}_j(t), \quad i \in E.$$

Da wir uns nur mit Prozessen beschäftigen, deren Abschluss sehr fern liegt, können wir den asymptotischen Ausdruck (Gleichung (9.10))

$$\mathrm{v}_i(t) = t\,g_i + \mathrm{v}_i, \; i \in E, \; t \to \infty.$$

10.2.1 Satz: Für die Werte v_i, $i \in E$, gilt die folgende Gleichung

(10.1) $$\sum_{j=1}^{N} a_{ij}\,g_j = 0, \; i \in E,$$

(10.2) $$q_i + \sum_{j=1}^{N} a_{i\,j}\, v_j - g_i = 0,\ i \in E$$

oder

$$v_i = \frac{1}{-a_{i\,i}} (q_i + \sum_{j \neq i} a_{i\,j}\, v_j - g_i),\ i \in E,$$

wobei

$$a_{i\,i} = -\sum_{j \neq i} a_{i\,j}.$$

▼

Beweis:

Umformen wir die Gleichungen (9.1) für $t \to \infty$,

$$\frac{d}{dt} (t\, g_i + v_i) = q_i + \sum_{j=1}^{N} a_{i\,j} (t\, g_j + v_j),\ i \in E$$

oder

(10.3) $$g_i = q_i + t \sum_{j=1}^{N} a_{i\,j}\, g_j + \sum_{j=1}^{N} a_{i\,j}\, v_j.$$

Wenn die Gleichungen (10.3) für alle grossen t gelten sollen, dann erhalten wir die beiden Systeme von linearen algebraischen Gleichungen

$$\sum_{j=1}^{N} a_{i\,j}\, g_j = 0,\ i \in E,$$

$$g_i = q_i + \sum_{j=1}^{N} a_{i\,j}\, v_j,\ i \in E.$$

Aus der letzten Gleichung folgt unmittelbar

$$g_i = q_i + a_{i\,i}\, v_i + \sum_{j \neq i} a_{i\,j}\, v_j$$

oder

$$v_i = \frac{1}{-a_{i\,i}} (q_i + \sum_{j \neq i} a_{i\,j}\, v_j - g),\ i \in E.$$

10.2.2 Bemerkung: Die Gleichungen (10.1) und (10.2) sind amalog zu (xii) und (xiii) (siehe Abschnitt 6.2) im diskreten Prozess.

Die Lösung der Gleichungen (10.1) ergibt den Gewinn für jeden Zustand in Abhängigkeit von den Gewinnen der ergodischen Klassen im Prozess.

Der relative Wert eines Zustandes in jeder Klasse wird gleich Null gesetzt, und die Gleichungen (10.2) werden verwendet, um die verbleibenden relativen Werte und die Gewinne der ergodischen Klassen zu bestimmen.

10.3 Die Politik-Verbesserung

Wir nehmen an, dass wir eine Politik haben, die optimal ist, wenn t Zeiteinheiten verbleiben, und dass diese Politik zu erwartende Gesamterlöse $v_i(t)$, $i \in E$, hat. Wenn wir wissen wollen, welche Politik zu befolgen ist, wenn mehr Zeit als t zur Verfügung steht, dann sehen wir aus den Gleichungen (9.1)

$$\frac{d}{dt} v_i(t) = q_i + \sum_{j=1}^{N} a_{i\,j}\, v_j(t),\ i \in E,$$

dass wir unsere Zuwachsrate von $v_i(t)$ maximieren können, in dem wir

(10.4) $$q_i^k + \sum_{j=1}^{N} a_{ij}^k v_j(t)$$

in bezug auf die Strategien k im Zustand i maximieren. Wenn t gross ist, können wir

$$v_j(t) = t\, g_j + v_j,\ j \in E,$$

verwenden und erhalten

$$q_i^k + \sum_{j=1}^{N} a_{ij}^k \left(t\, g_j + v_j\right)$$

oder

(10.5) $$q_i^k + \sum_{j=1}^{N} a_{ij}^k v_j + t \sum_{j=1}^{N} a_{ij}^k g_j$$

als die Grösse, die im i-ten Zustand maximiert werden soll. Für grosse t wird der Ausdruck (10.5) maximiert durch die Strategie, die die Gewinntestgrösse

(10.6) $$\sum_{j=1}^{N} a_{ij}^k g_j$$

maximiert, wobei die Gewinne der alten Politik eingesetzt werden.

Wenn jedoch alle Strategien den Gleichen Wert vom Ausdruck (10.6) ergeben, oder wenn eine Gruppe von Strategien denselben maximalen Wert ergibt, dann entscheidet man sich für die Strategie, die die Werttestgrösse

(10.7) $$q_i^k + \sum_{j=1}^{N} a_{ij}^k v_j$$

maximiert, wobei die relativen Werte der alten Politik verwendet werden.

Die Relativwerte können für den Werttest benützt werden, weil eine konstante Differenz die Entscheidungen innerhalb einer Klasse nicht beeinflusst.

Der allgemeine Iterationszyklus wird in nächstem Satz dargestellt.

10.3.1 Satz: Der Iterationszyklus für die Ermittlung der optimalen Politik,

Schritt 0 (Initialisierung): Fixieren wir eine Anfangspolitik **d**

Schritt 1 (Wertbestimmung): Man verwende a_{ij} und q_i für die gegebene Politik und löse

$$\sum_{j=1}^{N} a_{ij}\, g_j = 0\ ,\ i \in E,$$

$$q_i + \sum_{j=1}^{N} a_{ij}\, v_j - g_i = 0,\ i \in E$$

oder

$$v_i = \frac{1}{-a_{ii}} \left(q_i + \sum_{j \neq i} a_{ij}\, v_j - g_i\right),\ i \in E$$

für alle relativen Werte v_i, $i \in E$, und g_i, indem man den Wert eines v_i in jeder ergodischen Klasse gleich Null setzt.

Schritt 2 (**Verbesserung der Politik**): Für jeden Zustand $i \in E$ bestimmt man die Strategie k' der neuen Politik **d'**, die

$$\sum_{j=1}^{N} a_{ij}^k g_j$$

maximiert, indem man die Gewinne der vorhergehenden Politik verwendet. Diese Strategie ist die neue Entscheidung im i-ten Zustand.

Wenn

$$\sum_{j=1}^{N} a_{ij}^k g_j$$

für alle Strategien gleich ist, oder wenn mehrere Strategien gleich gut sind gemäss diesem Test, muss die Entscheidung auf Grund der relativen Werte anstatt des Gewinnes getroffen werden. Demzufolge bestimme man, wenn der Gewinntest misslingen sollte, die Strategie k', die

$$q_i^k + \sum_{j=1}^{N} a_{i\,j}^k \, v_j$$

maximiert, wobei die relativen Werte der vorangehenden Politik benutzt werden.

Damit hat man die neue Entscheidung im i-ten Zustand bestimmt.

Schritt 3 (**Prüfung von Konvergenz**): Der Iterationszyklus endet, wenn die Politiken **d** und **d'** bei zwei aufeinanderfolgenden Iterationen stimmen überein. Sonst muss den Schritt 1 mit den neuen Werten wiederholt werden.

▼

Beweis:

Er stimmt völlig überein mit dem Satz 6.3.1 für den diskreten Fall und hat auch einen analogen Beweis. Die Regeln über den Beginn und das Ende des Verfahrens bleiben unverändert.

10.4 Die Markov-Kette mit einer ergodischen Klasse

Wenn, wie es üblich ist, alle möglichen Politiken des Prozesses einer ergodischen Klasse entsprechen, so kann das rechnerische Verfahren beträchtlich vereinfacht werden. Da alle Zustände in jeder Markov-Kette die Gleiche Gewinne g haben, verlangt die Wertbestimmung nur die Lösung der Gleichungen (10.2)

$$q_i + \sum_{j=1}^{N} a_{i\,j} \, v_j - g_i = 0, \;\; i \in E$$

oder

$$v_i = \frac{1}{-a_{i\,i}} \left(q_i + \sum_{j \neq i} a_{i\,j} \, v_j - g_i \right), \;\; i \in E,$$

wobei v_N gleich Null gesetzt wird. Die Lösung für g und die übrigen v_i, $i \in E \backslash \{N\}$ wird benutzt, um eine verbesserte Politik zu bestimmen.

Multiplikation der Gleichungen (10.2) mit der Grenzwahrscheinlichkeit π_i, $i \in E$, und Summierung über i ergeben das bereits bekannte Resultat

$$g = \sum_{i=1}^{N} \pi_i \, q_i.$$

Die Politik-Verbesserung wird einfach: Für jeden Zustand $i \in E$ bestimme man die Strategie k, die

$$q_i^k + \sum_{j=1}^{N} a_{i\,j}^k \, v_j$$

maximiert, wobei die relativen Werte der vorherigen Politik verwendet werden. Diese Strategie wird die neue Entscheidung im i-ten Zustand. Eine neue Politik ist bestimmt, wenn dieses Verfahren für jeden Zustand durchgeführt worden ist.

Der Iterationszyklus für kontinuierliche Systeme mit einer ergodischen Klasse wird in nächstem Satz dargestellt.

10.4.1 Satz: Der Iterationszyklus für die Ermittlung der optimalen Politik,

Schritt 0 (Initialisierung): Fixieren wir eine Anfangspolitik **d**

Schritt 1 (Wertbestimmung): Man verwende $a_{i\,j}$ und q_i für die gegebene Politik und löse

$$q_i + \sum_{j=1}^{N} a_{i\,j} \, v_j - g = 0, \;\; i \in E$$

oder

$$v_i = \frac{1}{-a_{i\,i}} (q_i + \sum_{j \neq i} a_{i\,j}\, v_j - g),\ i \in E$$

für alle relativen Werte v_i, $i \in E$, und g, unter der Annahme, dass z.B. $v_N = 0$ ist.

Schritt 2 (**Verbesserung der Politik**): Für jeden Zustand $i \in E$ stelle man die Strategie k' der neuen Politik **d**' fest, welche

$$q_i^k + \sum_{j=1}^{N} a_{i\,j}^k\, v_j$$

maximiert unter Verwendung der relative Werte v_i, $i \in E$, der vorherigen Politik **d**.

k' ergibt dann die neue Entscheidung im i-ten Zustand, $q_i^{k'}$ wird q_i , und $a_{i\,j}^{k'}$ wird $a_{i\,j}$.

Schritt 3 (**Prüfung von Konvergenz**): Der Iterationszyklus endet, wenn die Politiken **d** und **d**' bei zwei aufeinanderfolgenden Iterationen stimmen überein. Sonst muss den Schritt 1 mit den neuen Werten wiederholt werden.

▼

Beweis: Der Beweis der Eigenschaften des Iterationszyklus für den kontinuierlichen Fall ist dem Beweis für den diskreten Fall sehr ähnlich.

Betrachten wir zwei Politiken A und B. Die Politik-Verbesserung hat die Politik B als Nachfolger der Politik A hervorgebracht. Deshalb wissen wir, dass

$$q_i^B + \sum_{j=1}^{N} a_{i\,j}^B\, v_j^A \geq q_i^A + \sum_{j=1}^{N} a_{i\,j}^A\, v_j^A,\ i \in E$$

oder

$$\gamma_i = q_i^B + \sum_{j=1}^{N} a_{i\,j}^B\, v_j^A - q_i^A - \sum_{j=1}^{N} a_{i\,j}^A\, v_j^A, \tag{10.8}$$

und

$$\gamma_i \geq 0$$

ist, Aus den Gleichungen der Wertbestimmung folgt

$$q_i^B + \sum_{j=1}^{N} a_{i\,j}^B\, v_j^B - g^B = 0,\ i \in E, \tag{10.9}$$

$$q_i^A + \sum_{j=1}^{N} a_{i\,j}^A\, v_j^A - g^A = 0,\ i \in E. \tag{10.10}$$

Wenn die Gleichung (10.10) von der Gleichung (10.9) subtrahiert wird, und wenn die Gleichung (10.8) benutzt wird, um

$$q_i^B - q_i^A$$

zu eliminieren, erhalten wir

$$\gamma_i + \sum_{j=1}^{N} a_{i\,j}^B (v_j^B - v_j^A) - (g^B - g^A) = 0. \tag{10.11}$$

Sei

$$g^\Delta = g^B - g^A \text{ und } v_i^\Delta = v_j^B - v_j^A.$$

Dann wird die Gleichung (10.11) zu

$$\gamma_i + \sum_{j=1}^{N} a_{i\,j}^B\, v_j^\Delta - g^\Delta = 0,\ i \in E. \tag{10.12}$$

Die Gleichungen (10.12) sind die Gleichungen der Wertbestimmung, die in Differenzen anstatt in absoluten werten geschrieben sind. Wir kennen die Lösung

$$(10.13)\quad g^{\Delta} = \sum_{i=1}^{N} \pi_i^B \gamma_i,$$

wobei π_i^B die Grenzwahrscheinlichkeit des Zustandes i unter der Politik B ist. Da alle

$$\pi_i^B \geq 0$$

und alle

$$\gamma_i \geq 0$$

sind, folgt

$$g^{\Delta} \geq 0.$$

Insbesondere wird g^B grösser als g^A sein, wenn ein Zuwachs der Testgrösse

$$q_i^k + \sum_{j=1}^{N} a_{ij}^k v_j$$

in irgendeinem Zustand i erzielt wird, der unter der Politik B nicht-transient ist.

Der Beweis dafür, dass der Iterationszyklus gegen die optimale Politik konvergieren muss, ist derselbe wie im Kapitel 4 für den diskreten Fall.

Dieser Satz ist völlig analog zum Satz 4.4.1 für diskrete Prozesse mit einer ergodischen Klasse.

10.4.2 Bemerkung: Man beachte, dass, wenn die Iteration in der Politik-Verbesserung mit allen $v_i = 0$, $i \in E$, gestartet wird, die zuerst gewählte Politik diejenige ist , die die Erlösrate in jedem Zustand maximiert. Diese Politik entschpricht der Politik, die den zu erwartenden unmittelbaren Erlös für diskrete Prozesse maximiert.

10.5 Das Dilemma des Werkmeisters

Nun willen wir das Problem des Werkmeisters lösen.

Zustand i	Strategie k	a_{i1}^k	a_{i2}^k	Erlösrate q_i^k
1 Laufende Anlage	1 (gewöhnliche Unterhalt)	−5	5	6
1	2 (teurer Unterhalt)	−2	2	4
2 Anlage ausser Betrieb	1 (Reparatur in eigener Werkstatt)	4	−4	−3
2	2 (Reparatur ausserhalb)	7	−7	−5

Welcher Wartungs- und welcher Instandsetzungsdienst wird den grössten Erlös pro Zeiteinheit einbringen? Da alle Politiken im System einer ergodischen Klasse entsprechen, kann das vereinfachte Verfahren (Satz 10.4.1) angewendet werden.

Schritt 0. Als unsere Anfangspolitik wählen wir diejenige, die die Erlösrate für jeden Zustand maximiert. Das ist die Politik, die aus dem normalen Unterhaltsservice und einigen Reparaturen besteht. Für diese Politik gilt

$$\mathbf{d} = \begin{pmatrix} 1 \\ 1 \end{pmatrix},\ \mathbf{A} = \begin{pmatrix} -5 & 5 \\ 4 & -4 \end{pmatrix},\ \mathbf{q} = \begin{pmatrix} 6 \\ -3 \end{pmatrix}.$$

Schritt 1. Die Wertbestimmungsgleichungen

$$q_i + \sum_{j=1}^{N} a_{ij} v_j - g = 0,\ i \in E$$

lauten

$$6 - 5\,v_1 + 5\,v_2 - g = 0,$$

$$-3 + 4\,v_1 - 4\,v_2 - g = 0.$$

Die Lösung dieser Gleichungen mit $v_2 = 0$ ist

$$g = 1,\ v_1 = 1,\ v_2 = 0$$

```
Solve[{6 - 5 v1 - g == 0, -3 + 4 v1 - g == 0}, {v1, g}]
```

```
{{v1 → 1, g → 1}}
```

Schritt 2. Um eine Politik mit höherem Gewinn zu finden, führen wir die Politik-Verbesserung durch folgende Tabelle

Zustand	Strategie $d(i) = k$	$q_i^k + \sum_{j=1}^{M} a_{i\,j}^k\, v_j$
1	1	$6 - 5 \cdot 1 = 1$
1	2	$4 - 2 \cdot 1 = 2 \leftarrow$
2	1	$-3 + 4 \cdot 1 = 1$
2	2	$-5 + 7 \cdot 1 = 2 \leftarrow$

Die zweite Strategie in jedem Zustand ergibt also eine bessere Politik. Es stellt sich heraus, dass die Politik mit der kostspieligeren Wartung und mit den auswärtigen Reparaturen vorteilhafter ist als die mit den normalen Diensten.

Schritt 0. Für die neue Politik haben wir

$$\mathbf{d} = \begin{pmatrix} 2 \\ 2 \end{pmatrix},\ \mathbf{A} = \begin{pmatrix} -2 & 2 \\ 7 & -7 \end{pmatrix},\ \mathbf{q} = \begin{pmatrix} 4 \\ -5 \end{pmatrix}$$

Schritt 1. wir werten diese Politik aus untrer anwendung der Gleichungen

$$q_i + \sum_{j=1}^{N} a_{i\,j}\, v_j - g = 0,\ \ i \in E$$

und erhalten

$$4 - 2\,v_1 + 2\,v_2 - g = 0,$$

$$-5 + 7\,v_1 - 7\,v_2 - g = 0.$$

Die Lösung dieser Gleichungen mit $v_2 = 0$ ist

$$g = 2,\ v_1 = 1,\ v_2 = 0.$$

```
Solve[{4 - 2 v1 - g == 0, -5 + 7 v1 - g == 0}, {v1, g}]
```

```
{{v1 → 1, g → 2}}
```

Man beachte, dass der Gewinn grösser ist als zuvor.

Schritt 2. Nun müssten wir die Politik-Verbesserung durchführen, um zu sehen, ob wir noch eine bessere Politik finden können. Da sich jedoch die Werte übereinstimmend nicht geändert haben, würde die Politik-Verbesserung wiederum die Politik

$$\mathbf{d} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

ergeben.

Schritt 3. Da wir zweimal hintereinander die gleiche Politik erhalten haben, muss sie die optimale Politik sein.

Somit sollte der Werkmeister den kostespieligeren Unterhalt wählen und die Reparaturen nach auswärts vergeben. Auf diese Art erhäht er im durchschnitt seinen Profit von \$1 auf \$2 pro Stunde. Man beachte, dass der Werkmeister bereit sein sollte, \$1 für eine sofortige Reparatur zu zahlen, da

$$v_1 - v_2 = 1$$

ist. Es steht dem Leser frei die Politiken

$$\mathbf{d} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ und } \mathbf{d} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

zu untersuchen, um sich zu überzeugen, dass die Erlöse pro Stunde niedriger sind als bei der optimalen Politik.

10.6 Eine Bemerkung zu den praktischen Berechnungen

Wir haben gesehen, dass die Lösung des kontinuierlichen entscheidungsprozesses ungefähr ebensoviel Rechenaufwand erfordert, wie die Lösung des entsprechenden diskreten Prozesses. In der Tat sind diese zwei Prozesstypen rechnerisch äquivalent, so dass für die Lösung beider Probleme das gleiche Computer-Programm verwendet werden kann. Um dies zu erläutern, wollen wir die Wertbestimmungsgleichungen für den diskreten Prozess (Gleichungen (xii) und (xiii), Abschnitt 6.2) noch einmal hinschreiben

(xii) $$g_i = \sum_{j=1}^{N} p_{i\,j}\, g_j\,,\ i \in E,$$

(xiii) $$v_i = q_i + \sum_{j=1}^{N} p_{i\,j}\, v_j - g_i,\ \ i \in E.$$

Diese Gleichungen können folgendemassen geschrieben werden

$$\sum_{j=1}^{N} \left(p_{i\,j} - \delta_{i\,j}\right) g_j = 0,$$

$$q_i + \sum_{j=1}^{N} \left(p_{i\,j} - \delta_{i\,j}\right) v_j - g_i = 0,$$

wobei $\delta_{i\,j}$ das Kronecker-Symbol ist, d.h.

$$\delta_{i\,j} = 1, \text{ wenn } i = j$$

$$\delta_{i\,j} = 0, \text{ wenn } i \neq j.$$

Wenn wir nun

$$a_{i\,j} = p_{i\,j} - \delta_{i\,j}$$

setzen, erhalten wir

$$\sum_{j=1}^{N} a_{i\,j}\, g_j = 0,$$

$$q_i + \sum_{j=1}^{N} a_{i\,j}\, v_j - g_i = 0.$$

Da sind die Wertbestimmungsgleichungen (Gleichungen (10.1) und (10.2), Abschnitt 10.2) für den kontinzuierlichen entscheidungsprozess. folglich können wir, wenn wir ein Programm für die Lösung der Gleichungen (xii) und (xiii) für den diskreten Prozess haben, dieses Programm für die Lösung des kontinuierlichen Prozess, der durch die Matrix A beschrieben wird, durch Transfromieren der Übergangsraten zu “Pseudo”-Übergangswahrscheinlichkeiten gemäss der Relation

$$p_{i\,j} = a_{i\,j} + \delta_{i\,j}$$

verwenden.

10.6.1 Bemerkung: Wenn im Computer-Programm angenommen wir, dass

$$0 \le p_{i\,j} \le 1$$

ist, ist es notwendig, $a_{i\,j}$ so zu normieren, dass

$$-1 \le a_{i\,i} \le 0.$$

Was die Politikverbesserung anbetrifft, so maximieren wir im diskreten Fall entweder

$$\sum_{j=1}^{N} p_{i\,j}^{k}\, g_j \text{ oder } q_i^k + \sum_{j=1}^{N} p_{i\,j}^{k}\, \mathrm{v}_j$$

bezüglich aller Strategien k im Zustand $i \in E$.

Unsere Entscheidungen würden unverändert bleiben, wenn wir statt dessen

$$\sum_{j=1}^{N} a_{i\,j}^{k}\, g_j \text{ und } q_i^k + \sum_{j=1}^{N} a_{i\,j}^{k}\, \mathrm{v}_j.$$

Das sind jedoch die Testgrössen für die Politik-Verbesserung beim kontinuierlichen Prozess. Demzufolge kann eine für den diskreten Prozess programmierte Politik-Verbesserung für den kontinuierlichen Prozess verwendet werden, wenn die Transformation

$$p_{i\,j}^{k} = a_{i\,j}^{k} + \delta_{i\,j}$$

durchgeführt wird.

Die diskreten und kontinuierlichen entscheidungsprozesse sind somit rechnerisch äquivalent. Das gleiche Computer-Programm kann nach einer einfachen Datentransformation für beide Prozesse verwendet werden.