Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

MC SYLLABUS 39

COLLOQUIUM STOCHASTISCHE SPELEN

O.J. VRIEZE
G.L. WANROOIJ

MATHEMATISCH CENTRUM AMSTERDAM 1978

AMS(MOS) subject classification scheme (1970): 90DXX, 90D05, 90D10, 90D15

ISBN 90 6196 167 X

VOORWOORD

Deze syllabus is gebaseerd op de stof gepresenteerd tijdens het colloquium "Stochastische spelen". Dit maandelijks colloquium werd gehouden van oktober 1975 tot juni 1976.

Na de inleidende begrippen, waarin een gemeenschappelijk kader wordt geformuleerd voor de op zichzelf losstaande hoofdstukken, en een korte beschouwing over matrixspelen (hoofdstuk I) komen vervolgens de artikelen van SHAPLEY [57] (hoofdstuk II), TAKAHASHI [62] (hoofdstuk III), EVERETT [13] (hoofdstuk IV), GILETTE [18] en HOFFMAN & KARP [22] (hoofdstuk V) aan de orde. In hoofdstuk VI worden enige algoritmen voor het oplossen van stochastische spelen behandeld. In hoofdstuk VII worden resultaten van het onderzoek van BEWLEY & KOHLBERG genoemd [3], [4] en [5]. Deze in 1976 ververschenen artikelen zijn niet tijdens het colloquium aan de orde gekomen, maar vormen een belangrijke ontwikkeling op het gebied van de stochastische spelen. Andere recente literatuur, die niet behandeld is, is samengevat in hoofdstuk VIII.

Tijdens het colloquium is door H. VAN DER WHEEL (Universiteit van Amsterdam) gesproken over endogene prijsstellingstheorie. Belangstellenden worden verwezen naar het proefschrift van VAN DER WHEEL [75].

J. VAN DER WAL (Technische Hogeschool Eindhoven) presenteerde enkele van zijn resultaten met betrekking tot successieve approximatie in eindige stoppende stochastische spelen, waarvan in sektie 6.3 een samenvatting is opgenomen.

Wij zijn van mening, dat de inhoud van deze syllabus door het commentaar van H.C. TIJMS (Vrije Universiteit Amsterdam) aanmerkelijk is verbeterd. Hiervoor zijn wij hem zeer erkentelijk.

O.J. Vrieze
G.L. Wanrooij

INHOUD blz.

HOOFDSTUK I

INLEIDENDE BEGRIPPEN

1.1. Notaties, definities en enkele lemma's

Een twee-persoons niet-coöperatief stochastisch spel, notate Γ, is een dynamisch systeem, dat kan worden beschreven door middel van een zestal: $\Gamma = (S,A_1,A_2,g_1,g_2,p)$. S, A_1 en A_2 zijn niet lege verzamelingen; g_1, g_2 en p zijn afbeeldingen. Deze spelkomponenten hebben de volgende betekenis:

S : de toestandsruimte; in deze syllabus wordt steeds verondersteld dat de verzameling S aftelbaar is.

A_1: de verzameling van mogelijke akties voor speler 1. De verzameling van toegelaten akties voor speler 1 in toestand $s \in S$ (notatie $A_1(s)$) is in het algemeen toestandsafhankelijk. Een element van $A_1(s)$ noteren we als $a_1(s)$.

A_2: idem als A_1, maar dan voor speler 2.

g_1: de afbeelding g_1: $S \times A_1 \times A_2 \to \mathbb{R}$ gedefinieerd voor alle triples (s,a_1,a_2) waarvoor $(a_1,a_2) \in A_1(s) \times A_2(s)$ is de uitbetalingsfunctie voor speler 1.

g_2: idem als g_1, maar dan voor speler 2.

p : de overgangskansenafbeelding. Laat $\mathcal{P}$ de verzameling van alle maten μ op S zijn met $\mu(S) \leq 1$ (hier rijtjes van niet-negatieve getallen met som hoogstens 1), dan is p een afbeelding p: $S \times A_1 \times A_2 \to \mathcal{P}$, gedefinieerd voor alle triples (s,a_1,a_2), waarvoor $(a_1,a_2) \in$ $\in A_1(s) \times A_2(s)$. Notatie: $p(S_1|s,a_1,a_2)$ met $S_1 \subset S$ is het gewicht dat de maat behorende bij (s,a_1,a_2) toekent aan S_1.

Een partijtje van het spel verloopt als volgt: Er wordt een begintoestand $s_0 \in S$ gespecificeerd; de beide spelers kiezen onafhankelijk een aktie $a_i(s_0) \in A_i(s_0)$, $i = 1,2$; speler i krijgt een uitbetaling $g_i(s_0,a_1(s_0),a_2(s_0))$, $i = 1,2$; vervolgens springt het spel naar een nieuwe toestand, bepaald door het kansmechanisme $p(\cdot|s_0,a_1(s_0),a_2(s_0))$, enz. Als

$p(S|s_0,a_1(s_0),a_2(s_0)) < 1$ dan is $1 - p(S|s_0,a_1(s_0),a_2(s_0))$ de kans dat het spel stopt als in toestand s_0 de spelers de akties $a_1(s_0) \in A_1(s_0)$ en $a_2(s_0) \in A_2(s_0)$ kiezen.

Een spel Γ is dus eigenlijk een verzameling van spelen, namelijk voor iedere begintoestand vindt er een ander spel plaats. Het is echter gebruikelijk in de theorie der stochastische spelen om deze verzameling van spelen als één spel op te vatten. Een reden hiervoor is dat er gezocht wordt naar strategieën voor de beide spelers met bepaalde kenmerken, die dezelfde zijn voor iedere begintoestand.

Al naar gelang de eigenschappen van de spelkomponenten kunnen we een aantal specifieke spelen onderscheiden.

DEFINITIE 1.1.1. Een *stoppend spel* is een spel waarbij $p(S|s,a_1,a_2) < 1$ voor alle $s \in S$ en alle $(a_1,a_2) \in A_1(s)\times A_2(s)$.

DEFINITIE 1.1.2. Een *niet-stoppend spel* is een spel waarbij $p(S|s,a_1,a_2) = 1$ voor alle $s \in S$ en alle $(a_1,a_2) \in A_1(s)\times A_2(s)$.

Merk op dat er spelen zijn die noch stoppend noch niet-stoppend zijn.

DEFINITIE 1.1.3. Een *eindig spel* is een spel waarbij zowel S, A_1 als A_2 eindige verzamelingen zijn.

DEFINITIE 1.1.4. Een *twee-persoons nulsomspel* is een spel waarbij $g_1(s,a_1,a_2) + g_2(s,a_1,a_2) = 0$ voor alle $s \in S$ en alle $(a_1,a_2) \in A_1(s)\times A_2(s)$.

DEFINITIE 1.1.5. Een *éénstapsspel* is een spel waarbij $p(S|s,a_1,a_2) = 0$ voor alle $s \in S$ en alle $(a_1,a_2) \in A_1(s)\times A_2(s)$.

DEFINITIE 1.1.6. Een *matrixspel* is een eindig, éénstaps, twee-persoons nulsomspel, waarbij S uit één element bestaat.

DEFINITIE 1.1.7. Een *recursief spel* is een twee-persoons nulsomspel, waarbij er een deelverzameling $S_1 \subset S$ bestaat zodanig dat

1°. $s \in S_1 \Rightarrow A_1(s)$ en $A_2(s)$ bestaan beide uit 1 element en $p(S|s,a_1(s),a_2(s)) = 0$.

2°. $s \in S\backslash S_1 \Rightarrow g_1(s,a_1(s),a_2(s)) = g_2(s,a_1(s),a_2(s)) = 0$ en $p(S|s,a_1(s),a_2(s)) = 1$, alle $(a_1,a_2) \in A_1(s)\times A_2(s)$.

We gaan nu het begrip strategie invoeren. Daarvoor moeten er eerst een aantal andere begrippen gedefinieerd worden.

DEFINITIE 1.1.8. Een *gemengde aktie* (notatie $\pi_i(s)$) voor speler i in toestand s is een kansverdeling op zijn aktieruimte $A_i(s)$, $i = 1,2$.

Als $A_i(s)$ hoogstens aftelbaar veel elementen bevat, dan is $\pi_i(s)$ een rijtje $\pi_i(1,s), \pi_i(2,s), \ldots$, z.d.d. $\pi_i(a_i,s) \geq 0$, $a_i \in A_i(s)$ en $\sum_{a_i \in A_i(s)} \pi_i(a_i,s) = 1$. Als $A_i(s)$ een overaftelbare verzameling is, dan beschouwen we een σ-algebra $\mathcal{A}_i(s)$ bij $A_i(s)$, waarop $\pi_i(s)$ dan als kansmaat gedefinieerd is.

Laat $\Pi_i(s)$ de verzameling van alle gemengde akties voor speler i in toestand s gedefinieerd op de σ-algebra $\mathcal{A}_i(s)$. De σ-algebra $\mathcal{A}_i(s)$ kan zo gekozen worden dat alle één-punts deelverzamelingen van $A_i(s)$ ertoe behoren. De aktie $a_i(s) \in A_i(s)$ kan dan worden opgevat als de kansmaat die kans 1 toekent aan het punt $a_i(s)$. Een aktie $a_i(s) \in A_i(s)$ wordt wel een zuivere aktie genoemd.

Als speler i in toestand s een gemengde aktie $\pi_i(s) \in \Pi_i(s)$ speelt, dan voert hij eerst een loting uit, waarbij de kansen verdeeld zijn volgens de kansmaat $\pi_i(s)$ om te bepalen welke zuivere aktie hij zal nemen.

Laat $A(s) = A_1(s) \times A_2(s)$ en $\Pi(s) = \Pi_1(s) \times \Pi_2(s)$, $s \in S$. Een element van $\Pi(s)$ noteren we als $\pi(s)$. $\Pi(s)$ is dan de verzameling van produkt-kansmaten gedefinieerd op de produkt σ-algebra $\mathcal{A}_1(s) \times \mathcal{A}_2(s)$ (d.w.z. de kleinste σ-algebra gedefinieerd bij $A(s)$ die $\mathcal{A}_1(s) \times \mathcal{A}_2(s)$ omvat).

DEFINITIE 1.1.9. De *historie* van het spel op tijdstip t (notatie h_t) is $h_t = (s_0, a_1(s), a_2(s_0), s_1, a_1(s_1), a_2(s_1), \ldots, s_{t-1}, a_1(s_{t-1}), a_2(s_{t-1}))$ als s_τ respektievelijk $a_i(s_\tau)$ de toestand respektievelijk de aktie van speler i, $i = 1,2$, op tijdstip τ geweest is, $\tau = 0,1,\ldots,t-1$ (h_0 is een lege string).

DEFINITIE 1.1.10. Een spel met *perfekte herinnering* is een spel waarbij iedere speler op ieder tijdstip t de historie h_t kent.

In deze syllabus worden uitsluitend spelen met perfekte herinnering beschouwd.

DEFINITIE 1.1.11. Een *gedragsstrategie* (voor speler i (notatie π_i)) is een voorschrift dat aan ieder tijdstip t, iedere toestand s_t en iedere historie h_t een element $\pi_i(t,s_t,h_t) \in \Pi_i(s_t)$ toevoegt, waarbij s_t de toestand op tijdstip t weergeeft.

Laat Π_i de verzameling van gedragsstrategieën voor speler i zijn. In deze syllabus beperken we ons tot de klasse Π_i van strategieën. De klasse

Π_i is niet de meest algemene klasse van strategieën. Als we het stochastisch spel zouden weergeven in de zogenaamde uitgebreide vorm m.b.v. een gerichte graaf (zie J. VON NEUMANN & O. MORGENSTERN [41]), dan verkrijgen we een boom van oneindige diepte. Hierop zouden we zuivere en gemengde strategieën kunnen definiëren in de zin van KUHN [32] en AUMANN [1]. De klasse van strategieën die we aldus verkrijgen is ruimer dan de boven gedefinieerde klasse Π_i. Echter AUMANN [1] heeft voor een deelklasse van spelen bewezen, dat iedere gemengde strategie, gedefinieerd op het spel in uitgebreide vorm, een equivalente gedragsstrategie bezit, waarbij twee strategieën voor een speler equivalent zijn, als tegen een willekeurige strategie van de andere speler en voor iedere begintoestand de verwachting van de uitbetaling op tijdstip t voor beide strategieën gelijk zijn, $t = 0,1,\ldots$.

Alle in deze syllabus te behandelen stochastische spelen vallen binnen de deelklasse van spelen waarvoor Aumann's resultaat geldig is, zodat het beperken tot de klasse Π_i van gedragsstrategieën voor speler i geen beperking van de algemeenheid inhoudt. Een gedragsstrategie zullen we in het vervolg kortweg strategie noemen.

DEFINITIE 1.1.12. Een strategie heet *semi-Markov* als deze voor iedere t wat de historie betreft alleen afhankelijk is van s_0 (notatie $\pi_i(t,s_t,s_0)$).

DEFINITIE 1.1.13. Een strategie heet *Markov* of *geheugenloos* als deze voor iedere t onafhankelijk is van de historie (notatie $\pi_i(t,s_t)$).

DEFINITIE 1.1.14. Een strategie heet *stationair* als deze voor iedere t onafhankelijk is van zowel t als van de historie (notatie $\pi_i(s_t)$).

Een gegeven begintoestand $s_0 \in S$ en een paar strategieën $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$ legt een stochastisch proces vast met betrekking tot de toestanden die achtereenvolgens worden doorlopen en de akties, die de spelers in deze toestanden kiezen. Laat $p^t(s_t|s_0,\pi_1,\pi_2)$ de kans zijn dat op tijdstip t het systeem zich in toestand s_t bevindt als s_0 de begintoestand is en de beide spelers respektievelijk $\pi_1 \in \Pi_1$ en $\pi_2 \in \Pi_2$ spelen. Laat $P^t_{i,s_t}(H|s_t,s_0,\pi_1,\pi_2) = \Pr\{a_i(s_t) \in H \mid s_t,s_0,\pi_1,\pi_2\}$ voor $H \in A_i(s_t)$ de kans zijn dat speler i op tijdstip t in toestand s_t bij begintoestand s_0 een aktie behorende tot H kiest, als de beide spelers respektievelijk de strategieën $\pi_1 \in \Pi_1$ en $\pi_2 \in \Pi_2$ spelen.

In het volgende veronderstellen we steeds dat zowel $g_1(s,\cdot,\cdot)$ als $g_2(s,\cdot,\cdot)$ meetbare functies zijn t.o.v. $A_1(s) \times A_2(s)$ voor alle $s \in S$.

Als de spelers respektievelijk de strategieën $\pi_1 \in \Pi_1$ en $\pi_2 \in \Pi_2$ spelen, dan is voor begintoestand s_0 de verwachte uitbetaling aan speler i op tijdstip t (notatie $V_i^t(s_0,\pi_1,\pi_2)$) nu te schrijven als

$$V_i^t(s_0,\pi_1,\pi_2) =$$

$$= \sum_{s_t \in S} \left\{ \int_{A_1(s_t)} \int_{A_2(s_t)} g_i(s_t,a_1,a_2) P_{1s_t}^t(da_1 | s_t,s_0,\pi_1,\pi_2) P_{2s_t}^t(da_2 | (s_t,s_0,\pi_1,\pi_2) \right\} \cdot$$

$$\cdot\, p^t(s_t | s_0,\pi_1,\pi_2).$$

In deze syllabus komen drie kriteria aan de orde.

1°. Verdisconteerde opbrengstenmodel.

Een verdisconteringsfactor $\beta \in [0,1)$ is gespecificeerd. De uitbetaling aan speler i tot en met tijdstip t onder $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$ en $s_0 \in S$ is na verdiscontering gelijk aan

$$W_i^t(s_0,\pi_1,\pi_2) = \sum_{\tau=0}^{t} \beta^\tau\, V_i^\tau(s_0,\pi_1,\pi_2).$$

De totale verdisconteerde uitbetaling over een oneindige horizon is dan gelijk aan

$$W_i(s_0,\pi_1,\pi_2) = \lim_{t\to\infty} W_i^t(s_0,\pi_1,\pi_2) = \lim_{t\to\infty} \sum_{\tau=0}^{t} \beta^\tau V_i^\tau(s_0,\pi_1,\pi_2),$$

indien deze limiet bestaat.

2°. Totale opbrengstenmodel.

De opbrengsten op de verschillende tijdstippen worden bij elkaar opgeteld. De uitbetaling tot en met tijdstip t aan speler i onder $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$ en $s_0 \in S$ is dan gelijk aan

$$X_i^t(s_0,\pi_1,\pi_2) = \sum_{\tau=0}^{t} V_i^\tau(s_0,\pi_1,\pi_2).$$

De totale verwachte uitbetaling over een oneindige horizon is dan gelijk aan

$$X_i(s_0,\pi_1,\pi_2) = \lim_{t\to\infty} X_i^t(s_0,\pi_1,\pi_2) = \lim_{t\to\infty} \sum_{\tau=0}^{t} V_i^\tau(s_0,\pi_1,\pi_2),$$

indien deze limiet bestaat.

Merk op dat een verdisconteerd model opgevat kan worden als een totale opbrengstenmodel door de overgangskansen als volgt te transformeren: $p(s'|s,a_1,a_2)$ wordt $\beta \cdot p(s'|s,a_1,a_2)$, $\forall (a_1,a_2) \in A_1(s) \times A_2(s)$, $\forall s \in S$.

3°. Gemiddelde opbrengstenmodel.

De gemiddelde uitbetaling per stap na t tijdstippen onder $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$ en $s_0 \in S$ is gelijk aan

$$Y_i^t(s_0,\pi_1,\pi_2) = \frac{1}{t+1} \sum_{\tau=0}^{t} V_i^{\tau}(s_0,\pi_1,\pi_2).$$

De gemiddelde verwachte uitbetaling over een oneindige horizon is dan

$$Y_i(s_0,\pi_1,\pi_2) = \lim_{t\to\infty} Y_i^t(s_0,\pi_1,\pi_2) = \lim_{t\to\infty} \frac{1}{t+1} \sum_{\tau=0}^{t} V_i^{\tau}(s_0,\pi_1,\pi_2),$$

indien deze limiet bestaat.

Beschouwen we een stochastisch spel met een bepaald kriterium, dan nemen we in het vervolg steeds aan dat de limieten van de desbetreffende t-staps uitbetalingen bestaan. Merk op, dat dit voor nulsomspelen inhoudt, dat de verwachte uitbetalingen voor de beide spelers elkaars tegengestelde zijn.

De volgende definitie geldt voor alle drie kriteria.

DEFINITIE 1.1.15. Een ε-*evenwichtspunt van strategieën* voor $\varepsilon \geq 0$ is een paar $(\pi_1^{\varepsilon},\pi_2^{\varepsilon})$ waarvoor geldt

$$V_1(s,\pi_1^{\varepsilon},\pi_2^{\varepsilon}) \geq V_1(s,\pi_1,\pi_2^{\varepsilon})-\varepsilon \quad \forall\pi \in \Pi_1,\ \forall s \in S$$

en

$$V_2(s,\pi_1^{\varepsilon},\pi_2^{\varepsilon}) \geq V_2(s,\pi_1^{\varepsilon},\pi_2)-\varepsilon \quad \forall\pi_2 \in \Pi_2,\ \forall s \in S,$$

waarbij $V_i(s,\pi_1,\pi_2)$ de totale verwachte opbrengst (verdisconteerd, totaal of gemiddeld) is voor speler i bij begintoestand s en strategieën $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$. Een 0-evenwichtspunt wordt kortweg evenwichtspunt genoemd.

Een spel kan geen, één of meerdere (ε-)evenwichtspunten bezitten. In het algemeen bepalen verschillende evenwichtspunten verschillende opbrengsten voor de spelers. Dit is niet het geval voor de twee-persoons nulsomspelen, zoals uit onderstaand lemma blijkt.

LEMMA 1.1.16. *Zijn voor een twee-persoons nulsomspel* (π_1^*,π_2^*) *en* (π_1^{**},π_2^{**}) *twee evenwichtspunten, dan geldt* $V_1(s,\pi_1^*,\pi_2^*) = V_1(s,\pi_1^{**},\pi_2^{**})$.

BEWIJS. Uit de definitie van evenwichtspunt volgt:

$$V_1(s,\pi_1^*,\pi_2^*) \geq V_1(s,\pi_1^{**},\pi_2^*) = -V_2(s,\pi_1^{**},\pi_2^*) \geq -V_2(s,\pi_1^{**},\pi_2^{**}) =$$

$$= V_1(s,\pi_1^{**},\pi_2^{**}).$$

Evenzo

$$V_1(s,\pi_1^{**},\pi_2^{**}) \geq V_1(s,\pi_1^*,\pi_2^*),$$

waarmee het lemma bewezen is. □

GEVOLG 1.1.17. *Zijn voor een twee-persoons nulsomspel* (π_1^*,π_2^*) *en* (π_1^{**},π_2^{**}) *twee evenwichtspunten, dan vormen ook* (π_1^*,π_2^{**}) *en* (π_1^{**},π_2^*) *twee evenwichtspunten.*

Uit lemma 1.1.16 blijkt dat de volgende definities (vergelijk ook lemma 1.1.20) zinvol zijn. Deze definities zijn op ieder van de drie genoemde kriteria van toepassing.

DEFINITIE 1.1.18. Een twee-persoons nulsomspel wordt *strikt bepaald* genoemd als er een functie $V: S \to \mathbb{R}$ bestaat, zodanig dat

$$\sup_{\pi_1} \inf_{\pi_2} V_1(s,\pi_1,\pi_2) = \inf_{\pi_2} \sup_{\pi_1} V_1(s,\pi_1,\pi_2) = V(s), \qquad \forall s \in S.$$

De functie V wordt de *waardefunctie* van het spel genoemd. Als voor $\varepsilon \geq 0$ $\pi_1^\varepsilon(\pi_2^\varepsilon)$ een strategie voor speler 1 (speler 2) is, waarvoor geldt

$$V_1(s,\pi_1^\varepsilon,\pi_2) \geq V(s) - \varepsilon, \ \forall \pi_2 \in \Pi_2, \ \forall s \in S$$

$(V_2(s,\pi_1,\pi_2^\varepsilon) \geq -V(s) - \varepsilon, \ \forall \pi_1 \in \Pi_1, \ \forall s \in S)$, dan wordt π_1^ε (π_2^ε) ε-optimaal genoemd voor speler 1 (speler 2). 0-optimaal wordt kortweg optimaal genoemd.

Uit definitie 1.1.18 blijkt, dat, als π_1^ε (π_2^ε) ε-optimaal is voor speler 1 (speler 2), speler 1 (speler 2) zich een uitbetaling van minstens $V(s) - \varepsilon$ (minstens $-V(s) - \varepsilon$) kan garanderen voor begintoestand $s \in S$ door π_1^ε (π_2^ε) te spelen, als $V(\cdot)$ de waardefunctie van het spel is.

LEMMA 1.1.19. *Voor ieder twee-persoons nulsomspel geldt voor alle* $s \in S$

$$\inf_{\pi_2} \sup_{\pi_1} V_1(s,\pi_1,\pi_2) \geq \sup_{\pi_1} \inf_{\pi_2} V_1(s,\pi_1,\pi_2).$$

BEWIJS. Uit $\sup\limits_{\pi_1} V_1(s,\pi_1,\pi_2) \geq V_1(s,\pi_1,\pi_2)$ voor alle $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$ blijkt

$$\inf_{\pi_2} \sup_{\pi_1} V_1(s,\pi_1,\pi_2) \geq \inf_{\pi_2} V_1(s,\pi_1,\pi_2) \text{ voor alle } \pi_1 \in \Pi_1,$$

waaruit de bewering in het lemma volgt. □

LEMMA 1.1.20. *Een twee-persoons nulsom stochastisch spel is dan en slechts dan strikt bepaald als minstens één van de volgende twee uitspraken waar is:*
1°. *het spel bezit een evenwichtspunt;*
2°. *het spel bezit voor iedere* $\varepsilon > 0$ *een* ε*-evenwichtspunt.*

BEWIJS. a). Laat het spel strikt bepaald zijn en $V(\cdot)$ de waardefunctie zijn. Uit definitie 1.1.18 blijkt dan dat voor alle s geldt

$$\sup_{\pi_1} \inf_{\pi_2} V_1(s,\pi_1,\pi_2) = \inf_{\pi_2} \sup_{\pi_1} V_1(s,\pi_1,\pi_2) = V(s).$$

Hieruit volgt dat er voor alle $\varepsilon > 0$ en alle $s \in S$ een paar $(\pi_1^{\varepsilon s},\pi_2^{\varepsilon s})$ bestaat, zodanig dat

$$V_1(s,\pi_1^{\varepsilon s},\pi_2) \geq V(s) - \tfrac{1}{2}\varepsilon, \qquad \text{alle } \pi_2 \in \Pi_2$$

en

$$V_1(s,\pi_1,\pi_2^{\varepsilon s}) \leq V(s) + \tfrac{1}{2}\varepsilon, \qquad \text{alle } \pi_1 \in \Pi_2.$$

Voor het spel met begintoestand s is, wat $\pi_i^{\varepsilon s}$ betreft alleen datgene van belang, wat $\pi_i^{\varepsilon s}$ voorschrijft bij begintoestand s. We kunnen nu uit $\{\pi_i^{\varepsilon s} \mid s \in S\}$ een strategie π_i^{ε} als volgt samenstellen: $\pi_i^{\varepsilon}(t,s_t,h_t) = \pi_i^{\varepsilon s_0}(t,s_t,h_t)$ als $h_t = (s_0,a_1(s_0),a_2(s_0),\ldots)$. Gemakkelijk is in te zien dat ook voor $(\pi_1^{\varepsilon},\pi_2^{\varepsilon})$ geldt:

$$V_1(s,\pi_1^{\varepsilon},\pi_2) \geq V(s) - \tfrac{1}{2}\varepsilon, \qquad \text{alle } \pi_2 \in \Pi_2$$

(1.1) en

$$V_1(s,\pi_1,\pi_2^{\varepsilon}) \leq V(s) + \tfrac{1}{2}\varepsilon, \qquad \text{alle } \pi_1 \in \Pi_1.$$

Uit (1.1) volgt $|V_1(s,\pi_1^{\varepsilon},\pi_2^{\varepsilon}) - V(s)| \leq \tfrac{1}{2}\varepsilon$, zodat voor het paar $(\pi_1^{\varepsilon},\pi_2^{\varepsilon})$ geldt, dat voor elke $s \in S$

$$(1.2) \quad \begin{cases} V_1(s,\pi_1,\pi_2^{\varepsilon}) \leq V(s)+\tfrac{1}{2}\varepsilon \leq V_1(s,\pi_1^{\varepsilon},\pi_2^{\varepsilon}) + \varepsilon \qquad \text{alle } \pi_1 \in \Pi_1 \\ \text{en} \\ V_2(s,\pi_1^{\varepsilon},\pi_2) = -V_1(s,\pi_1^{\varepsilon},\pi_2) \leq -V(s)+\tfrac{1}{2}\varepsilon \leq -V_1(s,\pi_1^{\varepsilon},\pi_2^{\varepsilon}) + \varepsilon = V_2(s,\pi_1^{\varepsilon},\pi_2^{\varepsilon}) + \varepsilon, \\ \qquad\qquad\qquad\qquad \text{alle } \pi_2 \in \Pi_2. \end{cases}$$

Uit (1.2) blijkt dat het paar $(\pi_1^\varepsilon,\pi_2^\varepsilon)$ ε-evenwichtspunt is.

b) Laat nu minstens één van de beide uitspraken 1° en 2° waar zijn. Laat 1° waar zijn, dan bestaat er een paar (π_1^*,π_2^*), z.d.d.

$$(1.3)\qquad V_1(s,\pi_1,\pi_2^*) \le V_1(s,\pi_1^*,\pi_2^*) \le V_1(s,\pi_1^*,\pi_2), \quad \forall(\pi_1,\pi_2)\in\Pi_1\times\Pi_2,\ \forall s\in S.$$

Uit (1.3) kunnen we afleiden

$$(1.4)\qquad \inf_{\pi_2}\sup_{\pi_1} V_1(s,\pi_1,\pi_2) \le V_1(s,\pi_1^*,\pi_2^*) \le \sup_{\pi_1}\inf_{\pi_2} V_1(s,\pi_1,\pi_2),\ \forall s\in S.$$

Uit (1.4) en lemma 1.1.19 volgt nu dat het spel strikt bepaald is en dat $V_1(\cdot,\pi_1^*,\pi_2^*)$ de waardefunctie is.

Laat 2° waar zijn. Het bewijs van dit gedeelte van de bewering is eveneens te vinden in TIJS [64] en in HORDIJK, VRIEZE & WANROOIJ [24]. Laat voor $\varepsilon > 0$ $(\pi_1^\varepsilon,\pi_2^\varepsilon)$ een ε-evenwichtspunt, dus

$$(1.5)\qquad V_1(s,\pi_1,\pi_2^\varepsilon)-\varepsilon \le V_1(s,\pi_1^\varepsilon,\pi_2^\varepsilon) \le V_1(s,\pi_1^\varepsilon,\pi_2)+\varepsilon,\quad \forall(\pi_1,\pi_2)\in\Pi_1\times\Pi_2.$$

Laat ε_n, $n = 1,2,\ldots$ een rij positieve getallen z.d.d. $\lim_{n\to\infty}\varepsilon_n = 0$. Dan geldt, gebruik makend van (1.5) voor ieder paar natuurlijke getallen (i,j):

$$V_1(s,\pi_1^{\varepsilon_j},\pi_2^{\varepsilon_j}) - \varepsilon_i-\varepsilon_j \le V_1(s,\pi_1^{\varepsilon_j},\pi_2^{\varepsilon_i}) - \varepsilon_i \le V_1(s,\pi_1^{\varepsilon_i},\pi_2^{\varepsilon_i}) \le$$
$$\le V_1(s,\pi_1^{\varepsilon_i},\pi_2^{\varepsilon_j}) + \varepsilon_i \le V_1(s,\pi_1^{\varepsilon_j},\pi_2^{\varepsilon_j}) + \varepsilon_i+\varepsilon_j.$$

Hieruit volgt:

$$|V_1(s,\pi_1^{\varepsilon_i},\pi_2^{\varepsilon_i}) - V_1(s,\pi_1^{\varepsilon_j},\pi_2^{\varepsilon_j})| \le \varepsilon_i+\varepsilon_j,\qquad \forall s\in S,$$

zodat de rij $V_1(s,\pi_1^{\varepsilon_n},\pi_2^{\varepsilon_n})$, $n = 1,2,\ldots$ convergeert, $\forall s\in S$. Laat $V(s) = \lim_{n\to\infty} V_1(s,\pi_1^{\varepsilon_n},\pi_2^{\varepsilon_n})$. Voor iedere $\varepsilon > 0$ bestaat er dan een n, z.d.d. $\varepsilon_n < \varepsilon$ en tevens

$$(1.6)\qquad |V_1(s,\pi_1^{\varepsilon_n},\pi_2^{\varepsilon_n}) - V(s)| \le \varepsilon,\qquad \forall s\in S.$$

Uit (1.5) en (1.6) volgt nu $V_1(s,\pi_1^{\varepsilon_n},\pi_2) \ge V_1(s,\pi_1^{\varepsilon_n},\pi_2^{\varepsilon_n}) - \varepsilon_n \ge V(s) - 2\varepsilon$, voor alle $\pi_2\in\Pi_2$, waaruit we kunnen afleiden:

$$\sup_{\pi_1} \inf_{\pi_2} V_1(s,\pi_1,\pi_2) \geq V(s)-2\varepsilon, \qquad \forall s \in S.$$

Analoog kunnen we uit (1.5) en (1.6) afleiden

$$\inf_{\pi_2} \sup_{\pi_1} V_1(s,\pi_1,\pi_2) \leq V(s)+2\varepsilon, \qquad \forall s \in S,$$

zodat de kombinatie met lemma 1.1.19 oplevert, dat $V(\cdot)$ de waarde van het spel is. □

In het volgende hoofdstuk zullen we voorwaarden geven waaronder een twee-persoons nulsomspel een waarde bezit. Daarbij zullen we een aantal klassieke resultaten voor matrixspelen nodig hebben, die in de volgende sektie behandeld worden.

1.2. Matrixspelen

Uit definitie 1.1.6 blijkt, dat een matrixspel gerepresenteerd kan worden d.m.v. een matrix G van m rijen en n kolommen, waarbij m het aantal zuivere akties voor speler 1 is en n het aantal zuivere akties voor speler 2. Het getal g_{ij} op de (i,j)-de plaats in de matrix is de uitbetaling aan speler 1 als speler 1 zuivere aktie i kiest en speler 2 zuivere aktie j. De uitbetaling aan speler 2 is dan $-g_{ij}$.

Een gemengde aktie voor speler 1 (notatie π_1) kan gezien worden als een rij-vektor: $\pi_1 = (\pi_{11},\pi_{12},\ldots,\pi_{1m})$, met $\pi_{1i} \geq 0$, alle i en $\sum_{i=1}^{m} \pi_{1i} = 1$. Evenzo is een gemengde aktie voor speler 2 (notatie π_2) een kolom-vektor $\pi_2 = (\pi_{21},\pi_{22},\ldots,\pi_{2n})^T$, met $\pi_{2j} \geq 0$, alle j en $\sum_{j=1}^{n} \pi_{2j} = 1$. Door een gemengde aktie voor speler 1 als een rij-vektor en voor speler 2 als een kolom-vektor te noteren vermijden we in het navolgende steeds het getransponeerd teken te moeten schrijven.

Merk op, dat de verzameling van alle gemengde strategieën voor speler 1 (notatie Π_1) kan worden gezien als een kompakte, konvexe deelverzameling van de $\mathbb{R}^m$. Evenzo voor speler 2, maar nu in de $\mathbb{R}^n$ (notatie Π_2).

Als speler 1 aktie $\pi_1 \in \Pi_1$ speelt en speler 2 $\pi_2 \in \Pi_2$ dan is de verwachte uitbetaling aan speler 1 (notatie $V_1(\pi_1,\pi_2)$) gelijk aan $V_1(\pi_1,\pi_2) = \pi_1 G \pi_2$. De verwachte uitbetaling aan speler 2 (notatie $V_2(\pi_1,\pi_2)$) is dan $V_2(\pi_1,\pi_2) = -\pi_1 G \pi_2$.

<u>LEMMA 1.2.1</u>. *Een paar akties (π_1^*,π_2^*) is dan en slechts dan evenwichtspunt*

voor het matrixspel G *als geldt:*

$$\pi_1 G\pi_2^* \le \pi_1^* G\pi_2^* \le \pi_1^* G\pi_2, \qquad \forall(\pi_1,\pi_2) \in \Pi_1\times\Pi_2.$$

Als er een evenwichtspunt (π_1^*,π_2^*) *bestaat, dan is de waarde van het spel* $\pi_1^* G\pi_2^*$.

<u>BEWIJS</u>. Het eerste deel van het lemma volgt direkt uit de definitie 1.1.15 van evenwichtspunt en het feit dat

$$\pi_1 G\pi_2 = V_1(\pi_1,\pi_2) = -V_2(\pi_1,\pi_2), \qquad \forall(\pi_1,\pi_2) \in \Pi_1\times\Pi_2.$$

Het tweede gedeelte volgt direkt uit lemma 1.1.20. □

<u>LEMMA 1.2.2</u>. *Een matrixspel bezit dan en slechts dan een waarde* V *als geldt:*

$$\max_{\pi_1}\min_{\pi_2} \pi_1 G\pi_2 = \min_{\pi_2}\max_{\pi_1} \pi_1 G\pi_2 = V.$$

<u>BEWIJS</u>. Het bestaan van een waarde voor een spel is equivalent met het strikt bepaald zijn van het spel hetgeen equivalent is met

$$\sup_{\pi_1}\inf_{\pi_2} \pi_1 G\pi_2 = \inf_{\pi_2}\sup_{\pi_1} \pi_1 G\pi_2 = V \qquad \text{(definitie 1.1.18).}$$

We moeten dus laten zien dat sup inf vervangen mag worden door max min en inf sup door min max.

Daar $\pi_1 G\pi_2$ continu is op $\Pi_1\times\Pi_2$ en daar Π_1 (als deelverzameling van $\mathbb{R}^m$) en Π_2 (als deelverzameling van $\mathbb{R}^n$) kompakte verzamelingen zijn blijkt achtereenvolgens

$$\inf_{\pi_2} \pi_1 G\pi_2 = \min_{\pi_2} \pi_1 G\pi_2, \qquad \forall\pi_1 \in \Pi_1,$$

$$\min_{\pi_2} \pi_1 G\pi_2 \text{ is continu op } \Pi_1$$

en

$$\sup_{\pi_1}\min_{\pi_2} \pi_1 G\pi_2 = \max_{\pi_1}\min_{\pi_2} \pi_1 G\pi_2.$$

Op dezelfde manier kunnen we laten zien dat inf sup vervangen mag worden door min max. □

De waarde van een matrixspel G wordt in het vervolg aangeduid als val G.

STELLING 1.2.3. *Voor ieder matrixspel* G *geldt*

$$\max_{\pi_1} \min_{\pi_2} \pi_1 G \pi_2 = \min_{\pi_2} \max_{\pi_1} \pi_1 G \pi_2 = \text{val } G.$$

Deze stelling is voor het eerst bewezen door J. VON NEUMANN (1928) [40]. Nadien zijn hiervoor nog vele bewijzen in de literatuur verschenen. We zullen hier een bewijs afkomstig van J. NASH (1950) [39] geven.

BEWIJS. Het bewijs verloopt als volgt. We gaan een afbeelding

$$T\colon \Pi_1 \times \Pi_2 \to \Pi_1 \times \Pi_2$$

definiëren, die minstens 1 dekpunt bezit. De afbeelding T zal zo gekozen worden dat ieder dekpunt van T overeenkomt met een evenwichtspunt van het spel G.
Noteer de i-de rij van G als $G_{i\cdot}$.
Noteer de j-de kolom van G als $G_{\cdot j}$.
Definieer voor $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$ en voor $i = 1,\dots,m$; $j = 1,\dots,n$:

$$c_i(\pi_1,\pi_2) = \max(0, G_{i\cdot}\pi_2 - \pi_1 G \pi_2), \tag{1.7}$$

$$d_j(\pi_1,\pi_2) = \max(0, \pi_1 G \pi_2 - \pi_1 G_{\cdot j}). \tag{1.8}$$

T wordt nu gedefinieerd als $T(\pi_1,\pi_2) = (\pi_1^0,\pi_2^0)$ met

$$\pi_{1i}^0 = \frac{\pi_{1i} + c_i(\pi_1,\pi_2)}{1 + \sum_{k=1}^m c_k(\pi_1,\pi_2)}, \qquad i = 1,\dots,m \tag{1.9}$$

en

$$\pi_{2j}^0 = \frac{\pi_{2j} + d_j(\pi_1,\pi_2)}{1 + \sum_{\ell=1}^n d_\ell(\pi_1,\pi_2)}, \qquad j = 1,\dots,n. \tag{1.10}$$

Men kan direkt nagaan dat $\pi_{1i}^0 \geq 0$, alle i en $\sum_{i=1}^m \pi_{1i}^0 = 1$, d.w.z. $\pi_1^0 \in \Pi_1$. Evenzo $\pi_2^0 \in \Pi_2$.

We zullen nu aantonen dat de afbeelding $T\colon \Pi_1 \times \Pi_2 \to \Pi_1 \times \Pi_2$ een continue afbeelding is en minstens 1 dekpunt bezit. De continuïteit van T volgt direkt uit de definities (1.7), (1.8), (1.9) en (1.10). Uit het feit dat Π_1 en Π_2 kompakte konvexe verzamelingen zijn volgt direkt dat ook $\Pi_1 \times \Pi_2$ kompakt konvex is. Op grond van de dekpuntsstelling van Brouwer volgt nu dat T minstens 1 dekpunt bezit.

We tonen tenslotte aan, dat (π_1^*, π_2^*) dan en slechts dan dekpunt van T is als (π_1^*, π_2^*) evenwichtspunt voor het spel G is.

a) Stel (π_1^*, π_2^*) is een evenwichtspunt voor G. Dan geldt volgens lemma 1.2.1:

$$\pi_1 G \pi_2^* \le \pi_1^* G \pi_2^* \le \pi_1^* G \pi_2, \qquad \text{voor alle } (\pi_1, \pi_2) \in \Pi_1 \times \Pi_2.$$

Kiezen we in het bijzonder $\pi_1 = (0,0,\ldots,\pi_{1i}=1,\ldots,0)$ en $\pi_2 = (0,0,\ldots,\pi_{2j}=1,\ldots,0)$ dan volgt $G_{i.}\pi_2^* \le \pi_1^* G \pi_2^* \le \pi_1^* G_{.j}$, voor alle i,j. Dus $c_i(\pi_1^*, \pi_2^*) = 0$ en $d_j(\pi_1^*, \pi_2^*) = 0$ voor alle i,j, zodat uit (1.9) en (1.10) volgt dat $T(\pi_1^*, \pi_2^*) = (\pi_1^*, \pi_2^*)$.

b) Laat nu (π_1^*, π_2^*) een dekpunt van T zijn. Uit de relatie $\pi_1^* G \pi_2^* = \sum_{i=1}^m \pi_{1i}^*(G_{i.}\pi_2^*)$ en het feit dat $\pi_{1i}^* \ge 0$ met $\sum_{i=1}^m \pi_{1i}^* = 1$ volgt dat er tenminste 1 index i_0 bestaat met $\pi_{1i_0}^* > 0$ en $G_{i.}\pi_2^* \le \pi_1^* G \pi_2^*$.
Voor deze i_0 geldt dan $c_{i_0}(\pi_1^*, \pi_2^*) = 0$ en vanwege (π_1^*, π_2^*) dekpunt geldt tevens

$$\pi_{1i_0}^* = \frac{\pi_{1i_0}^*}{1 + \sum_{k=1}^m c_k(\pi_1^*, \pi_2^*)}.$$

Aangezien $\pi_{1i_0}^* > 0$ volgt hieruit $\sum_{i=1}^m c_i(\pi_1^*, \pi_2^*) = 0$ en daar $c_i(\pi_1^*, \pi_2^*) \ge$ ≥ 0 voor alle i kunnen we concluderen, dat $c_i(\pi_1^*, \pi_2^*) = 0$ voor alle i.

Uit (1.7) volgt nu

$$G_{i.}\pi_2^* \le \pi_1^* G \pi_2^*, \qquad i = 1,\ldots,m. \tag{1.11}$$

Laat $\pi_1 \in \Pi_1$. Vermenigvuldigen we beide leden van (1.11) met $\pi_{1i} \ge 0$ en sommeren we over i, dan volgt m.b.v. het feit dat $\sum \pi_{1i} = 1$ dat

$$\pi_1 G \pi_2^* \le \pi_1^* G \pi_2^*, \qquad \text{voor alle } \pi_1 \in \Pi_1. \tag{1.12}$$

Op analoge manier kunnen we afleiden dat

$$\pi_1^* G \pi_2^* \le \pi_1^* G \pi_2, \qquad \text{voor alle } \pi_2 \in \Pi_2. \tag{1.13}$$

Uit (1.12), (1.13) en lemma 1.2.1 volgt nu dat het paar (π_1^*, π_2^*) een evenwichtspunt voor het spel G is. We kunnen nu concluderen, dat het matrixspel minstens 1 evenwichtspunt bezit. De lemma's 1.2.1 en 1.2.2 leveren nu het gestelde in de stelling. □

Uit deze stelling blijkt niet alleen dat ieder matrixspel een waarde heeft maar tevens, dat de beide spelers optimale strategieën bezitten.

<u>LEMMA 1.2.4</u>. *Zijn* G_1 *en* G_2 *twee* $m\times n$-*spelmatrices en geldt* $g_{1ij} \le g_{2ij} + k$ *voor alle* i,j *dan is* val $G_1 \le$ val $G_2 + k$.

<u>BEWIJS</u>. Laat π_2^* een optimale strategie zijn voor speler 2 in spel G_2, dan voor alle i:

$$G_{1i.}\pi_2^* \le G_{2i.}\pi_2^* + k \le \text{val } G_2 + k.$$

De strategie π_2^* garandeert speler 2 voor spel G_1 dus tegen iedere zuivere aktie van speler 1 een verwachting van de uitbetaling aan speler 1 van hoogstens val $G_2 + k$. M.a.w. val $G_1 \le$ val $G_2 + k$. □

<u>OPMERKING 1.2.5</u>. Als geldt $g_{1ij} < g_{2ij} + k$ voor alle i,j, dan val $G_1 <$ $<$ val $G_2 + k$.

<u>LEMMA 1.2.6</u>. *Zijn* G_1 *en* G_2 *twee* $m\times n$-*spelmatrices dan geldt*

$$\text{val}\,|G_1 - \text{val } G_2| \le \max_{i,j} |g_{1ij} - g_{2ij}|.$$

<u>BEWIJS</u>. Er geldt $g_{1ij} \le g_{2ij} + \max_{i,j}|g_{1ij} - g_{2ij}|$ alle i,j. Uit lemma 1.2.4 volgt dan

$$\text{val } G_1 - \text{val } G_2 \le \max_{i,j}|g_{1ij} - g_{2ij}|. \tag{1.14}$$

Op grond van symmetrie geldt eveneens

$$\text{val } G_2 - \text{val } G_1 \le \max_{i,j}|g_{1ij} - g_{2ij}|. \tag{1.15}$$

Kombineren van (1.14) en (1.15) geeft de bewering. □

HOOFDSTUK II

HET STOCHASTISCH SPEL VAN SHAPLEY

2.1. Inleiding

In zijn beroemd artikel [57] beschouwt SHAPLEY een eindig, stoppend, twee-persoons nulsomspel onder het totale opbrengstenkriterium. Hij toont aan, dat dit spel binnen de klasse der Markov-strategieën strikt bepaald is (en dus een waarde heeft) en dat de beide spelers optimale stationaire strategieën bezitten.

Wij zullen de afleiding van deze resultaten geven, waarbij we tevens zullen laten zien dat de optimale strategieën voor de beide spelers binnen de klasse van de Markov-strategieën ook optimaal zijn binnen de klasse van alle (gedrags-)strategieën. Daartoe zullen we eerst het t-staps stochastisch spel beschouwen met een eindig aantal beslissingstijdstippen.

2.2. Het t-staps stochastisch spel

Beschouw een stochastisch spel van Shapley. Laat N het eindige aantal toestanden zijn en laat m_k respektievelijk n_k het eindige aantal zuivere akties voor speler 1 respektievelijk speler 2 in toestand k zijn, $k = 1,2,\dots,N$. Als in toestand k de spelers 1 en 2 respektievelijk zuivere akties i en j kiezen, dan vindt er een direkte uitbetaling plaats van $g(k,i,j)$ aan speler 1 door speler 2 en is de kans dat het spel naar toestand ℓ overgaat $p(\ell|k,i,j)$, $\ell = 1,2,\dots,N$. Dus

$$g_1(k,i,j) = -g_2(k,i,j) = g(k,i,j), \qquad \text{alle } i, j \text{ en } k.$$

De situatie in toestand k kan symbolisch worden weergegeven d.m.v. een $m_k \times n_k$-matrix G_k, waarbij op de (i,j)-de plaats van G_k staat:

$$g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)\, G_\ell .$$

Met $\Gamma(k)$ wordt bedoeld het specifieke stochastische spel, waarbij toestand k de begintoestand is.

We beschouwen een stoppend stochastisch spel, d.w.z. we werken onder de volgende aanname:

$$H_0: \qquad p_0 \stackrel{\text{def}}{=} \min_{k,i,j} \left(1 - \sum_{\ell=1}^{N} p(\ell|k,i,j)\right) > 0.$$

Merk op dat $(1-p_0)^t$ een bovengrens is voor de kans, dat het spel na t stappen nog niet geëindigd is. Derhalve, vanwege $p_0 > 0$, eindigt het spel met kans 1 in een eindig aantal stappen. Merk tevens op, dat de totale verwachte uitbetaling bestaat en voor elke strategie begrensd wordt door M/p_0, waarbij M een getal is met $|g(k,i,j)| \leq M$, alle k, i en j.

DEFINITIE 2.2.1. *Een t-staps twee-persoons nulsom stochastisch spel met nabetalingsfunctie* R is een stochastisch spel dat stopt op tijdstip t (als het daarvoor al niet gestopt is) met een nabetaling R(s) aan speler 1 en -R(s) aan speler 2 als de eindtoestand op tijdstip t toestand s is.

Met $\Gamma^t(k,R)$ bedoelen we het specifieke t-staps twee-persoons nulsom stochastisch spel met nabetalingsfunctie R, waarbij k de begintoestand van het spel is, $t = 0,1,2,\ldots$. Met $\Gamma^t(R) = \{\Gamma^t(k,R) \mid k = 1,\ldots,N\}$ kunnen we dan het t-staps twee-persoons nulsom stochastisch spel met nabetalingsfunctie R aanduiden.

Een Markov-strategie voor speler 1 (speler 2) voor het spel $\Gamma^t(R)$ (notatie π_1^t (π_2^t)) kan gezien worden als $\pi_1^t = \{\pi_1(k,\tau) \mid k = 1,\ldots,n;$ $\tau = 0,1,\ldots,t-1\}$ $(\pi_2^t = \{\pi_2(k,\tau) \mid k = 1,\ldots,N; \tau = 0,1,\ldots,t-1\})$ waarbij $\pi_1(k,\tau) = (\pi_{11}(k,\tau),\ldots,\pi_{1m_k}(k,\tau))$ $(\pi_2(k,\tau) = (\pi_{21}(k,\tau),\ldots,\pi_{2n_k}(k,\tau)))$ een gemengde aktie is voor speler 1 (speler 2) in toestand k op tijdstip τ.

Met $\text{val}\{a_{ij}\}$ bedoelen we de waarde van een matrixspel waarvan het (i,j)-de element gelijk aan a_{ij} is.

STELLING 2.2.2. *Ieder t-staps stochastisch spel van Shapley met nabetalingsfunctie* R *is strikt bepaald. De waarde-vektor kan worden gevonden d.m.v. de volgende recurrente betrekking:*

$$\text{val } \Gamma^\tau(k,R) = \text{val}\left\{g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)\cdot\text{val } \Gamma^{\tau-1}(\ell,R)\right\},$$

$$k = 1,\ldots,N;$$

$$\tau = 1,\ldots,t,$$

waarbij val $\Gamma^0(k,R) \overset{\text{def}}{=} R(k)$.

De strategie $\pi_1^{*t} = \{\pi_1^*(k,\tau) \mid k = 1,2,\ldots,N;\ \tau = 0,1,\ldots,t-1\}$ $(\pi_2^{*t} = $ $= \{\pi_2^*(k,\tau) \mid k = 1,\ldots,N;\ \tau = 0,1,\ldots,t-1\}$ *voor speler 1 (speler 2), waarbij* $\pi_1^*(k,\tau)$ $(\pi_2^*(k,\tau))$ *een optimale gemengde aktie is voor speler 1 (speler 2) in het matrixspel* $\{g(k,i,j) + \sum_{\ell=1}^N p(\ell|k,i,j)\ \mathrm{val}\ \Gamma^{t-\tau-1}(\ell,R)\}$, $k =$ $= 1,2,\ldots,N;\ \tau = 0,1,\ldots,t-1$, *is een optimale Markov-strategie binnen de klasse van alle strategieën voor speler 1 (speler 2) in het spel* $\Gamma^t(R)$.

<u>BEWIJS</u>. Noteer de totale verwachte uitbetaling in het t-staps stochastisch spel $\Gamma^t(R)$ bij strategieën π_1 en π_2 als $V^t(k,\pi_1,\pi_2)$ voor begintoestand k. Het bewijs verloopt d.m.v. induktie naar t.

a) Voor t = 1 hebben we alleen te maken met de verzameling matrixspelen

$$\left\{\{g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)\ R(\ell)\} \mid k = 1,\ldots,N\right\}$$

en ieder van deze spelen bezit een waarde en optimale strategieën voor de beide spelers (stelling 1.2.3).

b) Veronderstel dat de bewering juist is voor $\tau = 1,2,\ldots,t-1$. We laten nu zien dat de bewering dan ook juist is voor $\tau = t$. Laat voor het (t-1)-staps spel π_1^{*t-1} en π_2^{*t-1} optimale Markov-strategieën zijn voor speler 1 respektievelijk speler 2 in de zin van de bewering in de stelling. Dan is $\bar{V}^{t-1} = (V^{t-1}(1,\pi_1^{*t-1},\pi_2^{*t-1}),\ldots,V^{t-1}(N,\pi_1^{*t-1},\pi_2^{*t-1}))$ de waarde-vektor van $\Gamma^{t-1}(R)$.

Beschouw nu de verzameling matrixspelen:

$$\left\{\{g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)\cdot V^{t-1}(\ell,\pi_1^{*t-1},\pi_2^{*t-1})\} \mid k=1,2,\ldots,N\right\}.$$

Laat $\bar{V}^* = (V_1^*,\ldots,V_N^*)$, waarbij V_k^* de waarde is van het k-de matrixspel en laat $\pi_1^*(k)$ $(\pi_2^*(k))$ hierbij een optimale aktie voor speler 1 (speler 2) zijn. Definieer voor het t-stapsspel de Markov-strategie π_1^{*t} voor speler 1 als volgt:

π_1^{*t} schrijft op tijdstip 0 $\pi_1^*(k)$ voor in toestand k en
op tijdstip τ $\pi_1^{*t-1}(k,\tau-1)$ in toestand k, $\tau = 1,\ldots,t-1$.

Laat voor een willekeurige strategie π_2 voor speler 2 voor het t-stapsspel $\pi_2(0,k) = (\pi_{21}(0,k),\ldots,\pi_{2n_k}(0,k)$ het voorschrift op tijdstip 0 in toestand k en laat π_{2kij} de strategie voor speler 2 voor $\Gamma^{t-1}(R)$ afgeleid uit π_2 als volgt: π_{2kij} schrijft op tijdstip 0 voor in toestand ℓ, wat

π_2 op tijdstip 1 in toestand ℓ voorschrijft bij historie $h_1 = (k,i,j)$, etc.

Er geldt dan voor $k = 1,2,\dots,N$:

$$\min_{\pi_2} V^t(k,\pi_1^{*t},\pi_2) =$$

$$(2.1)\quad = \min_{\pi_2} \sum_{j=1}^{n_k} \pi_{2j}(0,k) \sum_{i=1}^{m_k} \pi_{1i}^*(k)\left\{g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)V^{t-1}(\ell,\pi_1^{*t-1},\pi_{2kij})\right\}$$

$$(2.2)\quad \ge \min_{\pi_2} \sum_{j=1}^{n_k} \pi_{2j}(0,k) \sum_{i=1}^{m_k} \pi_{1i}^*(k)\left\{g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)V^{t-1}(\ell,\pi^{*t-1},\pi_2^{*t-1})\right\}$$

$$(2.3)\quad = \mathrm{val}\left\{g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)V^{t-1}(\ell,\pi_1^{*t-1},\pi_2^{*t-1})\right\} = V_k^*.$$

De overgang van (2.2) naar (2.3) is gebaseerd op het optimaal zijn van $\pi_1^*(k)$ voor speler 1 in het matrixspel

$$\left\{g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)V^{t-1}(\ell,\pi_1^{*t-1},\pi_2^{*t-1})\right\}.$$

We hebben nu dus

$$(2.4)\qquad \min_{\pi_2} V^t(k,\pi_1^{*t},\pi_2) \ge V_k^*, \qquad k = 1,\dots,N.$$

Als we de Markov-strategie π_2^{*t} voor speler 2 definiëren als: π_2^{*t} schrijft op tijdstip 0 $\pi_2^*(k)$ in toestand k voor en op tijdstip τ in toestand k $\pi^{*t-1}(k,\tau-1)$ dan kunnen we op dezelfde manier laten zien dat geldt:

$$(2.5)\qquad \max_{\pi_1} V^t(k,\pi_1,\pi_2^{*t}) \le V_k^*, \qquad k = 1,\dots,N,$$

(2.4) en (2.5) kombineren geeft:

$$(2.6)\qquad V^t(k,\pi_1,\pi_2^{*t}) \le V_k^* = V^t(k,\pi_1^{*t},\pi_2^{*t}) \le V^t(k,\pi_1^{*t},\pi_2),\ k=1,\dots,N.$$

Het paar (π_1^{*t},π_2^{*t}) vormt dus een evenwichtspunt, zodat uit lemma 1.1.20 blijkt dat het spel $\Gamma^t(R)$ strikt bepaald is. Verder is π_1^{*t} (π_2^{*t}) een optimale strategie voor speler 1 (speler 2) en is $(V^t(1,\pi_1^{*t},\pi_2^{*t})),\dots,$ $V^t(N,\pi_1^{*t},\pi_2^{*t}))$ de waarde-vektor. Uit (2.3) blijkt de recurrente betrekking voor $\tau = t$. Uit de constructiewijze van π_1^{*t} en π_2^{*t} en de aanname dat π_1^{*t-1} en π_2^{*t-1} voldoen aan de bewering in de stelling, blijkt dat π_1^{*t} en π_2^{*t} voldoen aan de bewering in de stelling.

Hiermee is de inductiestap voltooid, waarmee de stelling bewezen is. □

2.3. Het stochastisch spel van Shapley

We gaan nu bewijzen dat een stochastisch spel van Shapley een waarde-vektor bezit en dat de beide spelers optimale stationaire strategieën ter beschikking staan.

De totale verwachte uitbetaling voor het spel startend in toestand k als speler 1 strategie π_1 speelt en speler 2 speelt π_2 noteren we als $V_k(\pi_1,\pi_2)$. Laat $\bar{V}(\pi_1,\pi_2) = (V_1(\pi_1,\pi_2),\dots,V_N(\pi_1,\pi_2))$.

DEFINITIE 2.3.1. De $m_k \times n_k$-matrixspelen $G_k(\bar{v})$, waarbij het (i,j)-de element gelijk is aan $g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)v_\ell$ en $\bar{v} \in \mathbb{R}^N$ worden de *dummy spelen* bij de vektor $\bar{v}$ genoemd, $k = 1,\dots,N$.

STELLING 2.3.2. *Het stochastisch spel van Shapley is strikt bepaald ten opzichte van het totale opbrengstenmodel. De waarde-vektor kan gevonden worden als de unieke oplossing van het stelsel* $\phi_k = \text{val } G_k(\bar{\phi})$, $k = 1,\dots,N$ *met* $\bar{\phi} = (\phi_1,\dots,\phi_N)$.

BEWIJS. Definieer de afbeelding $T\colon \mathbb{R}^N \to \mathbb{R}^N$ als $T\bar{\alpha} = \bar{\beta}$ met $\beta_k = \text{val } G_k(\bar{\alpha})$. Beschouw op $\mathbb{R}^N$ de norm $\|\bar{\alpha}\| = \max_k |\alpha_k|$; dan is $\mathbb{R}^N$ een volledige metrische ruimte onder deze norm.

We laten eerst zien dat de afbeelding T een contractieafbeelding is onder de norm $\|\cdot\|$. Voor alle $\bar{\alpha},\bar{\beta} \in \mathbb{R}^N$ vinden we met behulp van lemma 1.2.6:

$$\begin{aligned}
\|T\bar{\alpha}-T\bar{\beta}\| &= \max_k |\text{val } G_k(\bar{\alpha}) - \text{val } G_k(\bar{\beta})| \\
&\le \max_{k,i,j} \Big|\sum_{\ell=1}^{N} p(\ell|k,i,j)(\alpha_\ell-\beta_\ell)\Big| \\
&\le \max_{k,i,j} \Big|\sum_{\ell=1}^{N} p(\ell|k,i,j)\Big| \max_\ell |\alpha_\ell-\beta_\ell| \\
&= (1-p_0)\|\bar{\alpha}-\bar{\beta}\|,
\end{aligned}$$

waarmee bewezen is dat T een contractieafbeelding is met contractiefactor $\le 1-p_0$.

Uit de Banach-Picard vaste puntenstelling (bijv. L. SCHWARTZ [55], p.141) volgt nu dat T een uniek dekpunt, zeg $\bar{v}^*$, heeft waarbij voor elke $\bar{\alpha} \in \mathbb{R}^N$ geldt $\lim_{n\to\infty} \|T^n\bar{\alpha}-\bar{v}^*\| = 0$, waaruit volgt $\lim_{n\to\infty} T^n\bar{\alpha} = \bar{v}^*$. Hierbij is $T^n\bar{\alpha} = T(T^{n-1}\bar{\alpha})$, $n = 2,3,\ldots$.

Beschouw nu het t-stapsspel $\Gamma^t(\bar{0}) = \{\Gamma^t(k,\bar{0}) \mid k = 1,\ldots,N\}$. $\bar{0}$ is een N-vektor met iedere komponent 0. Uit de definitie van dummyspelen, de definitie van T en stelling 2.2.2 (recurrente betrekkingen) blijkt dat $T^t(\bar{0})$ de waarde-vektor van $\Gamma^t(\bar{0})$ is.

Als nu in het specifieke stochastische spel $\Gamma(k)$ speler 1 een strategie speelt, die, wat de eerste t-stappen betreft, overeenstemt met een optimale strategie voor $\Gamma^t(k,\bar{0})$ en willekeurig daarna, dan krijgt hij een totale verwachte uitbetaling van tenminste $(T^t(\bar{0}))_k - \varepsilon_t$, waarbij $\varepsilon_t = (1-p_0)^t \frac{M}{p_0}$ en $M = \max_{k,i,j}|g(k,i,j)|$, tegen iedere strategie van speler 2. Daar $\varepsilon_t \to 0$ en $(T^t(\bar{0}))_k \to v_k^*$ als $t \to \infty$ zien we dat er voor iedere $\varepsilon > 0$ en elke k een $\pi_1^{\varepsilon k}$ bestaat zodanig dat

$$V_k(\pi_1^{\varepsilon k},\pi_2) \geq v_k^* - \varepsilon, \qquad \text{voor alle } \pi_2,\ k = 1,\ldots,N,$$

waaruit volgt

$$\sup_{\pi_1} \inf_{\pi_2} V_k(\pi_1,\pi_2) \geq v_k^*, \qquad k = 1,\ldots,N. \tag{2.8}$$

Analoog vinden we

$$\inf_{\pi_2} \sup_{\pi_1} V_k(\pi_1,\pi_2) \leq v_k^*, \qquad k = 1,\ldots,N. \tag{2.9}$$

Uit (2.8), (2.9) en lemma 1.1.19 volgt:

$$\sup_{\pi_1} \inf_{\pi_2} V_k(\pi_1,\pi_2) = \inf_{\pi_2} \sup_{\pi_1} V_k(\pi_1,\pi_2) = v_k^*, \qquad k = 1,\ldots,N. \tag{2.10}$$

Het spel is dus strikt bepaald en heeft waarde-vektor $\bar{v}^*$, die gelijk is aan de unieke oplossing van het stelsel

$$\phi_k = (T\bar{\phi})_k = \text{val } G_k(\bar{\phi}), \qquad k = 1,\ldots,N. \qquad \square$$

<u>STELLING 2.3.3.</u> *Laat voor een stochastisch spel van Shapley* $\bar{v}^*$ *de waarde-vektor zijn. De stationaire strategieën* $\pi_1^* = (\pi_1^*(1),\ldots,\pi_1^*(N))$ *respektievelijk* $\pi_2^* = (\pi_2^*(1),\ldots,\pi_2^*(N))$ *zijn optimaal voor speler 1 respektievelijk*

speler 2 dan en slechts dan als $\pi_1^*(k)$ *respektievelijk* $\pi_2^*(k)$ *overeenstemt met een optimale aktie in het dummy matrixspel* $G_k(\bar{v})$ *voor speler 1 respektievelijk speler 2.*

<u>BEWIJS</u>. a) Veronderstel dat $\pi_1^*(k)$ respektievelijk $\pi_2^*(k)$ overeenstemt met een optimale aktie in het dummy matrixspel $G_k(\bar{v})$ voor speler 1 respektievelijk speler 2. Beschouw het t-stapsspel $\Gamma^t(\bar{v}^*) = \{\Gamma^t(k,\bar{v}^*) \mid k = 1,\ldots,N\}$. Aangezien $v_k^* = \text{val } G_k(\bar{v}^*) = (T\bar{v}^*)_k$ en aangezien val $\Gamma^t(\bar{v}^*) = T^t(\bar{v}^*)$ zien we dat val $\Gamma^t(\bar{v}^*) = \bar{v}^*$, $t = 1,2,\ldots$. Een optimale Markov-strategie voor speler 1 in $\Gamma^t(\bar{v}^*)$ overeenkomstig stelling 2.2.2 is dan de strategie die op ieder tijdstip τ in toestand k voorschrijft de aktie $\pi_1^*(k)$ te spelen, $\tau = 0,1,\ldots,t-1$; $k = 1,\ldots,N$.

Speelt speler 1 nu de stationaire strategie $\pi_1^* = (\pi_1^*(1),\ldots,\pi_1^*(N))$ in Γ, dan geldt voor iedere t, dat zijn totale verwachte uitbetaling na t stappen minstens $\bar{v}^* - (1-p_0)^t\|\bar{v}^*\|$ is tegen iedere strategie van speler 2. Vanaf tijdstip t is de totale verwachte uitbetaling aan speler 1 minstens $-(1-p_0)^t\cdot\frac{M}{p_0}$, voor iedere toestand op tijdstip t en tegen iedere strategie van speler 2.

We zien dus dat de strategie π_1^* speler 1 verzekert van een totale verwachte opbrengst

$$\bar{v}^* - (1-p_0)^t(\|\bar{v}^*\| + \frac{M}{p_0}). \tag{2.11}$$

(2.11) is geldig voor iedere t en daar $\lim_{t\to\infty}(1-p_0)^t = 0$ blijkt dat $\inf_{\pi_2} \bar{V}(\pi_1^*,\pi_2) \geq \bar{v}^*$, zodat π_1^* optimaal is voor speler 1. Evenzo kunnen we laten zien $\sup_{\pi_1} \bar{V}(\pi_1,\pi_2^*) \leq \bar{v}^*$.

b) Laat $\pi_1^* = (\pi_1^*(1),\ldots,\pi_1^*(N))$ een optimale stationaire strategie zijn voor speler 1. Laat π_2 een willekeurige stationaire strategie voor speler 2 zijn. De totale verwachte uitbetaling bij π_1^* en π_2 voldoet aan het stelsel

$$V_k(\pi_1^*,\pi_2) = \sum_{i=1}^{m_k} \pi_{1i}^*(k) \sum_{j=1}^{n_k} \pi_{2j}(k)\left\{g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)V_\ell(\pi_1^*,\pi_2)\right\}, \tag{2.12}$$

$$k = 1,\ldots,N.$$

Veronderstel dat er toestanden k zijn waarvoor $\pi_1^*(k)$ niet optimaal is in $G_k(\bar{v}^*)$. Zonder de algemeenheid te schaden veronderstellen we dat dit de toestanden 1 t/m N_1 zijn. Speler 2 heeft dan een stationaire strategie

$\pi_2^0 = (\pi_2^0(1),\dots,\pi_2^0(N))$, z.d.d.:

$$(2.13)\qquad v_k^* > \sum_{i=1}^{m_k} \pi_{1i}^*(k) \sum_{j=1}^{n_k} \pi_{2j}^0(k)\left\{g(k,i,j)+\sum_{\ell=1}^{N} p(\ell|k,i,j)v_\ell^*\right\},\ k = 1,\dots,N_1$$

en

$$(2.14)\qquad v_k^* = \sum_{i=1}^{m_k} \pi_{1i}^*(k) \sum_{j=1}^{n_k} \pi_{2j}^0(k)\left\{g(k,i,j)+\sum_{\ell=1}^{N} p(\ell|k,i,j)v_\ell^*\right\},\ k = N_1+1,\dots,N.$$

(2.13) en (2.14) aftrekken van (2.12) voor π_2^0 geeft

$$(2.15)\qquad V_k(\pi_1^*,\pi_2^0)-v_k^* < \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} \pi_{1i}^*(k)\cdot\pi_{2j}^0(k) \sum_{\ell=1}^{N} p(\ell|k,i,j)\{V_\ell(\pi_1^*,\pi_2^0)-v_\ell\},$$

$$k = 1,\dots,N_1$$

$$(2.16)\qquad V_k(\pi_1^*,\pi_2^0)-v_k^* = \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} \pi_{1i}^*(k)\cdot\pi_{2j}^0(k) \sum_{\ell=1}^{N} p(\ell|k,i,j)\{V_\ell(\pi_1^*,\pi_2^0)-v_\ell\}$$

$$k = N_1+1,\dots,N.$$

Daar $\sum_{i=1}^{m_k} \pi_{1i}^*(k) = 1$, $\sum_{j=1}^{n_k} \pi_{2j}^0(k) = 1$ en $0 \le \sum_{\ell=1}^{N} p(\ell|k,i,j) \le 1-p_0 < 1$ en daar $V_k(\pi_1^*,\pi_2^0) \ge v_k^*$, alle k ($\pi_1^*$ is optimaal), blijkt uit (2.15) en (2.16) dat $\max_k(V_k(\pi_1^*,\pi_2^0)-v_k^*) = 0$ (anders is er een k waarvoor linkerlid groter is dan rechterlid) en dus $V_k(\pi_1^*,\pi_2^0) = v_k^*$, alle k, maar dit is in tegenspraak met (2.15). M.a.w. de veronderstelling dat er toestanden k zijn waarvoor $\pi_1^*(k)$ niet optimaal is, is niet juist. □

OPMERKING 2.3.4. Het stochastisch spel van Shapley met rationale coëfficiënten $g(k,i,j)$ en $p(\ell|k,i,j)$ hoeft niet noodzakelijkerwijs een rationale waarde te hebben. Dit in tegenstelling tot het Markov-beslissingsprobleem met eindig aantal toestanden en eindig aantal akties in iedere toestand.

VOORBEELD. $\Gamma = \{\Gamma_1\}$; beide spelers hebben 2 zuivere akties ter beschikking.

$$\Gamma_1: \begin{pmatrix} 4 & 1+\tfrac{1}{2}\Gamma_1 \\ 2 & 1+\tfrac{3}{4}\Gamma_1 \end{pmatrix}.$$

De unieke oplossing van de vergelijking

$$\phi = \mathrm{val}\begin{pmatrix} 4 & 1+\tfrac{1}{2}\phi \\ 2 & 1+\tfrac{3}{4}\phi \end{pmatrix}$$

wordt gegeven door $v^* = \sqrt{8}$.

OPMERKING 2.3.5. In hoofdstuk I bij de bespreking van de diverse kriteria hebben we reeds opgemerkt dat een niet-stoppend spel onder het verdisconteerde kriterium opgevat kan worden als een speciaal geval van het stoppend spel onder het totale opbrengstenkriterium. Namelijk alle stopkansen zijn in iedere toestand onder ieder paar akties gelijk aan $1 - \beta$. We zien dus dat alle resultaten afgeleid in dit hoofdstuk ook van toepassing zijn op spelen onder het verdisconteerde opbrengstenkriterium.

OPMERKING 2.3.6. De resultaten in dit hoofdstuk kunnen gegeneraliseerd worden tot spelen waarbij voor ieder paar zuivere stationaire strategieën voor de beide spelers geldt dat de bijbehorende matrix P van overgangskansen "transient" is, d.w.z. $\lim_{t\to\infty} P^t = 0$, waarbij $P^t = P(P^{t-1})$, $t = 2,3,\ldots$. Dit kan dan op dezelfde manier als VEINOTT [65] gedaan heeft voor Markov-beslissingsproblemen.

OPMERKING 2.3.7. Een stochastisch spel van Shapley met perfekte informatie is een spel, waarbij in toestand k óf $m_k = 1$ óf $n_k = 1$, $k = 1,\ldots,N$. Uit stelling 2.3.3 volgt dan, dat voor een spel met perfekte informatie beide spelers optimale zuivere stationaire strategieën bezitten.

OPMERKING 2.3.8. Als $n_k = 1$, $k = 1,\ldots,N$, dus de tweede speler heeft in iedere toestand maar 1 aktie ter beschikking, waardoor hij in wezen geëlimineerd wordt, dan hebben we te maken met een Markov-beslissingsprobleem voor speler 1. Voor dit geval heeft speler 1 optimale zuivere stationaire strategieën (bijv. HOWARD [25]). Merk op dat dit probleem een speciaal geval is van opmerking 2.3.7.

HOOFDSTUK III

TWEE-PERSOONS NULSOMSPELEN
MET OVERAFTELBARE AKTIERUIMTEN

3.1. Inleiding

In dit hoofdstuk behandelen we de klasse van nulsomspelen met eindige toestandsruimte, pre-kompakte aktieruimten voor de beide spelers en de overgangskansen zijn zodanig dat het spel stoppend is. Deze klasse van spelen is voor het eerst bestudeerd door M. TAKAHASHI in zijn artikel *Stochastic games with infinitely many strategies* [62]. Als kriterium beschouwt hij het totale opbrengstenkriterium.

In zijn artikel bewijst Takahashi het strikt bepaald zijn van dit spel binnen de klasse der stationaire strategieën onder bepaalde voorwaarden op de uitbetalingsfuncties en de overgangskansen. In het laatste gedeelte van zijn artikel laat hij zien dat het spel ook strikt bepaald is binnen de klasse der Markov-strategieën en dat in beide gevallen de waarde gelijk is.

Ten opzichte van het stochastisch spel van Shapley is de enige verandering het overaftelbaar zijn van de zuivere aktieruimten. Opgemerkt dient te worden dat de klasse van spelen met aftelbare aktieruimten (en niet eindig) niet binnen de klasse van spelen ligt, die in dit hoofdstuk behandeld wordt.

In de sekties 3.2 en 3.3 worden Takahashi's resultaten gepresenteerd. In sektie 3.4 breiden we deze resultaten uit tot de klasse van (gedrags-) strategieën, waarbij gebruik gemaakt wordt van enkele resultaten uit de Markov-beslissingstheorie.

3.2. Eénstapsspelen, dummyspelen

In deze sektie leiden we enkele resultaten betreffende éénstapsspelen (spelen in normale vorm) en dummyspelen af.

Een niet-coöperatief twee-persoons nulsomspel in normale vorm is een triplet (A_1,A_2,g), waarbij A_1 (A_2) de zuivere aktieruimte voor speler 1 (speler 2) en $g\colon A_1\times A_2 \to \mathbb{R}$ de uitbetalingsfunctie voor speler 1; de uitbetalingsfunctie voor speler 2 is dan $-g$.

DEFINITIE 3.2.1. Laten A_1 en A_2 twee verzamelingen zijn. Laat $f\colon A_1\times A_2 \to \mathbb{R}$ een uniform begrensde reëelwaardige functie zijn. De functie δ op $A_1\times A_1$ gedefinieerd als

$$\delta^f(a_{11},a_{12}) = \sup_{a_2\in A_2} |f(a_{11},a_2) - f(a_{12},a_2)| \qquad \text{voor alle } (a_{11},a_{12}) \in A_1\times A_1, \tag{3.1}$$

noemen we de *intrinsieke metriek* op A_1 bij (A_2,f).
Evenzo is de intrinsieke metriek op A_2 bij (A_1,f) gedefinieerd als

$$\delta^f(a_{21},a_{22}) = \sup_{a_1\in A_1} |f(a_1,a_{21}) - f(a_1,a_{22})| \qquad \text{voor alle } (a_{21},a_{22})\in A_2\times A_2.$$

Onder deze intrinsieke metrieken zijn A_1 en A_2 pseudo-metrische ruimtes. Als namelijk $\delta^f(a_{11},a_{12}) = 0$, dan is niet noodzakelijkerwijs $a_{11} = a_{12}$. Echter A_i, $i = 1,2$, kan gereduceerd worden tot een metrische ruimte $\tilde{A}_i$, $i = 1,2$ door die punten samen te laten vallen die afstand 0 hebben. Deze reductie heeft geen invloed op de speelmogelijkheden voor de beide spelers, hetgeen als volgt blijkt. Laat voor $a_i \in A_i$ $\tilde{a}_i(a_i) \in \tilde{A}_i$ de representant van de verzameling van punten die met a_i samenvallen. Laat $\tilde{f}$ de restrictie van f tot $\tilde{A}_1\times\tilde{A}_2$, dan geldt

$$\tilde{f}(\tilde{a}_1(a_1),\tilde{a}_2(a_2)) = f(a_1,a_2) \qquad \text{voor alle } (a_1,a_2) \in A_1\times A_2.$$

In het vervolg nemen we aan, dat (A_i,δ^f), $i = 1,2$ metrische ruimtes zijn, eventueel verkregen na reductie.

Laat $\mathcal{A}_i$, de kleinste σ-algebra die alle open deelverzamelingen van (A_i,δ^f) omvat en laat Π_i de verzameling van alle kansmaten gedefinieerd op $\mathcal{A}_i$, $i = 1,2$.

Laat $\mathcal{A}$ de produkt σ-algebra van $\mathcal{A}_1$ en $\mathcal{A}_2$, d.w.z. de kleinste σ-algebra bij $A_1\times A_2$ die $\mathcal{A}_1\times\mathcal{A}_2$ omvat. $\Pi_1\times\Pi_2$ is dan de verzameling van alle produktkansmaten op $\mathcal{A}$. Merk op dat zowel $\mathcal{A}_i$, Π_i als $\mathcal{A}$ afhankelijk zijn van f.

Beschouw nu het twee-persoons nulsomspel in normale vorm (A_1,A_2,g). Als bovenstaand kunnen we definiëren:

$$(A_1,\delta^g),(A_2,\delta^g),\mathcal{A}_1,\mathcal{A}_2,\Pi_1,\Pi_2,\mathcal{A}.$$

De verzameling Π_i is dan de verzameling van gemengde akties voor speler i, $i = 1,2$.

<u>LEMMA 3.2.2.</u> *Als* $g\colon A_1\times A_2 \to \mathbb{R}$ *een uniform begrensde meetbare functie is t.o.v.* $\mathcal{A}$ *en als speler 1* $\pi_1 \in \Pi_1$ *en speler 2* $\pi_2 \in \Pi_2$ *speelt, dan is de verwachte uitbetaling (aan speler 1) te schrijven als:*

$$\begin{aligned} g(\pi_1,\pi_2) &= \int_{A_1\times A_2} g(a_1,a_2)d(\pi_1(a_1)\times\pi_2(a_2)) \\ &= \int_{A_1}\left\{\int_{A_2} g(a_1,a_2)d\pi_2(a_2)\right\}d\pi_1(a_1) = \\ &= \int_{A_2}\left\{\int_{A_1} g(a_1,a_2)d\pi_1(a_1)\right\}d\pi_2(a_2). \end{aligned}$$

<u>BEWIJS</u>. Uit de uniforme begrensdheid van g volgt

$$\int_{A_1\times A_2} |g(a_1,a_2)|\,|d(\pi_1\times\pi_2)| < \infty$$

en daar $g(\cdot,\cdot)$ meetbaar is t.o.v. $\mathcal{A}$ geldt dat g integreerbaar is op $A_1\times A_2$, zodat toepassen van de stelling van Fubini de bewering in het lemma geeft. □

Over de vraag wanneer g meetbaar is t.o.v. $\mathcal{A}$, doet het volgende lemma een uitspraak.

<u>LEMMA 3.2.3.</u> *Als* $g\colon A_1\times A_2 \to \mathbb{R}$ *begrensd is en minstens één van de verzamelingen* (A_1,δ^g) *en* (A_2,δ^g) *separabel is, dan is* $g(\cdot,\cdot)$ *meetbaar t.o.v.* $\mathcal{A}$.

<u>BEWIJS</u>. Het bewijs van deze bewering is nogal technisch van aard. We volstaan met een verwijzing naar WALD [71], pag. 34 en 35. □

Met een pre-kompakte metrische ruimte bedoelen we een metrische ruimte waarvoor iedere rij een Cauchy deelrij bezit. Het pre-kompakt zijn van een metrische ruimte is equivalent met: Voor iedere $\varepsilon > 0$ is er een eindige deelverzameling die ε-dicht ligt in de ruimte, d.w.z. er bestaat een natuurlijk getal N_ε, z.d.d. er N_ε punten te vinden zijn, waarvoor de bijbehorende N_ε sfeertjes met straal ε de ruimte overdekken (zie KELLY [29]; de metrische versie van stelling 32, pag. 198).

STELLING 3.2.4. *Als voor een spel* (A_1,A_2,g) $g: A_1\times A_2 \to \mathbb{R}$ *uniform begrensd is en* (A_1,δ^g) *pre-kompakt, dan geldt*

a) (A_2,δ^g) *is pre-kompakt;*

b) (A_1,A_2,g) *is strikt bepaald, d.w.z.*

$$\sup_{\pi_1}\inf_{\pi_2} g(\pi_1,\pi_2) = \inf_{\pi_2}\sup_{\pi_1} g(\pi_1,\pi_2).$$

BEWIJS. Een pre-kompakte metrische ruimte is separabel, zodat (A_1,δ^g) separabel is. Uit de lemma's 3.2.3 en 3.2.2 blijkt nu dat $g(\pi_1,\pi_2)$ ééndduidig bepaald is voor iedere $(\pi_1,\pi_2) \in \Pi_1\times\Pi_2$.

a) Het pre-kompakt zijn van (A_1,δ^g) is equivalent met, voor iedere $\varepsilon > 0$ bestaat er een eindige deelverzameling $\alpha(\varepsilon) \subset A_1$ die ε-dicht ligt in A_1. Neem $\varepsilon_1 > 0$ vast en noteer $\alpha(\varepsilon_1)$ als α_1. Vervangen we bij de definitie van de intrinsieke metriek op A_2 de verzameling A_1 door α_1, dan verkrijgen we een nieuwe metriek, die we noteren als $\delta^g_{\alpha_1}$. Daar α_1 een eindige verzameling is en g uniform begrensd volgt d.m.v. de definitie van $\delta^g_{\alpha_1}$ dat $(A_2,\delta^g_{\alpha_1})$ pre-kompakt is. Kies $\varepsilon_2 > 0$ en laat $\alpha_2(\varepsilon_2) \subset A_2$ de eindige deelverzameling die ε_2-dicht ligt in A_2 t.o.v. de metriek $\delta^g_{\alpha_1}$. Uit

$$|\delta^g(a_{21},a_{22}) - \delta^g_{\alpha_1}(a_{21},a_{22})| \le 2\varepsilon_1 \qquad \text{voor alle } (a_{21},a_{22}) \in A_2\times A_2,$$

volgt dat de verzameling $\alpha_2(\varepsilon_2)$ $(\varepsilon_2+2\varepsilon_1)$-dicht ligt in A_2 t.o.v. de metriek δ^g en daar ε_1 en ε_2 willekeurig positief zijn volgt bewering a.

b) Kies $\varepsilon > 0$. Laat $A_{11},\ldots,A_{1k}$ een meetbare partitie van A_1, zodanig dat de diameter van A_{1i} niet groter is dan ε. Laat $A_{21},\ldots,A_{2\ell}$ een meetbare partitie van A_2, zodanig dat de diameter van A_{2j} niet groter is dan ε. Kies $a_{1i} \in A_{1i}$ willekeurig, $i = 1,\ldots,k$ en kies $a_{2j} \in A_{2j}$ willekeurig, $j = 1,\ldots,\ell$ en laat $\alpha_1 = \{a_{11},\ldots,a_{1k}\}$ en $\alpha_2 = \{a_{21},\ldots,a_{2\ell}\}$. Associeer nu met een kansmaat $\pi_1 \in \Pi_1$ de kansmaat $\pi_1^{\alpha_1} \in \Pi_1$ gedefinieerd als $\pi_1^{\alpha_1}(a_{1i}) = \pi_1(A_{1i})$, $i = 1,\ldots,k$. Evenzo, associeer met $\pi_2 \in \Pi_2$ de kansmaat $\pi_2^{\alpha_2}$ gedefinieerd als $\pi_2^{\alpha_2}(a_{2j}) = \pi_2(A_{2j})$, $j = 1,\ldots,\ell$.

Dan geldt:

$$(3.2) \qquad |g(\pi_1,\pi_2) - g(\pi_1^{\alpha_1},\pi_2)| \le \varepsilon, \qquad \text{voor alle } (\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$$

$$(3.3) \qquad |g(\pi_1,\pi_2) - g(\pi_1,\pi_2^{\alpha_2})| \le \varepsilon, \qquad \text{voor alle } (\pi_1,\pi_2) \in \Pi_1 \times \Pi_2.$$

Uit (3.3) volgt:

$$(3.4) \qquad \sup_{\pi_1^{\alpha_1}} \inf_{\pi_2^{\alpha_2}} g(\pi_1^{\alpha_1},\pi_2^{\alpha_2}) \le \sup_{\pi_1^{\alpha_1}} \inf_{\pi_2} g(\pi_1^{\alpha_1},\pi_2) + \varepsilon$$

$$\le \sup_{\pi_1} \inf_{\pi_2} g(\pi_1,\pi_2) + \varepsilon.$$

Uit (3.2) volgt:

$$(3.5) \qquad \inf_{\pi_2^{\alpha_2}} \sup_{\pi_1^{\alpha_1}} g(\pi_1^{\alpha_1},\pi_2^{\alpha_2}) \ge \inf_{\pi_2^{\alpha_2}} \sup_{\pi_1} g(\pi_1,\pi_2^{\alpha_2}) - \varepsilon$$

$$\ge \inf_{\pi_2} \sup_{\pi_1} g(\pi_1,\pi_2) - \varepsilon.$$

Daar α_1 en α_2 eindige verzamelingen zijn, blijkt uit stelling 1.2.3 voor matrixspelen dat geldt

$$(3.6) \qquad \inf_{\pi_2^{\alpha_2}} \sup_{\pi_1^{\alpha_1}} g(\pi_1^{\alpha_1},\pi_2^{\alpha_2}) = \sup_{\pi_1^{\alpha_1}} \inf_{\pi_2^{\alpha_2}} g(\pi_1^{\alpha_1},\pi_2^{\alpha_2}).$$

Kombineren van (3.4), (3.5) en (3.6) geeft

$$(3.7) \qquad \sup_{\pi_1} \inf_{\pi_2} g(\pi_1,\pi_2) \ge \inf_{\pi_2} \sup_{\pi_1} g(\pi_1,\pi_2) - 2\varepsilon.$$

Daar $\varepsilon > 0$ willekeurig is volgt bewering b) nu uit (3.7) en lemma 1.1.19. □

Bovenstaand bewijs is grotendeels identiek aan de bewijzen van de stellingen 2.1 en 2.2 uit WALD [71].

LEMMA 3.2.5. *Als zowel* (A_1,δ^g) *als* (A_1,δ^f) *pre-kompakt zijn, dan is ook* (A_1,δ^{g+cf}) *pre-kompakt, voor alle* $c \in \mathbb{R}$.

BEWIJS. Uit de definitie van δ^{g+cf} volgt:

$$\delta^{g+cf}(a_{11},a_{12}) \leq \delta^{g}(a_{11},a_{12}) + |c|\delta^{f}(a_{11},a_{12}), \tag{3.8}$$

$$\text{alle } (a_{11},a_{12}) \in A_1 \times A_1.$$

Laat nu $\{a_{1n}\}$ een willekeurige rij in (A_1,δ^{g+cf}). Dan is er een deelrij $\{a'_{1n}\}$ die zowel Cauchy rij is in (A_1,δ^{f}) als Cauchy rij in (A_1,δ^{g}), maar dan volgt met (3.8):

$$\lim_{\substack{n\to\infty \\ m\to\infty}} \delta^{g+cf}(a'_{1n},a'_{1m}) \leq \lim_{\substack{n\to\infty \\ m\to\infty}} \delta^{g}(a'_{1n},a'_{1m}) + |c| \lim_{\substack{n\to\infty \\ m\to\infty}} \delta^{f}(a'_{1n},a'_{1m}) = 0,$$

zodat $\{a'_{1n}\}$ ook Cauchy rij is in (A_1,δ^{g+cf}). □

LEMMA 3.2.6. *Zijn* f *en* g *twee uniform begrensde functies op* $A_1 \times A_2$ *en zijn* (A_1,δ^{f}) *en* (A_1,δ^{g}) *pre-kompakt, dan is het spel* $(A_1,A_2,g+cf)$ *strikt bepaald, met* $c \in \mathbb{R}$.

BEWIJS. Kombineren van lemma 3.2.5 en stelling 3.2.4 geeft de bewering in het lemma. □

We gaan nu de dummyspelen definiëren bij het stochastisch spelmodel, zoals bestudeerd door Takahashi.

Laat N het aantal toestanden van het spel zijn. Laat g_k: $A_1(k) \times A_1(k) \to \mathbb{R}$ de uitbetalingsfunctie in toestand k zijn. Laat

$$p_0 = \inf_{k,a_1(k),a_2(k)} \left(1 - \sum_{\ell=1}^{N} p(\ell|k,a_1(k),a_2(k))\right). \tag{3.9}$$

Takahashi werkt onder de volgende aannames:

H1: De verzameling $(A_1(k),\delta^{g_k})$ is pre-kompakt, $k = 1,\ldots,N$.

H2: De verzameling $(A_1(k),\delta^{p(\ell|k,\cdot,\cdot)})$ is pre-kompakt, $\ell = 1,\ldots,N$; $k = 1,\ldots,N$.

De dummyspelen worden op dezelfde manier gedefinieerd als in hoofdstuk II bij het stochastische spel van Shapley (definitie 2.3.1). De dummyspelen bij de N-vektor $\bar{v} = (v_1,\ldots,v_N)$ zijn de N éénstapsspelen $G_k(\bar{v}) = (A_1(k),A_2(k),g_k(\cdot,\cdot) + \sum_{\ell=1}^{N} p(\ell|k,\cdot,\cdot)v_\ell)$, $k = 1,\ldots,N$.
Uit de aannames H1 en H2 en lemma 3.2.6 volgt dat de dummyspelen strikt bepaald zijn voor iedere N-vektor $\bar{v}$.

DEFINITIE 3.2.7. Een vektor $\bar{v}^* = (v_1^*,\dots,v_N^*)$ wordt *waarde-vektor* van de dummyspelen genoemd, als geldt:

$$v_k^* = \text{val } G_k(\bar{v}^*), \qquad k = 1,\dots,N.$$

STELLING 3.2.8. *De dummyspelen bezitten een unieke waarde-vektor.*

BEWIJS. Definieer de afbeelding T: $\mathbb{R}^N \to \mathbb{R}^N$ als volgt: $T\bar{v} = \bar{w}$, met $w_k = \text{val } G_k(\bar{v})$, $k = 1,\dots,N$.
Analoog als in het bewijs van stelling 2.3.2 gaan we aantonen dat de afbeelding T een uniek dekpunt bezit, waarmee de stelling bewezen is. Daartoe merken we op dat de lemma's 1.2.4 en 1.2.5 uitgebreid kunnen worden tot de klasse van éénstapsspelen met pre-kompakte aktieruimten. Als norm op de $\mathbb{R}^N$ nemen we $\|\bar{v}\| = \max_k |v_k|$. Voor alle $\bar{v},\bar{w} \in \mathbb{R}^N$ geldt

$$\begin{aligned}
\|T\bar{v} - T\bar{w}\| &= \max_k \left|\text{val } G_k(\bar{v}) - \text{val } G_k(\bar{w})\right| \\
&\le \sup_{k,a_1(k),a_2(k)} \Big| g_k(a_1(k),a_2(k)) + \sum_{\ell=1}^{N} p(\ell|k,a_1(k),a_2(k))v_\ell - \\
&\qquad - g_k(a_1(k),a_2(k)) - \sum_{\ell=1}^{N} p(\ell|k,a_1(k),a_2(k))w_\ell \Big| \\
&= \sup_{k,a_1(k),a_2(k)} \Big| \sum_{\ell=1}^{N} p(\ell|k,a_1(k),a_2(k))\cdot(v_\ell - w_\ell) \Big| \\
&\le (1-p_0)\|\bar{v}-\bar{w}\|.
\end{aligned}$$

T is dus een contractieafbeelding, zodat de Banach-Picard vaste puntenstelling de existentie van een uniek dekpunt levert. □

3.3. Het stochastisch spel van Takahashi

In eerste instantie beschouwen we alleen stationaire strategieën. Voor $\pi_1 = (\pi_1(1),\dots,\pi_1(N))$ met $\pi_1(k)$ een gemengde aktie in toestand k, $k = 1,\dots,N$, als stationaire strategie voor speler 1 en $\pi_2 = (\pi_2(1),\dots,\pi_2(N))$ als stationaire strategie voor speler 2 noteren we de verwachte uitbetaling op tijdstip t bij begintoestand k als $V^t(k,\pi_1,\pi_2)$. De totale verwachte uitbetaling bij begintoestand k onder π_1 en π_2 noteren we als $V(k,\pi_1,\pi_2)$. Laat $\bar{V}^t(\pi_1,\pi_2) = (V^t(1,\pi_1,\pi_2),\dots,V^t(N,\pi_1,\pi_2))$ en $\bar{V}(\pi_1,\pi_2) = (V(1,\pi_1,\pi_2),\dots,V(N,\pi_1,\pi_2))$.

Op grond van aanname H_2 kunnen we bij iedere π_1 en π_2 m.b.v. de lemma's 3.2.2 en 3.2.3 de overgangskansenmatrix $P(\pi_1,\pi_2)$ samenstellen waarvan het (k,ℓ)-de element de waarde

$$p(\ell|k,\pi_1(k),\pi_2(k)) = \int_{A_1(k)} \int_{A_2(k)} p(\ell|k,a_1,a_2)d\pi_1(k)(a_1)d\pi_2(k)(a_2)$$

heeft, $k = 1,\ldots,N$; $\ell = 1,\ldots,N$. We korten $P(\pi_1,\pi_2)$ wel eens af tot P, indien duidelijk is bij welke (π_1,π_2) P behoort.

$\bar{V}^t(\pi_1,\pi_2)$ kunnen we nu schrijven als:

$$\text{(3.10)} \qquad \bar{V}^t(\pi_1,\pi_2) = P^t\cdot\bar{g}(\pi_1,\pi_2), \qquad t = 0,1,\ldots ,$$

waarbij $\bar{g}(\pi_1,\pi_2) = (g_1(\pi_1(1),\pi_2(1)),\ldots,g_N(\pi_1(N),\pi_2(N)))$ en $P^0 = I$. Uit (3.9) en (3.10) volgt

$$\text{(3.11)} \qquad |\bar{V}^t(k,\pi_1,\pi_2)| \leq (1-p_0)^t\cdot M \quad \text{met } M = \sup_{k,a_1(k),a_2(k)} |g_k(a_1(k),a_2(k))|,$$

zodat

$$\text{(3.12)} \qquad |V(k,\pi_1,\pi_2)| \leq \lim_{t\to\infty} \sum_{\tau=0}^{t} |V^\tau(k,\pi_1,\pi_2)| \leq \frac{M}{p_0}, \; k = 1,\ldots,N.$$

M.b.v. (3.11) en (3.12) volgt nu

$$\begin{aligned}\bar{V}(\pi_1,\pi_2) &= \sum_{t=0}^{\infty} P^t\cdot\bar{g}(\pi_1,\pi_2) = P^0\bar{g}(\pi_1,\pi_2) + \sum_{t=1}^{\infty} P^t\cdot\bar{g}(\pi_1,\pi_2)\\ &= \bar{g}(\pi_1,\pi_2) + P\sum_{t=0}^{\infty} P^t\cdot\bar{g}(\pi_1,\pi_2)\\ &= \bar{g}(\pi_1,\pi_2) + P\cdot\bar{V}(\pi_1,\pi_2).\end{aligned}$$

Dus

$$\text{(3.13)} \qquad \bar{V}(\pi_1,\pi_2) = \bar{g}(\pi_1,\pi_2) + P\bar{V}(\pi_1,\pi_2), \text{ alle } \pi_1 \text{ en } \pi_2.$$

STELLING 3.3.1. *Een twee-persoons nulsom stoppend stochastisch spel onder de aannames H1 en H2 is strikt bepaald binnen de klasse der stationaire strategieën. De waarde-vektor is gelijk aan de waarde-vektor* $\bar{v}^*$ *van de dummyspelen. Voor* $\varepsilon \geq 0$ *komt een stelsel* ε*-optimale strategieën voor de dummyspelen* $G_k(\bar{v}^*)$, $k = 1,\ldots,N$, *voor speler* i *overeen met een* ε/p_0*-optimale strategie voor het stochastische spel voor speler* i, i = 1,2.

<u>BEWIJS</u>. Laat $\bar{v}^*$ de waarde-vektor van de dummyspelen zijn. Laat $\Pi_i(k)$ de verzameling van gemengde akties in toestand k voor speler i en $\Pi_i = \times_{k=1}^{N} \Pi_i(k)$ de verzameling van stationaire strategieën voor speler i, i = 1,2. Laat $h_k(\pi_1(k),\pi_2(k),\bar{v}^*) = g_k(\pi_1(k),\pi_2(k)) + \sum_{\ell=1}^{N} p(\ell|k,\pi_1(k),\pi_2(k))\cdot \cdot v_\ell^*$ en $\bar{h}(\pi_1,\pi_2,\bar{v}^*) = (h_1(\pi_1(1),\pi_2(1),\bar{v}^*),\ldots,h_N(\pi_1(N),\pi_2(N),\bar{v}^*))$, zodat

$$(3.14) \qquad \bar{h}(\pi_1,\pi_2,\bar{v}^*) = \bar{g}(\pi_1,\pi_2) + P(\pi_1,\pi_2)\cdot\bar{v}^*.$$

(3.14) aftrekken van (3.13) geeft

$$(3.15) \qquad \bar{V}(\pi_1,\pi_2) - \bar{v}^* = \bar{h}(\pi_1,\pi_2,\bar{v}^*) - \bar{v}^* + P(\pi_1,\pi_2)(\bar{V}(\pi_1,\pi_2) - \bar{v}^*).$$

Herhaald invullen van het rechterlid van (3.15) als vervanging voor $\bar{V}(\pi_1,\pi_2) - \bar{v}^*$ in het rechterlid geeft

$$(3.16) \qquad \bar{V}(\pi_1,\pi_2) - \bar{v}^* = \sum_{\tau=0}^{t} P^\tau(\pi_1,\pi_2)(\bar{h}(\pi_1,\pi_2,\bar{v}^*) - \bar{v}^*) + P^{t+1}(\pi_1,\pi_2)(\bar{V}(\pi_1,\pi_2) - \bar{v}^*).$$

Neem $\varepsilon > 0$ willekeurig. Laat $\pi_1^\varepsilon = (\pi_1^\varepsilon(1),\ldots,\pi_1^\varepsilon(N))$ zodanig dat $\pi_1^\varepsilon(k)$ ε-optimaal is voor speler 1 in het k-de dummyspel $G_k(\bar{v}^*)$, k = 1,...,N. Dan geldt:

$$(3.17) \qquad \bar{h}(\pi_1^\varepsilon,\pi_2,\bar{v}^*) \geq \bar{v}^* - \varepsilon\cdot\bar{1}, \qquad \text{alle } \pi_2 \in \Pi_2.$$

(3.17) invullen in (3.16), $t \to \infty$ laten gaan en gebruik maken van (3.9) geeft

$$(3.18) \qquad \bar{V}(\pi_1^\varepsilon,\pi_2) - \bar{v}^* \geq \sum_{\tau=0}^{\infty} (1-p_0)^\tau(-\varepsilon)\cdot\bar{1} = \frac{-\varepsilon}{p_0}\cdot\bar{1}, \qquad \text{alle } \pi_2 \in \Pi_2.$$

En daar $\varepsilon > 0$ willekeurig volgt uit (3.18)

$$(3.19) \qquad \sup_{\pi_1\in\Pi_1} \inf_{\pi_2\in\Pi_2} \bar{V}(\pi_1,\pi_2) \geq \bar{v}^*.$$

Analoog kunnen we laten zien dat voor de strategie $\pi_2^\varepsilon = (\pi_2^\varepsilon(1),\ldots,\pi_2^\varepsilon(N))$ met $\pi_2^\varepsilon(k)$ ε-optimaal in het k-de dummyspel geldt

$$(3.20) \qquad \bar{V}(\pi_1,\pi_2^\varepsilon) \leq \bar{v}^* + \frac{\varepsilon}{p_0}\cdot\bar{1}, \text{ alle } \pi_1 \in \Pi_1,$$

zodat

$$(3.21) \qquad \inf_{\pi_2\in\Pi_2}\ \sup_{\pi_1\in\Pi_1}\ \bar{V}(\pi_1,\pi_2) \le \bar{v}^*.$$

(3.19) en (3.21) gekombineerd met lemma 1.1.19 geeft nu het strikt bepaald zijn:

$$\inf_{\pi_2\in\Pi_2}\ \sup_{\pi_1\in\Pi_1}\ \bar{V}(\pi_1,\pi_2) = \sup_{\pi_1\in\Pi_1}\ \inf_{\pi_2\in\Pi_2}\ \bar{V}(\pi_1,\pi_2) = \bar{v}^*.$$

Uit $\bar{v}^*$ is waarde-vektor en het feit dat (3.18) en (3.20) ook gelden voor $\varepsilon = 0$ volgt direkt de tweede bewering in de stelling. □

<u>OPMERKING 3.3.2</u>. J. VON NEUMANN [41] heeft een twee-persoons nulsomspel eerlijk genoemd als de waarde nul is. Als de spelen $(A_1(k),A_2(k),g_k)$ eerlijk zijn, $k = 1,\dots,N$, dan is ook het stochastische spel eerlijk onafhankelijk van de overgangskansen (waarde-vektor van de dummyspelen is de nul-vektor).

De volgende twee lemma's, die we niet zullen bewijzen, zijn overgenomen uit WALD [71] (stelling 2.3, pag. 39 en stellingen 2.4 en 2.5, pag. 40) en dienen als voorbereiding op de stellingen 3.3.5 en 3.3.6.

<u>LEMMA 3.3.3</u>. *Voor een spel in normale vorm* (A_1,A_2,g), *waarbij* (A_1,δ^g) *prekompakt is, bestaan er voor iedere* $\varepsilon > 0$ *eindige verzamelingen* $\alpha_1 \subset A_1$ *en* $\alpha_2 \subset A_2$, *z.d.d. de waarde van* (A_1,A_2,g) *hoogstens* ε *verschilt van de waarde van ieder van de volgende drie spelen:*

$$(\alpha_1,A_2,g), \quad (A_1,\alpha_2,g) \quad \textit{en} \quad (\alpha_1,\alpha_2,g).$$

<u>LEMMA 3.3.4</u>. *Voor een spel in normale vorm* (A_1,A_2,g), *waarbij* A_2 *een eindige verzameling is, bestaat er voor iedere* $\varepsilon > 0$ *een verzameling* $\alpha_1 \subset A_1$, *z.d.d.* α_1 *hoogstens evenveel elementen bevat als* A_2 *en de waarde van* (α_1,A_2,g) *hoogstens* ε *verschilt van de waarde van* G. *Indien tevens* (A_1,δ^g) *kompakt is, kan* α_1 *zo gekozen worden, dat de waarden van de spelen* (A_1,A_2,g) *en* (α_1,A_2,g) *gelijk zijn.*

<u>STELLING 3.3.5</u>. *Voor een twee-persoons nulsom stoppend stochastisch spel* Γ *onder de aannames* H1 *en* H2 *bestaan er voor iedere* $\varepsilon > 0$ *eindige verzamelingen* $\alpha_i(k) \subset A_i(k)$, $i = 1,2$; $k = 1,\dots,N$, *z.d.d. de waarde van* Γ *hoogstens* ε *verschilt van de waarde van ieder van de volgende drie spelen:*

Γ_1: *het spel waarbij speler 1 in toestand* k $\alpha_1(k)$ *als zuivere aktieruimte heeft en speler 2* $A_2(k)$, $k = 1,2,\ldots,N$.

Γ_2: *het spel waarbij speler 1 in toestand* k $A_1(k)$ *als zuivere aktieruimte heeft en speler 2* $\alpha_2(k)$, $k = 1,2,\ldots,N$.

Γ_3: *het spel waarbij speler 1 in toestand* k $\alpha_1(k)$ *als zuivere aktieruimte heeft en speler 2* $\alpha_2(k)$, $k = 1,2,\ldots,N$.

<u>BEWIJS</u>. We bewijzen de stelling alleen voor Γ_1. Voor Γ_2 en Γ_3 verlopen de bewijzen op dezelfde manier.

Laat $\bar{v}^*$ de waarde-vektor van Γ zijn. Neem $\varepsilon > 0$. Op grond van lemma 3.3.3 kunnen we eindige verzamelingen $\alpha_1(k) \subset A_1(k)$, $k = 1,\ldots,N$ kiezen z.d.d. de waarde van het spel $\tilde{G}_k(\bar{v}^*) = (\alpha_1(k), A_2(k), h_k(\cdot,\cdot,\bar{v}^*))$ hoogstens $\frac{1}{2}\varepsilon\, p_0$ verschilt van de waarde van het k-de dummyspel $G_k(\bar{v}^*) = (A_1(k), A_2(k), h_k(\cdot,\cdot,\bar{v}^*))$.

Noteer een stationaire strategie voor speler 1 die in toestand k uitsluitend gewicht legt op $\alpha_1(k)$, $k = 1,\ldots,N$ als $\pi_1^{\alpha_1}$ en laat $\tilde{\pi}_1^{\alpha_1} = (\tilde{\pi}_1^{\alpha_1}(1),\ldots,\tilde{\pi}_1^{\alpha_1}(N))$ zodanig zijn, dat $\tilde{\pi}_1^{\alpha_1}(k)$ een $\frac{1}{2}\varepsilon\, p_0$-optimale strategie voor speler 1 in $\tilde{G}_k(\bar{v}^*)$ is. Dan geldt:

$$h_k(\tilde{\pi}_1^{\alpha_1}(k), \pi_2(k), \bar{v}^*) \geq \mathrm{val}\ \tilde{G}_k(\bar{v}^*) - \tfrac{1}{2}\varepsilon\, p_0 \geq \mathrm{val}\ G_k(\bar{v}^*) - \varepsilon\, p_0$$

$$= \bar{v}^* - \varepsilon\, p_0, \quad \text{alle } \pi_2(k) \in \Pi_2(k);\ k = 1,\ldots,N.$$

M.a.w. $\tilde{\pi}_1^{\alpha_1}$ komt overeen met een stelsel $\varepsilon\, p_0$-optimale strategieën voor de dummyspelen, zodat volgens stelling 3.3.1 $\tilde{\pi}_1^{\alpha_1}$ een ε-optimale strategie voor speler 1 in het oorspronkelijke stochastische spel is. Maar dan geldt

$$\sup_{\pi_1} \inf_{\pi_2} \bar{V}(\pi_1,\pi_2) \geq \sup_{\pi_1^{\alpha_1}} \inf_{\pi_2} \bar{V}(\pi_1^{\alpha_1},\pi_2) \geq \inf_{\pi_2} \bar{V}(\tilde{\pi}_1^{\alpha_1},\pi_2) \tag{3.22}$$

$$\geq \bar{v}^* - \varepsilon\cdot\bar{1} = \sup_{\pi_1} \inf_{\pi_2} \bar{V}(\pi_1,\pi_2) - \varepsilon\cdot\bar{1}.$$

Uit (3.22) volgt nu de bewering in de stelling:

$$\left| \sup_{\pi_1} \inf_{\pi_2} \bar{V}(\pi_1,\pi_2) - \sup_{\pi_1^{\alpha_1}} \inf_{\pi_2} \bar{V}(\pi_1^{\alpha_1},\pi_2) \right| \leq \varepsilon\cdot\bar{1}. \qquad \square$$

STELLING 3.3.6. *Voor een twee-persoons nulsom stoppend stochastisch spel onder de aannames H1 en H2 en zodanig dat* $A_2(k)$, $k = 1,\ldots,N$ *een eindige verzameling is, bestaan er deelverzamelingen* $\alpha_1(k) \subset A_1(k)$, $k = 1,\ldots,N$, *zodanig dat* $\alpha_1(k)$ *hoogstens evenveel elementen bevat als* $A_2(k)$, *terwijl de waarde van het spel hoogstens* ε *verandert als speler 1 zich tot de verzamelingen* $\alpha_1(k)$, $k = 1,\ldots,N$ *beperkt.*

BEWIJS. Het bewijs verloopt op dezelfde manier als het bewijs van stelling 3.3.5, met dien verstande dat nu lemma 3.3.4 gebruikt moet worden. □

Tot nu toe werden de topologische condities opgelegd op de verzamelingen $(A_1(k),\delta^{g_k})$, enz. In het volgende starten we met kompakte topologische ruimtes, waarop de functies $g_k(\cdot,\cdot)$ en $p(\ell|k,\cdot,\cdot)$ continu zijn.

Het volgende lemma, waarvan het bewijs gevonden kan worden in VRIEZE [66], dient als schakel om deze spelen met behulp van de methode van Takahashi op te lossen.

LEMMA 3.3.7. *Laat* A_1 *en* A_2 *kompakte topologische ruimtes zijn, die voldoen aan het eerste axioma van aftelbaarheid (KELLY [29], pag.50) en laat* $f\colon A_1 \times A_2 \to \mathbb{R}$ *continu zijn in de produkttopologie. Dan zijn de metrische verzamelingen* (A_1,δ^f) *en* (A_2,δ^f) *kompakt (eventueel na reductie) en is de functie* f *continu t.o.v. deze metrieken.*

STELLING 3.3.8. *Een twee-persoons nulsom stoppend stochastisch spel, met zuivere aktieruimtes* $A_1(k)$ *en* $A_2(k)$, $k = 1,\ldots,N$, *die kompakt topologisch zijn en voldoen aan het eerste axioma van aftelbaarheid en waarbij* $g_k(\cdot,\cdot)$ *en* $p(\ell|k,\cdot,\cdot)$ *continue functies zijn op* $A_1(k)\times A_2(k)$ *in de produkttopologie,* $k = 1,\ldots,N$; $\ell = 1,\ldots,N$, *is strikt bepaald binnen de klasse der stationaire strategieën. Tevens bezitten beide spelers optimale stationaire strategieën.*

BEWIJS. Uit lemma 3.3.7 blijkt dat de ruimtes $(A_i(k),\delta^{g_k})$ en $(A_i(k),\delta^{p(\ell|k,\cdot,\cdot)})$ kompakt zijn, $i = 1,2$; $k = 1,\ldots,N$; $\ell = 1,\ldots,N$. Tevens is $g_k(\cdot,\cdot)$ als continue functie op een kompakte verzameling uniform begrensd, zodat aan de voorwaarden in stelling 3.3.1 voldaan is, waaruit blijkt dat het spel strikt bepaald is.

Eveneens uit stelling 3.3.1 volgt dat voor iedere $\varepsilon > 0$ de beide spelers ε-optimale stationaire strategieën bezitten. Laat $\bar{v}^*$ de waarde-vektor zijn en laat voor $k = 1,\ldots,N$:

$$h_k(\pi_1(k),\pi_2(k),\bar{v}^*) = g_k(\pi_1(k),\pi_2(k)) + \sum_{\ell=1}^{N} p(\ell|k,\pi_1(k),\pi_2(k))\cdot v_\ell^*,$$

$$\text{alle } (\pi_1(k),\pi_2(k)) \in \Pi_1(k)\times\Pi_2(k).$$

Merk op dat de restrictie $\tilde{h}_k(\cdot,\cdot,\bar{v}^*)$ van $h_k(\cdot,\cdot,\bar{v}^*)$ tot $A_1(k)\times A_2(k)$ (de gedegenereerde kansmaten) continu is, zodat volgens lemma 3.3.7

$$(A_1(k),\delta^{\tilde{h}_k(\cdot,\cdot,\bar{v}^*)})$$

kompakt is. Hieruit volgt dat $\Pi_1(k)$ kompakt is in de zwakke topologie. (PARTHASARATHY [44], pag. 39), zodat eveneens $\Pi_1 = X_{k=1}^N \Pi_1(k)$ kompakt is in de produkttopologie.

Laat $<\varepsilon_n>$ een net van getallen z.d.d. $\varepsilon_n > 0$ en $\lim \varepsilon_n = 0$. Laat $\pi_1^{\varepsilon_n}$ een strategie voor speler 1 zijn, z.d.d. $h_k(\pi_1^{\varepsilon_n}(k),\pi_2(k),\bar{v}^*) \geq v_k^* - \varepsilon_n$, alle $\pi_2(k) \in \Pi_2(k)$. Dan bezit het net $<\pi_1^{\varepsilon_n}>$ een deelnet $<\pi_1^{\varepsilon_{n'}}>$, dat naar een element van Π_1 convergeert, zeg π_1^*.

Voor vaste $a_2(k)$ is $\tilde{h}_k(\cdot,a_2(k),\bar{v}^*)$ continu op $(A_1(k),\delta^{h_k(\cdot,\cdot,\bar{v}^*)})$, zodat $\lim h_k(\pi_1^{\varepsilon_{n'}}(k),a_2(k),\bar{v}^*) = h_k(\pi_1^*(k),a_2(k),\bar{v}^*)$ en daar $h_k(\pi_1^{\varepsilon_{n'}}(k),a_2(k),\bar{v}^*) \geq$ $\geq v_k^* - \varepsilon_{n'}$, alle $a_2(k) \in A_2(k)$ volgt hieruit

$$h_k(\pi_1^*(k),a_2(k),\bar{v}^*) \geq v_k^*, \qquad \text{alle } a_2(k) \in A_2(k)$$

en dit is equivalent met

$$h_k(\pi_1^*(k),\pi_2(k),\bar{v}^*) \geq v_k^*, \qquad \text{alle } \pi_2(k) \in \Pi_2(k).$$

π_1^* komt dus overeen met een stelsel optimale strategieën voor de dummyspelen, zodat π_1^* volgens stelling 3.3.1 optimaal is in het stochastische spel voor speler 1. □

<u>STELLING 3.3.9</u>. *Voor een twee-persoons nulsom stoppend stochastisch spel met* $A_2(k)$ *een eindige verzameling en* $A_1(k)$ *een kompakte topologische verzameling, die aan het eerste axioma van aftelbaarheid voldoet, terwijl* $g_k(\cdot,\cdot)$ *en* $p(\ell|k,\cdot,\cdot)$ *continue functies zijn t.o.v. de produkttopologie,* $k = 1,\ldots,N$; $\ell = 1,\ldots,N$, *bestaan er eindige verzamelingen* $\alpha_1(k) \subset A_1(k)$, *z.d.d.* $\alpha_1(k)$ *hoogstens evenveel elementen bevat als* $A_2(k)$, $k = 1,\ldots,N$ *en zodanig dat de waarde van het spel gelijk blijft, als speler 1 zich tot* $\alpha_1(k)$ *beperkt.*

<u>BEWIJS</u>. Uit stelling 3.3.8 blijkt dat het spel strikt bepaald is. Laat $\bar{v}^*$

de waarde-vektor zijn en beschouw de dummyspelen $(A_1(k), A_2(k), h_k(\cdot,\cdot,\bar{v}^*))$, $k = 1,\ldots,N$. Uit lemma 3.3.4 volgt dat er eindige verzamelingen $\alpha_1(k) \subset A_1(k)$ bestaan, z.d.d. $\alpha_1(k)$ hoogstens evenveel elementen bevat als $A_2(k)$ en zodanig dat de waarde van $(\alpha_1(k), A_2(k), h_k(\cdot,\cdot,\bar{v}^*))$ v_k^* blijft. Maar dit betekent dat de waarde-vektor van de dummyspelen bij het stochastische spel, waarin $A_1(k)$ door $\alpha_1(k)$ vervangen wordt, $\bar{v}^*$ is, zodat de waarde-vektor van het spel zelf ook $\bar{v}^*$ is (stelling 3.3.1). □

Tot nu toe hebben we ons beperkt tot de klasse der stationaire strategieën. In de laatste sektie van zijn artikel laat Takahashi zien, dat het stochastische spel, zoals vastgelegd in stelling 3.3.1, ook strikt bepaald is binnen de klasse der Markov-strategieën en dat de waarde dezelfde blijft. We zullen in het kort schetsen op welke manier Takahashi dit aantoont.

Bij een spel Γ, dat aan de condities van stelling 3.3.1 voldoet, wordt een nieuw spel $\tilde{\Gamma}$ gedefinieerd als volgt:

$\tilde{S} = \{1,2,3,\ldots\}$

$\tilde{A}_i(tN+k) = A_i(k)$, $k = 1,\ldots,N$; $t = 0,1,\ldots$ (N is aantal toestanden van het oorspronkelijke spel)

$\tilde{g}_{tN+k}(\cdot,\cdot) = g_k(\cdot,\cdot)$, $k = 1,\ldots,N$; $t = 0,1,\ldots$

$\tilde{p}((t+1)N+\ell \mid tN+k,\cdot,\cdot) = p(\ell \mid k,\cdot,\cdot)$, $\ell = 1,\ldots,N$; $k = 1,\ldots,N$; $t = 0,1,\ldots$

Merk op dat, als het spel $\tilde{\Gamma}$ zich in een van de toestanden $tN+k$, $k = 1,\ldots,N$, bevindt, dat het spel dan met kans 1 maar één van de toestanden $t(N+1)+\ell$, $\ell = 1,\ldots,N$ springt.

Takahashi laat zien dat er een één-één afbeelding is tussen de verzameling van de stationaire strategieën voor het spel $\tilde{\Gamma}$ en de verzameling van Markov-strategieën voor het spel Γ en tevens dat de bijbehorende verwachte opbrengsten dezelfde zijn.

Alhoewel het aantal toestanden van $\tilde{\Gamma}$ niet eindig is, laat Takahashi vervolgens zien dat zijn methode ook op het spel $\tilde{\Gamma}$ van toepassing is. Hij komt tot een oneindig dimensionale waardevektor $\tilde{\bar{v}}^*$, waarbij $(\tilde{v}^*(tN+1),\ldots, \tilde{v}^*(tN+N)) = \bar{v}^*$, $t = 0,1,\ldots$, met $\bar{v}^*$ waardevektor van Γ. Terug vertalen van dit resultaat naar het spel Γ geeft het gestelde.

In de volgende sektie laten we zien dat de resultaten van Takahashi ook geldig zijn binnen de klasse der gedragsstrategieën.

3.4. Uitbreiding naar historie-afhankelijke strategieën

Voor de begrippen historie van het spel en gedragsstrategie verwijzen we naar de definities 1.1.9 en 1.1.11.

Laat $SA_1A_2 = \{(k,a_1,a_2) \mid (a_1,a_2) \in A_1(k) \times A_2(k),\ k \in \{1,\ldots,N\}\}$. $\times_{\tau=1}^{t} SA_1A_2$ is dan de verzameling van alle mogelijke histories van het spel op tijdstip t, $t = 0,1,2,\ldots$, waarbij $\times_{\tau=1}^{0} SA_1A_2 \stackrel{\text{def}}{=} \phi$. Laat

$$\Pi_i^H = \{\pi_i \mid \pi_i \text{ is gedragsstrategie en voor alle t en k is } \pi_i(t,k,\cdot) \text{ een Borelmeetbare afbeelding van } \times_{\tau=1}^{t} SA_1A_2 \text{ naar } \Pi_i(k)\}.$$

We beperken ons hier tot de deelklassen van gedragsstrategieën Π_1^H en Π_2^H. In VRIEZE [67] is een methode ontwikkeld waarbij alle gedragsstrategieën beschouwd worden.

Op elementaire wijze kan nagegaan worden dat voor ieder paar $(\pi_1,\pi_2) \in \Pi_1^H \times \Pi_2^H$ de totale verwachte verdisconteerde opbrengst bestaat. Merk op, dat zowel alle stationaire strategieën als alle Markov-strategieën voor speler i tot Π_i^H behoren.

Laat $\tilde{\pi}_2 \in \Pi_2^H$ een stationaire strategie zijn. Laat voor $\pi_1 \in \Pi_1^H$ en historie $h_1 \in SA_1A_2$ de strategie $\pi_{1h_1} \in \Pi_1^H$ gedefinieerd zijn als

$$\pi_{1h_1}(t,k,h_t) = \pi_1(t+1,k,h'_{t+1}),$$

waarbij h'_{t+1} de historie (h_1,h_t) is, $t = 0,1,\ldots$, $k = 1,\ldots,N$, $h_t \in \times_{\tau=1}^{t} SA_1A_2$.

Op soortgelijke wijze als (3.13) afgeleid is kunnen we nu laten zien dat voor $\tilde{\pi}_2$ en iedere $\pi_1 \in \Pi_1^H$ geldt:

$$(3.23) \qquad V(k,\pi_1,\tilde{\pi}_2) = \int_{A_1(k)} \left\{ \int_{A_2(k)} g_k(a_1,a_2) + \sum_{\ell=1}^{N} p(\ell|k,a_1,a_2) \cdot V(\ell,\pi_{1ka_1a_2},\tilde{\pi}_2)\, d\tilde{\pi}_2(k)(a_2) \right\} d\pi_1(0,k,h_0)(a_1),$$

waarbij h_0 de lege string is. Laat

$$(3.24) \qquad W_k(\tilde{\pi}_2) = \sup_{\pi_1 \in \Pi_1^H} V(k,\pi_1,\tilde{\pi}_2).$$

LEMMA 3.4.1. *De vektor* $\bar{W}(\tilde{\pi}_2) = (W_1(\tilde{\pi}_2),\dots,W_N(\tilde{\pi}_2))$ *voldoet aan*

$$W_k(\tilde{\pi}_2) = \sup_{a_1\in A_1(k)} \left\{g_k(a_1,\tilde{\pi}_2) + \sum_{\ell=1}^{N} p(\ell|k,a_1,\tilde{\pi}_2)W_\ell(\tilde{\pi}_2)\right\},$$

$$k = 1,\dots,N.$$

BEWIJS. Neem $\varepsilon > 0$ willekeurig, en laat π_1^ε z.d.d. $V(k,\pi_1^\varepsilon,\tilde{\pi}_2) \ge W_k(\tilde{\pi}_2) - \varepsilon$, $k = 1,\dots,N$, zodat gebruik makend van (3.23) en (3.24) we krijgen

$$W_k(\tilde{\pi}_2)-\varepsilon \le V(k,\pi_1^\varepsilon,\tilde{\pi}_2)$$

$$= \int_{A_1(k)} \left\{ \int_{A_2(k)} \left\{g_k(a_1,a_2)+\sum_{\ell=1}^{N} p(\ell|k,a_1,a_2)V(\ell,\pi^\varepsilon_{1ka_1a_2},\tilde{\pi}_2)\right\}d\tilde{\pi}_2(k)(a_2)\right\} \cdot$$
$$\cdot\, d\pi_1^\varepsilon(0,k,h_0)(a_1)$$

$$\le \int_{A_1(k)} \left\{ \int_{A_2(k)} \left\{g_k(a_1,a_2)+\sum_{\ell=1}^{N} p(\ell|k,a_1,a_2)W_\ell(\tilde{\pi}_2)\right\}d\tilde{\pi}_2(k)(a_2)\right\}d\pi_1^\varepsilon(0,k,h_0)(a_1)$$

$$= \int_{A_1(k)} \left\{g_k(a_1,\tilde{\pi}_2)+\sum_{\ell=1}^{N} p(\ell|k,a_1,\tilde{\pi}_2)W_\ell(\tilde{\pi}_2)\right\}d\pi_1^\varepsilon(0,k,h_0)(a_1)$$

$$\le \sup_{a_1\in A_1(k)} \left\{g_k(a_1,\tilde{\pi}_2) + \sum_{\ell=1}^{N} p(\ell|k,a_1,\tilde{\pi}_2)W_\ell(\tilde{\pi}_2)\right\}.$$

Daar $\varepsilon > 0$ willekeurig is volgt hieruit

$$(3.25)\qquad W_k(\tilde{\pi}_2) \le \sup_{a_1\in A_1(k)} \left\{g_k(a_1,\tilde{\pi}_2) + \sum_{\ell=1}^{N} p(\ell|k,a_1,\tilde{\pi}_2)W_\ell(\tilde{\pi}_2)\right\}.$$

Neem nu π_1^ε, zodanig dat voor alle $h_1 \in SA_1A_2$ en $k \in \{1,\dots,N\}$ geldt

$$(3.26)\qquad V(k,\pi^\varepsilon_{1h_1},\tilde{\pi}_2) \ge W_k(\tilde{\pi}_2) - \varepsilon.$$

Gebruik makend van (3.26), (3.24) en (3.23) krijgen we:

$$(3.27)\qquad W_k(\tilde{\pi}_2) \ge V(k,\pi_1^\varepsilon,\tilde{\pi}_2) =$$

$$= \int\limits_{A_1(k)} \left\{ \int\limits_{A_2(k)} \left\{ g_k(a_1,a_2) + \sum_{\ell=1}^{N} p(\ell|k,a_1,a_2) V(\ell,\pi^{\varepsilon}_{1ka_1a_2},\tilde{\pi}_2) \right\} d\tilde{\pi}_2(k)(a_2) \right\} \cdot$$

$$\cdot\ d\pi_1(0,k,h_0)(a_1)$$

$$\leq \int\limits_{A_1(k)} \left\{ \int\limits_{A_2(k)} \left\{ g_k(a_1,a_2) + \sum_{\ell=1}^{N} p(\ell|k,a_1,a_2)(W_\ell(\tilde{\pi}_2)-\varepsilon) \right\} d\tilde{\pi}_2(k)(a_2) \right\} d\pi_1(0,k,h_0)(a_1)$$

$$\geq \int\limits_{A_1(k)} \left\{ g_k(a_1,\tilde{\pi}_2) + \sum_{\ell=1}^{N} p(\ell|k,a_1,\tilde{\pi}_2)(W_\ell(\tilde{\pi}_2)-\varepsilon) \right\} d\pi_1(0,k,h_0)(a_1).$$

(3.27) is geldig voor iedere kansmaat $\pi_1(k,h_0) \in \Pi_1(k)$. Beschouw de gedegenereerde kansmaten, dan gaat (3.27) over in

$$W_k(\tilde{\pi}_2) \geq g_k(a_1,\tilde{\pi}_2) + \sum_{\ell=1}^{N} p(\ell|k,a_1,\tilde{\pi}_2) W_\ell(\tilde{\pi}_2) - \varepsilon(1-p_0),$$

$$\text{alle } a_1 \in A_1(k)$$

en daar $\varepsilon > 0$ willekeurig is volgt hieruit

$$(3.28) \qquad W_k(\tilde{\pi}_2) \geq \sup_{a_1 \in A_1(k)} \left\{ g_k(a_1,\tilde{\pi}_2) + \sum_{\ell=1}^{N} p(\ell|k,a_1,\tilde{\pi}_2) W_\ell(\tilde{\pi}_2) \right\}.$$

(3.27) en (3.28) kombineren geeft nu de bewering in het lemma. □

<u>STELLING 3.4.2.</u> *Voor iedere stationaire strategie $\tilde{\pi}_2$ voor speler 2 geldt*

$$\sup_{\pi_1 \in \Pi^H_1} V(k,\pi_1,\tilde{\pi}_2) = \sup_{\bar{a}_1 \in X^N_{\ell=1} A_1(\ell)} V(k,\bar{a}_1,\tilde{\pi}_2), \qquad k = 1,\ldots,N,$$

waarbij $\bar{a}_1$ opgevat dient te worden als de zuivere stationaire strategie voor speler 1, die in toestand ℓ $a_1(\ell)$ voorschrijft.

<u>BEWIJS.</u> Neem een stationaire strategie $\tilde{\pi}_2$ voor speler 2 vast en beschouw voor speler 1 het volgende Markov-beslissingsprobleem:

> N toestanden, genummerd 1 t/m N; in toestand k is de zuivere aktieruimte gelijk aan $A_1(k)$; als in toestand k aktie $a_1(k) \in A_1(k)$ gekozen wordt, dan is de direkte opbrengst $g_k(a_1(k),\tilde{\pi}_2(k))$ en zijn de overgangskansen $p(\ell|k,a_1(k),\tilde{\pi}_2(k))$.

De stationaire strategieën voor speler 1 in bovenstaand Markov-beslissingsprobleem corresponderen op één-één wijze met de stationaire strategieën voor speler 1 in het oorspronkelijke stochastische spel.
Laat $\tilde{V}_k(\pi_1)$ de totale verwachte opbrengst voor speler 1 in het Markov-

beslissingsprobleem als bij stationaire strategie π_1 speelt en k de starttoestand is. Door uit te schrijven kan nagegaan worden dat

$$(3.29) \qquad \tilde{V}_k(\pi_1) = V(k,\pi_1,\tilde{\pi}_2).$$

Lemma 3.4.1 zegt dat $\bar{W}(\tilde{\pi}_2)$ voldoet aan de optimaliteitsvergelijking voor bovenstaand Markov-beslissingsprobleem. Uit DENARDO [10] (stelling 3; zie ook zijn voorbeeld 3) volgt dan, dat $\bar{W}(\tilde{\pi}_2)$ de optimale opbrengst is voor bovenstaand Markov-beslissingsprobleem en tevens dat $\bar{W}(\tilde{\pi}_2)$ gevonden kan worden door het supremum te nemen over de zuivere stationaire strategieën, zodat we met (3.29) krijgen:

$$W_k(\tilde{\pi}_2) = \sup_{\bar{a}_1 \in \times_{\ell=1}^N A_1(\ell)} \tilde{V}_k(\bar{a}_1) = \sup_{\bar{a}_1 \in \times_{\ell=1}^N A_1(\ell)} V(k,\bar{a}_1,\tilde{\pi}_2), \qquad k = 1,\dots,N. \quad \square$$

Uit stelling 3.4.2 nu blijkt dat tegen een stationaire strategie van speler 2, speler 1 niet beter heeft dan zijn klasse van stationaire strategieën, zodat een ε-optimale stationaire strategie voor speler 2 binnen de klassen der stationaire strategieën ook ε-optimaal is binnen de klassen Π_1^H en Π_2^H.

HOOFDSTUK IV

RECURSIEVE SPELEN

4.1. Eindige recursieve spelen

De eindige recursieve spelen worden gegeven door de definities 1.1.3 en 1.1.7. De verzameling toestanden kunnen wij splitsen in S_1, een verzameling toestanden waarin het spel stopt (met kans 1) en $S \backslash S_1$, een verzameling toestanden waarin geen betaling plaatsvindt en waarin het spel niet stopt.

VOORBEELD 4.1.1.

$$\Gamma_1 : \begin{bmatrix} \frac{1}{2}\Gamma_1 + \frac{1}{2}\Gamma_2 & \frac{1}{4}\Gamma_1 + \frac{3}{4}\Gamma_3 \\ \frac{3}{4}\Gamma_1 + \frac{1}{4}\Gamma_3 & \frac{1}{2}\Gamma_1 + \frac{1}{2}\Gamma_2 \end{bmatrix}$$

$$\Gamma_2 : [1] \qquad \Gamma_3 : [0]$$

In dit voorbeeld geldt $S_1 = \{2,3\}$. Ten behoeve van een kompaktere notatie kunnen wij de toestanden S_1 weglaten. Overgang naar de toestanden S_1 wordt vervangen door stoppen van het spel. De uitbetaling die in de toestanden S_1 plaatsvindt wordt één periode eerder gedaan onder de voorwaarde dat het spel stopt. Na deze veranderingen kunnen we voorbeeld 4.1.1. noteren als:

$$\Gamma_1 : \begin{bmatrix} \frac{1}{2} + \frac{1}{2}\Gamma_1 & \frac{1}{4}\Gamma_1 \\ \frac{3}{4}\Gamma_1 & \frac{1}{2} + \frac{1}{2}\Gamma_1 \end{bmatrix}$$

Bij het stochastische spel van Shapley noteerden wij als in toestand k akties i en j genomen worden de gevolgen als

$$g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)\Gamma_\ell$$

Nu noteren wij:

$$s(k,i,j)\ g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)\Gamma_\ell.$$

waarbij $\sum_{\ell=1}^{N} p(\ell|k,i,j) + s(k,i,j) = 1$, $s(k,i,j)$ de kans dat het spel stopt en $g(k,i,j)$ is de verwachte betaling van speler 2 aan speler 1 die alleen plaatsvindt als het spel stopt. Doordat er nu paren van strategieën kunnen zijn waarbij het spel een positieve kans heeft niet te stoppen moeten wij aan een oneindig durende spelrealisering een uitbetaling toekennen en kiezen daarvoor de uitbetaling nul.

VOORBEELD 4.1.2.

$$\Gamma\colon [\Gamma \quad 1]$$

Dit spel met één toestand heeft waarde nul. De unieke optimale strategie voor speler 2 is het altijd kiezen van zijn eerste aktie met kans 1 waardoor het spel niet stopt.

VOORBEELD 4.1.3. Dit voorbeeld wordt in eerste instantie gegeven door een "game of survival". De "games of survival" zijn generalisaties van "gambler's ruin" problemen en worden behandeld door MILNOR & SHAPLEY [38]. Dit voorbeeld en het volgende voorbeeld zijn afkomstig van HAUSNER [20]. De spelers spelen het nulsomspel:

$$\begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}$$

en herhalen dit totdat één der spelers geen geld meer heeft. Beide spelers starten met 1 gulden. Het doel van speler 1 is speler 2 blut te spelen. Het doel van speler 2 is te overleven. Dit spel kan worden geschreven als het volgende recursieve spel met 1 toestand:

$$\Gamma\colon \begin{bmatrix} \Gamma & 1 \\ 1 & 0 \end{bmatrix}.$$

De betaling aan speler 1 is 0 als hij verliest of bij niet stoppend spel. De betaling aan speler 1 is 1 als hij wint.

VOORBEELD 4.1.4. Dit voorbeeld is ook gegeven als een "game of survival". Het matrixspel is nu:

$$\begin{bmatrix} 2 & -1 \\ -2 & 1 \end{bmatrix}.$$

Het totale kapitaal is 4 gulden. Er zijn 3 mogelijke toestanden n.l. speler 1 heeft 1,2 of 3 gulden in zijn bezit. De doeleinden van de spelers zijn als in voorbeeld 4.1.3. Dit spel is equivalent met een recursief spel met 3 toestanden:

$$\Gamma_1: \begin{bmatrix} \Gamma_3 & 0 \\ 0 & \Gamma_2 \end{bmatrix} \quad \Gamma_2: \begin{bmatrix} 1 & \Gamma_1 \\ 0 & \Gamma_3 \end{bmatrix} \quad \Gamma_3: \begin{bmatrix} 1 & \Gamma_2 \\ \Gamma_1 & 1 \end{bmatrix}.$$

In toestand i bezit speler 1 i gulden. De betaling in dit recursieve spel is hetzelfde als in voorbeeld 4.1.3.

VOORBEELD 4.1.5. Dit voorbeeld is afgeleid van de "Blotto games" waarbij militair potentiëel moet worden toegewezen aan diverse mogelijke slagvelden. Kolonel Blotto voert het bevel over een kamp met 3 éénheden. Zijn opdracht is een vijandelijk kamp met 2 éénheden te elimineren. Gedekt door de nachtelijke duisternis kan Blotto met 0,1,2 of 3 éénheden naar het vijandelijke kamp trekken. De vijand kan in dezelfde nacht met 0,1 of 2 éénheden aanvallen. De achterblijvende éénheden hebben een verdedigende functie. Wanneer aanvallende éénheden ten opzichte van verdedigende éénheden in de meerderheid zijn valt het kamp. Als geen van beide kampen valt dan biedt de volgende nacht dezelfde mogelijkheden. De betaling aan Blotto is 1 als hij het vijandelijke kamp verovert zonder zijn eigen kamp te verliezen en -1 als hij (onder welke omstandigheden dan ook) zijn eigen kamp verliest. Dit is een recursief spel met één toestand:

$$\Gamma: \begin{bmatrix} \Gamma & \Gamma & \Gamma \\ \Gamma & \Gamma & 1 \\ \Gamma & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}.$$

De theorie van de recursieve spelen waarmee wij deze voorbeelden kunnen oplossen is behandeld door EVERETT [13]. In een artikel van ORKIN [43] worden eindige recursieve spelen behandeld, hierin staan echter geen resultaten die niet reeds in EVERETT voorkomen.

4.2. Notaties en definities

De recursieve spelen in het artikel van EVERETT hebben een eindige toestandsruimte, de aktieruimten voor beide spelers zijn willekeurige verzamelingen. Alle toegestane randomisaties worden als elementen van de aktieruimte beschouwd. De aktieruimte voor speler 1 in toestand k noteren wij als $\Pi_1(k)$ voor speler 2 als $\Pi_2(k)$. Als $\pi_1(k) \in \Pi_1(k)$ en $\pi_2(k) \in \Pi_2(k)$ als aktie worden gekozen noteren wij de gevolgen als:

$$s(k,\pi_1(k),\pi_2(k)) \cdot g(k,\pi_1(k),\pi_2(k)) + \sum_{\ell=1}^{N} p(\ell \mid (k,\pi_1(k),\pi_2(k))\Gamma_\ell .$$

Of als duidelijk is welke akties $\pi_1(k)$ en $\pi_2(k)$ zijn bedoeld ook wel kortweg:

$$s_k\, g_k + \sum_{\ell=1}^{N} p_{k\ell}\Gamma_\ell .$$

Wij zullen uitsluitend Markov-strategieën (definitie 1.1.13) beschouwen en deze van nu af aan strategieën noemen. Een strategie π_1 voor speler 1 is een rij $\pi_1^0, \pi_1^1, \pi_1^2, \ldots$, waarbij $\pi_1^t = (\pi_1^t(1),\ldots,\pi_1^t(N))$ met $\pi_1^t(k) \in \Pi_1(k)$. Π_1 is de verzameling van strategieën voor speler 1. Evenzo voor speler 2. Gegeven een paar strategieën π_1 en π_2 gedraagt het systeem zich als een niet homogene Markovketen. Wij noteren P^t voor de matrix met als (k,ℓ)'de element $p_{k\ell}^t = p(\ell \mid k,\pi_1^t(k),\pi_2^t(k))$, S^t voor de diagonaal matrix met als (k,k)'de element $s_k^t = s(k,\pi_1^t(k),\pi_2^t(k)$ en $\bar{g}^t = (g_1^t,\ldots,g_N^t)^T$ met $g_k^t = g(k,\pi_1^t(k),\pi_2^t(k))$. De éénheids matrix noteren wij met I en zullen een leeg produkt van matrices, b.v. $\Pi_{i=0}^{-1}P^i$, ook als éénheidsmatrix opvatten. Voor gegeven strategieën (π_1,π_2) geldt dat de k'de komponent van de vektor $(\Pi_{j=0}^{t-1}P^j)\; S^t\; \bar{g}^t$ gelijk is aan de verwachte opbrengst op tijdstip t bij begintoestand k.

$$\bar{E}^t(\pi_1,\pi_2) = \sum_{i=0}^{t} (\Pi_{j=0}^{i-1}P^j)\; S^i\; \bar{g}^i$$

heeft als k'de komponent de verwachting van de opbrengst tot en met tijdstip t bij begintoestand k. Als de betaling $g(k_1,\pi_1(k),\pi_2(k))$ begrensd is kunnen wij de totaal verwachtte opbrengst schrijven als:

$$\overline{E}(\pi_1,\pi_2) = \lim_{t\to\infty} \overline{E}^t(\pi_1,\pi_2) = \sum_{i=0}^{\infty} (\Pi_{j=0}^{i-1} P^j)\ S^i\ \overline{g}^i.$$

Hiermee is tevens de waarde nul toegekend aan een niet stoppende spelrealisering.

Bij iedere vektor $\overline{W} = (w_1,\ldots,w_N)^T$ en iedere begintoestand k construeren wij een dummyspel $G_k(\overline{W})$ dat ontstaat door de symbolische betaling $s_k g_k + \sum_{\ell=1}^N p_{k\ell}\ \Gamma_\ell$ te vervangen door de betaling $s_k g_k + \sum_{\ell=1}^N p_{k\ell} w_\ell$.

AANNAME 4.2.1. *Wij veronderstellen dat de dummyspelen $G_k(\overline{W})$ strikt bepaald zijn.*

Zoals bij het stochastische spel van Shapley definiëren wij een afbeelding $\overline{W} \to T(\overline{W})$, waarbij $T(\overline{W})$ de vektor is met als k'de komponent de waarde van het spel $G_k(\overline{W})$.

DEFINITIE 4.2.2. Een vektor W is *bereikbaar voor speler 1* als geldt: $T(\overline{W})_k > w_k$ als $w_k > 0$ en $T(\overline{W})_k \geq w_k$ als $w_k \leq 0$ voor alle k. Een vektor $\overline{W}$ is *bereikbaar voor speler 2* als geldt: $T(\overline{W})_k < w_k$ als $w_k < 0$ en $T(\overline{W})_k \leq w_k$ als $w_k \geq 0$ voor alle k. De verzameling van vektoren bereikbaar voor speler 1 noteren wij als C_1, voor speler 2 als C_2.

Uit de definitie blijkt dat alleen de nul-vektor voor beide spelers bereikbaar zou kunnen zijn.

4.3. Betekenis van bereikbaarheid

STELLING 4.3.1.

(a) *Als $\overline{W} \in C_1$ dan bestaat er bij iedere $\varepsilon > 0$ een $\pi_{1\varepsilon} \in \Pi_1$ zodanig dat:*

$$\overline{E}(\pi_{1\varepsilon},\pi_2) \geq \overline{W} - \varepsilon\overline{1} \quad \textit{voor alle } \pi_2 \in \Pi_2.$$

(b) *Als $\overline{W} \in C_2$ dan bestaat er bij iedere $\varepsilon > 0$ een $\pi_{2\varepsilon} \in \Pi_2$ zodanig dat:*

$$\overline{E}(\pi_1,\pi_{2\varepsilon}) \leq \overline{W} + \varepsilon\overline{1} \quad \textit{voor alle } \pi_1 \in \Pi_1.$$

BEWIJS. Laat $\overline{W} \in C_1$ dan bestaat er vanwege de eindige toestandsruimte een $\gamma > 0$ zodanig dat $\gamma \leq T(\overline{W})_k - w_k$ voor alle k waarvoor $w_k > 0$.

Gegeven $\varepsilon > 0$ construeren wij de strategie $\pi_{1\varepsilon}$ als volgt: kies δ zodanig dat $0 < \delta < \min(\gamma,\varepsilon)$ en vervolgens voor iedere toestand k.

(1) Als $G_k(\overline{W})$ een optimale strategie $\pi_1^*(k)$ voor speler 1 heeft laat dan $\pi_{1\varepsilon}^t(k) = \pi_1^*(k)$.

(2) Als $G_k(\overline{W})$ geen optimale strategie voor speler 1 heeft en $w_k > 0$ laat $\pi_{1\varepsilon}^t(k) = \tilde{\pi}_1(k)$, waarbij $\tilde{\pi}_1(k)$ een δ-optimale strategie is voor speler 1 in $G_k(\overline{W})$.

(3) Als $G_k(\overline{W})$ geen optimale strategie voor speler 1 heeft en $w_k \leq 0$ laat dan $\pi_{1\varepsilon}^t(k) = \pi_1^t(k)$, waarbij $\pi_1^t(k)$ een δ^t optimale strategie is voor speler 1 in $G_k(\overline{W})$ en $\delta^t = (\tfrac{1}{2})^{t+1}\delta$.

Deze constructie is voor toestanden die aan (1) voldoen voor iedere $\varepsilon > 0$ hetzelfde en $\pi_{1\varepsilon}$ is stationair in de toestanden die aan (1) of (2) voldoen. Wij zullen aantonen dat deze $\pi_{1\varepsilon}$ aan het gestelde in (a) voldoet. Laat in het spel $G_k(\overline{W})$ de aktie $\pi_{1\varepsilon}^t(k)$ worden gekozen door speler 1 en een willekeurige aktie $\pi_2^t(k)$ door speler 2 en schrijf voor de verwachte opbrengst in het spel $G_k(\overline{W})$

$$L_k^t = s_k^t\, g_k^t + \sum_{\ell=0}^{N} p_{k\ell}^t w_\ell .$$

Het is eenvoudig in te zien dat

(a) als (1) geldt en $w_k > 0$ dan

$$L_k^t \geq T(\overline{W})_k \geq w_k + \gamma > w_k + \gamma - \delta$$

(b) als (1) geldt en $w_k \leq 0$ dan

$$L_k^t \geq T(\overline{W})_k \geq w_k > w_k - \delta^t$$

(c) als (2) geldt, dan is $w_k > 0$ en dus

$$L_k^t \geq T(\overline{W})_k - \delta \geq w_k + \gamma - \delta$$

(d) als (3) geldt, dan is $w_k \leq 0$ en

$$L_k^t \geq T(\overline{W})_k - \delta^t \geq w_k - \delta^t.$$

Derhalve

$$s_k g_k^t + \sum_{\ell=1}^{N} p_{k\ell}^t w_\ell \geq \begin{cases} w_k + \gamma - \delta & \text{als } w_k > 0 \\ w_k - \delta^t & \text{als } w_k \leq 0 \end{cases} .$$

Wij noteren $\overline{\mu}$ en $\overline{\delta^t}$ voor de vektoren met als k'de komponent

$$\mu_k = \begin{cases} \gamma - \delta & \text{als } w_k > 0 \\ 0 & \text{als } w_k \leq 0 \end{cases}$$

$$\delta_k^t = \begin{cases} 0 & \text{als } w_k > 0 \\ \delta^t & \text{als } w_k \leq 0 \end{cases} .$$

Dan geldt $S^t\overline{g}^t + P^t\,\overline{W} \geq \overline{W} + \overline{\mu} - \delta^t$ en dus $S^t\overline{g}^t \geq (I-P^t)\,\overline{W} + \overline{\mu} - \delta^t$. Omdat de ongelijkheid door vermenigvuldiging met matrices met niet negatieve elementen behouden blijft geldt:

$$\overline{E}^t = \sum_{i=0}^{t} (\Pi_{j=0}^{i-1} P^j)\; S^i\overline{g}^i \geq \sum_{i=0}^{t} (\Pi_{j=0}^{i-1} P^j)\; [(I-P^i)\overline{W}+\mu-\overline{\delta}^i].$$

Vanwege

$$\{\sum_{i=0}^{t} (\Pi_{j=0}^{i-1} P^j)[I-P^i]\}\;\overline{W} = \{I-P^o+P^o-P^oP^1+\ldots-P^oP^1\ldots P^t\}\;\overline{W} =$$

$$= \overline{W} - (\Pi_{i=0}^{t} P^i)\;\overline{W}.$$

vinden wij

$$\overline{E}^t \geq \overline{W} - (\Pi_{i=0}^{t} P^i)\;\overline{W} + \sum_{i=0}^{t} (\Pi_{j=0}^{i-1} P^j)\;\overline{\mu} - \sum_{i=0}^{t} (\Pi_{j=0}^{i-1} P^j)\;\overline{\delta}^i.$$

Laat $\tau = \max w_i/(\gamma-\delta)$ indien dit positief is en anders $\tau = 0$ (alle $w_i \leq 0$). Als $w_k \leq 0$ dan $\tau\mu_k = \tau.0 = 0 \geq w_k$. Als $w_k > 0$ dan $\tau\mu_k = (w_\ell/(\gamma-\delta))/(\gamma-\delta) = w_\ell$, waarbij ℓ een index is met $w_\ell = \max w_i$. Dus in beide gevallen geldt $\tau\mu_k \geq w_k$. Derhalve $\tau\overline{\mu} \geq \overline{W}$. Dit heeft tot gevolg dat:

$$\sum_{i=0}^{t} (\Pi_{j=0}^{i-1} P^j)\;\overline{\mu} - (\Pi_{i=0}^{t} P^i)\;\overline{W} \geq$$

$$\sum_{i=0}^{t} (\Pi_{j=0}^{i-1} P^j)\;\overline{\mu} - \tau(\Pi_{i=0}^{t} P^i)\;\overline{\mu} = \sum_{i=0}^{t} \overline{\mu}^{i-1} - \tau\overline{\mu}^t,$$

waarbij $\overline{\mu}^t = (\Pi_{i=0}^t P^i)\ \overline{\mu}$, voor $t = -1,0,1,\ldots$.

Als $\sum_{i=0}^\infty \mu_k^{i-1}$ niet eindig is dan geldt vanwege $0 \le \mu_k^t \le \gamma - \delta$ dat $\sum_{i=0}^t \mu_k^{i-1} \ge \tau\mu_k^t$ als t groot genoeg is. Als $\sum_{i=0}^\infty \mu_k^{i-1}$ eindig is, laat dan $\sum_{i=0}^\infty \mu_k^{i-1} = S$ en kies $\varepsilon = S/(1+\tau)$. Als t groot genoeg is dan geldt: $S - \sum_{i=0}^t \mu_k^{i-1} \le \varepsilon$. Hieruit volgt dat $\sum_{i=0}^t \mu_k^{i-1} \ge S - \varepsilon = S - S/(1+\tau) = S\tau/(1+\tau)$ en $\tau\mu_k^t \le \tau S/(1+\tau)$. Dus bestaat er een m zodanig dat

$$\sum_{i=0}^{t} \overline{\mu}^{i-1} \ge \tau\overline{\mu}^t \quad \text{als } t \ge m.$$

Ofwel

$$-(\Pi_{i=0}^t P^i)\ \overline{W} + \sum_{i=0}^{t} (\Pi_{j=0}^{i-1} P^j)\ \overline{\mu} \ge 0 \quad \text{als } t \ge m.$$

Kombineren wij dit resultaat met de ongelijkheid (4.1) dan levert dit voor $t \ge m$

$$\text{(4.2)} \qquad \overline{E}^t \ge \overline{W} - \sum_{i=0}^{t} (\Pi_{j=0}^{i-1} P^j)\ \overline{\delta}^i .$$

Er geldt:

$$(\Pi_{j=0}^{i-1} P^j)\ \overline{\delta}^i \le (\Pi_{j=0}^{i-1} P^j)\ \delta^i\ \overline{1} \le \delta^i\ \overline{1}$$

en dus

$$\lim_{t\to\infty} \sum_{i=0}^{t} (\Pi_{j=0}^{i-1} P^j)\ \overline{\delta}^i \le \lim_{t\to\infty} \sum_{i=0}^{t} \delta^i\ \overline{1} = (\sum_{i=0}^{\infty} \delta^i)\ \overline{1} =$$

$$= (\sum_{i=0}^{\infty} (\tfrac{1}{2})^{i+1}\delta)\overline{1} = \delta\overline{1}.$$

Merk op dat $\lim_{t\to\infty} \sum_{i=0}^t (\Pi_{j=0}^{i-1} P^j)\ \overline{\delta}^i$ bestaat want de som is voor alle komponenten begrensd en alle termen zijn positief. Uit (4.2) en $\delta \le \varepsilon$ volgt

$$\overline{E}(\pi_{1\varepsilon},\pi_2) = \lim_{t\to\infty} \overline{E}^t \ge \overline{W} - \lim_{t\to\infty} \sum_{i=0}^{t} (\Pi_{j=0}^{i-1} P^j)\ \overline{\delta}^i \ge$$

$$\ge \overline{W} - \delta\overline{1} \ge \overline{W} - \varepsilon\overline{1} \quad \text{voor alle } \pi_2 \in \Pi_2,$$

waarmee gedeelte (a) bewezen is. Het bewijs van (b) verloopt op dezelfde manier door het konstrueren van een strategie $\pi_{2\varepsilon}$ voor speler 2. □

OPMERKING 4.3.2. Als voor speler 1 optimale strategieën bestaan voor $G_k(\overline{W})$ voor alle k, dan is de gekonstrueerde strategie $\pi_{1\varepsilon}$, speel in iedere toestand optimaal voor $G_k(\overline{W})$, onafhankelijk van ε en dus geldt:

$$\overline{E}(\pi_{1\varepsilon},\pi_2) \geq \overline{W} \quad \text{voor alle } \pi_2 \in \Pi_2.$$

Evenzo voor speler 2.

OPMERKING 4.3.3. De strategie $\pi_{1\varepsilon}$ is stationair in de toestanden k waarvoor een optimale strategie bestaat voor speler 1 voor $G_k(\overline{W})$ of waarvoor $w_k > 0$. De strategie $\pi_{2\varepsilon}$ is stationair in de toestanden waarvoor een optimale strategie bestaat voor speler 2 voor $G_k(\overline{W})$ of waarvoor $w_k < 0$.

4.4 De kritieke vektor

DEFINITIE 4.4.1. Een vektor $\overline{V}$ heet een *kritieke vektor* van het recursieve spel als voor iedere omgeving $\mathcal{O}(\overline{V})$ van $\overline{V}$ geldt $\mathcal{O}(\overline{V}) \cap C_1 \neq \phi$ en $\mathcal{O}(\overline{V}) \cap C_2 \neq \phi$.

STELLING 4.4.2. *Als een recursief spel een kritieke vektor $\overline{V}$ heeft geldt: Bij iedere $\varepsilon > 0$ bestaan er strategieën $\pi_{1\varepsilon} \in \Pi_1$ en $\pi_{2\varepsilon} \in \Pi_2$ zodanig dat*

$$\overline{E}(\pi_{1\varepsilon},\pi_2) \geq \overline{V} - \varepsilon\overline{1} \quad \textit{voor alle } \pi_2 \in \Pi_2$$

$$\overline{E}(\pi_1,\pi_{2\varepsilon}) \leq \overline{V} + \varepsilon\overline{1} \quad \textit{voor alle } \pi_1 \in \Pi_1$$

$\pi_{1\varepsilon}$ is stationair in de toestanden k waarvoor $v_k > 0$ en $\pi_{2\varepsilon}$ is stationair in de toestanden k waarvoor $v_k < 0$. In toestanden k waarvoor geldt dat alle dummy spelen $G_k(\overline{W})$ optimale strategieen voor beide spelers hebben zijn zowel $\pi_{1\varepsilon}$ als $\pi_{2\varepsilon}$ stationair.

BEWIJS. Laat $\delta = \min\{\varepsilon/2, v_k | v_k > 0\}$ en $S_\delta(\overline{V})$ een sfeer met straal δ en middelpunt $\overline{V}$. Kies $\overline{W} \in S_\delta(\overline{V}) \cap C_1$, door de keuze van δ zorgen wij ervoor dat als $v_k > 0$ dan ook $w_k > 0$. Konstrueer nu voor deze vektor $\overline{W}$ een strategie $\pi_{1\varepsilon/2}$ zoals in het bewijs van stelling 4.3.1. Deze strategie is stationair in de toestanden die in de stelling zijn genoemd, want als $v_k > 0$ dan is ook $w_k > 0$ en

$$\overline{E}(\pi_{1\varepsilon/2},\pi_2) \geq \overline{W} - \varepsilon/2\ \overline{1} \geq \overline{V} - \varepsilon\overline{1} \quad \text{voor alle } \pi_2 \in \Pi_2.$$

Het bewijs voor het bestaan van een strategie $\pi_{2\varepsilon}$ verloopt evenzo. □

OPMERKING 4.4.3. Uit het bewijs van stelling 4.2.2. blijkt dat stelling 4.3.1.(a) niet alleen geldt voor elementen van C_1, maar ook voor ophopingspunten van C_1. Als voor alle toestanden k alle dummy spelen $G_k(\overline{W})$ optimale strategieën voor beide spelers bezitten en als het spel tevens een kritieke vektor bezit, hoeft niet te gelden: Er bestaat een $\pi_1^* \in \Pi_1$ zodanig dat

$$\overline{E}(\pi_1^*,\pi_2) \geq \overline{V} \quad \text{voor alle } \pi_2 \in \Pi_2.$$

LEMMA 4.4.4. *Voor een recursief spel geldt:*

$$T(\overline{W}) - \delta\overline{1} \leq T(\overline{W}-\delta\overline{1}) \leq T(\overline{W}) \leq T(\overline{W}+\delta\overline{1}) \leq T(\overline{W}) + \delta\overline{1}.$$

voor alle $\overline{W}$ *en alle* $\delta > 0$.

BEWIJS: Laat $\pi_{1\varepsilon}(k)$ een ε-optimale strategie voor speler 1 in $G_k(\overline{W})$ en $\pi_{2\varepsilon}(k)$ een ε-optimale strategie voor speler 2 in $G_k(\overline{W}-\delta\overline{1})$ dan geldt:

$$\text{val } G_k(\overline{W}) - \delta - \varepsilon \leq s_k g_k + \sum_{\ell=1}^{N} p_{k\ell} w_\ell - \delta \leq$$

$$s_k g_k + \sum_{\ell=1}^{N} p_{k\ell}(w_\ell - \delta) \leq \text{val } G_k(\overline{W}-\delta\overline{1}) + \varepsilon.$$

Dit kan voor alle $\varepsilon > 0$, dus

$$\text{val } G_k(\overline{W}) - \delta \leq \text{val } G_k(\overline{W}-\delta\overline{1}).$$

Laat $\pi_{1\varepsilon}(k)$ een ε-optimale strategie voor speler 1 in $G_k(\overline{W}-\delta\overline{1})$ en $\pi_{2\varepsilon}(k)$ een ε-optimale strategie voor speler 2 in $G_k(\overline{W})$ dan geldt:

$$\text{val } G_k(\overline{W}-\delta\overline{1}) - \varepsilon \leq s_k g_k + \sum_{\ell=1}^{N} p_{k\ell}(w_\ell - \delta)$$

$$\leq s_k g_k + \sum_{\ell=1}^{N} p_{k\ell} w_\ell \leq \text{val } G_k(\overline{W}) + \varepsilon.$$

Dit kan voor alle $\varepsilon > 0$, dus

$$\text{val } G_k(\overline{W}-\delta\overline{1}) \leq \text{val } G_k(\overline{W}).$$

De afleiding van de overige relaties gaat evenzo. □

STELLING 4.4.5.

(a) *Er is hoogstens één kritieke vektor.*

(b) *Stel dat* $\overline{V}$ *een kritieke vektor is, dan geldt* $\overline{W} \leq \overline{V}$ *voor alle* $\overline{W} \in C_1$ *en* $\overline{W} \geq \overline{V}$ *voor alle* $\overline{W} \in C_2$.

(c) $T(\overline{V}) = \overline{V}$.

BEWIJS:

(a) Laat $\overline{V}$ en $\overline{V}^*$ kritieke vektoren van het recursieve spel zijn. Laat $\varepsilon > 0$. Kies $\pi_{1\varepsilon}$ zodanig dat $\overline{E}(\pi_{1\varepsilon},\pi_2) \geq \overline{V} - \varepsilon\overline{1}$ voor alle $\pi_2 \in \Pi_2$ en $\pi_{2\varepsilon}$ zodanig dat $\overline{E}(\pi_1,\pi_{2\varepsilon}) \leq \overline{V}^* + \varepsilon\overline{1}$ voor alle $\pi_1 \in \Pi_1$ dan geldt $\overline{V} - \varepsilon\overline{1} \leq \overline{E}(\pi_{1\varepsilon},\pi_{2\varepsilon}) \leq \overline{V}^* + \varepsilon\overline{1}$. Dit kan voor alle $\varepsilon > 0$, dus $\overline{V} \leq \overline{V}^*$; analoog $\overline{V}^* \leq \overline{V}$ en dus $\overline{V} = \overline{V}^*$.

(b) Laat $\overline{W} \in C_1$ en $\varepsilon > 0$. Kies $\pi_{1\varepsilon} \in \Pi_1$ zodanig dat $\overline{E}(\pi_{1\varepsilon},\pi_2) \geq \overline{W} - \varepsilon\overline{1}$ voor alle $\pi_2 \in \Pi_2$ en $\pi_{2\varepsilon} \in \Pi_2$ zodanig dat $\overline{E}(\pi_1,\pi_{2\varepsilon}) \leq \overline{V} + \varepsilon\overline{1}$ voor alle $\pi_1 \in \Pi_1$ dan geldt $\overline{W} - \varepsilon\overline{1} \leq \overline{E}(\pi_{1\varepsilon},\pi_{2\varepsilon}) \leq \overline{V} + \varepsilon\overline{1}$. Dit kan voor alle $\varepsilon > 0$, dus $\overline{W} \leq \overline{V}$. De tweede bewering in (b) kan op dezelfde manier worden bewezen.

(c) Uit de definitie van een kritieke vektor volgt dat bij elke $\varepsilon > 0$ $\overline{W}_1 \in C_1$ en $\overline{W}_2 \in C_2$ bestaan met $\overline{W}_1 \geq \overline{V} - \varepsilon\overline{1}$ en $\overline{W}_2 \leq \overline{V} + \varepsilon\overline{1}$. Dan vinden wij met behulp van lemma 4.4.4 de definitie van bereikbaarheid en gedeelte (b) van deze stelling dat

$$T(\overline{V}) - \varepsilon\overline{1} \leq T(\overline{V}-\varepsilon\overline{1}) \leq T(\overline{W}_1) \leq \overline{W}_1 \leq \overline{V} \leq \overline{W}_2 \leq T(\overline{W}_2) \leq T(\overline{V}+\varepsilon\overline{1})$$
$$\leq T(\overline{V}) + \varepsilon\overline{1}.$$

Aangezien wij ε willekeurig klein kunnen kiezen geldt $T(\overline{V}) = \overline{V}$. □

Bij ieder recursief spel $\overline{\Gamma} = (\Gamma_1,\ldots,\Gamma_N)$ kunnen wij gereduceerde spelen $\overline{\Gamma}_q(\overline{W}_{q^c})$ konstrueren, waarbij $q \subset \{1,\ldots,N\}$ en q^c het komplement van q is. De toestanden in het gereduceerde spel zijn de toestanden in q. $\overline{W}_{q^c}$ is een vektor met een komponent voor ieder element van q^c. Een overgang naar een toestand in q^c in het oorspronkelijke spel wordt vervangen door een betaling gelijk aan de komponent van $\overline{W}_{q^c}$ die bij de toestand behoort waarna het spel stopt. De betaling van dit gereduceerde spel in toestand $k \in q$

kunnen wij schrijven als:

$$s_k g_k + \sum_{\ell \in q^c} p_{k\ell} w_\ell + \sum_{\ell \in q} p_{k\ell} \Gamma_\ell .$$

Wij beschouwen voor konstante ν het gereduceerde spel $\overline{\Gamma}_q(\nu \overline{1}_q c)$. Als het deelspel met begintoestand $k \in q$ een waarde heeft zullen wij deze waarde noteren als $v_k(\nu)$.

<u>LEMMA 4.4.6</u>. *Als $\alpha > 0$ en $\beta < 0$ zodanig dat $\beta < g(k,\pi_1(k),\pi_2(k)) < \alpha$ voor alle $k \in q$, $\pi_1(k) \in \Pi_1(k)$ en $\pi_2(k) \in \Pi_2(k)$ en $v_k(\nu)$ bestaat voor alle $k \in q$ en alle ν dan geldt:*

(a) *Als $\nu \in [\alpha,\beta]$ dan $\beta \le v_k(\nu) \le \alpha$.*

(b) *Voor alle $\delta > 0$ en voor alle ν geldt: $v_k(\nu) - \delta \le v_k(\nu-\delta) \le v_k(\nu) \le v_k(\nu+\delta) \le v_k(\nu) + \delta$ en als bovendien $v_k(\nu) \ne \nu$ dan $v_k(\nu) - \delta < v_k(\nu-\delta)$ en $v_k(\nu+\delta) < v_k(\nu) + \delta$ voor alle k.*

<u>BEWIJS</u>. Wij beschouwen het deelspel met begintoestand $k \in q$ en zullen de onderindex k weglaten. De waarde van dit spel is $v(\nu)$. Voor ieder paar strategieën $\pi_1 \in \Pi_1$ en $\pi_2 \in \Pi_2$ kan de verwachte betaling geschreven worden in de vorm $E(\pi_1,\pi_2) = (1-S)E + S\nu$, waarbij S de kans is dat één der betalingen ν plaatsvindt; E is de verwachting van de overige betalingen gegeven dat niet één der betalingen ν plaatsvindt. Vanwege $\beta \le E \le \alpha$ en $\beta \le \nu \le \alpha$ geldt $(1-S)\beta + S\beta \le (1-S)E + S\nu \le (1-S)\alpha + S\alpha$. Hieruit volgt $\beta \le E(\pi_1,\pi_2) \le \alpha$ voor alle $\pi_1 \in \Pi_1$ en $\pi_2 \in \Pi_2$ en $\beta \le v(\nu) \le \alpha$, waarmee (a) bewezen is. Omdat $v(\nu)$ bestaat geldt dat er bij iedere $\varepsilon > 0$ $\pi_{1\varepsilon} \in \Pi_1$ en $\pi_{2\varepsilon} \in \Pi_2$ bestaan zodanig dat:

$$E(\pi_{1\varepsilon},\pi_2) = (1-\delta)E + S\nu \ge v(\nu)-\varepsilon \qquad \text{voor alle } \pi_2 \in \Pi_2. \tag{4.3}$$

en

$$E(\pi_1,\pi_{2\varepsilon}) = (1-S)E + S\nu \le v(\nu)+\varepsilon \qquad \text{voor alle } \pi_1 \in \Pi_1. \tag{4.4}$$

Wij bekijken nu het effect van deze strategieën in het gereduceerde spel, waarbij betaling ν vervangen is door $\nu - \delta$. Ook hier nemen wij het deelspel met begintoestand k en noteren de waarde van het spel $v(\nu-\delta)$. Uit (4.4) volgt:

$$E(\pi_1,\pi_{2\varepsilon}) = (1-S)E + S(\nu-\delta) \leq v(\nu) + \varepsilon - S\delta \leq v(\nu) + \varepsilon \qquad \text{voor alle } \pi_1 \in \Pi_1.$$

Wij kunnen $\varepsilon > 0$ willekeurig klein kiezen, zodat $v(\nu-\delta) \leq v(\nu)$. Uit (4.3) volgt:

$$(4.5) \qquad E(\pi_{1\varepsilon},\pi_2) = (1-S)E + S(\nu-\delta) \geq v(\nu) - \varepsilon - S\delta \quad \text{voor alle } \pi_2 \in \Pi_2.$$

Dit kan voor alle $\varepsilon > 0$ en omdat $S \leq 1$ geldt $v(\nu-\delta) \geq v(\nu) - \delta$. Laat nu $v(\nu-\delta) = v(\nu) - \delta$. Er bestaat een strategie $\pi^*_{2\varepsilon}$ die ε-optimaal is in het spel met eindbetaling $\nu - \delta$.

$$(4.6) \qquad E(\pi_1,\pi^*_{2\varepsilon}) = (1-S)E + S(\nu-\delta) \leq v(\nu-\delta) + \varepsilon = v(\nu) - \delta + \varepsilon \qquad \text{voor alle } \pi_1 \in \Pi_1.$$

Wij zorgen er nu voor dat $\beta < \nu - \delta$ en $\alpha > \nu$; eventueel door het kiezen van een kleinere β en een grotere α. Laat nu het strategieënpaar $\pi_{1\varepsilon},\pi^*_{2\varepsilon}$ worden gespeeld in het spel met eindbetaling $\nu - \delta$, waarbij $\pi_{1\varepsilon}$ een ε-optimale strategie is in het spel met eindbetaling ν. Uit (4.6) en $\beta \leq E$ volgt $(1-S^*)\beta + S^*(\nu-\delta) \leq v(\nu) - \delta + \varepsilon$ en vanwege $\nu - \delta - \beta > 0$ geldt:

$$(4.7) \qquad S^* \leq \frac{v(\nu) - \delta + \varepsilon - \beta}{\nu - \delta - \beta},$$

waarbij S^* de kans is dat het spel stopt onder $\pi_{1\varepsilon},\pi^*_{2\varepsilon}$. Uit (4.3) volgt $(1-S^*)\alpha + S^*\nu \geq v(\nu) - \varepsilon$ en vanwege $\alpha < \nu$ vinden wij:

$$(4.8) \qquad S^* \leq \frac{\alpha - v(\nu) + \varepsilon}{\alpha - \nu}.$$

Uit (4.5) en (4.7) volgt:

$$(4.9) \qquad E(\pi_{1\varepsilon},\pi^*_{2\varepsilon}) \geq v(\nu) - \varepsilon - \delta \left(\frac{v(\nu)-\delta+\varepsilon-\beta}{\nu - \delta - \beta}\right).$$

Uit (4.5) en (4.8) volgt:

$$(4.10) \qquad E(\pi_{1\varepsilon},\pi^*_{2\varepsilon}) \geq v(\nu) - \varepsilon - \delta \left(\frac{\alpha-v(\nu)+\varepsilon}{\alpha - \nu}\right).$$

Uit (4.6) en (4.9) volgt:

$$v(\nu-\delta) + \varepsilon \geq E(\pi_{1\varepsilon},\pi^*_{2\varepsilon}) \geq v(\nu) - \varepsilon - \delta(\frac{v(\nu-\delta)+\varepsilon-\beta}{\nu-\delta-\beta}).$$

Dit kan voor alle $\varepsilon > 0$ en dus geldt:

$$v(\nu-\delta) \geq v(\nu) - \delta(\frac{v(\nu)-\delta-\beta}{\nu-\delta-\beta}).$$

Omdat $v(\nu-\delta) = v(\nu) - \delta$ volgt $(v(\nu)-\delta-\beta)/(\nu-\delta-\beta) \geq 1$ en dus:

(4.11) $\quad v(\nu) \geq \nu$.

Evenzo volgt uit (4.6) en (4.10) dat $v(\nu-\delta) \geq v(\nu)- \delta(\alpha-v(\nu))/(\alpha-\nu)$ dus $(\alpha-v(\nu))/(\alpha-\nu) \geq 1$ en dus ook:

(4.12) $\quad v(\nu) \leq \nu$.

Uit (4.11) en (4.12) volgt $v(\nu) = \nu$. Wij bewezen dat uit $v(\nu-\delta) = v(\nu) - \delta$ volgt dat $v(\nu) = \nu$. Hieruit volgt dat als $v(\nu) \neq \nu$ dan geldt $v(\nu) - \delta < v(\nu-\delta)$. De ongelijkheden voor $\nu - \delta$ zijn hiermee bewezen. Voor $\nu + \delta$ verloopt het bewijs evenzo. □

Laat $\overline{U} = (u_1,\ldots,u_N)$ en $\overline{V} = (v_1,\ldots,v_N)$ vektoren zijn dan noteren wij:

$\overline{U} \overset{\circ}{\geq} \overline{V}$ als geldt $u_k > v_k$ of $(u_k \geq v_k$ en $v_k \leq 0)$ voor $k = 1,\ldots,N$.

$\overline{U} \overset{\circ}{\leq} \overline{V}$ als geldt $u_k < v_k$ of $(u_k \leq v_k$ en $v_k \geq 0)$ voor $k = 1,\ldots,N$.

Dus $\overline{W}$ is bereikbaar voor speler 1 als geldt $T(\overline{W}) \overset{\circ}{\geq} \overline{W}$ en $\overline{W}$ is bereikbaar voor speler 2 als geldt $T(\overline{W}) \overset{\circ}{\leq} \overline{W}$.

<u>LEMMA 4.4.7</u>. *Laat $\overline{V}$ de kritieke vektor zijn van een recursief spel $\overline{\Gamma}$ en laat $\overline{\Gamma}_q(\overline{V}_{q^c})$ het gereduceerde spel dat ontstaat uit $\overline{\Gamma}$ door de overgang naar een toestand $k \in q^c$ te vervangen door betaling v_k. Dit gereduceerde spel heeft als kritieke vektor $\overline{V}$ beperkt tot de elementen van q.*

<u>BEWIJS</u>: Bij iedere $\varepsilon > 0$ bestaat er een $\overline{W} \in S_\varepsilon(\overline{V}) \cap C_1$. Vanwege $\overline{W} \in C_1$ geldt $T(\overline{W}) \overset{\circ}{\geq} \overline{W}$. Uit stelling 4.4.5 (b) volgt dat $\overline{W} \leq \overline{V}$ en dus ook $\overline{W}_{q^c} \leq \overline{V}_{q^c}$.

Uit lemma 4.4.4 volgt dat val $G_k(\overline{V}_q c, \overline{W}_q) \geq$ val $G_k(\overline{W}_q c, \overline{W}_q)$ = val $G_k(\overline{W}) \overset{\circ}{\geq} w_k$ voor alle $k \in q$. Dus voor de waardetransformatie T_q van het gereduceerde spel geldt: $T_q(\overline{W}_q) \overset{\circ}{\geq} \overline{W}_q$ en dus is $\overline{W}_q$ bereikbaar voor speler 1 in het gereduceerde spel. Tevens geldt $\overline{W}_q \in S_\varepsilon(\overline{V}_q)$. Dus iedere sfeer om $\overline{V}_q$ bevat een vektor bereikbaar voor speler 1. Evenzo bevat iedere sfeer om V_q een vektor bereikbaar voor speler 2. $\overline{V}_q$ is dus de kritieke vektor van het gereduceerde spel. □

STELLING 4.4.8. *Ieder recursief spel heeft een kritieke vektor.*

BEWIJS: Wij bewijzen de stelling m.b.v. volledige inductie naar het aantal toestanden in het recursieve spel. Laat $N = 1$, wij spreken dan van een simpel recursief spel (zoals de voorbeelden 4.1.2,4.1.3 en 4.1.5). Als $T(W) > W$ of ($T(W) = W$ en $W \leq 0$) dan is W bereikbaar voor speler 1. Als $T(W) < W$ of ($T(W) = W$ en $W > 0$) dan is W bereikbaar voor speler 2. Dus $C_1 \cup C_2 = \mathbb{R}$ (de verzameling van de reële getallen). Laat $\alpha > 0$ een bovengrens en $\beta < 0$ een ondergrens voor de mogelijke betalingen zijn dan volgt uit lemma 4.4.6 (a) dat $T(\alpha) \leq \alpha$ en dus is α bereikbaar voor speler 2 en tevens dat $T(\beta) \geq \beta$ en dus is β bereikbaar voor speler 1. Vanwege $C_1 \neq \emptyset$, $C_2 \neq \emptyset$ en de samenhang van $\mathbb{R}$ bestaat er een $V \in \mathbb{R}$ die zowel ophopingspunt van C_1 als van C_2 is. Deze V is dus de kritieke vektor. Wij veronderstellen dat alle recursieve spelen met $N \leq n$ een kritieke vektor hebben en zullen bewijzen dat alle recursieve spelen met $N = n + 1$ een kritieke vektor hebben. Laat $\overline{\Gamma}$ een recursief spel met $n + 1$ toestanden en construeer een gereduceerd spel door 1 toestand k weg te laten en overgangen vanuit andere toestanden naar k te vervangen door een betaling ν. Wij zullen dit gereduceerde spel noteren als $\overline{\Gamma}_r(\nu)$. Volgens de inductieveronderstelling heeft dit spel een kritieke vektor $\overline{V}_r(\nu)$. Beschouw nu het spel: $G_k(\overline{V}_r(\nu), \nu)$, dit spel is strikt bepaald en heeft dus een waarde:

$$\tilde{V}(\nu) = \text{val } G_k(\overline{V}_r(\nu), \nu). \tag{4.13}$$

Uit lemma 4.4.4 volgt:

$$\text{val } G_k(\overline{V}_r(\nu), \nu) \geq \text{val } G_k(\overline{V}_r(\nu) - \delta,\ \nu - \delta) \quad \text{voor alle } \delta > 0. \tag{4.14}$$

Uit lemma 4.4.6 (b) volgt $\overline{V}_r(\nu-\delta) \geq \overline{V}_r(\nu) - \delta$ voor alle $\delta > 0$ en dus ook:

$$\text{val } G_k(\overline{V}_r(\nu)-\delta,\nu-\delta) \le \text{val } G_k(\overline{V}_r(\nu-\delta),\nu-\delta), \quad \text{voor alle } \delta > 0. \tag{4.15}$$

Tevens volgt uit lemma 4.4.6 (b) $\overline{V}_r(\nu) \ge \overline{V}_r(\nu-\delta)$ en dus:

$$\text{val } G_k(\overline{V}_r(\nu-\delta),\nu-\delta) \le \text{val } G_k(\overline{V}_r(\nu),\nu) \quad \text{voor alle } \delta > 0. \tag{4.16}$$

Uit (4.14), (4.15) en (4.16) volgt:

$$\tilde{V}(\nu) - \delta \le \tilde{V}(\nu-\delta) \le \tilde{V}(\nu) \quad \text{voor alle } \delta > 0. \tag{4.17}$$

Evenzo geldt:

$$\tilde{V}(\nu) \le \tilde{V}(\nu+\delta) \le \tilde{V}(\nu) + \delta \quad \text{voor alle } \delta > 0. \tag{4.18}$$

Als $\alpha > 0$ een bovengrens en $\beta < 0$ een ondergrens is voor de betalingen in $\overline{\Gamma}$ en $\beta \le \nu \le \alpha$ dan volgt m.b.v. lemma 4.4.6 (a) dat $\overline{\beta} \le \overline{V}_r(\nu) \le \overline{\alpha}$. Dus zijn α en β ook bovengrens respectievelijk ondergrens van de betaling in $G_k(\overline{V}_r(\nu),\nu)$ en derhalve $\beta \le \tilde{V}(\nu) \le \alpha$. Kombineren wij dit met (4.17) en (4.18) dan volgt dat $\nu \to \tilde{V}(\nu)$ een continue afbeelding is van het gesloten interval $[\beta,\alpha]$ in zichzelf. Volgens de dekpuntstelling van Brouwer heeft $\tilde{V}(\nu)$ op $[\beta,\alpha]$ tenminste één dekpunt. De verzameling dekpunten is een gesloten verzameling. Laat ν^* het dekpunt zijn met de kleinste absolute waarde. Wij zullen bewijzen dat $\overline{V} = [\overline{V}_r(\nu^*),\nu^*]$ de kritieke vektor is van $\overline{\Gamma}$. Wij kiezen nu $\varepsilon > 0$ vast en zullen aantonen dat $S_\varepsilon(\overline{V})$ een vektor bevat die bereikbaar is voor speler 1 en onderscheiden daartoe de gevallen $\nu^* > 0$ en $\nu^* \le 0$.

geval $\nu^* > 0$.

Laat ν zodanig dat $0 \le \nu < \nu^*$ en $\delta = \nu^* - \nu$. Uit (4.17) volgt $\tilde{V}(\nu) = \tilde{V}(\nu^*-\delta) \ge \tilde{V}(\nu^*) - \delta = \nu^* - \delta = \nu$, maar $\tilde{V}(\nu) \ne \nu$ want ν^* is het dekpunt met de kleinste absolute waarde. Dus geldt:

$$\tilde{V}(\nu) > \nu \quad \text{als } 0 \le \nu < \nu^* \tag{4.19}$$

Laat $\nu^\varepsilon = \max\{0,\nu^*-\varepsilon/3\}$ dan geldt dat $\tilde{V}(\nu^\varepsilon) > \nu^\varepsilon$ en kies δ zodanig dat

$$\tilde{V}(\nu^\varepsilon) > \nu^\varepsilon + \delta \text{ en } 0 < \delta < \varepsilon/2. \tag{4.20}$$

Beschouw het gereduceerde spel met eindbetaling ν^{ε}. Dit spel heeft op grond van de inductieveronderstelling een kritieke vektor $\overline{V}_r(\nu^{\varepsilon})$. Er bestaat dus een $\overline{W}_r \in S_{\delta}(\overline{V}_r(\nu^{\varepsilon}))$ bereikbaar voor speler 1, dus

(4.21) $\qquad \mathrm{val}\, G_i(\overline{W}_r,\nu^{\varepsilon}) > (\overline{W}_r)_i$ of $(\mathrm{val} G_i(\overline{W}_r,\nu^{\varepsilon}) \geq (\overline{W}_r)_i$ en $(\overline{W}_r)_i \leq 0)$
voor alle $i \neq k$.

Vanwege $\overline{W}_r > \overline{V}_r(\nu^{\varepsilon}) - \delta$ en lemma 4.4.4 geldt:

$$\mathrm{val}\, G_k(\overline{W}_r,\nu^{\varepsilon}) \geq \mathrm{val}\, G_k(\overline{V}_r(\nu^{\varepsilon})-\delta,\nu^{\varepsilon}-\delta) \geq \mathrm{val}\, G_k(\overline{V}_r(\nu^{\varepsilon}),\nu^{\varepsilon}) - \delta.$$

Uit (4.20) en (4.13) volgt dat $\mathrm{val}\, G_k(\overline{V}_r(\nu^{\varepsilon}),\nu^{\varepsilon}) - \delta > \nu^{\varepsilon}$ en dus

(4.22) $\qquad \mathrm{val}\, G_k(\overline{W}_r,\nu^{\varepsilon}) > \nu^{\varepsilon}$.

Laat $\overline{W} = (\overline{W}_r,\nu^{\varepsilon})$ dan geldt op grond van (4.21) en (4.22) $T(\overline{W}) \overset{\circ}{\geq} \overline{W}$ en dus is $\overline{W}$ bereikbaar voor speler 1 in $\overline{\Gamma}$. Uit $\nu^{*} - \nu^{\varepsilon} < \varepsilon/2$ en lemma 4.4.6 (b) volgt $\overline{V}_r(\nu^{*}) - \overline{V}_r(\nu^{\varepsilon}) < \varepsilon/2$. Omdat $\delta < \varepsilon/2$ geldt $\overline{W}_r \in S_{\delta}(\overline{V}_r(\nu^{\varepsilon})) \subset S_{\varepsilon/2}(\overline{V}_r(\nu^{\varepsilon}))$ en $\overline{W}_r \in S_{\varepsilon}(\overline{V}_r(\nu^{*}))$ en derhalve $\overline{W} \in S_{\varepsilon}(\overline{V})$.

geval $\nu^{*} \leq 0$.

Wij noteren weer $\overline{V} = (v_1,\ldots,v_N) = [\overline{V}_r(\nu^{*}),\nu^{*}]$. Laat nu $p \subset \{1,\ldots,N\}$ zodanig dat

(4.23) $\qquad v_i \neq \nu^{*} \Longleftrightarrow i \in p$.

$\overline{V}_r(\nu^{*})$ is de kritieke vektor van het gereduceerde spel met eindbetaling ν^{*} en dus volgens stelling 4.4.5 (c) een dekpunt van de waarde transformatie van dit spel. Er geldt dus:

(4.24) $\qquad \mathrm{val}\, G_i(\overline{V}_r(\nu^{*}),\nu^{*}) = v_i \qquad$ voor alle $i \neq k$.

Omdat ν^{*} een dekpunt van de afbeelding $\nu \to \widetilde{V}(\nu)$ is geldt dit ook voor toestand k.

(4.25) $\qquad \mathrm{val}\, G_k(\overline{V}_r(\nu^{*}),\nu^{*}) = \widetilde{V}(\nu^{*}) = \nu^{*}$.

Uit (4.23), (4.24) en (4.25) volgt

(4.26) $\quad \text{val } G_i(\overline{V}_r(\nu^*),\nu^*) = \nu^* \quad$ voor alle $i \in p^c$.

Beschouw nu het gereduceerde spel $\overline{\Gamma}_p(\nu\overline{1}_{p}c)$; de toestanden van dit spel zijn de elementen van p en overgangen naar elementen van p^c zijn vervangen door een eindbetaling ν. Wij noteren de kritieke vektor van dit spel als $\tilde{\overline{V}}_p(\nu)$. Het spel $\overline{\Gamma}_r\nu^*$, waarbij alleen toestand k is verwijderd, heeft als kritieke vektor $\overline{V}_r(\nu^*)$. Noteren wij de vektor $\overline{V}$ beperkt tot elementen van p als $\overline{V}_p$ dan geldt op grond van lemma 4.4.7:

(4.27) $\quad \tilde{\overline{V}}_p(\nu^*) = \overline{V}_p$.

Laat $\varepsilon > 0$ en $\nu = \nu^* - \varepsilon$ dan volgt uit lemma 4.4.6 (b) dat $\tilde{\overline{V}}_p(\nu) = \tilde{\overline{V}}_p(\nu^*-\varepsilon) \geq \tilde{\overline{V}}_p(\nu^*) - \varepsilon$. Omdat op grond van (4.27) en (4.23) $\tilde{\overline{V}}_p(\nu^*)$ geen komponent gelijk aan ν^* heeft geldt $\tilde{\overline{V}}_p(\nu) > \tilde{\overline{V}}_p(\nu^*) - \varepsilon\overline{1}_p$, ofwel:

(4.28) $\quad \tilde{\overline{V}}_p(\nu^*) - \tilde{\overline{V}}_p(\nu) < \varepsilon\overline{1}_p$.

Er bestaat dus een δ zodanig dat $0 < \delta < \varepsilon$ en $\tilde{\overline{V}}_p(\nu^*) - \tilde{\overline{V}}_p(\nu) < \varepsilon\overline{1}_p - \delta\overline{1}_p$ en derhalve

(4.29) $\quad \tilde{\overline{V}}_p(\nu) - \delta\overline{1}_p > \tilde{\overline{V}}_p(\nu^*) - \varepsilon\overline{1}_p$.

Kies nu $\overline{W}_p \in S_\delta(\tilde{\overline{V}}_p(\nu))$ zodanig dat

(4.30) $\quad \text{val } G_i(\nu\overline{1}_{p}c,\overline{W}_p) \overset{\circ}{\geq} (\overline{W}_p)_i \quad$ voor alle $i \in p$

; dit kan omdat $\tilde{\overline{V}}_p(\nu)$ kritieke vektor is in het spel $\overline{\Gamma}_p(\nu\overline{1}_{p}c)$. Schrijven we nu (4.26) als:

$$\text{val } G_i(\nu^*\overline{1}_{p}c,\tilde{\overline{V}}_p(\nu^*)) = \nu^* \quad \text{voor alle } i \in p^c$$

en passen lemma 4.4.4 toe, dan volgt:

$$\text{val } G_i((\nu^*-\varepsilon)\,\overline{1}_{p}c,\ \tilde{\overline{V}}_p(\nu^*) - \varepsilon\overline{1}_p) \geq \nu^* - \varepsilon \quad \text{voor alle } i \in p^c$$

en omdat $\nu = \nu^* - \varepsilon$ geldt:

$$\text{val } G_i(\nu\bar{1}_p c,\ \tilde{\bar{V}}_p(\nu^*) - \varepsilon\bar{1}_p) \ge \nu \quad \text{voor alle } i \in p^c$$

vanwege (4.29) en lemma 4.4.4 volgt:

$$\text{val } G_i(\nu\bar{1}_p c, \overset{\approx}{\bar{V}}_p(\nu) - \delta\bar{1}_p) \ge \nu \quad \text{voor alle } i \in p^c. \tag{4.31}$$

Uit (4.31), lemma 4.4.4 en de keuze van $\bar{W}_p$ volgt:

$$\text{val } G_i(\nu\bar{1}_p c, \bar{W}_p) \ge \nu \quad \text{voor alle } i \in p^c. \tag{4.32}$$

Schrijven wij nu $\bar{W} = (\nu\bar{1}_p c, \bar{W}_p)$ dan vormen omdat $\nu \le 0$, (4.39) en (4.32) samen de bewering dat val $G_i(\bar{W}) \overset{\circ}{\ge} W_i$ voor alle i, dus $\bar{W}$ is bereikbaar voor speler 1. Uit (4.27) en (4.28) volgt $\|\tilde{\bar{V}}_p(\nu) - \bar{V}_p\| < \varepsilon$. Kombineren wij dit met $\bar{W}_p \in S_\delta(\tilde{\bar{V}}_p(\nu))$ en $\delta < \varepsilon$ dan volgt dat $\|\bar{W}_p - \bar{V}_p\| < 2\varepsilon$ en vanwege $\|\nu\bar{1}_p c - \nu^*\bar{1}_p c\| < \varepsilon$ dus ook $\|\bar{W} - \bar{V}\| < 2\varepsilon$. Dus ook in geval $\nu^* \le 0$ geldt dat iedere omgeving van $\bar{V}$ een vektor $\bar{W}$ bereikbaar voor speler 1 bevat.

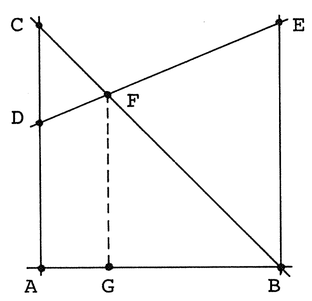
Evenzo kan worden bewezen dat iedere omgeving van $\bar{V}$ een vektor bevat die bereikbaar is voor speler 2. □

4.5. Opmerkingen over de voorbeelden

bij voorbeeld 4.1.3

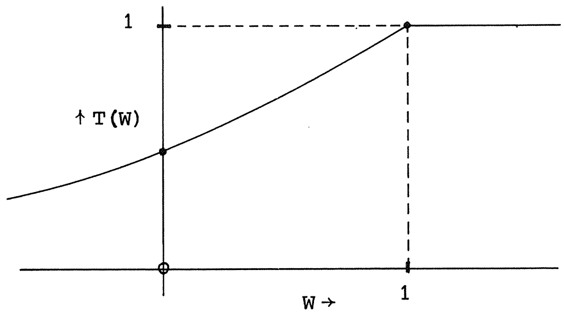
Beschouw het spel $G(W) = \begin{bmatrix} W & 1 \\ 1 & 0 \end{bmatrix}$.

Laat $T(W) = \text{val } G(W)$. Als $W \ge 1$ dan $T(W) = 1$. Als $W < 1$ dan kunnen wij $T(W)$ als volgt bepalen. Horizontaal zetten wij de gemengde strategieën van speler 2 uit op een interval van lengte 1. De afstand tot punt A is de kans waarmee speler 2 aktie 2 speelt.



EB : CD = 1:(1-W)

De verwachte betaling aan speler 1, als deze zijn eerste aktie kiest, wordt weergegeven door lijnstuk DE; als hij zijn tweede aktie kiest door lijnstuk CB. De lengte van FG is de waarde van het spel; dus $T(W) = 1/(2-W)$.



Er is slechts één dekpunt van de waardetransformatie, n.l. voor $W = 1$. Dit is dus tevens de kritieke vektor, in dit geval met slechts één komponent, van het recursieve spel. $C_1 = (-\infty,1)$ is bereikbaar voor speler 1. $C_2 = [1,\infty)$ is bereikbaar voor speler 2, immers de kritieke vektor is positief en dus bereikbaar voor speler 2. Speelt speler 2 een strategie die optimaal is in het matrixspel

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

dan is dit een optimale strategie voor speler 2 in het recursieve spel, dus iedere strategie is optimaal voor speler 2. Speelt speler 1 optimaal voor het matrixspel

$$\begin{bmatrix} 1-\varepsilon & 1 \\ 1 & 0 \end{bmatrix}$$

dan is dit een ε-optimale strategie voor speler 1 in het recursieve spel. Deze strategie bereikt zelfs meer dan de verwachte waarde $1-\varepsilon$, want $W = 1-\varepsilon$ en dus $T(W) = 1/(1+\varepsilon)$. Er bestaat geen optimale strategie voor speler 1.

bij voorbeeld 4.1.4

Beschouw de volgende spelen:

$$G_1(\overline{W}) = \begin{bmatrix} w_3 & 0 \\ 0 & w_2 \end{bmatrix}, \; G_2(\overline{W}) = \begin{bmatrix} 1 & w_1 \\ 0 & w_3 \end{bmatrix} \text{ en } G_3(\overline{W}) = \begin{bmatrix} 1 & w_2 \\ w_1 & 1 \end{bmatrix}.$$

Er blijkt een uniek dekpunt te bestaan voor de waarde-transformatie T. Dit is tevens de kritieke vektor $\overline{V} = (1-\sqrt{2}/2, 1/2, \sqrt{2}/2)$. Omdat $\overline{V} > 0$ is $\overline{V}$ bereikbaar voor speler 2. Het invullen van $\overline{V}$ in de matrixspelen en het oplossen daarvan levert een optimale strategie voor speler 2. De volledige oplossing is te vinden in LUCE & RAIFFA [35], blz. 469.

bij voorbeeld 4.1.5

Beschouw het spel:

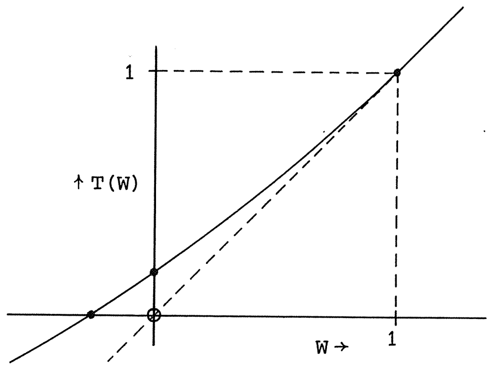
$$G(W) = \begin{bmatrix} W & W & W \\ W & W & 1 \\ W & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}.$$

Als $W \geq 1$ dan domineert rij 1 de rijen 2,3 en 4 dus $T(W) = \text{val } G(W) = W$. Als $W < 1$ dan domineert rij 2 rij 1; de waarde van $G(W)$ kunnen wij bepalen door het oplossen van het volgende l.p. probleem. Maximaliseer λ onder de voorwaarden

$$\begin{aligned} Wx_1 + Wx_2 + x_3 &\geq \lambda \\ Wx_1 + x_2 - x_3 &\geq \lambda \\ x_1 - x_2 - x_3 &\geq \lambda \\ x_1 + x_2 + x_3 &= 1 \\ x_1 \,,\; x_2 \,,\; x_3 &\geq 0 \end{aligned} \quad .$$

De maximale waarde van λ is

$$(-W^2+4W+1)/(W^2-4W+7)$$



$[1,\infty)$ is de verzameling dekpunten van de waarde-transformatie T. De kritieke vektor is $V = 1$ en is dus bereikbaar voor speler 2; $1 - \varepsilon$ met $\varepsilon > 0$ is bereikbaar voor speler 1. Dus $C_1 = (-\infty,1)$ en $C_2 = [1,\infty)$. Speelt speler 2 optimaal voor het matrixspel:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}$$

dan speelt speler 2 optimaal in het recursieve spel. Dus iedere strategie is optimaal voor speler 2. Speelt speler 1 optimaal voor het matrixspel:

$$\begin{bmatrix} 1-\varepsilon & 1-\varepsilon & 1-\varepsilon \\ 1-\varepsilon & 1-\varepsilon & 1 \\ 1-\varepsilon & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}$$

dan speelt speler 1 ε-optimaal in het recursieve spel. Als $\delta > 0$ klein genoeg is dan is een strategie van de vorm $(0,1-\delta-\delta^2,\delta,\delta^2)$ ε-optimaal voor speler 1. Bij kleine ε zal Blotto vrijwel met kans 1 aktie 2 spelen; de verwachte speelduur kan dan zeer groot worden.

HOOFDSTUK V

NIET STOPPENDE STOCHASTISCHE SPELEN

5.1. Inleiding

Wij zullen ons in dit hoofdstuk beperken tot twee-persoons nulsomspelen en zoals in het model van Shapley tot een eindige toestandsruimte en tot een eindig aantal akties voor beide dpelers in alle toestanden. We gebruiken dan ook dezelfde notaties: toestanden $1,\dots,N$; akties voor speler 1 in toestand k: $1,\dots,m_k$; akties voor speler 2 in toestand k: $1,\dots,n_k$.
In het model van Shapley geldt:

$$(5.1) \qquad \sum_{\ell=1}^{N} p(\ell|k,i,j) < 1 \quad \text{voor alle } i,j \text{ en } k.$$

In dit model nemen wij aan:

$$(5.2) \qquad \sum_{\ell=1}^{N} p(\ell|k,i,j) \leq 1 \quad \text{voor alle } i,j \text{ en } k.$$

Echter het is geen beperking in (5.2) aan te nemen dat:

$$(5.3) \qquad \sum_{\ell=1}^{N} p(\ell|k,i,j) = 1 \quad \text{voor alle } i,j, \text{ en } k.$$

Dit kan altijd bereikt worden door een extra absorberende toestand $N + 1$ in te voeren met slechts één mogelijke aktie voor beide spelers en met opbrengst 0 zodat:

$$(5.4) \qquad p(N+1|k,i,j) = 1 - \sum_{\ell=1}^{N} p(\ell|k,i,j).$$

VOORBEELD 5.1.1. Beschouw het volgende eenpersoons stochastische spel (Markov beslissings proces)

$$\Gamma : \begin{bmatrix} 2 + \Gamma \\ 1 + \Gamma \\ 3 + \frac{1}{2}\Gamma \end{bmatrix}$$

door het toevoegen van een toestand zoals beschreven kunnen wij dit spel schrijven als:

$$\Gamma_1 : \begin{bmatrix} 2 + \Gamma_1 \\ 1 + \Gamma_1 \\ 3 + \frac{1}{2}\Gamma_1 + \frac{1}{2}\Gamma_2 \end{bmatrix} \qquad \Gamma_2 : [\Gamma_2].$$

Wordt in dit spel altijd met kans 1 aktie 1 gekozen dan is de totaal verwachter opbrengst ∞. Wordt altijd met kans 1 aktie 2 gekozen dan is de totaal verwachte opbrengst ook ∞. Het kriterium van de totaal verwachte opbrengst maakt dus geen onderscheid tussen deze twee strategieën. Echter de gemiddelde opbrengst per tijdéénheid is in het eerste geval 2 in het tweede geval 1.

Laat $V_1^t(k,\pi_1,\pi_2)$ de verwachte opbrengst op tijdstip t voor speler 1. als k de begintoestand is en strategieën π_1 en π_2 worden gespeeld. Dan is de gemiddelde opbrengst van speler 1 gegeven k, π_1 en π_2:

$$(5.5) \qquad Y_1(k,\pi_1,\pi_2) = \limsup_{t\to\infty} \frac{1}{t+1} \sum_{\tau=0}^{t} V_1^\tau(k,\pi_1,\pi_2).$$

<u>VOORBEELD 5.1.2.</u> ("the big match"). Iedere dag kiest speler 2 aktie 1 of 2, speler 1 probeert de keuze van 2 te voorspellen. Als hij goed raadt krijgt hij een punt. Dit herhaalt zich zolang speler 1 aktie 1 voorspelt. Als speler 1 goed raadt op de dag dat hij aktie 2 voorspelt dan krijgt hij iedere volgende dag een punt; als hij op deze dag fout raad dan krijgt hij iedere volgende dag niets. Dit is het volgende spel met toestand 1 als starttoestand.

$$\Gamma_1 : \begin{bmatrix} 1 + \Gamma_1 & \Gamma_1 \\ \Gamma_2 & 1 + \Gamma_3 \end{bmatrix}$$

$$\Gamma_2 : [\Gamma_2] \qquad \Gamma_3 : [1 + \Gamma_3] .$$

Voor speler 1 wordt een stationaire strategie volledig bepaald door $\pi_1^1(1)$, de kans waarmee hij in toestand 1 aktie 1 speelt; voor speler 2 evenzo door $\pi_2^1(1)$. Identificeren we de kansen $\pi_1^1(1)$ en $\pi_2^1(1)$ met de bijbehorende stationaire strategieën dan kunnen we voor de gemiddelde opbrengst schrijven:

(5.6) $$\begin{aligned} &Y_1(\pi_1^1(1),\pi_2^1(1)) = 1 - \pi_1^1(1) \quad \text{als } \pi_1^1(1) \neq 1 \text{ en} \\ &Y_1(1,\pi_2^1(1)) = \pi_2^1(1). \end{aligned}$$

Dit spel heeft geen evenwichtspunt van stationaire strategieën want:

$$\max_{\pi_1^1(1)} \min_{\pi_2^1(1)} Y_1(\pi_1^1(1),\pi_2^1(1)) =$$

$$\max_{\pi_1^1(1)} Y_1(\pi_1^1(1), \text{ als } \pi_1^1(1) = 1 \text{ dan } 0, \text{ anders } 1) = 0$$

en

$$\min_{\pi_2^1(1)} \max_{\pi_1^1(1)} Y_1(\pi_1^1(1),\pi_2^1(1)) =$$

$$\min_{\pi_2^1(1)} Y_1(\text{als } \pi_2^1(1) \geq \tfrac{1}{2} \text{ dan } 1, \text{ anders } \neq 1, \pi_2^1(1)) = \tfrac{1}{2}\,.$$

Dit voorbeeld is uitgewerkt in BLACKWELL & FERGUSON [8]; zij laten zien dat de waarde van het spel $\frac{1}{2}$ is en geven ε-optimale strategieën voor speler 1. Een optimale strategie voor speler 1 kan niet worden gevonden. Laat n.l. π_1 een willekeurige strategie zijn voor speler 1. Wij onderscheiden voor π_1 de volgende twee mogelijkheden:

(i) Er is geen tijdstip waarop, als speler 2 op de tijdstippen 1 t/m t-1 aktie 2 heeft gekozen voor 't eerst met positieve kans aktie 2 wordt gespeeld.

(ii) Er bestaat wel zo'n tijdstip.

Geldt (i) en speelt speler 2 altijd met kans 1 aktie 2 dan is de gemiddelde opbrengst 0. Geldt (ii) en speelt speler 2 t-1 maal aktie 2 éénmaal aktie 1 en daarna aktie 1 met kans $\frac{1}{2}$ dan is de gemiddelde opbrengst $(1-\varepsilon).\frac{1}{2} + \varepsilon.0 = \frac{1}{2} - \frac{1}{2}\varepsilon$, waarbij ε de positieve kans is waarmee speler 1 op tijdstip t aktie 2 speelt. In beide gevalllen geldt dat de gemiddelde opbrengst kleiner is dan $\frac{1}{2}$. De strategie π_1 is dus niet optimaal. Dit voorbeeld is op een niet principiële wijziging na afkomstig van GILETTE [18].

Voorts bevat het artikel van GILETTE [18] de volgende twee uitspraken:

I Stochastische spelen met volledige informatie (d.w.z. in iedere toestand heeft slechts één der spelers meer dan één aktie ter beschikking) bezitten een evenwichtspunt van zuivere stationaire strategieën.

II Stochastische spelen waarbij voor ieder paar zuivere stationaire strategieën geldt dat de matrix van overgangswaarschijnlijkheden één aperiodieke kernfuik vormt bezitten een evenwichtspunt van stationaire strategieën.

Echter voor het bewijzen van beide uitspraken gebruikt GILETTE een onjuiste uitbreiding van een stelling van HARDY & LITTLEWOOD. De onjuistheid werd aangetoond door LIGGETT & LIPPMAN [34], en zij gaven een nieuw bewijs voor stochastische spelen met volledige informatie. De argumentatie in dit bewijs is echter niet volledig; een eenvoudiger bewijs voor een iets sterkere stelling werd gegeven door FEDERGRUEN [14], blz. 20. Ook de uitspraak over het tweede type stochastische spelen blijkt juist; in een artikel van HOFFMAN & KARP [22] wordt op andere wijze een sterker resultaat afgeleid. Zij veronderstellen een irreducibel stochastisch spel (d.w.z. voor ieder paar zuivere stationaire strategieën geldt dat de matrix van overgangswaarschijnlijkheden één kernfuik vormt). In de volgende twee paragrafen zullen wij hun artikel behandelen.

5.2. Het Markov beslissingsproces

In deze paragraaf worden enige resultaten op het gebied van het éénpersoons stochastische spel, meer bekend onder de naam Markov beslissingsproces, samengevat. Ook in deze paragraaf veronderstellen wij dat irreducibiliteit geldt en omdat er slechts één speler is noteren we $g(k,i)$, $p(\ell|k,i)$ enz. Het bewijs van de nu volgende stelling kan worden gevonden in DERMAN [11].

STELLING 5.2.1. *Voor ieder Markov beslissingsproces bestaat er een optimale stationaire strategie.*

Is π een stationaire strategie dan kunnen wij de gemiddelde opbrengst als volgt schrijven:

$$(5.1) \qquad Y(\pi) = \sum_{k=1}^{N} \phi_k(\pi) \sum_{i=1}^{m_k} g(k,i)\, \pi^i(k),$$

waarbij $\phi_1(\pi),\ldots,\phi_N(\pi)$ de unieke positieve invariante kansverdeling is bij strategie π. Vanwege de irreducibiliteit is de waarde $Y(\pi)$ voor iedere begintoestand hetzelfde. Voor de overgangswaarschijnlijkheden behorende bij een stationaire strategie π, $p_{k\ell}(\pi)$ geldt:

$$(5.2) \qquad p_{k\ell}(\pi) = \sum_{i=1}^{m_k} \pi^i(k)\ p(\ell|k,i).$$

De bij π behorende invariante kansverdeling kan men vinden door het oplossen van het stelsel

$$(5.3) \qquad \phi_\ell(\pi) = \sum_{k=1}^{N} p_{k\ell}(\pi)\ \phi_k(\pi) \qquad \ell = 1,\ldots,N.$$

$$(5.4) \qquad \sum_{k=1}^{N} \phi_k(\pi) = 1.$$

Vanwege de irreducibiliteit geldt:

$$(5.5) \qquad \phi_k(\pi) > 0 \quad k = 1,\ldots,N.$$

Wij schrijven nu:

$$(5.6) \qquad x^i(k) = \phi_k(\pi)\ \pi^i(k)$$

$x^i(k)$ is de fraktie van het aantal tijdstippen waarop onder strategie π toestand k optreedt en aktie i wordt gekozen.

$$(5.7) \qquad Y(\pi) = \sum_{k=1}^{N} \sum_{i=1}^{m_k} x^i(k)\ g(k,i).$$

Vanwege $\pi^i(k) \geq 0$ en (5.5) geldt:

$$(5.8) \qquad x^i(k) \geq 0 \quad k = 1,\ldots,N \quad i = 1,\ldots,m_k.$$

Vanwege $\sum_{i=1}^{m_k} \pi^i(k) = 1$ geldt:

$$(5.9) \qquad \sum_{i=1}^{m_k} x^i(k) = \phi_k(\pi).$$

Uit (5.4) en (5.9) volgt:

$$(5.10) \qquad \sum_{k=1}^{N} \sum_{i=1}^{m_k} x^i(k) = 1.$$

Uit (5.2), (5.3), (5.6) en (5.9) volgt:

$$(5.11)\qquad \sum_{i=1}^{m_\ell} x^i(\ell) - \sum_{k=1}^{N}\sum_{i=1}^{m_k} x^i(k)\ p(\ell|k,i) = 0 \qquad \ell = 1,\dots,N.$$

Hierbij is $\sum_{i=1}^{m_\ell} x^i(\ell)$ de fractie van het aantal tijdstippen waarop het systeem zich in toestand ℓ bevindt. Bij iedere stationaire strategie π vinden we uit (5.2), (5.3), (5.4), (5.5) en (5.6) de $x_i(k)$ $k = 1,\dots,N$; $i = 1,\dots,m_k$ die voldoen aan (5.8), (5.10) en (5.11). Andersom behoort bij iedere verzameling $x_i(k)$'s die voldoet aan (5.8), (5.10) en (5.11) een stationaire strategie die bekend kan worden uit (5.6) en (5.9). Het vinden van een optimale stationaire strategie kan dus worden geschreven als ℓ.p. probleem: maximaliseer (5.7) onder de voorwaarden (5.8), (5.10) en (5.11). Dit ℓ.p. probleem zullen wij als probleem I aanduiden; het hiermee duale probleem als probleem II.

De vergelijkingen (5.11) zijn lineair afhankelijk, want optellen levert:

$$\sum_{\ell=1}^{N}\Big[\sum_{i=1}^{m_\ell} x^i(\ell) - \sum_{k=1}^{N}\sum_{i=1}^{m_k} x^i(k)p(\ell|k,i)\Big] =$$

$$\sum_{\ell=1}^{N}\sum_{i=1}^{m_\ell} x^i(\ell) - \sum_{k=1}^{N}\sum_{i=1}^{m_k} x^i(k) \sum_{\ell=1}^{N} p(\ell|k,i) = 0.$$

De coëfficiënten van de vergelijkingen (5.10) en (5.11) vormen dus een matrix met hoogstens rang N. Een toegelaten basisoplossing voor probleem I heeft hoogstens N variabelen ongelijk nul. Uit (5.5) en (5.9) volgt: Voor een toegelaten oplossing geldt dat er voor iedere $k = 1,\dots,N$ minstens één $x^i(k) > 0$ is. Dus voor iedere toegelaten basisoplossing is precies één der $x^i(k) > 0$. Dit betekent volgens (5.6) en (5.9) dat de bijbehorende $\pi^i(k) = 1$. Er is dus een zuivere stationaire strategie.

Beschouw nu het duale probleem (probleem II): minimaliseer q onder de voorwaarden:

$$(5.12)\qquad q + v_k \geq g(k,i) + \sum_{\ell=1}^{N} p(\ell|k,i)\ v_\ell \qquad k = 1,\dots,N \quad i = 1,\dots,m_k.$$

Als $(q,v_1,\dots,v_N)$ een toegelaten oplossing is en c een willekeurige konstante dan is ook $(q,v_1+c,\dots,v_N+c)$ een toegelaten oplossing.

LEMMA 5.2.2. *Voor iedere optimale oplossing* $(q,v_1,\dots,v_N)$ *van probleem* II *geldt:*

$$(5.13) \qquad q + v_k = \max_{i=1,\dots,m_k} \; g(k,i) + \sum_{\ell=1}^{N} p(\ell|k,i)v_\ell \qquad k = 1,\dots,N.$$

BEWIJS: Bij iedere $k \in \{1,\dots,N\}$ bestaat er een i_k zodanig dat $X^{i_k}(k) > 0$. De corresponderende verschilvariabele in probleem II is gelijk aan nul, dus

$$q + v_k = g(k,i_k) + \sum_{\ell=1}^{N} p(\ell|k,i_k)v_\ell. \quad \square$$

STELLING 5.2.3. *Als* $(q,v_1,\dots,v_N)$ *en* $(q',v_1',\dots,v_N')$ *beiden voldoen aan* (5.13) *dan geldt* $q = q'$ *en bestaat er een konstante* c *zodanig dat* $v_k - v_k' = c$ voor $k = 1,\dots,N$.

BEWIJS: Bij iedere $k \in \{1,\dots,N\}$ bestaat er een i_k zodanig dat:

$$q + v_k = g(k,i_k) + \sum_{\ell=1}^{N} p(\ell|k,i_k)v_\ell$$

en bovendien geldt:

$$q' + v_k' \geq g(k,i_k) + \sum_{\ell=1}^{N} p(\ell|k,i_k)v_\ell'.$$

Hieruit volgt:

$$(5.14) \qquad (q'-q) + (v_k'-v_k) \geq \sum_{\ell=1}^{N} p(\ell|k,i_k)(v_\ell'-v_\ell).$$

Stel $q' - q < 0$ dan volgt:

$$v_k' - v_k > \sum_{\ell=1}^{N} p(\ell|k,i_k)(v_\ell'-v_\ell) \geq \min_\ell (v_\ell'-v_\ell), \qquad k = 1,\dots,N$$

hetgeen onmogelijk is. Verwisseling van $(q,\bar{v})$ en $(q',\bar{v}')$ levert de onjuistheid van $q - q' < 0$. Dus geldt $q' = q$; invullen in (5.14) levert:

$$\bar{v}' - \bar{v} \geq P(\bar{v}'-\bar{v}), \qquad \text{waarbij}$$

P een matrix met op plaats (k,ℓ) $p(\ell|k,i_k)$, maar omdat P irreducibel is moet gelden $v_k' - v_k = c$ voor $k = 1,\dots,N$. $\square$

Als wij een oplossing van (5.13) hebben bepaald, waarvoor b.v. het eindige iteratieve algorithme van Howard kan worden gebruikt dan is probleem II opgelost.

5.3. Het twee-persoons nulsomspel

Laat speler 2 een stationaire strategie π_2 spelen, dan wordt speler 1 geconfronteerd met een Markov beslissingsproces waarbij:

$$g(k,i) = \sum_{j=1}^{n_k} \pi_2^{\,j}(k)\ g(k,i,j) \text{ en}$$

$$(\ell|k,i) = \sum_{j=1}^{n_k} \pi_2^{\,j}(k)\ p(\ell|k,i,j).$$

Wij kunnen het hierbij behorende probleem II oplossen; de minimale waarde van q noteren wij als $q(\pi_2)$.

De stationaire strategieën van speler 2 vormen een kompakte verzameling in de $\sum_{k=1}^{N} n_k$ dimensionale Euclidische ruimte. Met gebruikmaking van de irreducibiliteit bewijzen HOFFMAN & KARP dat $q(\pi_2)$ een continue functie is van π_2. Er bestaat dus een stationaire strategie $\pi_2^{\,*}$ voor speler 2 zodanig dat $q(\pi_2^{\,*}) \le q(\pi_2)$ voor alle (ook niet stationaire) strategieën π_2 voor speler 2. Wij zullen aantonen dat $\pi_2^{\,*}$ een optimale strategie is voor speler 2 in het stochastische spel.

Beschouw het Markov beslissingsproces waarmee speler 1 wordt geconfronteerd als speler 2 de strategie $\pi_2^{\,*}$ van te voren zou aankondigen.

$$g(k,i) = \sum_{j=1}^{n_k} g(k,i,j)\ \pi_2^{\,*j}(k) \text{ en} \tag{5.15}$$

$$p(\ell|k,i) = \sum_{j=1}^{n_k} p(\ell|k,i,j)\ \pi_2^{\,*j}(k). \tag{5.16}$$

Laat $q^*, v_1^{\,*}, \ldots, v_N^{\,*}$ een optimale oplossing van het bijbehorende probleem II zijn. Wij noteren $G_k(v_1,\ldots,v_N)$ voor het matrixspel met als element op plaats (i,j,):

$$g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)v_\ell.$$

<u>STELLING 5.3.1.</u>

$$q^* + v_k^{\,*} = \text{val } G_k(v_1^{\,*},\ldots,v_N^{\,*}) \quad \textit{voor } k = 1,\ldots,N.$$

<u>BEWIJS</u>: Uit lemma 5.2.2 volgt:

$$(5.17) \qquad q^* + v_k^* = \max_{i=1,\dots,m_k} [(g(k,i) + \sum_{\ell=1}^{N} p(\ell|k,i)v_\ell^*] \text{ voor } k = 1,\dots,N.$$

Uit (5.15) en (5.16) volgt:

$$q^* + v_k^* = \max_{i=1,\dots,m_k} [\sum_{j=1}^{n_k} (g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)v_\ell^*) \; \pi_2^{*j}(k)]$$

$$\text{voor } k = 1,\dots,N.$$

Wij moeten dus aantonen dat $(\pi_2^{*1}(k),\dots,\pi_2^{*n_k}(k))$ een optimale strategie is voor speler 2 in het matrixspel $G_k(v_1^*,\dots,v_N^*)$ voor $k = 1,\dots,N$. Stel er is een toestand h, waarvoor speler 2 een betere strategie $(\pi_2^{o1}(h),\dots,\pi_2^{on_h}(h))$ heeft, dan geldt:

$$(5.18) \qquad q^* + v_h^* > \max_{i=1,\dots,m_h} \sum_{j=1}^{n_h} (g(h,i,j) + \sum_{\ell=1}^{N} p(\ell|h,i,j)v_\ell^*)\pi_2^{oi}(h) =$$

$$\max_{i=1,\dots,m_h} [(g^o(h,i) + \sum_{\ell=1}^{N} p^o(\ell|h,i)v_\ell^*],$$

waarbij

$$(5.19) \qquad g^o(h,i) = \sum_{j=1}^{n_h} g(h,i,j) \; \pi_2^{oj}(h)$$

en

$$(5.20) \qquad p^o(\ell|h,i) = \sum_{j=1}^{n_h} p(\ell|h,i,j) \; \pi_2^{oj}(h).$$

Laat π_2^0 de stationaire strategie voor speler 2 die in alle toestanden $k \neq h$ gelijk is aan π_2^* en in toestand h: $(\pi_2^{o1}(h),\dots,\pi_2^{on_h}(h))$. Het bij π_2^o behorende probleem II verkrijgen we door $g(k,i)$ en $p(\ell|k,i)$ voor $k \neq h$ te berekenen volgens (5.15) en (5.16) en $g(h,i)$ en $p(\ell|h,i)$ volgens (5.19) en (5.20). Uit (5.17) en (5.18) volgt dat $(q^*,v_1^*,\dots,v_n^*)$ een toegelaten oplossing is van probleem II, maar omdat niet aan het stelsel vergelijkingen in lemma 5.2.2 is voldaan, is het geen optimale oplossing. Dus $q(\pi_2^0) < q^* = q(\pi_2^*)$, tegenspraak. □

<u>OPMERKING 5.3.2.</u> Als speler 2 stationaire strategie π_2^* speelt dan verzekert hij zich van een gemiddelde opbrengst $-q(\pi_2^*) = -q^*$ tegen iedere (op grond van stelling 5.2.1 ook niet stationaire) strategie van speler 1.

STELLING 5.3.3. *Als* $q + v_k = \mathrm{val}\ G_k(v_1,\dots,v_N)$ *en* $q^* + v_k^* = \mathrm{val}\ G_k(v_1^*,\dots,v_N^*)$ *voor* $k = 1,\dots,N$ *dan geldt* $q = q^*$ *en er bestaat een getal* c *zodanig dat* $v_k - v_k^* = c$ *voor* $k = 1,\dots,N$.

BEWIJS: Laat $(\pi_2^1(k),\dots,\pi_2^{n_k}(k))$ een optimale strategie voor speler 2 in in het matrixspel $G_k(v_1,\dots,v_N)$ en laat $(\pi_1^1(k),\dots,\pi_1^{m_k}(k))$ een optimale strategie voor speler 1 in het matrixspel $G_k(v_1^*,\dots,v_N^*)$, dan geldt:

$$\mathrm{val}\ G_k(v_1,\dots,v_N) = q + v_k \geq$$
$$\sum_{i=1}^{m_k}\sum_{j=1}^{n_k}(g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)v_\ell)\ \pi_1^i(k)\ \pi_2^j(k)$$

en

$$\mathrm{val}\ G_k(v_1^*,\dots,v_N^*) = q^* + v_k^* \leq$$
$$\sum_{i=1}^{m_k}\sum_{j=1}^{n_k}(g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)v_\ell^*)\ \pi_1^i(k)\ \pi_2^j(k).$$

Hieruit volgt:

$$(q-q^*) + (v_k-v_k^*) \geq$$
$$\sum_{i=1}^{m_k}\sum_{j=1}^{n_k}\sum_{\ell=1}^{n} p(\ell|k,i,j)(v_\ell-v_\ell^*)\ \pi_1^i(k)\ \pi_2^j(k) =$$
$$\sum_{\ell=1}^{N}(\sum_{i=1}^{m_k}\sum_{j=1}^{n_k} p(\ell|k,i,j)\pi_1^i(k)\pi_2^j(k))\ (v_\ell-v_\ell^*)\ \text{voor } k = 1,\dots,N.$$

Hierna verloopt het bewijs hetzelfde als het bewijs van stelling 5.2.3. □

Als wij de rol van beide spelers verwisselen vinden wij een stationaire strategie π_1^* voor speler 1 die hem verzekert van een gemiddelde opbrengst $q(\pi_1^*)$ tegen iedere, ook niet stationaire, strategie van speler 2. Voor de oplossing van het bij π_1^* behorende probleem II geldt eveneens stelling 5.3.1. Volgens stelling 5.3.3 is de oplossing van $q + v_k = \mathrm{val}\ G_k(v_1,\dots,v_N)$ op een konstante voor de $v_1,\dots,v_N$ na uniek. Dus $q(\pi_1^*) = q^*$ en hieruit volgt dat π_1^* en π_2^* optimale stationaire strategieën in het stochastische spel zijn.

HOOFDTUK VI

SUCCESSIEVE APPROXIMATIE IN TWEE-PERSOONS NULSOMSPELEN

6.1. Inleiding

In dit hoofstuk worden een aantal successieve approximatiemethoden gegeven, met behulp waarvan eindige twee-persoons nulsomspelen opgelost kunnen worden. Oplossen wil zeggen het vinden van boven- en ondergrenzen voor de waarde van het spel en ε-optimale strategieën voor de beide spelers.

In sektie 6.2 worden twee methoden voor het gemiddelde opbrengsten kriterium gepresenteerd. De eerst is de successieve approximatiemethode van HOFFMAN & KARP [22] en de tweede is die van POLLATSCHEK & AVI-ITZHAK [49].

In sektie 6.3. komt een verzameling samenhangende algoritmen van VAN DER WAL [70] voor spelen onder het verdisconteerde opbrengsten kriterium aan de orde. Twee algoritmen uit deze verzamenling waren reeds eerder bekend, namelijk de successieve approximatiemethode van SHAPLEY [57], d.w.z. het herhaald uitvoeren van de waarde-transformatie en de verdisconteerde versie van de successieve approximatiemethode van HOFFMAN & KARP [22], (uitgewerkt in RAO,CHANDRASEKARAN & NAIR [51]). Tevens wordt in sektie 6.3. de methode van POLLATSCHEK & AVI-ITZHAK [49] voor het verdisconteerde model gegeven.

6.2. Algoritmen voor het oplossen van Niet-stoppende spelen

Het algoritme van HOFFMAN & KARP [22].

Evenals in hoofstuk 5 zullen wij veronderstellen dat voor ieder paar zuivere stationaire strategieën de matrix van overgangswaarschijnlijkheden uit één kernfuik bestaat. Uit stelling 5.3.3 blijkt dat een oplossing q^*, $v_1^*,\dots,v_N^*$ van de betrekkingen:

$$(6.1) \qquad q + v_k = \text{val } G_k(v_1,\dots,v_N) \qquad k = 1,\dots,N$$

op een konstante na voor de $v_1,\dots,v_N$ uniek is. Het gevolg hiervan is, dat q^* de waarde van het spel is bij het gemiddelde opbrengstkriterium. Uit het bewijs van stelling 5.3.1 blijkt dat het op ieder tijdstip in toestand k spelen van een optimale strategie voor het matrixspel $G_k(v_1^*,\dots,v_N^*)$ een optimale stationaire strategie is voor het stochastische spel. Het oplossen van het stochastische spel is hiermee teruggebracht tot het oplossen van de betrekkingen (6.1). Het algoritme is als volgt:

stap 1: kies een stationaire strategie voor speler 2: $(\pi_2)^0$; $r = 0$.

stap 2: bepaal de oplossing $q^r, v_k^r, \dots, v_N^r$ van:

$$(6.2)\begin{cases} q^r + v_k^r = \max\limits_{1\le i\le m_k} \left[\sum\limits_{j=1}^{n_k} \left(g(k,i,j) + \sum\limits_{\ell=1}^{N} p(\ell|k,i,j)v^r\right)(\pi_2^j(k))^r \right] \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{voor } k = 1,\dots,N. \\ v_N = 0 \end{cases}$$

stap 3: construeer $(\pi_2)^{r+1}$, een stationaire strategie voor speler 2 waarvoor geldt: $(\pi_2(k))^{r+1}$ is een optimale strategie in het matrixspel $G_k(v_1^r,\dots,v_N^r)$ voor $k = 1,\dots,N$; $r = r + 1$ en terug naar stap 2.

Voor het oplossen van stap 2 kan Howard's algoritme voor het oplossen van Markov beslissingsprocessen worden gebruikt. De convergentie van het algoritme staat vermeld in de volgende stelling.

STELLING 6.2.1. *De rij* $q^0, q^1, \dots$ *en de rijen* $v_k^0, v_k^1, \dots$ *voor* $k = 1,\dots,N$ *convergeren. Noteer de limieten respectievelijk als* $q^*, v_1^*, \dots, v_N^*$ *dan geldt:*

$$q^* + v_k^* = \text{val } G_k(v_1^*,\dots,v_N^*) \qquad k = 1,\dots,N.$$

BEWIJS: Uit (6.2) en de constructie van $(\pi_2)^{r+1}$ in stap 3 volgt dat $q^r, v_1^r, \dots, v_N^r$ een toegelaten oplossing is van het probleem II, zoals geconstrueerd in hoofdstuk 5, behorende bij de strategie $(\pi_2)^{r+1}$. De optimale oplossing is $q^{r+1}, v_1^{r+1}, \dots, v_N^{r+1}$, dus $q^{r+1} \le q^r$ voor alle r. De rij $q^0, q^1, \dots$ is dus niet stijgend.

Laat π_1 een zuivere stationaire strategie voor speler 1 en π_2 een stationaire strategie voor speler 2 dan heeft het stelsel:

$$(6.3)\qquad \begin{cases} q + v_k = \sum_{j=1}^{n_k} (g(k,\pi_1(k),j) + \sum_{\ell=1}^{N} p(\ell|k,\pi_1(k),j)v_\ell)\ \pi_2^{\,j}(k) \\ \qquad\qquad\qquad\qquad\qquad\qquad k = 1,\dots,N \\ v_N = 0 \end{cases}$$

vanwege de eis dat de matrix van overgangswaarschijnlijkheden irreducibel is, een unieke oplossing die continu afhangt van π_2.

De stationaire strategieën vormen een kompakte verzameling, dus bij vaste π_1 is de verzameling mogelijke oplossingen van stelsel (6.3) die men verkrijgt door π_2 te variëren een kompakte verzameling. Er zijn eindig veel zuivere stationaire strategieën π_1. De verzameling oplossingen van stelsels van de vorm (6.3) is dus kompakt. De vektoren $(q^r,v_1^r,\dots,v_N^r)$ $r = 0,1,2,\dots$ zijn allen oplossingen van stelsels van deze vorm en liggen dus in een kompakte verzameling. De rij van deze vektoren heeft dus een convergente deelrij. Laat deze rij convergeren naar $q^+,v_1^+,\dots,v_N^+$. De rij van strategieën corresponderende met deze deelrij: $((\pi_2(1))^{r+1},\dots,(\pi_2(N))^{r+1})$ heeft een convergente deelrij, want de stationaire strategieën vormen een kompakte verzameling. Wij noteren deze limiet als $(\pi_2)^+ = ((\pi_2)^+,\dots,(\pi_2(N))^+)$. $(\pi_2)^+$ is een optimale strategie van het matrixspel $G_k(v_1^+,\dots,v_N^+)$, want de uitbetaling van een matrixspel is een continue functie van de gebruikte strategieën en de coëfficiënten in de matrix.

Uit (6.3) en de constructie van $(\pi_2)^r$ in stap 3 volgt:

$$q^r + v_k^r \geq \text{val } G_k(v_1^r,\dots,v_N^r) =$$

$$\max_{i=1,\dots,m_k} \sum_{j=1}^{n_k} (g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)v_\ell^r)(\pi_2^{\,j}(k))^r$$

$$\text{voor } k = 1,\dots,N.$$

De waarde van een matrixspel is een continue functie van de coëfficiënten en dus geldt:

$$(6.4) \qquad q^+ + v_k^+ \geq \text{val } G_k(v_1^+,\ldots,v_N^+) =$$

$$\max_{i=1,\ldots,m_k} \sum_{j=1}^{n_k} (g(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)v_\ell^+)(\pi_2^{\ j}(k))^+$$

$$\text{voor } k = 1,\ldots,N.$$

Stel voor zekere k geldt:

$$q^+ + v_k^+ > \text{val } G_k(v_1^+,\ldots,v_N^+).$$

Uit (6.4) volgt dat $q^+,v_1^+,\ldots,v_N^+$ een toegelaten oplossing is van het probleem II dat behoort bij de strategie $(\pi_2)^+$ van speler 2. Uit lemma 5.2.2 volgt dat deze oplossing geen optimale oplossing is. Dus $q((\pi_2)^+) < q^+$ hetgeen op grond van de constatering dat de rij $q^0,q^1,\ldots$ een dalende rij is onmogelijk is.

Uit deze tegenspraak kunnen we concluderen:

$$q^+ + v_k^+ = \text{val } G_k(v_1^+,\ldots,v_N^+) \qquad k = 1,\ldots,N.$$

Omdat de oplossing van (6.3) uniek is moet iedere convergente deelrij dezelfde limiet hebben; de rij is dus convergent. □

Het algoritme van POLLATSCHECK & AVI-ITZHAK [49].

Kies $v_N = 0$ en schrijf het stelsel (6.1) als:

$$\text{val } G_1(v_1,\ldots,v_{N-1},0) - v_1 - q = 0$$

$$\text{val } G_{N-1}(v_1,\ldots,v_{N-1},0) - v_{N-1} - q = 0$$

$$\text{val } G_N(v_1,\ldots,v_{N-1},0) - q = 0.$$

Dit stelsel van N vergelijkingen heeft $v_1,\ldots,v_{N-1},q$ als onbekenden. Voor de oplossing van dit stelsel wordt de Newton-Raphson methode gebruikt. Noteer het stelsel vergelijkingen kortweg als:

$$f_1(X_1,\dots,X_N) = 0$$

$$f_N(X_1,\dots,X_N) = 0$$

dan is de methode als volgt:

stap 1: kies $X_1^0,\dots,X_N^0$ willekeurig
$r = 0$

stap 2: bereken $X_1^{r+1},\dots,X_N^{r+1}$ uit:

$$\begin{pmatrix} X_1^{r+1} \\ \vdots \\ X_N^{r+1} \end{pmatrix} = \left[\left(\frac{\partial f_k}{\partial X_\ell}\right)_{X_1^r,\dots,X_N^r}\right]^{-1} \begin{pmatrix} f_1(X_1^r,\dots,X_N^r) \\ \vdots \\ f_N(X_1^r,\dots,X_N^r) \end{pmatrix} - \begin{pmatrix} X_1^r \\ \vdots \\ X_N^r \end{pmatrix}$$

$r = r + 1$ en herhaal stap 2.

POLLATSCHECK & AVI-ITZHAK bewijzen:

$$\frac{\partial G_k(v_1,\dots,v_N)}{\partial v_\ell} = \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} \pi_1^i(k)\ \pi_2^j(k)\ p(\ell|k,i,j)$$

voor $k = 1,\dots,N$,

waarbij $\pi_1(k)$ en $\pi_2(k)$ optimale strategieën zijn voor het matrixspel $G_k(v_1,\dots,v_N)$, voor zover deze partiële afgeleiden bestaan. Er zijn slechts eindig veel punten waar deze afgeleiden niet bestaan. De afgeleiden die nodig zijn voor het uitvoeren van de Newton-Raphson methode kunnen met behulp van de hierboven gegeven partiële afgeleiden worden berekend. Een bewijs voor de convergentie van deze methode wordt niet gegeven.

Het algoritme van RIOS & YAÑEZ [52]

Laat val $\Gamma^t(k,0)$ de waarde van het t-staps stochastische spel startend in toestand k met nabetaling 0 zijn. Noteer:

$$q^t = \frac{1}{N} \sum_{k=1}^{N} (\text{val } \Gamma^{t+1}(k,0) - \text{val } \Gamma^t(k,0)).$$

Onder de voorwaarde dat bij alle paren akties de matrices van overgangswaarschijnlijkheden uitsluitend positieve elementen bevatten bewijzen RIOZ & YAÑEZ het volgende: $\lim_{t\to\infty} q^t$ en $\lim_{t\to\infty}$ (val $\Gamma^t(k,0)-tq^t$) bestaan voor $k = 1,\dots,N$. Noteer $q = \lim_{t\to\infty} q^t$ en $v_k = \lim_{t\to\infty}$ (val $\Gamma^t(k,0)-tq^t$) voor $k = 1,\dots,N$ dan geldt voor $q, v_1,\dots,v_N$ de relatie (6.1). val $\Gamma^t(k,0)$ kan worden gevonden door de recurrente betrekkingen van stelling 2.2.2 te gebruiken.

6.3. Successieve approximatiemethoden voor verdisconteerde stochastische spelen.

De inhoud van deze sektie is een afspiegeling van de inhoud van een rapport van Van der Wal [69].

Er wordt een voorbeeld gepresenteerd, waaruit blijkt dat het algoritme van POLLATSCHEK & AVI-ITZHAK [49] voor verdisconteerde spelen niet noodzakelijkerwijs convergeert (zoals RAO, CHANSRASEKARAN & NAIR [51] beweerden bewezen te hebben).

Bovendien wordt er een verzameling successieve approximatie algoritmen behandeld, waarin na iedere waarde transformatie operatie een strategie van de minimaliserende speler gefixeerd wordt en de waarde van het resulterende Markov-beslissingsprobleem voor de maximaliserende speler benaderd wordt door een aantal standaard successieve approximatie operaties.

We beschouwen een niet-stoppend eindig twee-persoons nulsomspel (zie definities 1.1.2, 1.1.3 en 1.1.4). De uitbetalingen worden verdisconteerd en de verdisconteringsfaktor is $\beta \in [0,1)$. De notaties zijn gelijk aan die in de voorgaande sekties. De verzameling van alle stationaire strategieën voor speler 1 (speler 2) noteren we als Π_1 (Π_2). Spelen de spelers $(\pi_1,\pi_2) \in \Pi_1\times\Pi_2$, dan wordt de totale verwachte verdisconteerde opbrengst genoteerd als $\overline{V}(\pi_1,\pi_2) = (V_1(\pi_1,\pi_2),\dots,V_N(\pi_1,\pi_2))$, waarbij $V_k(\pi_1,\pi_2)$ bij begintoestant k behoort.

Uit het artikel van Shapley [57] (hoofdstuk 2; zie ook opmerking 2.3.7) blijk dat bovengenoemd spel een waarde heeft en dat de beide spelers ε-optimale strategieën bezitten binnen de klasse der stationaire strategieën voor alle $\varepsilon \geq 0$. (deze strategieën zijn tevens ε-optimaal binnen de klasse van alle (gedrags-) strategieën). Laat $\overrightarrow{v}^* = (v_1^*,\dots,v_N^*)$ de waardevektor zijn.

Laat voor $\overline{v} \in \mathbb{R}^N$ $G_k(\overline{v})$, $k = 1,2,\dots,N$ de dummy-spelen zijn bij de N-vektor $\overline{v}$. (zie definitie 2.3.1). Merk op dat een stelsel akties

$\pi_1(k)$, $k = 1,2,\ldots,N$ voor speler 1 voor de dummy-spelen, waarbij $\pi_1(k) = (\pi_1(k,1),\ldots,\pi_1(k,m_k))$ een gemengde aktie is in dummy-spel k, opgevat kan worden als een stationaire strategie voor het stochastische spel en omgekeerd.

Definieer de volgende afbeeldingen.

$$L : \Pi_1 \times \Pi_2 \times \mathbb{R}^N \to \mathbb{R}^N, \quad U : \Pi_2 \times \mathbb{R}^N \to \mathbb{R}^N \quad \text{en } T: \mathbb{R}^N \to \mathbb{R}^N$$

als

$$(L(\pi_1,\pi_2)\bar{v})_k = \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} \pi_1(k,i) \cdot \pi_2(k,j)\{g(k,i,j) + \beta \sum_{\ell=1}^{N} p(\ell|k,i,j)v_\ell\}$$

$$k = 1,2,\ldots,N; \ \forall(\pi_1,\pi_2,\bar{v}) \in \Pi_1 \times \Pi_2 \times \mathbb{R}^N.$$

(6.5) $$U(\pi_2,\bar{v}) = \max_{\pi_1 \in \Pi_1} L(\pi_1,\pi_2)\bar{v}, \qquad \forall(\pi_2,\bar{v}) \in \Pi_2 \times \mathbb{R}^N$$

(6.6) $$T\bar{v} = \max_{\pi_1 \in \Pi_1} \min_{\pi_2 \in \Pi_2} L(\pi_1,\pi_2)\bar{v}, \qquad \forall\bar{v} \in \mathbb{R}^N$$

max en max min worden hierbij componentsgewijs genomen.

Merk op dat $L(\pi_1,\pi_2)\bar{v}$ de uitbetalingsvektor is bij de dummy-spelen als speler 1 respectievelijk speler 2 het stelsel akties $\pi_1 = (\pi_1(1),\ldots,\pi_1(N))$ respectievelijk $\pi_2 = (\pi_2(1),\ldots,\pi_2(N))$ kiest. $U(\pi_2,\bar{v})$ is de maximaal te behalen uitbetaling voor speler 1 in de dummy-spelen als speler 2 π_2 speelt en $T\bar{v}$ is de waardevektor van de dummy-spelen $G_k(\bar{v})$, $k = 1,2,\ldots,N$. T is gelijk aan Shapley's waarde transformatie operator.

We zullen de volgende eigenschappen van deze afbeeldingen gebruiken.

Eigenschap 6.3.1. $L(\pi_1,\pi_2)(\cdot)$, $U(\pi_2,\cdot)$ *en* $T(\cdot)$ *zijn monotoon op* $\mathbb{R}^N$.

Eigenschap 6.3.2. $L(\pi_1,\pi_2)(\cdot)$ *en* $T(\cdot)$ *zijn strikte contraktie afbeeldingen met betrekking tot de maximum norm in* $\mathbb{R}^N$ *(die we noteren als* $\|\cdot\|$*) met contraktiestraal* β.

Eigenschap 6.3.3. $L(\pi_1,\pi_2)(\cdot)$ *heeft als uniek dekpunt* $\bar{v}(\pi_1,\pi_2)$.

Eigenschap 6.3.4. $T(\cdot)$ *heeft als uniek dekpunt* $\bar{v}^*$.

Eigenschap 6.3.1 kan rechtstreeks bewezen worden. De eigenschappen 6.3.2, 6.3.3 en 6.3.4 staan bewezen in [70] (zie ook hoofdstuk 2).

Het algoritme van POLLATSCHEK & AVI-ITZHAK [49] kan met behulp van bovenstaande afbeeldingen als volgt worden geformuleerd.

Agoritme 6.3.5.

stap 1: Laat $\bar{v}_0 = (0,0,\dots,0) \in \mathbb{R}^N$

stap 2: Bepaal voor $n = 0,1,\dots$ strategieën $\pi_1^{n+1} \in \Pi_1$ en $\pi_2^{n+1} \in \Pi_2$ en een vektor $\bar{v}_{n+1}$, z.d.d. voor $\forall(\pi_1,\pi_2) \in \Pi_1\times\Pi_2$ geldt $L(\pi_1,\pi_2^{n+1})\bar{v}_n \le L(\pi_1^{n+1},\pi_2^{n+1})\bar{v}_n (=T\bar{v}_n) \le L(\pi_1^{n+1},\pi_2)\bar{v}_n$ en $\bar{v}_{n+1} = \bar{V}(\pi_1^{n+1},\pi_2^{n+1})$ = dekpunt van de afbeelding $L(\pi_1^{n+1},\pi_2^{n+1})(\cdot)$.

We geven een spel waarvoor het algoritme van Pollatschek en Avi-Itzhak niet convergeert.

VOORBEELD 6.3.6. Het spel heeft twee toestanden. In toestand 1 staan beide spelers twee zuivere akties ter beschikking en in toestand 2 beiden één. Uitbetalingen: $g(1,1,1) = 3$, $g(1,1,2) = 6$, $g(1,2,1) = 2$, $g(1,2,2) = 1$ en $g(2,1,1) = 0$. Overgangskansen: $p(1|1,1,1) = 1$, $p(1|1,1,2) = \frac{1}{3}$, $p(1|1,2,1) = 1$, $p(1|1,2,2) = 1$, $p(1|2,1,1) = 0$ en steeds is $p(2|k,i,j) = 1 - p(1|k,i,j)$. Laat $\beta = \frac{3}{4}$. Merk op dat een stationaire strategie $\pi_1(\pi_2)$ volledig gekarakteriseerd is door de kans $\alpha_1(\alpha_2)$ die speler 1 (speler 2) op aktie 1 in toestand 1 legt. Een paar stationaire strategieën (π_1,π_2) wordt dan ook weergegeven door het paar (α_1,α_2) met $\alpha_1 = \pi_1(1,1)$ en $\alpha_2 = \pi_2(1,1)$. Laat $\bar{v}_0 = (0,0)$. Het eerste dummy-spel is dan $\begin{pmatrix}3 & 6\\ 2 & 1\end{pmatrix}$ en het tweede (0). Men ziet dat het eerste dummy-spel een evenwichtspunt van zuivere akties bezit en wel als beide spelers hun eerste aktie kiezen; dus $(\pi_1^1,\pi_2^1) = (1,1)$. Dan is $\bar{v}_1 = \bar{V}(\pi_1^1,\pi_2^1) = (12,0)$. Om π_1^2 en π_2^2 te vinden moeten we het matrixspel

$$\begin{pmatrix}3 + \frac{3}{4}.1.12 & 6 + \frac{3}{4}.\frac{1}{3}.12\\ 2 + \frac{3}{4}.1.12 & 1 + \frac{3}{4}.1.12\end{pmatrix} = \begin{pmatrix}12 & 9\\ 11 & 10\end{pmatrix}$$

oplossen. Ook nu is er weer een evenwichtspunt van zuivere akties, namelijk als beide spelers hun 2de aktie kiezen, dus $(\pi_1^2,\pi_2^2) = (0,0)$; dan is $\bar{v}_2 = \bar{V}(\pi_1^2,\pi_2^2) = (4,0)$. In de volgende stap moeten we het matrixspel

$$\begin{pmatrix}3 + \frac{3}{4}.1.4 & 6 + \frac{3}{4}.\frac{1}{3}.4\\ 2 + \frac{3}{4}.1.4 & 1 + \frac{3}{4}.1.4\end{pmatrix} = \begin{pmatrix}6 & 7\\ 5 & 4\end{pmatrix}$$

oplossen. Het blijkt dat $(\pi_1^3,\pi_2^3) = (1,1)$ en $\bar{v}_3 = \bar{V}(\pi_1^3,\pi_2^3) = (12,0)$. Aldus doorgaande vinden we

$$\bar{v}_{2n-1} = (12,0) \text{ en } \bar{v}_{2n} = (4,0), \quad n = 1,2,\ldots .$$

M.a.w. het algoritme 6.3.5 convergeert niet voor dit spel.

<u>OPMERKING 6.3.7.</u> De numerieke resultaten in [49] suggereren, dat het toch zinvol kan zijn het algoritme toe te passen, indien niet voldaan wordt aan de voldoende condities voor convergentie van het algoritme. (Deze conditie is $\max_k \sum_{\ell=1}^{N} (\max_{i,j} p(\ell|k,i,j) - \min_{i,j} p(\ell|k,i,j)) < 1 - \max_{k,i,j} \sum_{\ell=1}^{N} p(\ell|k,i,j))$.

We presenteren nu een verzameling strategie verbeterende waarde benaderende algoritmen, waarin na iedere strategie verbeterende stap de strategie van de minimaliserende speler gefixeerd wordt en de waarde van het resulterende Markov-beslissingsprobleem benaderd wordt door een aantal approximatiestappen. We laten zien, dat ieder algoritme een ε-interval voor de waarde van het verdisconteerde stochastische spel geeft en ε-optimale strategieën voor de beide spelers. Laat $\bar{e} = (1,1,\ldots,1) \in \mathbb{R}^N$.

Beschouw voor iedere integer λ tussen 1 en ∞ (grenzen ingesloten) het volgende algoritme.

<u>Algoritme (λ) 6.3.8.</u>

stap 1: Bepaal een vektor $\bar{v}_0$ met de eigenschap $T\bar{v}_0 \leq \bar{v}_0$
kies $\varepsilon > 0$ en $n := 0$.

stap 2: (strategie verbeterende stap)
Bepaal $T\bar{v}_n$ en een strategie π_2^{n+1} voor speler 2, die voldoet aan
$L(\pi_1,\pi_2^{n+1})\bar{v}_n \leq T\bar{v}_n$, $\forall \pi_1 \in \Pi_1$.
Als $\max_k\{(T\bar{v}_n)_k - (\bar{v}_n)_k\} - \min_k\{(T\bar{v}_n)_k - (\bar{v}_n)_k\} \leq \varepsilon\frac{(1-\beta)}{\beta}$
Ga naar STOP.

stap 3: (waarde benaderende stap)
Bepaal $\bar{v}_{n+1} = U^{\lambda}(\pi_2^{n+1},\bar{v}_n)$; $\quad n := n+1$.
Ga naar STAP 2.
(met $U^{\lambda}(\pi_2^{n+1},\bar{v}_n)$ wordt bedoeld het λ keer toepassen van de operator $U(\pi_2^{n+1},\cdot)$ op $\bar{v}_n$).

STOP $\quad T\bar{v}_n + \frac{\beta}{1-\beta}\{\min_k\{(T\bar{v}_n)_k - (\bar{v}_n)_k\}\cdot\bar{e} \leq \bar{v} \leq T\bar{v}_n + \frac{\beta}{1-\beta}\max_k\{(T\bar{v}_n)_k - (\bar{v}_n)_k\}\bar{e}$.
De strategieën π_1^{n+1} en π_2^{n+1}, die voldoen
$L(\pi_1,\pi_2^{n+1})\bar{v}_n \leq L(\pi_1^{n+1},\pi_2^{n+1})\bar{v}_n \leq L(\pi_1^{n+1},\pi_2)\bar{v}_n$, $\forall(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$
zijn ε-optimaal voor respectievelijk speler 1 en speler 2.

Merk op dat voor $\lambda = 1$ we hebben $\bar{v}_n = T\bar{v}_{n-1}$, zodat $\bar{v}_n$ reeds in stap 2 bepaald is.

We zullen nu laten zien dat dit algoritme na een eindig aantal berekeningen stopt en dat de beweringen onder STOP juist zijn.

LEMMA 6.3.9. $\bar{v}^* \le \bar{v}_k \le T\bar{v}_{k-1} \le \bar{v}_{k-1}$, $\quad k = 1,2,\ldots$.

BEWIJS: Het bewijs is gebaseerd op volledige inductie.
$k = 1$. Daar π_2^1 optimaal is voor speler 2 voor de dummy-spelen $G_k(\bar{v}_0)$, $k = 1,2,\ldots,N$ volgt:

$$(6.7) \qquad U(\pi_2^1,\bar{v}_0) = T\bar{v}_0 \le \bar{v}_0.$$

Uit (6.7) en de monotonie van $U(\pi_2^1,.)$ (eigenschap 6.3.1) volgt:

$$(6.8) \qquad v_1 = U^\lambda(\pi_2^1,\bar{v}_0) \le U(\pi_2^1,\bar{v}_0) = T\bar{v}_0 \le \bar{v}_0.$$

Uit de monotonie van T (eigenschap 6.3.1) en (6.7) volgt $T^r(\bar{v}_0) \le \bar{v}_0$, $r = 1,2,\ldots$, zodat m.b.v. $\bar{v}^* = \lim_{r\to\infty} T^r\bar{v}_0$ blijkt:

$$(6.9) \qquad \bar{v}^* \le T^r\bar{v}_0 \le \bar{v}_0, \quad r = 0,1,2,\ldots .$$

Uit de definities van $U(\pi_2^1,\cdot)$ en T((6.5) en (6.6)) blijkt dat:

$$(6.10) \qquad U(\pi_2^1,\bar{v}) \ge T\bar{v}, \ \forall\bar{v} \in \mathbb{R}^N,$$

zodat gebruik makend van (6.9)

$$(6.11) \qquad \bar{v}_1 = U^\lambda(\pi_2^1,\bar{v}_0) \ge T^\lambda\bar{v}_0 \ge \bar{v}_0.$$

(6.8), (6.9) en (6.11) kombineren levert de bewering in het lemma voor $k = 1$.

Veronderstel dat de bewering waar is voor $k = 1,2,\ldots,n$ en we laten zien dat de bewering ook waar is voor $k = n + 1$. Met behulp van deze veronderstelling, de monotonie van $U(\pi_2,\cdot)$ en $T(\cdot)$ en (6.10) volgt nu:

$$(6.12) \qquad U(\pi_2^{n+1},\bar{v}_n) = T\bar{v}_n = T(U^\lambda(\pi_2^n,\bar{v}_{n-1})) \le U^{\lambda+1}(\pi_2^n,\bar{v}_{n-1}) \le$$
$$U^\lambda(\pi_2^n,\bar{v}_{n-1}) = \bar{v}_n.$$

Dus $U(\pi_2^{n+1},\bar{v}_n) \le \bar{v}_n$, waaruit volgt:

$$\bar{v}_{n+1} = U^{\lambda}(\pi_2^{n+1},\bar{v}_n) \le U(\pi_2^{n+1},\bar{v}_n) \le \bar{v}_n \,. \tag{6.13}$$

Uit $\bar{v}_n \ge \vec{v}^*$ en (6.10) tenslotte volgt:

$$\bar{v}_{n+1} = U^{\lambda}(\pi_2^{n+1},\bar{v}_n) \ge T^{\lambda}\bar{v}_n \ge T\vec{v}^* = \vec{v}^* . \tag{6.14}$$

Kombineren van (6.12), (6.13) en (6.14) geeft de bewering voor $k = n+1$. □

<u>GEVOLG 6.3.10</u>. $\lim_{n\to\infty} \bar{v}_n = \vec{v}^*$.

<u>BEWIJS</u>. Uit lemma 6.3.9 zien we $\vec{v}^* \le \bar{v}_n \le T\bar{v}_{n+1} \ldots \le T^n\bar{v}_0$ en daar $\lim_{n\to\infty} T^n\bar{v}_0 = \vec{v}^*$ volgt $\lim_{n\to\infty} \bar{v}_n = \vec{v}^*$. □

Uit gevolg 6.3.10 kunnen we nu concluderen dat algoritme (λ) 6.3.8 na een eindig aantal iteraties stopt. Aangezien voor $\lambda < \infty$ het aantal berekeningen in iedere iteratie eindig is en voor $\lambda = \infty$ we in *stap* 3 Howard's strategie iteratie algoritme kunnen gebruiken, waarvoor ook een eindig aantal berekeningen nodig zijn, mogen we nu concluderen dat voor iedere λ het algoritme (λ) 6.3.8 na een eindig aantal berekeningen stopt.

We gaan nu bewijzen dat de beweringen onder STOP correct zijn. Laat $\mu_n = \min_k\{(T\bar{v}_n)_k-(\bar{v}_n)_k\}$ en $\nu_n = \max_k\{(T\bar{v}_n)_k-(\bar{v}_n)_k\}$. Merk op dat uit lemma 6.3.9 blijkt dat zowel $\mu_n \le 0$ als $\nu_n \le 0$.

<u>STELLING 6.3.11</u>. $T\bar{v}_n + \frac{\beta}{1-\beta}\,\mu_n\bullet\bar{e} \le \vec{v}^* \le T\bar{v}_n + \frac{\beta}{1-\beta}\,\nu_n\bullet\bar{e}$, $\quad n = 0,1,2,\ldots$.

<u>BEWIJS</u>. Uit de definitie van ν_n volgt $T\bar{v}_n \le \bar{v}_n + \nu_n\bullet\bar{e}$, zodat

$$T^r\bar{v}_n \le T^{r-1}(\bar{v}_n+\nu_n\,\bar{e}) = T^{r-1}\bar{v}_n + \beta^{r-1}\nu_n\bullet\bar{e} \le \ldots \le \tag{6.15}$$

$$\le T\bar{v}_n + (\beta+\ldots+\beta^{r-1})\nu_n\bullet\bar{e} \,.$$

$\vec{v}^* \le T^r\bar{v}_n$, alle r, kombineren met (6.15) voor $r \to \infty$ geeft

$$\vec{v}^* \le T\bar{v}_n + \frac{\beta}{1-\beta}\,\nu_n\bullet\bar{e} \,.$$

Evenzo kunnen we laten zien $\vec{v}^* \ge T\bar{v}_n + \frac{\beta}{1-\beta}\,\mu_n\bullet\bar{e}$. □

Uit stelling 6.3.11 blijkt dat, als $\nu_n - \mu_n \le \frac{\varepsilon(1-\beta)}{\beta}$, we een 2ε interval voor $\vec{v}^*$ gevonden hebben.

<u>STELLING 6.3.12</u>. *Als* $\nu_n - \mu_n \le \frac{\varepsilon(1-\beta)}{\beta}$ *dan zijn* π_1^{n+1} *en* π_2^{n+1} *ε-optimale strategieën binnen de klasse van alle strategieën.*

<u>BEWIJS</u>. Laat π_2 een willekeurige stationaire strategie voor speler 2 zijn. Er geldt dan:

$$L(\pi_1^{n+1},\pi_2)\bar{v}_n \ge L(\pi_1^{n+1},\pi_2^{n+1})\bar{v}_n = T\bar{v}_n \ge \bar{v}_n + \mu_n\cdot\bar{e},$$

zodat

$$(6.16)\quad L^r(\pi_1^{n+1},\pi_2)\bar{v}_n \ge L^{r-1}(\pi_1^{n+1},\pi_2)(\bar{v}_n+\mu_n\bar{e}) = L^{r-1}(\pi_1^{n+1},\pi_2)\bar{v}_n + \beta^{r-1}\mu_n\cdot\bar{e} \ge$$
$$\ge \dots \ge L(\pi_1^{n+1},\pi_2)\bar{v}_n + (\beta+\dots+\beta^{r-1})\mu_n\cdot\bar{e} \ge$$
$$\ge T\bar{v}_n + \sum_{i=1}^{r-1}\beta^i\cdot\mu_n\cdot\bar{e}$$
$$\ge T\bar{v}_n + \sum_{i=1}^{\infty}\beta^i\cdot\mu_n\cdot\bar{e} = T\bar{v}_n + \frac{\beta}{1-\beta}\mu_n\cdot\bar{e} \qquad r = 1,2,\dots\,.$$

Eigenschap 6.3.3 geeft:

$$(6.17)\qquad \bar{V}(\pi_1^{n+1},\pi_2) = \lim_{r\to\infty} L^r(\pi_1^{n+1},\pi_2)\bar{v}_n$$

(6.16) en (6.17) kombineren met stelling 6.3.11 geeft:

$$\bar{V}(\pi_1^{n+1},\pi_2) \ge T\bar{v}_n + \frac{\beta}{1-\beta}\mu_n\cdot\bar{e} \ge \vec{v}^* + \frac{\beta}{1-\beta}(\mu_n-\nu_n)\cdot\bar{e},$$

zodat $\nu_n - \mu_n \le \frac{\varepsilon(1-\beta)}{\beta}$ impliceert

$$(6.18)\qquad \bar{V}(\pi_1^{n+1},\pi_2) \ge \vec{v}^* - \varepsilon\bar{e},\ \forall\pi_2 \in \Pi_2.$$

Uit (6.18) blijk, dat π_1^{n+1} ε-optimaal is t.o.v. alle stationaire strategieën van speler 2. In hoofdstuk 3 (stelling 3.4.2) is aangetoond, dat voor een stochastisch spel van Takahashi een speler tegen een vaste stationaire strategie van de andere speler niet beter heeft dan zijn klasse van stationaire strategieën. Het onderhavig model is een speciaal geval van een spel van Takahashi; zodat we met behulp van stelling 3.4.2 zien dat π_1^{n+1} ook ε-optimaal is t.o.v alle strategieën van speler 2. □

OPMERKINGEN.

Van de verzameling successieve approximatie algoritmen, zoals gedefinieerd onder algoritme (λ) 6.3.8 zijn twee reeds bekend. Namelijk voor $\lambda = 1$ hebben we de methode van Shapley met verbeterde grenzen en voor $\lambda = \infty$ verkrijgen we het algoritme van HOFFMAN & KARP [22].

In het algoritme (λ) 6.3.8 is het aantal successieve approximatie operaties in *stap* 3 constant in iedere iteratie. Dit is echter onnodig beperkend. Men kan gemakkelijk inzien, dat als we in *stap* 3 $\bar{v}_{n+1} = U^{\lambda}(\pi_2^n, \bar{v}_n)$ vervangen door $\bar{v}_{n+1} = U^{\lambda_n}(\pi_2^n, \bar{v}_n)$ het resulterende algoritme dezelfde eigenschappen heeft als algoritme (λ) 6.3.8. Dit is interessant daar gedurende de eerste paar iteraties het dekpunt van $U(\pi_2^n, \cdot)$ een relatief aanzienlijke afstand tot $\vec{v}^*$ kan hebben. Het is dan dus niet erg zinvol een grote waarde van λ te gebruiken. Na een aantal iteraties zal het dekpunt van $U(\pi_2^n, \cdot)$ dichter bij $\vec{v}^*$ komen te liggen, zodat het zinvoller wordt een grotere waarde van λ te gebruiken.

Een speciaal geval treedt op als één van de spelers slechts één zuivere äktie ter beschikking heeft in iedere toestand. In dit geval is het stochastisch spel niets anders dan een Markov-beslissingsprobleem en de verzameling algoritmen (λ) 6.3.8 zijn nu precies de verzameling algoritmen die door VAN NUNEN [42] beschreven zijn.

Een van de voorwaarden in het model is $\sum_{\ell=1}^{N} p(\ell|k,i,j) = 1$, voor alle k,i, en j. Dit kunnen we afzwakken tot $0 \le \sum_{\ell=1}^{N} p(\ell|k,i,j) \le 1$. Het bewijs dat het algoritme ook nu na een eindig aantal berekeningen stopt is identiek. Echter om een interval voor $\vec{v}^*$ te vinden en boven- en ondergrenzen voor $\bar{V}(\pi_1^{n+1}, \pi_2)$ en $\bar{V}(\pi_1, \pi_2^{n+1})$ te bepalen, zullen we rekening moeten houden met de kleinste en grootste waarde van $\sum_{\ell=1}^{N} p(\ell|k,i,j)$. (Zie bijvoorbeeld PORTEUS [50], die dit geval voor Markov-beslissingsprobleem beschouwt; voor stochastische spelen hebben we echter een gelijksoortige situatie.)

HOOFDSTUK VII

ASYMPTOTISCHE THEORIE VAN STOCHASTISCHE SPELEN

7.1. Inleiding

In dit hoofdstuk worden enige resultaten van BEWLEY en KOHLBERG genoemd ([3],[4] en [5]). Bewijzen zullen wij niet vermelden. Zij houden zich bezig met het eindige twee-persoons stochastische spel en bestuderen het asymptotisch gedrag van de t-staps versie als $t \to \infty$ en het asymptotisch gedrag van de met rentevoet r verdisconteerde versie van het spel als $r \downarrow 0$. De verdisconteringsfactor β is dan $1/(1+r)$.

De waardevektor van het t-stapsspel noteren wij als $\bar{V}^t$ en de waardevektor van het met rentevoet r verdisconteerde spel als $\bar{V}(r)$. De gemiddelde opbrengst per stap in het t-staps spel is gelijk aan $\frac{1}{t}\bar{V}^t$. In het met rentevoet r verdisconteerde spel heeft betaling van $r\bar{V}(r)/(1+r)$ op ieder tijdstip de huidige waarde $\bar{V}(r)$. BEWLEY en KOHLBERG tonen aan dat $\lim_{t\to\infty} \frac{1}{t}\bar{V}_t$ en $\lim_{r\downarrow 0} r\bar{V}(r)/(1+r) = \lim_{r\downarrow 0} r\bar{V}(r)$ bestaan en gelijk zijn. Een belangrijke tussenstap voor het bereiken van dit resultaat is de volgende stelling.

<u>STELLING 7.1.1</u>. *Bij ieder stoachastisch spel bestaat er een geheel getal* M *z.d.d.*

(i) $\bar{V}(r)$ *heeft een reeksontwikkeling van de vorm*

$$\bar{V}(r) = \bar{a}^{-M} r^{-1} + \bar{a}^{-M-1} r^{-(M-1)/M} + a^{-M-2} r^{-(M-2)/M} + \dots, \tag{7.1}$$

waarbij $\bar{a}^{-i} \in \mathbb{R}^N$ *voor alle* i *en* $r > 0$ *voldoende klein is.*

(ii) *Er bestaan optimale stationaire strategieën voor speler* 1 *in het met rentevoet* r *verdisconteerde spel gegeven door de volgende vektoren voor* $k = 1,\dots,N$.

$$\pi_1^k(r) = \pi_1^{0k} + \pi_1^{1k} r^{1/M} + \pi_1^{2k} r^{2/M} + \dots, \tag{7.2}$$

waarbij $\pi_1^k(r) \in \mathbb{R}^{m_k}$ *een kansvektor is,* $\pi_1^{ik} \in \mathbb{R}^{m_k}$ *voor alle* i *en* $r > 0$ *voldoende klein is.*

(iii) *Een analoge uitspraak geldt voor speler* 2.

Voor Markov beslissingsprocessen (eenpersoons stochastische spelen) heeft deze stelling een eenvoudiger gedaante en is afkomstig van BLACKWELL [7]. In dit geval geldt $M = 1$ en $\pi_1^k(r) = \pi_1^{0k}$. Dus in de reeksontwikkelingen komen geen gebroken machten voor en een zelfde stationaire strategie (gegeven door π_1^{0k} voor $k = 1,\dots,N$) is optimaal voor alle rentevoeten $r > 0$ die voldoende klein zijn.

Het bestaan van $\lim_{r\downarrow 0} r\bar{V}(r)$ is een direkt gevolg van stelling 7.1.1.

Voor het met rentevoet r verdisconteerde spel bewees SHAPLEY (stelling 2.3.2) dat de waarde $\bar{V}(r) \in \mathbb{R}^N$ gelijk is aan de unieke oplossing van het stelsel vergelijkingen:

(7.3) $$\phi_k = \text{val } G_k(\bar{\phi}/(1+r)) \qquad k = 1,\dots,N.$$

Wij schrijven dit stelsel in vektornotatie:

(7.4) $$\bar{\phi} = \text{val } G(\bar{\phi}/(1+r)).$$

Deze vergelijking wordt de r-verdisconteringsvergelijking genoemd.

Stelling 7.1.1. vermeldt dat de oplossing $\bar{V}(r)$ als r voldoende klein van de vorm $\bar{a}^{-M}r^{-1} + \bar{a}^{-M-1}r^{-(M-1)/M} + \bar{a}^{-M-2}r^{-(M-2)/M} + \dots$ is.

Ofwel:

(7.5) $$\bar{a}^{-M}r^{-1} + \bar{a}^{-M-1}r^{-(M-1)/M} + \dots = \text{val } G\left(\frac{\bar{a}^{-M}r^{-1}+\bar{a}^{-M-1}r^{-(M-1)/M}+\dots}{1+r}\right)$$

als r voldoende klein is. Dit is een stelsel met overaftelbaar veel vergelijkingen n.l. N vergelijkingen voor iedere waarde van r. Dit stelsel wordt opgevat als stelsel met N vergelijkingen met als onbekenden $V_1(r),\dots,V_N(r)$ als functies van r. De verzameling functies van de voorgeschreven vorm zullen wij nader bekijken.

Wij beschouwen reeksen van de vorm $\sum_{i=i_0}^{\infty} a^{-i}r^{i/M}$, waarbij de sommatie convergeert als $r > 0$ voldoende klein is. Noteren we $\theta = \frac{1}{r}$ dan worden deze reeksen van de vorm $\sum_{i=i_0}^{\infty} a^{-i}(\theta^{-1})^{i/M} = \sum_{i=-\infty}^{i_0} a^{i}\theta^{i/M}$, waarbij de sommatie convergeert als θ voldoende groot is. De verzameling van deze reeksen noteren wij als F. Op F definiëren wij optelling en vermenigvuldiging op gelijke wijze als bij machtreeksen en wij noemen $\sum_{i=-\infty}^{i_0} a^i \theta^{i/M} \in F$ positief als er een I bestaat z.d.d. $a^I > 0$ en $a^i = 0$ voor alle $i > I$. Met deze definities blijkt F een geordend lichaam te zijn (Het lichaam van de reëele Puiseux reeksen).

Als $x = \sum_{i=-\infty}^{i_0} a^i \theta^{i/M} \in F$ en $t \in \mathbb{R}$ dan noteren wij:

(7.6) $$\sigma_t x = \sum_{i=-\infty}^{i_0} a^i t^{i/M}.$$

Er geldt:

(7.7) $$\sigma_t(x+y) = \sigma_t x + \sigma_t y,$$

(7.8) $$\sigma_t(xy) = \sigma_t x . \sigma_t y$$

en

(7.9) $$x > 0 \Longleftrightarrow \sigma_t x > 0 \quad \text{als t groot genoeg.}$$

Als $\bar{x} \in F^q$ dan $\sigma_t \bar{x} = (\sigma_t x_1, \ldots, \sigma_t x_q)$. BEWLEY en KOHLBERG bewijzen dat het lichaam F reëel gesloten is (d.w.z. er is geen geordende algebraïsche uitbreiding) en kunnen daarom het beginsel van TARSKI toepassen.

7.2. Het beginsel van Tarski

Allereerst enige definities:

DEFINITIE 7.2.1. Een geordend lichaam is *reëel gesloten* als er geen geordende algebraïsche uitbreiding bestaat.

DEFINITIE 7.2.2. Een *atomistische formule* is een uitdrukking van de vorm $p > 0$ of $p = 0$ waarbij p een veelterm is met één of meer variabelen en geheeltallige coëfficiënten.

DEFINITIE 7.2.3. Een *elementaire formule* is een uitdrukking die bestaat uit eindig veel atomistische formules verbonden door de symbolen $\wedge, \vee, \neg, \exists$ of $\forall$.

DEFINITIE 7.2.4. Een variabele in een elementaire formule is een *vrije variabele* als deze niet voorkomt voorafgegaan door $\exists$ of $\forall$.

DEFINITIE 7.2.5. Een *elementaire zin* is een elementaire formule zonder vrije variabelen.

Als de variabelen in een elementaire zin elementen van zeker geordend lichaam zijn is de zin waar of niet waar. De waarheid of onwaarheid van een elementaire formule met vrije variabelen kan afhankelijk zijn van de voor deze vrije variabelen ingevulde waarden. TARSKI [63] geeft een algoritme dat in een eindig aantal stappen onderzoekt of een elementaire zin waar of niet waar is over de reële getallen. In dit algoritme wordt uitsluitend gebruik gemaakt van het feit dat deze verzameling een reëel gesloten lichaam is.

GEVOLG 7.2.6. (Het beginsel van Tarski)
Als een elementaire zin waar is over de reële getallen is deze zin waar over ieder reëel gesloten lichaam (SEIDENBERG [56]).

De minimaxstelling voor matrixspelen (stelling 1.2.3) geldt in ieder geordend lichaam (WEYL [72]). Dus als A een matrix is van elementen uit geordend lichaam H dan is val $A \in H$ gedefiniëerd. De bewering val $A = y$ is een elementaire formule over H:

$$\exists \pi_{11},\ldots\pi_{1m},\pi_{21},\ldots,\pi_{2n} \in H \; (\pi_{11},\ldots,\pi_{1m},\pi_{21},\ldots,\pi_{2n} > 0 \wedge$$
$$\sum_{i=1}^{m} \pi_{1i} = 1 \wedge \sum_{j=1}^{n} \pi_{2j} = 1 \wedge \sum_{i=1}^{m} a_{ij}\pi_{1i} \geq y,\; j = 1,\ldots,n \wedge$$
$$\sum_{j=1}^{n} a_{ij}\pi_{2j} \leq y,\; i = 1,\ldots,m).$$

Voor een stochastisch spel met als betalingen en overgangskansen elementen van H is dus ook val $G(\bar{x}) = \bar{y}$ een elementaire formule over H. De vrije variabelen in deze formule zijn de betalingen, de overgangskansen en de komponenten van $\bar{x}$ en $\bar{y}$. SHAPLEY bewees de waarheid van de volgende elementaire zin als $H = \mathbb{R}$ (stelling 2.3.2)

"Voor alle waarden van $g(k,i,j)$ en $p(\ell|k,i,j) \geq 0$ z.d.d. $\sum_{\ell=1}^{N} p(\ell|k,i,j) = 1$ en $r > 0$ bestaat er een $\bar{\phi}$ z.d.d. $\bar{\phi} = \text{val } G\,(\bar{\phi}/(1+r))$."

Uit het beginsel van Tarski volgt dat deze elementaire zin juist is over F.

GEVOLG 7.2.7. *Er bestaat een* $\bar{x} \in F^N$ *z.d.d.*

$$\bar{x} = \text{val } G(\bar{x}/(1+r)). \tag{7.10}$$

deze vergelijking over F^N *wordt de limiet verdisconteringsvergelijking genoemd.*

De volgende stelling legt het verband tussen de val operatoren in de lichamen $\mathbb{R}$ en F.

STELLING 7.2.8. *Laat* A *een* $m \times n$ *matrix van elementen uit* F. $\sigma_t A$ *is de matrix met op plaats* i,j *het getal* $\sigma_t a_{ij}$. *Dan bestaat er een* t_0 *z.d.d. voor alle* $t > t_0$ *geldt:*

(i) $\text{val } \sigma_t A = \sigma_t \text{ val } A$

(ii) $\pi_1 \in F^m$ *optimaal voor speler* 1 *in* A $\Longleftrightarrow$ $\sigma_t \pi_1$ *is optimaal voor speler* 1 *in* $\sigma_t A$.

Hieruit kan de volgende stelling worden afgeleid.

STELLING 7.2.9. *Laat* $\bar{x} \in F^N$ *voldoen aan de limiet verdisconteringsvergelijking en laat* $\pi_1 = (\pi_1(1),\ldots,\pi_1(N))$ *met* $\pi_1(k) \in F^{m_k}$ *een optimale strategie voor speler* 1 *in het matrixspel* $G_k(\bar{x}/1+\theta^{-1}))$. *Als* $r > 0$ *voldoende klein dan geldt:*

(i) $\bar{V}(r) = \sigma_{r^{-1}}\bar{x}$

(ii) $\sigma_{r^{-1}}\pi_1$ *is optimaal voor speler* 1 *in het met rentevoet* r *verdisconteerde spel.*

Uit deze stelling tezamen met het feit dat de limiet verdisconteringsvergelijking een oplossing heeft wordt stelling 7.1.1 bewezen.

7.3. De limiet recursievergelijking

Wij definiëren de afbeelding $\tau\colon F \to F$ als volgt:

$$\tau x = \tau\left(\sum_{i=-\infty}^{i_0} a^i \theta^{i/M}\right) = \sum_{i=-\infty}^{i_0} a^i (\theta+1)^{i/M}, \tag{7.11}$$

waarbij $(\theta+1)^{i/M}$ wordt gedefinieerd door:

$$\theta^{i/M} + i\theta^{(i/M)-1} + \ldots \in F. \tag{7.12}$$

Als wij termen met gelijke machten van θ bij elkaar nemen verkrijgen wij een reeks van de vereiste vorm. Voor t voldoende groot (z.d.d. $\sigma_t x$ gedefinieerd is) geldt: $\sigma_t \tau x = \sigma_{t+1} x$. Vanwege de convergentie van $\sigma_t x$ convergeert ook $\sigma_t \tau x$ als t voldoende groot is, dus $\tau x \in F$. Als $\bar{x} \in F^q$ noteren wij $\tau\bar{x} = (\tau x_1, \ldots, \tau x_q)$. De vergelijking:

$$\tau\bar{x} = \text{val } G(\bar{x}) \quad \text{met } \bar{x} \in F^N \tag{7.13}$$

noemen wij de *limiet recursievergelijking*.

Als $x = \sum_{i=-\infty}^{i_0} a^i \theta^{i/M} \in F$ dan noteren wij ρx voor het getal I/M, waarbij I z.d.d. $a_I \neq 0$ en $a_i = 0$ voor alle $i > I$. Als $\bar{x} \in F^q$ noteren wij $\rho\bar{x} = \max\{\rho x_1, \ldots, \rho x_q\}$. De bewering $\bar{x} \in F^N$ en $\rho\bar{x} < \alpha$ schrijven wij ook wel als $\bar{x} = o(\theta^\alpha)$; de bewering $\bar{x} \in F^N$ en $\rho\bar{x} \leq \alpha$ ook wel als $\bar{x} = O(\theta^\alpha)$.

Het bewijs van het bestaan van $\lim_{t\to\infty} \frac{1}{t} \bar{V}^t$ en gelijkheid daarvan aan $\lim_{r\downarrow 0} r.\bar{V}(r)$ verloopt via de volgende stelling.

<u>STELLING 7.3.1.</u> *Als $\bar{x} \in F^N$ een oplossing is van de limiet verdisconteringsvergelijking dan geldt voor $\bar{y} = \bar{x}/(1+\theta^{-1})$*

$$\tau\bar{y} = \text{val } G(\bar{y}) + o(\theta^0). \tag{7.14}$$

BEWLEY & KOHLBERG geven een voobeeld van een stochastisch spel waarvoor de limiet verdisconteringsvergelijking geen oplossing heeft met uitsluitend geheeltallige machten van θ. Het hoofdresultaat van hun eerste artikel [3] formuleren wij nogmaals in de volgende stelling.

<u>STELLING 7.3.2.</u> $\lim_{r\downarrow 0} r.\bar{V}(r)$ *en* $\lim_{t\to\infty} \frac{1}{t} \bar{V}^t$ *bestaan en zijn gelijk.*

In een volgend artikel [4] worden de volgende stellingen bewezen.

<u>STELLING 7.3.3.</u> *Er bestaat een $\bar{x} \in F^N$ z.d.d.*

$$\tau\bar{x} = \text{val } G(\bar{x}) + o(\theta^{-1}). \tag{7.15}$$

Deze vergelijking wordt de *benaderende limiet recursievergelijking* genoemd.

<u>STELLING 7.3.4.</u> *Laat* $\bar{x} = \sum_{i=-\infty}^{i_0} \bar{a}^i \theta^{i/M} \in F^N$, *waarbij* $\bar{a}^i \in \mathbb{R}^N$ *voor alle i,*

een oplossing van de benaderende limiet recursievergelijking. Laat $\bar{x}^t = \sum_{i=1}^{i_0} \bar{a}^i\, t^{i/M} \in \mathbb{R}^N$ *voor* t *positief geheel. Dan bestaat er een* $B > 0$ *z.d.d.*

$$(7.16)\qquad \max_{k=1,\dots,N} |v_k^t - x_k^t| < B \log(t+1).$$

OPMERKING 7.3.5. Omdat $\max_{k=1,\dots,N} |v_k^t| \le t \max_{i,j,k} |g(k,i,j)|$ moet voor een oplossing $\bar{x} \in F^N$ van de benaderende limiet recursievergelijking gelden $\bar{x} = O(\theta^1)$.

Laat $\bar{x} \in F^N$ een oplossing zijn van de benaderende limiet recursievergelijking en voor $k = 1,\dots,N$, $\pi_1(k) \in F^{m_k}$ een optimale strategie voor speler 1 in het matrixspel $G_k(\bar{x})$. De komponenten van $\pi_1(k)$ zijn positief en sommeren tot 1. Wij kunnen dus schrijven $\pi_1(k) = \sum_{i=-\infty}^{0} \pi_{1i}(k)\, \theta^{i/M}$, waarbij $\pi_{1i}(k) \in \mathbb{R}^{m_k}$ voor alle i. Het blijkt dat ook $\sum_{i=-I}^{0} \pi_{1i}(k)\, \theta^{i/M}$ een kansvektor is in F^{m_k} voor alle $I \ge 0$. Laat nu

$$\hat{\pi}_1(k) = \sum_{i=-2M}^{0} \pi_{1i}(k)\, \theta^{i/M}$$

en laat

$$\hat{\pi}_1 = (\hat{\pi}_1(1),\dots,\hat{\pi}_1(N)).$$

Uit (7.7) en (7.9) volgt dat ook $\sigma_t\, \hat{\pi}_1(k)$ een kansvektor is in $\mathbb{R}^{m_k}$ als t voldoende groot is. Noteren wij tenslotte $v_k^t(\hat{\pi}_1)$ voor de minimale verwachte opbrengst van speler 1 in het t-stapsspel met begintoestand k als hij de volgende strategie speelt: kies op tijdstip t in toestand k aktie 1 als t zo klein is dat $\sigma_t(\hat{\pi}_1(k))$ niet gedefiniëerd is en kies anders volgens de kansverdeling $\sigma_t(\hat{\pi}_1(k))$, dan geldt de volgende stelling:

STELLING 7.3.6. *Er bestaat een* $B > 0$ *z.d.d.*

$$(7.17)\qquad v_k^t(\hat{\pi}_1) \ge v_k^t - B \log(t+1) \quad \textit{voor alle } t \textit{ en } k = 1,\dots,N.$$

Wij zullen nu de resultaten in het rapport [5] behandelen.

7.4. Kriteria voor het ∞-stapsspel

Het niet verdisconteerde ∞-staps stochastische spel kan dienen als model voor een spel dat gedurende een groot aantal stappen wordt gespeeld, terwijl het aantal stappen niet tevoren vast staat en geen reden is de toekomstige betalingen te verdisconteren. In zo'n spel kunnen wij diverse criteria voor de spelers formuleren die wij in twee klassen n.l. de <u>limietkriteria</u> en de <u>asymptotische kriteria</u> kunnen onderverdelen.

Bij de limietkriteria worden de spelers geacht als uitgangspunt te nemen dat het spel noch stopt noch verdisconteerd wordt. Deze kriteria geven een waardering aan voor kansverdelingen op de verzameling van oneindige rijen betalingen die het gevolg zijn van de gekozen strategieën. Laat $P_{\pi_1,\pi_2,k}$ de kansverdeling zijn over de rijen betalingen aan speler 1 als de strategieën π_1 en π_2 worden gekozen en k de begintoestand is. Een limietkriterium is een reëelwaardige functie (W) gedefiniëerd op de verzameling van kansverdelingen P. Strategie π_1 garandeert speler 1 bij begintoestand k en kriterium W de betaling V_k als geldt:

(7.18) $$\inf_{\pi_2} W\,(P_{\pi_1,\pi_2,k}) = V_k.$$

In kompakte notatie voor alle toestanden tezamen wordt dat: strategie π_1 garandeert speler 1 bij kriterium W betaling $\overline{V}$ als geldt:

(7.19) $$\inf_{\pi_2} W\,(P_{\pi_1,\pi_2}) = \overline{V}.$$

Wij noteren $E_{\pi_1,\pi_2,k}$ voor de verwachting m.b.t. de kansverdeling $P_{\pi_1,\pi_2,k}$ en E_{π_1,π_2} voor alle toestanden tezamen.

We onderscheiden de volgende limietkriteria, waarbij $\overline{g}^1, \overline{g}^2, \ldots$ de rij van betalingen is:

het *limiet verdisconteringskriterium:*

(7.20) $$W\,(P_{\pi_1,\pi_2}) = E_{\pi_1,\pi_2} \liminf_{r \downarrow 0} \frac{r}{1+r} \sum_{t=1}^{\infty} \frac{1}{(1+r)^{t-1}}\, \overline{g}^t$$

het *limiet verwachte verdisconteringskriterium:*

(7.21) $$W\,(P_{\pi_1,\pi_2}) = \liminf_{r \downarrow 0} E_{\pi_1,\pi_2} \frac{r}{1+r} \sum_{t=1}^{\infty} \frac{1}{(1+r)^{t-1}}\, \overline{g}^t$$

het *limiet gemiddelde kriterium:*

(7.22) $$W\,(P_{\pi_1,\pi_2}) = E_{\pi_1,\pi_2} \liminf_{t \to \infty} \frac{1}{t} \sum_{\tau=1}^{t} \overline{g}^\tau$$

het *limiet verwachte gemiddelde kriterium:*

$$(7.23)\qquad W\,(P_{\pi_1,\pi_2}) = \liminf_{t\to\infty} E_{\pi_1,\pi_2}\, \frac{1}{t}\sum_{\tau=1}^{t} \bar{g}^{\tau}.$$

Bij de asymptotische kriteria gaan de spelers er vanuit dat het spel stopt of verdisconteerd is, maar dat het aantal stappen t of de rentevoet r niet precies bekend is. In dit geval wordt gekeken naar de verwachte betaling per stap die een speler zich kan garanderen als t groot is uniform in t respectievelijk als $r > 0$ klein is uniform in r.

Wij onderscheiden twee asymptotische kriteria: Strategie π_1 garandeert speler 1 betaling $\bar{V}$ bij het *asymptotisch verdisconteringskriterium* als:

$$(7.24)\qquad \liminf_{r\downarrow 0}\, \inf_{\pi_2} E_{\pi_1,\pi_2} \frac{r}{1+r}\sum_{t=1}^{\infty} \frac{1}{(1+r)^{t-1}}\, \bar{g}^{t} \geq \bar{V}$$

en bij het *asymptotische gemiddelde kriterium* als:

$$(7.25)\qquad \liminf_{t\to\infty}\, \inf_{\pi_2} E_{\pi_1,\pi_2} \frac{1}{t}\sum_{\tau=1}^{t} \bar{g}^{\tau} \geq \bar{V}.$$

Analoge definities gelden voor speler 2 echter $\geq$ wordt $\leq$ en inf wordt vervangen door sup teneinde bij ieder kriterium een nulsomspel te behouden.

Als beide spelers een bepaald kriterium kiezen noemen wij het spel naar het kriterium. Dus b.v. *limiet verdisconteringsspel, limiet verwachte verdisconteringsspel* enz. Gegeven het kriterium worden de waarden en optimale strategieën op de gebruikelijke wijze gedefiniëerd.

Wij noteren $\bar{V}^{\infty} = \lim_{t\to\infty} \frac{1}{t}\bar{V}^{t}$. BEWLEY & KOHLBERG bewijzen de volgende stellingen:

STELLING 7.4.1. *In het asymptotisch verdisconteerde en het asymptotisch gemiddelde spel bestaat er geen strategie voor speler* t *die hem meer garandeert dan* $\bar{V}^{\infty}$.

STELLING 7.4.2. *Als een stationaire strategie* π_1 *speler* 1 $\bar{V} \in \mathbb{R}^N$ *garandeert voor één der* 6 *genoemde kriteria dan garandeert* π_1 *ook* $\bar{V}$ *voor de overige* 5 *kriteria.*

GEVOLG 7.4.3. *Als één der spelers een optimale stationaire strategie heeft* *voor één der 6 kriteria dan is deze strategie optimaal voor alle 6 kriteria.*

DEFINITIE 7.4.4. Een strategie is *optimaal in het niet verdisconteerde* ∞-*stapspel* als deze optimaal is voor alle 6 kriteria.

7.5. Uniforme optimaliteit

DEFINITIE 7.5.1. Een strategie is *uniform verdisconteerd optimaal* als deze optimaal is voor het verdisconteerde spel voor alle rentevoeten r die klein genoeg zijn.

STELLING 7.5.2. *Laat* $\bar{x} \in F^N$ *een oplossing van de limiet verdisconteringsvergelijking dan heeft speler 1 een uniform verdisconteerd optimale strategie dan en slechts dan als er voor* $k = 1,\dots,N$ *een optimale strategie bestaat voor speler 1 voor het spel* $G_k(\bar{x}/(1+\theta^{-1}))$ *waarvan alle komponenten reëel zijn (een reëele strategie). Als* $\pi_1(k) \in \mathbb{R}^{m_k}$ *voor* $k = 1,\dots,N$ *dan is de stationaire strategie* $\pi_1 = (\pi_1(1),\dots,\pi_1(N))$ *dan en slechts dan uniform verdisconteerd optimaal als* $\pi_1(k)$ *optimaal is in* $G_k(\bar{x}/(1+\theta^{-1}))$ *voor* $k = 1,\dots,N$.

OPMERKING 7.5.3. Als een speler een uniform verdisconteerd optimale strategie heeft dan heeft hij een stationaire uniform verdisconteerd optimale strategie. Als π_1 uniform verdisconteerd optimaal is voor speler 1 dan is $\hat{\pi}_1 = (\hat{\pi}_1(1),\dots,\pi_1(N))$ een stationaire uniform verdisconteerd optimale strategie als $\hat{\pi}_1(k)$ door π_1 wordt voorgeschreven in toestand k op één of ander tijdstip.

STELLING 7.5.4. *Als beide spelers uniform verdisconteerd optimale strategieën hebben en* $\bar{x}$ *is een oplossing van de limiet verdisconteringsvergelijking dan bezitten de komponenten van* $\bar{x}$ *geen gebroken machten van* θ. $\bar{x}$ *is dus van de vorm*

$$\sum_{i=-\infty}^{i_0} \bar{a}^i \theta^i .$$

DEFINITIE 7.5.5. Een strategie π_1 voor speler 1 heet *uniform t-staps optimaal* als de rij

$$V_k^t - \inf_{\pi_2} E_{\pi_1,\pi_2} \sum_{\tau=1}^{t} \bar{g}^\tau$$

begrensd is voor $k = 1,\dots,N$.

Als $\bar{x} = \bar{g}\theta + \bar{w}$ met $\bar{g},\bar{w} \in \mathbb{R}^N$ dan wordt de limiet recursievergelijking $\tau\bar{x} = \mathrm{val}\ G(\bar{x})$.

$$\bar{g}\theta + \bar{g} + \bar{w} = \mathrm{val}\ G(\bar{g}\theta+\bar{w}).$$

Deze vergelijking noemen wij de *gegeneraliseerde Howard vergelijking.*

<u>STELLING 7.5.6</u>. *Laat* $\bar{x} \in F^N$ *een oplossing van de limiet verdisconteringsvergelijking en laat*

$$\bar{x}/(1+\theta^{-1}) = \bar{a}^{M}\theta + \dots + \bar{a}^{1}\theta^{1/M} + \bar{a}^{0} + \bar{a}^{-1}\theta^{-1/M} + \dots \tag{7.26}$$

Als de spelers reële strategieën $\pi_1(k)$ *en* $\pi_2(k)$ *hebben die* val $G(\bar{x}/(1+\theta^{-1}) + O(\theta^0)$ *garanderen in het spel* $G_k(\bar{x}/(1+\theta^{-1}))$ *voor* $k = 1,\dots,N$ *dan geldt:*

(i) *De stationaire strategieen* $(\pi_i(1),\dots,\pi_i(N))$, $i = 1,2$, *zijn uniform t-staps optimaal en optimaal in het niet-verdisconteerde* ∞*-stapsspel.*

(ii) $\bar{a}^M\theta + \bar{a}^0$ *voldoet aan de gegeneraliseerde Howard vergelijking en* $\pi_1(k)$ *en* $\pi_2(k)$ *zijn optimaal in* $G_k(\bar{a}^M\theta+\bar{a}^0)$ *voor* $k = 1,\dots,N$.

Als beide spelers reële strategieën hebben die optimaal zijn in $G_k(\bar{x}/(1+\theta^{-1}))$ dan hebben zij beiden stationaire uniform verdisconteerd optimale strategieën. Bovendien bevat de oplossing van de limiet verdisconteringsvergelijking geen gebroken machten van θ.

<u>GEVOLG 7.5.7</u>. *Laat* $\bar{x} \in F^N$ *een oplossing van de limiet verdisconteringsvergelijking en de spelers hebben reële optimale strategieën* $\pi_1(k)$ *en* $\pi_2(k)$ *voor de spelen* $G_k(\bar{x}/(1+\theta^{-1}))$ *voor* $k = 1,\dots,N$ *dan geldt:*

(i) *De stationaire strategieën* $\pi_1 = (\pi_1(1),\dots,\pi_1(N))$ *en* $\pi_2 = (\pi_2(1),\dots,\pi_2(N))$ *zijn uniform verdisconteerd optimaal, uniform t-staps optimaal en optimaal in het niet verdisconteerde* ∞*-stapsspel.*

(ii) *De komponenten van* $\bar{x}$ *hebben geen gebroken machten van* θ.

(iii) *Als* $\bar{x}/(1+\theta^{-1}) = \bar{g}\theta + \bar{w} + O(\theta^0)$ *dan voldoet* $\bar{g}\theta + \bar{w}$ *aan de gegeneraliseerde Howard vergelijking en* $\pi_1(k)$ *en* $\pi_2(k)$ *zijn optimaal in* $G_k(\bar{g}\theta+\bar{w})$.

Met behulp van een voorbeeld wordt aangetoond dat de voorwaarden in

Stelling 7.5.6 voor het bestaan van uniform t-staps optimale strategieën niet noodzakelijk zijn. Een volgend voorbeeld is een spel waarin een speler geen optimale strategie heeft in het verdisconteerde ∞-stapsspel, maar toch de komponenten van de oplossing van de limiet verdisconteringsvergelijking alleen gehele machten van θ bezitten. Een derde voorbeeld laat zien dat er spelen bestaan die uniform t-staps optimale stragieën, maar geen uniform verdisconteerd optimale strategieën bezitten.

Veel reeds bekende resultaten voor Markov beslissingsprocessen met eindige aktie- en toestandsruimten blijken uit de stellingen van BEWLEY & KOHLBERG te volgen. Bekende resultaten op het gebied van de stochastische spelen kunnen eveneens m.b.v. deze stellingen worden bewezen, bijvoorbeeld:

(i) In een niet-verdisconteerd ∞-staps stochastisch spel met volledige informatie hebben beide spelers optimale stationaire strategieën (GILETTE [18] & LIGGETT & LIPPMAN [34]).

(ii) In een niet verdisconteerd ∞-staps irreducibel stochastisch spel hebben beide spelers optimale stationaire strategieën (GILETTE [18] en HOFFMAN & KARP [22]).

(iii) Als in een niet-verdisconteerd ∞-staps stochastisch spel de overgangskansen onafhankelijk zijn van de akties gekozen door één der spelers dan hebben beide spelers optimale strategieën (STERN [61]).

(iv) In een niet-verdisconteerd ∞-stapsspel waarin alle toestanden behalve de begintoestand absorberend zijn bestaat $\lim \frac{1}{t}\, \overline{V}^t$ (KOHLBERG [31]).

Laat A de collectie van niet-verdisconteerde ∞-staps stochastische spelen; B de spelen uit A met voor beide spelers ε-optimale strategieën; C de spelen uit B met voor beide spelers optimale strategieën en D de spelen uit C met voor beide spelers optimale stationaire strategieën. Dan volgt uit het voorbeeld van BLACKWELL & FERGUSON [8] dat B $\neq$ C. Uit een voorbeeld van STERN [61] volgt C $\neq$ D. BEWLEY & KOHLBERG vermelden de volgende onopgeloste problemen:

(i) Karakterisering van C.

(ii) Karakterisering van D.

(iii) Geldt A = B?

Van het volgende voorbeeld is niet bekend of het een element van B is.

VOORBEELD 7.5.8.

$$\Gamma_1 : \begin{bmatrix} \Gamma_2 & -1+\Gamma_1 \\ \Gamma_3 & \Gamma_2 \end{bmatrix} \qquad \Gamma_2 : \begin{bmatrix} \Gamma_1 & \Gamma_4 \\ 1+\Gamma_2 & \Gamma_1 \end{bmatrix}$$

$$\Gamma_3 : [\ -1 + \Gamma_3] \qquad \Gamma_4 : [\ 1 + \Gamma_4]$$

HOOFDSTUK VIII

OVERZICHT VAN RECENTE ONTWIKKELINGEN VOOR STOCHASTISCHE SPELEN

8.1. Inleiding

In de voorgaande hoofstukken is slechts een gedeelte van de literatuur over stochastische spelen aan de orde gekomen; voornamelijk resultaten gepubliceerd voor 1970. De ontwikkelingen na die tijd worden hieronder samengevat.

Deze ontwikkelingen kunnen in een aantal hoofdrichtingen worden opgesplitst.

a. Algoritmen om de waarde van (tee-persoons nulsom) spelen en (ε-) optimale stategieën voor de spelers te bepalen.

b. Waarde-existentie vragen voor eindige niet-stoppende twee-persoons nulsom spelen onder het gemiddelde kosten kriterium.

c. Het bestaan van (ε-) evenwichtspunten voor spelen, waarbij de toestandsruimte, spelersverzamelingen en aktieruimten steeds algemenere vormen aannemen.

d. Toepassingen van stochastische spelen.

In de volgende sekties worden voor ieder van deze gebieden de belangrijkste publikaties genoemd.

Dit overzicht pretendeert niet uitputtend te zijn, maar tracht een globaal inzicht te geven in recent ontwikkelingen op het gebied van de stochastische spelen. Een meer gedetailleerd overzicht kan worden gevonden in PARTHASARATHY & STERN [48].

8.2. Algoritmen

Alle tot nu toe ontwikkelde algoritmen op het gebied van de stochastische spelen zijn gebaseerd op de methode der successieve approximatie en hebben bijna allen betrekking op eindige twee-persoons nulsomspelen. In hoofdstuk 6 zijn de successieve approximatiemethoden van SHAPLEY [57], HOFFMAN & KARP [22], POLLATSCHEK & AVI-ITZHAK [49], RIOS & YANEZ [52] en VAN DER WAL [69] behandeld. Als aanvulling hierop noemen we:

a. VAN DER WAL [70] presenteert een verzameling algoritmen voor verdisconteerde (of stoppende) spelen, gebaseerd op overgangs-geheugenloze stoptijden, zoals geïntroduceerd door WESSELS [73]. Met behulp hiervan kunnen o.a. betere grenzen voor de Shapley methode gegeven worden. Een hiervan afgeleid algoritme lost spelen op waarvoor alle direkte opbrengsten positief zijn en de minimaliserende speler in iedere toestand een aktie heeft waardoor het spel onmiddellijk stopt.

b. PARTHASARATHY & RAGHAVAN [47] beschouwen spelen waarbij voor ieder paar strategieën van de beide spelers de matrix van overgangskansen uitsluitend afhankelijk is van speler 1. Onder deze aanname kunnen ze laten zien, dat voor het verdisconteerde model het vinden van de waarde en optimale strategieën voor de beide spelers geformuleerd kan worden als een lineair programmeringsprobleem, zodat er voor dit geval een eindig algoritme is.

c. KUSHNER & CHAMBERLAIN [33] beschouwen twee-persoons nulsomspelen met eindige toestandsruimte en kompakte aktieruimten voor beide spelers. Onder een aantal voorwaarden (1^0. $g(k_1,a_1,a_2) < \infty$ óf voor ieder paar strategieën eindigt het spel met kans $p > 0$ voor tijdstip N; 2^0. $g(k,a_1,a_2) > 0$; 3^0. de dummy-spelen bezitten een waarde voor alle $\bar{v} \in \mathbb{R}^N$; 4 . $p(\ell|k_1,a_1,a_2)$ en $g(k_1,a_1,a_2)$ zijn continu in a_1 en a_2) laten ze zien, dat successieve approximatie een benadering van de waarde en ε-optimale zuivere strategieën geeft voor het totale opbrengsten model.

8.3. Gemiddelde kosten kriterium

In de literatuur is reeds veel aandacht besteed aan de vraag of ieder eindig twee-persoons nulsomspel onder het gemiddelde kosten kriterium een waarde bezit. Tot nu toe is het bestaan van de waarde alleen voor speciale gevallen aangetoond. De artikelen van GILLETTE [18], LIGGET & LIPPMAN [34] en HOFFMAN & KARP [22], die in hoofstuk 5 aan de orde gekomen zijn, zijn hier voorbeelden van.

Een voorbeeld van GILLETTE [18] is door BLACKWELL & FERGUSON [8] uitgewerkt, waaruit blijkt dat speler 1 geen optimale strategie bezit (zie hoofdstuk 5); wel bezit het spel een waarde en geven ze ε-optimale historie afhankelijke strategieën voor speler 1. In HORDIJK, VRIEZE & WANROOIJ [24] wordt aangetoond, dat dit spel geen waarde bezit binnen de klasse der semi-Markov-strategieën. Alhoewel enigzins verstopt, geeft ook ZAMIR [78] een spel met een dergelijke eigenschap, waarbij gebruik gemaakt moet worden van het feit dat stochastische spelen, waarbij de beide spelers zich beperken tot de klasse der semi-Markov-strategieën opgevat kunnen worden als herhaalde spelen met niet-volledige informatie.

STERN [61] laat zien, dat, voor spelen, waarbij er een toestand k bestaat, z.d.d. voor ieder paar stationaire strategieën en vanuit iedere toestand ℓ er een positieve kans bestaat, dat toestand k bereikt wordt, het spel een waarde bezit en de beide spelers optimale stationaire strategieën hebben.

Tevens laat STERN [61] zien, dat, als voor ieder paar stationaire strategieën de matrix van overgangskansen uitsluitend van de strategie van speler 1 afhankelijk is, het spel een waarde bezit en de beide spelers optimale semi-Markov-strategieën bezitten.

In [15] geeft FEDERGRÜN twee typen spelen, waarvoor het bestaan van een waarde en optimale stationaire strategieën voor de beide spelers aangetoond wordt. In het eerste type is er een voorwaarde opgelegd op de struktuur van de Markov-ketens bij de overgangskansen matrices van de zuivere stationaire strategieën, waarvan het gevolg is dat de Markov-ketens bij ieder paar stationaire strategieën evenveel deelketens bezitten.

In het tweede type spel is de voorwaarde dat speler 1 voor iedere stationaire strategie π_2 van speler 2 een stationaire strategie π_1 bezit, zodanig dat voor ieder tweetal toestanden k en ℓ er een integer t bestaat met $p^t(\ell|k,\pi_1,\pi_2) > 0$ (de t-stapsoverganskans). Deze voorwaarde is een uitbreiding van de "communicating" eigenschap van BATHER [2].

In feite bewijst Federgrün het bovenstaande voor het N-persoonsspel. Hij bereikt zijn resultaten door eerst de verdisconteerde spelen te beschouwen en vervolgens te laten zien, dat $\lim_{\beta\uparrow 1} (1-\beta)V_\beta^*$ bestaat, waarbij V_β^* de waarde van het spel is bij verdisconteringsfactor β.

KOHLBERG [31] behandelt wat hij noemt "herhaalde spelen met absorberende toestanden". Hiermee bedoelt hij een twee-persoons nulsomspel met één doorgangstoestand. De andere toestanden zijn allen absorberend en beide spelers hebben in ieder van deze toestanden 1 aktie ter beschikking. Kohlberg laat zien dat dit spel een waarde heeft, die gevonden kan worden als de limiet van de waarden voor de t-stapsspelen als t naar ∞ gaat.

In [16] geeft FEDERGRÜN een successieve approximatie algoritme voor gemiddelde kosten spelen. Hiertoe introduceert hij een stelsel functionaal vergelijkingen, die gezien kunnen worden als uitbreidingen van het stelsel functionaal vergelijkingen ter oplossing van niet-verdisconteerde Markov-beslissingsproblemen. Het blijkt dat het bestaan van een oplossing van dit stelsel functionaal vergelijkingen een nodige (en niet voldoende) voorwaarde is voor het bestaan van een evenwichtspunt van stationaire strategieën. Het bestaan van een evenwichtspunt van stationaire strategieën blijkt een voldoende voorwaarde te zijn opdat het bovengenoemd algoritme convergeert.

Wellicht de belangrijkste artikelen op het gebied van de eindige twee-persoons nulsomspelen onder het gemiddelde kosten kriterium zijn de artikelen van BEWLEY & KOHLBERG [3], [4] en [5]. Een uitgebreide samenvatting hiervan is te vinden in hoofdstuk 7.

8.4. Spelen onder algemenere condities

We kunnen onderscheid maken tussen twee-persoons nulsomspelen en niet-nulsomspelen, alhoewel de resultaten voor deze laatste klasse van spelen natuurlijk ook op de eerste van toepassing zijn.

a. Twee-persoons nulsomspelen.

Wat de twee-persoons nulsomspelen betreft, noemen we de artikelen van WESSELS [74], GROENEWEGEN & WESSELS [19], KUSHNER & CHAMBERLAIN [33], MAITRA & PARTHASARATHY [36] en [37], PARTHASARATHY [45] en [46] en KAMERUD [27].

WESSELS [74] behandelt spelen, waarbij de toestandsruimte aftelbaar is en de aktieruimten voor de beide spelers eindig zijn in iedere toestand. De opbrengstfunctie is niet noozakelijkerwijs begrensd; echter er moet een positieve functie $\mu(\cdot)$ op S bestaan en getallen $M > 0$ en $\beta \in (0,1)$, z.d.d.

$$|g(s,a_1(s),a_2(s))| \le \frac{M}{\mu(s)}$$

en

$$\sum_{s'\in S} p(s'|s,a_1(s),a_2(s)) \cdot \frac{1}{\mu(s)} \le \frac{\beta}{\mu(s)} \; .$$

Onder deze condities bezit het spel een waarde onder het totale kosten kriterium en hebben de spelers optimale stationaire strategieën.

GROENEWEGEN & WESSELS [19] karakteriseren voor het model waarbij de toestandsruimte en de aktieruimten aftelbaar zijn, paren optimale strategieën. Als voorwaarde leggen ze op, dat onder ieder paar Markov-strategieën de totale verwachte opbrengst eindig is. Op twee manieren breiden ze de "conserving"-eigenschap voor strategieën uit de Markov-beslissingstheorie (geïntroduceerd door DUBINS & SAVAGE [12] en uitgewerkt door HORDIJK [23]) uit tot de stochastische spelen die respectievelijk "saddle-conserving-" eigenschap en "saddling-"eigenschap voor paren strategieën (van beide spelers één) genoemd wordt. Het blijkt dat paren optimale strategieën noodzakelijkerwijs "saddle-conserving" zijn, maar dat "saddle-conserving" niet voldoende is om een paar optimale strategieën te garanderen. Anderzijds is "saddling" wel voldoende; echter Groenewegen en Wessels geven een spel, waaruit blijkt, dat "saddling" niet nodig is.

De resultaten van KUSHNER & CHAMBERLAIN [33] (eindige toestandsruimte, kompakte aktieruimten) zijn reeds in sektie 8.2 genoemd.

In MAITRA & PARTHASARATHY [36] wordt verondersteld dat S, A_1 en A_2 kompakte metrische ruimten zijn. Onder de voorwaarden dat de uitbetalingsfunctie continue is op $S\times A_1\times A_2$ en dat $p(\cdot|s_n,a_{1n},a_{2n})$ (kansmaat op S) convergeert naar $p(\cdot|s_0,a_{10},a_{20})$ in de zwakke topologie op de verzameling van Borelkansmaten op S, als $(s_n,a_{1n},a_{2n}) \to (s_0,a_{10},a_{20})$, tonen ze het bestaan van een continue waardefunctie en meetbare optimale stationaire strategieën voor beide spelers aan onder het verdisconteerde opbrengsten kriterium.

In latere artikelen breidt PARTHASARATHY [45] en [46] dit model uit tot een model waarbij de toestandsruimte een Boreldeelverzameling van een volledige separabele metrische ruimte is en waarin de aktieruimten niet in alle toestanden hetzelfde hoeven te zijn. Wel worden er zwaardere continuïteitseisen aan de uitbetalingsfunctie en de overgangskansenfuncties opgelegd.

KAMERUD [27] behandelt positieve spelen. Voor S aftelbaar en A_1 en A_2

eindig toont hij aan, dat de functie $V(\cdot)$ waardefunctie van het spel is onder het totale opbrengstenkriterium als geldt $V(s) = \lim_{\beta\uparrow 1} V_\beta^*(s)$, $\forall s\in S$, waarbij $V_\beta^*(\cdot)$ de waardefunctie van het spel is onder het verdisconteerde opbrengsten kriterium. Tevens heeft speler 2 een optimale stationaire strategie en speler 1 een ε-optimale stationaire strategie (alleen afhankelijk begintoestand en lopende toestand).

Ook in MAITRA & PARTHASARATHY [37] worden positieve spelen beschouwd. In hun model zijn S_1, A_1 en A_2 kompakte metrische ruimten en is verondersteld, dat onder ieder paar strategieën de totale verwachte opbrengst uniform begrensd is. Onder dezelfde continuïteitseisen als in hun eerste artikel (zie boven MAITRA & PARTHASARATHY [36]) tonen ze het bestaan van een waarde, optimale stationaire strategieën voor speler 2 en ε-optimale stationaire strategieën voor speler 1 aan onder het totale kosten kriterium. Merkwaardigerwijs beschouwt Parthasarathy in een later artikel [45] dit zelfde probleem nog eens voor een eenvoudiger model, namelijk S_1, A_1 en A_2 eindig en verder dezelfde condities, alwaar echter het bewijs niet correct is.

b. Niet-nulsomspelen.

Stochastische twee-persoons niet-nulsomspelen zijn geïntroduceerd door ROGERS [53]. Hij beschouwt spelen met eindige toestands- en aktieruimten en bewijst voor het verdisconteerde model het bestaan van een evenwichtspunt. Ook onder het gemiddelde kosten kriterium toont hij voor hetzelfde model het bestaan van een evenwichtspunt aan onder de voorwaarde, dat voor ieder paar zuivere stationaire strategieën de Markov-keten behorende bij de matrix van overgangskansen één kernfuik bevat.

PARTHASARATHY [45] laat zien, dat ook voor een aftelbare toestandsruimte Roger's resultaat voor het verdisconteerde model geldig is.

SOBEL [59] heeft n-persoons stochastische spelen geïntroduceerd. Voor spelen met eindige toestandsruimte en eindige aktieruimten voor de diverse spelers heeft hij het bestaan van een evenwichtspunt aangetoond voor het verdisconteerde model binnen de klasse der stationaire strategieën. In hetzelfde artikel behandelt Sobel n-persoons stochastische spelen voor het gemiddelde kosten kriterium. Onder de voorwaarde, dat voor alle n-tallen zuivere stationaire strategieën de Markov-keten behorende bij de matrix van overgangskansen één kernfuik bevat, toont hij het bestaan van een evenwichtspunt aan. De betreffende stelling van Sobel is ruimer geformuleerd, doch het bewijs dekt alleen het bovenstaande.

De resultaten uit FEDERGRÜN [15] voor het n-persoonsspel met eindige toestands- en aktieruimten zijn reeds in sektie 8.3 genoemd. Onder strengere condities dan de aldaar genoemde op de Markovketenstructuur behorende bij de overgangskansenmatrices bij de n-tallen stationaire strategieën, toont FEDERGRÜN [15] het bestaan van een evenwichtspunt binnen de klasse der stationaire strategieën aan onder het gemiddelde kosten kriterium in spelen met aftelbare toestandsruimte en kompakte metrische aktieruimten en waarbij de uitbetalingsfunctie en de overgangskansenfuncties continu zijn in de gezamenlijke akties van de spelers. Dit resultaat wordt bereikt door eerst het verdisconteerde model te beschouwen, waarbij het bestaan van een evenwichtspunt aangetoond wordt, zonder dat gebruik gemaakt hoeft te worden van de voorwaarden op de ketenstructuur bij de verschillende n-tallen stationaire strategieën en vervolgens een rij verdisconteringsfactoren te kiezen, die naar 1 gaan.

Recentelijk heeft IDZIK [26] het twee-persoons verdisconteerde tweetraps stochastisch spel geïntroduceerd, waarmee bedoeld wordt een spel bepaald door de grootheden $S,A_1,A_2,g_1,g_2,p,\phi,\psi$; hierin hebben S,A_1,A_2,g_1,g_2 en p de gebruikelijke betekenis; ϕ: $S\times P_{A_1}\times P_{A_2} \to 2^{A_1}$ (verzameling van alle gesloten deelverzamelingen van A_1) en ψ: $S\times P_{A_1}\times P_{A_2} \to 2^{A_2}$ zijn Borel meetbare afbeeldingen (P_{A_1} en P_{A_2} zijn de verzamelingen van alle Borel kansmaten op A_1 respectievelijk A_2). In de eerste stap (de eerste trap) kiezen de spelers strategieën π_1 en π_2, waarbij π_1: $S \to P_{A_1}$ en π_2: $S \to P_{A_2}$ Borel-meetbare afbeeldingen zijn. Na de eerste stap volgt een normaal stochastisch spel (tweede trap) waarbij speler 1 in toestand $s \in S$ de aktieruimte $\phi(s,\pi_1(s),\pi_2(s))$ ter beschikking heeft en speler 2 $\psi(s,\pi_1(s),\pi_2(s))$. Idzik bewijst het bestaan van een evenwichtspunt voor S aftelbaar, A_1 en A_2 kompakt metrisch en de nodige continuïteitsvoorwaarden op g_1,g_2,p,ϕ en ψ.

Het verdisconteerde model is door VRIEZE [67] uitgebreid tot spelen waarbij aftelbare spelersverzamelingen toegelaten zijn. Onder verder dezelfde condities als in FEDERGRÜN [15] wordt het bestaan van een evenwichtspunt binnen de klasse der stationaire strategieën aangetoond, waarbij gebruik gemaakt wordt van een afbeelding, die gezien kan worden als een rechtstreekse uitbreiding van de afbeelding met behulp waarvan NASH [39] de minimax-stelling voor twee-persoonsnulsomspelen bewijst. In het tweede gedeelte van het artikel van Vrieze wordt aangetoond, dat dit evenwichtspunt binnen de klasse der stationaire strategieën ook evenwichtspunt is binnen de klasse van alle (gedrags-) strategieën.

Wat de n-persoons verdisconteerde spelen betreft, waarbij de toestandsruimte overaftelbaar is, zijn er in de literatuur meerdere pogingen gedaan om het bestaan van een evenwichtspunt onder geschikte continuïteitseisen aan te tonen. Zowel in SOBEL [60] als in FEDERGRÜN, e.a. [14] is een incorrect bewijs hiervoor gegeven, zodat dit vooralsnog een open probleem lijkt.

Zeer recent heeft WHITT [76] een geheel nieuwe aanpak gepresenteerd om dit probleem te attaqueren. Een spel met overaftelbare toestandsruimte kan worden benadered door een rij spelen met aftelbare toestandsruimte (waarvoor bestaan van evenwichtspunten bekend is) en wel zodanig, dat op den duur evenwichtspunten van de benaderende spelen overeenkomen met ε-evenwichtspunten van het oorspronkelijke spel. Op deze manier heeft Whitt, gebruik makend van het monotone contractie operator kader van DENARDO [10], het bestaan van ε-evenwichtspunten aangetoond voor verdisconteerde spelen met de toestandsruimte separabel metrisch, kompakt metrische aktieruimten, die tevens voor ieder van de n spelers uniform continu over de toestandsruimte is en waarbij de opbrengstfuncties en de overgangskansenfunctie aan de nodige continuïteitsvoorwaarden voldoen.

Tenslotte noemen we het begrip p-evenwichtspunt, zoals geïntroduceerd door HIMMELBERG, PARTHASARATHY, RAGHAVAN en VAN VLECK [21] voor twee-persoons niet-nulsomspelen. Voor p^* een kansmaat op de toestandsruimte wordt een p^*-evenwichtspunt gedefinieerd als een paar (π_1^*,π_2^*) waarvoor geldt:

$$p\{s \mid V_1(s,\pi_1^*,\pi_2^*) \geq V_1(s,\pi_1,\pi_2^*),\ \forall\pi_1 \text{ en } V_2(s,\pi_1^*,\pi_2) \geq V_2(s,\pi_1^*,\pi_2), \forall\pi_2^*\} = 1.$$

Ze bewijzen het bestaan van een p -evenwichtspaar binnen de klasse der stationaire strategieën voor verdisconteerde spelen, waarbij $S = [0,1]$, $A_1(s)$ en $A_2(s)$ zijn eindig, $g_n(s,a_1,a_2)$ heeft de gedaante $g_n(s,a_1,a_2) = h_n(s,a_1) + k_n(s,a_2)$ met $h_n(\cdot,a_1)$ en $k_n(\cdot,a_2)$ begrensde meetbare functies, $n = 1,2$ en $(p\cdot|s,a_1,a_2)$ kan worden geschreven als $p(\cdot|s,a_1,a_2) = [p_1(\cdot|s,a_1) + p_2(\cdot|s,a_2)]/_2$ waarbij p_1 en p_2 kansmaten zijn, die meetbaar in s zijn en tevens moet $p(\cdot|s,a_1,a_2)$ absoluut continu zijn met betrekking tot de kansmaat p^* voor iedere (s,a_1,a_2).

8.5. Toepassingen

Tot nu toe zijn er weinig toepassingen bekend van de theorie der stochastische spelen. De meeste toepassingen zijn bovendien nog enigszins theoretisch van aard. In hoeverre het werk van FILAR [17] hierop een uitzondering vormt is ons niet bekend. Een overzicht van de toepassingen der theorie der stochastische spelen, kan worden gevonden in VRIEZE [68], waarin met name de toepassingen op het gebied van de zoekmodellen uitgewerkt zijn (voor zoekmodellen zie bijv. SAKAGUCHI [54].)

Een van der eerste artikelen in de literatuur, waarin een toepassing van de stochastische spelen uitgewerkt is, is geschreven door CHARNESS en SCHROEDER [9]. Ze beschrijven het probleem van de onderzeeër, die een route moet bepalen om van een bepaald punt naar een ander punt te varen, waarbij de kosten geminimaliseerd moeten worden. Een onderdeel van de kosten is de kans op opsporing. Als tegenstander fungeert een opsporingsboot die, de zoekprocedure zoekt, die de kans op opsporing maximaliseert. Charness en Schroeder modelleren deze situatie als een eindig stoppend twee-persoons nulsomspel, waarvan ze de oplossing benaderen met behulp van een eindige rij lineaire programmerings problemen. Vervolgens modelleren ze het probleem als een eindigstapsspel.

Een toepassing in de economie is te vinden in het artikel van KIRMAN en SOBEL [30], waarin ze een dynamisch model van een oligopolie ontwikkelen. Iedere periode nemen n firma's gelijktijdig een beslissing ontrent de prijs voor hun product en de te produceren hoeveelheid. De toestand waarin het systeem verkeert, wordt bepaald door de voorraden van de firma's. De vraag in een bepaalde periode bij een bepaalde firma is een stochastische variabele, die afhangt van de gezamenlijk aangekondigde prijzen. De winst per periode voor een bepaalde firma is gelijk aan een opbrengstfunctie (akhankelijk van zijn geproduceerde hoeveelheid, zijn prijs en zijn vraag) verminderd met zijn lineaire produktiekosten. De opbrengsten worden verdisconteerd (iedere firma heeft eigen verdisconteringsfactor). Door de opbrengst gedeeltelijk over een periode te verschuiven verkrijgen Kirman en Sobel een uitdrukking voor de verwachte opbrengst per periode, die niet meer afhangt van de toestand, doch uitsluitend van de gekozen akties.

Tevens is de kansverdeling voor de overgang naar een nieuwe toestand onafhankelijk van de oude toestand. Op deze wijze is het oligopolie model geschreven als een verdisconteerd n-persoons stochastisch spel. De auteurs leggen vervolgens condities op aan het éénstapsspel, waardoor dat een evenwichtspunt van zuivere stationaire strategieën bezit. In dit geval geldt dat het herhaald spelen van de akties die door dit evenwichtspunt worden voorgeschreven een evenwichtspunt van zuivere stationaire strategieën vormt in het stochastische spel, mits de beginvoorraad maar niet groter is als het niveau waartoe men volgens deze strategie de voorraad moet aanvullen.

Andere toepassingen kunnen worden gevonden in SHUBIK en WHITT [58] en in ZACHRISSON [77].

REFERENTIES:

[1] AUMANN, R.J., *Mixed and Behaviour strategies in infinite extensive games, Advances in Game Theory,* Ann. Math. Studies no. 52, Princeton Univ. Press, Princeton (1964), 627-650.

[2] BATHER J., *Optimal decision procedures for finite Markov-chains I, II,* Adv. in Applied Prob. 5, (1973).

[3] BEWLEY, T. & E. KOHLBERG, *The asymptotic theory of stochastic games,* Mathematics of Operations Research 1, (1976), 197-208.

[4] BEWLEY, T. & E. KOHLBERG, *The asymptotic solution of a recursion equation arising in stochastic games,* Mathematics of Operations Research 1, (1976), 321-336.

[5] BEWLEY, T. & E. KOHLBERG, *On stochastic games with stationary optimal strategies,* Technical report No. 23, Harvard University (1976).

[6] BLACKWELL, D., *Discrete dynamic programming,* Annals of Math. Stat. 33, (1962), 719-726.

[7] BLACKWELL, D., *Discounted dynamic programming,* Ann. Math. Stat. 36, (1965), 226-235.

[8] BLACKWELL, D. & T.S. FERGUSON, *The big match,* Ann. Math. Stat. 39, (1968), 159-163.

[9] CHARNESS, A.M. & R.E. SCHROEDER, *On some stochastic tactical antisubmarine games,* Naval Res. Log. Quarterly 14, (1967), 291-312.

[10] DENARDO, E.V., *Contraction mappings in the theory underlying dynamic programming,* Siam Rev. 9, (1976), 165-177.

[11] DERMAN, C., *On seqential decisions and Markov-chains,* Management Science, (1962), 16-24.

[12] DUBINS, L. & L.J. SAVAGE, *How to gamble if you must,* Mc. Grawhill, New York, (1965).

[13] EVERETT, H., *Recursive games,* Ann. Math. Studies no 39, Princeton Univ. Press., Princeton, (1957), 15-45.

[14] FEDERGRÜN, A., in coöperation with O.J. VRIEZE & G.J. WANROOIJ, *On the existence of discounted and average return equilibrium policies in* N-*person stochastic games,* report BW 57/75, Mathematisch Centrum, Amsterdam, (1976).

[15] FEDERGRÜN, A., *On* N-*person games with denumerable state space,* report BW 67/76, Mathematisch Centrum, Amsterdam, (1976).

[16] FEDERGRÜN A., *On the functional equations in undiscounted and sensitive discounted stochastic games,* report BW 73/77, Mathematisch Centrum, Amsterdam, (1977).

[17] FILAR, J.A., *Markov games and the theory of duopoly,* M.S. dissertation submitted to Monash University, (1975).

[18] GILETTE, D., *stochastic games with zero-stop probabilities,* Ann. Math. Studies no. 39, Princeton University Press, Princeton, (1957), 179-187.

[19] GROENEWEGEN, L.P.J. & J. WESSELS, *On the relation between optimality and saddle conservation in Markov games,* Technological University, Eindhoven (Department of Mathematics), Memorandum COSOR 76-14, (1976).

[20] HANSNER, M., *Optimal strategies in games of survival,* Research Memorandum RM 776, The RAND corporation, (1952).

[21] HIMMELBERG, C.J., T. PARTHASARATHY, T.E.S. RAGHAVAN & F.S. VAN VLECK, *Existence of* p-*equilibrium and optimal strategies in stochastic games,* Proc. Amer. Math. Soc. vol. 60, (1976), 245-251.

[22] HOFFMAN, A.D. & R.M. KARP, *On non-terminating stochastic games,* Management Science 12, (1966), 359-370.

[23] HORDIJK, A., *Dynamic programming and Markov potential theory,* Mathematic Centre Tract no. 51, Mathematisch Centrum, Amsterdam (1974).

[24] HORDIJK, A., O.J. VRIEZE & G.L. WANROOIJ, *Semi-Markov strategies in stochastic games,* report BW 68/76, Mathematisch Centrum, Amsterdam, (1976).

[25] HOWARD, R., *Dynamic programming and Markov processes,* MIT Press, Cambridge, Massachusetts, (1960).

[26] IDZIK, A., *Two-stage non-coöperative discounted stochastic games,* Preprint, (1976).

[27] KAMERUD, D.B., *Repeated games and positive stochastic games,* Ph. D. dissertation submitted to the University of Minnesota, (1975).

[28] KAPLANSKY, I., *A contribution to Von Neumann's theory of games,* Annals of Mathematics 46, (1945), 474-479.

[29] KELLEY, J., *General Topology,* D. Von Nostrand Company, Inc. Princeton, (1955).

[30] KIRMAN, A.P. & M.J. SOBEL, *Dynamic oligopoly with inventories,* Econometrica 42, (1974), 279-287.

[31] KOHLBERG, E., *Repeated games with absorbing states,* Ann. Stat. 2, (1974), 724-738.

[32] KUHN, H., *Extensive games and the problem of information,* Ann. Math. Studies no. 28, Princeton Univ. Press, Princeton, (1953), 193-216.

[33] KUSHNER, H.J. & S.G. CHAMBERLAIN, *Finite state stochastic games: existence theorems and computational procedure,* IEEE, Trans Automatic Control AC-14, (1969), 248-255.

[34] LIGGETT, T.M. & S.A. LIPPMAN, *Stochastic games with perfect information and time average payoff,* Siam Rev. (1969), 604-607.

[35] LUCE, R. & H. RAIFFA, *Games and decisions,* John Wiley, New York, (1957).

[36] MAITRA, A. & T. PARTHASARATHY, *On stochastic games,* Jour. Opt. Theory and its Appl. 5, (1970), 289-300.

[37] MAITRA, A. & T. PARTHASARATHY, *On stochastic games II,* Jour. Opt. Theory and its Appl. 7, (1971), 154-160.

[38] MILNOR, J. & L.S. SHAPLEY, *Games of survival,* Annals of Mathematical Studies no. 39, Princeton University Press, Princeton, (1957), 47-78.

[39] NASH, J.F., *Equilibrium points in N-person games,* Proc. Nat. Acad. Sci. USA 36, (1950), 48-49.

[40] NEUMANN, J. von, *Zur Theory der Gesellschaftsspiele,* Mathematische Annalen 100, (1928), 295-320.

[41] NEUMAN, J. Von, & O. MORGENSTERN, *Theory of games and economic behaviour,* Princeton Univ. Press, Princeton, (1944).

[42] NUNEN, J.A.E.E. van, *A set of successive approximation methods for discounted Markov decision processes,* Zeitschrift für Operations Research, Vol. 20, (1976), 203-208.

[43] ORKIN, M., *Recursive matrix games,* Jour. Appl. Prob. 9, (1972) 813-820.

[44] PARTHASARATHY, K.R., *Probability measures on metric spaces,* Academic Press, New York, (1967).

[45] PARTHASARATHY, T., *Discounted and positive stochastic games,* Bull. Amer. Math. Soc. 77, (1971), 134-136.

[46] PARTHASARATHY, T., *Discounted, positive and non-cooperative stochastic games,* Int. Jour. Game Theory 2, (1973), 25-37.

[47] PARTHASARATHY, T. & T.E.S. RAGHAVAN, *Finite algorithms for stochastic games,* submitted for publication, (1976).

[48] PARTHASARATHY, & M. STERN, *Markov games - a survey,* Univ. of Illinois, Chicago and Analytic Services Inc., Virginia, (1977).

[49] POLLATSCHECK, M.A. & B. AVI-ITZHAK, *Algorithms for stochastic games with geometrical interpretation,* Management Science, Vol. 15, (1969), 399-415.

[50] PORTEUS, E.L., *Bounds and transformations for discounted finite Markov decision chains,* Operation Research, Vol. 23, (1975), 761-784.

[51] RAO, S.S., R. CHANDRASEKARAN & K.P.K. NAIR, *Algorithms for discounted stochastic games,* Jour. Opt. Theory and its Appl., Vol. 11, (1973), 627-637.

[52] RIOS, S. & I. YANEZ, *programmation sequentielle en concurrence,* Research Papers in Statistics, edited by F.N. David, John Wiley, New York, (1966), 289-299.

[53] ROGERS, P.D., *Non zero-sum stochastic games,* Report ORC 69-8, Operations Research Centre, Univ. of California, Berkely: Ph. D. dissertation (1969).

[54] SAKAGUCHI, M., *Two-sided search games,* J. of the Operations Research Society of Japan 16, (1972), 207-225.

[55] SCHWARTZ, L., *Topologie générale et analyse fonctionelle,* Hermann, Paris, (1970).

[56] SEIDENBERG, A., *A non decision method for elementary algebra,* Ann. of Math. 60, (1954), 365-374.

[57] SHAPLEY, L.S., *Stochastic games,* Proc. Nat. Acad. Sci. 39, (1953), 1095-1100.

[58] SHUBIK,M. & W. WHITT, *Fiat money in an economy with one non-durable good and no credit, Topics in differential games,* Edited by A. Blaquiere, North Holland Publishing Company, (1973).

[59] SOBEL, M.J., *Non-cooperative stochastic games,* Ann. Math. Stat. 42, (1971), 1930-1935.

[60] SOBEL, M.J., *Continuous stochastic games,* Jour. Appl. Prob. 10, (1973), 597-604.

[61] STERN, M., *On stochastic games with limiting average pay-off,* Ph. D. dissertation, submitted to the Univ. of Illinois, Circle Campus, Chicago, (1975).

[62] TAKAHASHI, M., *Stochastic games with infinitely many strategies,* Jour. Sci. Hiroshima Univ. Series A-I, Mathematics 26, (1962), 123-134.

[63] TARSKI, A., *A decision method for elementary algebra and geometry,* second ed., revised, Univ. of California Press, Berkely and Los Angeles, (1951).

[64] TIJS, S.H., *ε-Equilibrium point theorems for two-person games,* Mathematisch Instituut, Katholieke Universiteit Nijmegen, report 7629, (1976).

[65] VEINOTT, A.F., *Discrete dynamic programming with sensitive discount optimality criteria,* The Annals of Mathematical Statistics 40, (1969), 1635-1660.

[66] VRIEZE, O.J., *Non-cooperative countable games with compact action spaces,* report BW 65/76, Mathematisch Centrum, Amsterdam, (1976).

[67] VRIEZE, O.J., *The stochastic non-cooperative countable person game with countable state space and compact action spaces under the discounted payoff criterion,* report BW 66/76, Mathematisch Centrum, Amsterdam, (1976).

[68] VRIEZE, O.J., *Modelvorming met stochastische spelen,* rapport BC 19/77, Mathematisch Centrum, Amsterdam, (1977).

[69] WAL, J. van der, *The method of successive approximation for the discounted Markov-game,* Technological University Eindhoven, (Department of Mathematics), Memorandum COSOR 75-02, (1975).

[70] WAL, J. van der, *Successive approximation and discounted Markov games,* University of Technology Enschede, (Department of Applied Mathematics), Memorandum nr. 119, (1976).

[71] WALD, A., *Statistical decision functions,* Chelsea Publishing Company, New York, (1950).

[72] WEYL, H., *Elementary proof of a minimax theorem due to von Neumann,* Annals of Math. Studies 24, Princeton University Press, Princeton, (1950), 19-25.

[73] WESSELS, J., *Stopping times and Markov-programming,* to appear in Proceedings of 1974 - EMS-meeting and 7-th Prague Conference on Information Theory, Statistical decision functions and Random processes.

[74] WESSELS, J., *Markov games with unbounded rewards,* Technological University of Eindhoven (Department of Mathematics) Memorandum COSOR 76-05, (1976).

[75] WHEEL, H. van der, *Endogene prijstellingstheorie,* H.E. Stenfert Kroese B.V., Leiden, (1975).

[76] WHITT, W., *Representation and approximation of non-cooperative sequential games,* Operations Research Center, Bell Laboratories, Holmdel, New Jersey, (1977).

[77] ZACHRISSON, L.E., *Markov games,* Annals of Mathematical Studies no. 52, Princeton University Press, Princeton, (1964), 211-253.

[78] ZAMIR, S., *On the notion of values for games with infinitely many stages,* Ann. Stat. 1. (1973), 791-796.

UITGAVEN IN DE SERIE MC SYLLABUS

Onderstaande uitgaven zijn verkrijgbaar bij het Mathematisch Centrum, 2e Boerhaavestraat 49 te Amsterdam-1005, tel. 020-947272.

MCS 1.1 F. GÖBEL & J. VAN DE LUNE, *Leergang Besliskunde, deel 1: Wiskundige basiskennis*, 1965. ISBN 90 6196 014 2.

MCS 1.2 J. HEMELRIJK & J. KRIENS, *Leergang Besliskunde, deel 2: Kansberekening*, 1965. ISBN 90 6196 015 0.

MCS 1.3 J. HEMELRIJK & J. KRIENS, *Leergang Besliskunde, deel 3: Statistiek*, 1966. ISBN 90 6196 016 9.

MCS 1.4 G. DE LEVE & W. MOLENAAR, *Leergang Besliskunde, deel 4: Markovketens en wachttijden*, 1966. ISBN 90 6196 017 7.

MCS 1.5 J. KRIENS & G. DE LEVE, *Leergang Besliskunde, deel 5: Inleiding tot de mathematische besliskunde*, 1966. ISBN 90 6196 018 5.

MCS 1.6a B. DORHOUT & J. KRIENS, *Leergang Besliskunde, deel 6a: Wiskundige programmering 1*, 1968. ISBN 90 6196 032 0.

MCS 1.6b B. DORHOUT, J. KRIENS & J.TH. VAN LIESHOUT, *Leergang Besliskunde, deel 6b: Wiskundige programmering 2*, 1977. ISBN 90 6196 150 5.

MCS 1.7a G. DE LEVE, *Leergang Besliskunde, deel 7a: Dynamische programmering 1*, 1968. ISBN 90 6196 033 9.

MCS 1.7b G. DE LEVE & H.C. TIJMS, *Leergang Besliskunde, deel 7b: Dynamische programmering 2*, 1970. ISBN 90 6196 055 X.

MCS 1.7c G. DE LEVE & H.C. TIJMS, *Leergang Besliskunde, deel 7c: Dynamische programmering 3*, 1971. ISBN 90 6196 066 5.

MCS 1.8 J. KRIENS, F. GÖBEL & W. MOLENAAR, *Leergang Besliskunde, deel 8: Minimaxmethode, netwerkplanning, simulatie*, 1968. ISBN 90 6196 034 7.

MCS 2.1 G.J.R. FÖRCH, P.J. VAN DER HOUWEN & R.P. VAN DE RIET, *Colloquium Stabiliteit van differentieschema's, deel 1*, 1967. ISBN 90 6196 023 1.

MCS 2.2 L. DEKKER, T.J. DEKKER, P.J. VAN DER HOUWEN & M.N. SPIJKER, *Colloquium Stabiliteit van differentieschema's, deel 2*, 1968. ISBN 90 6196 035 5.

MCS 3.1 H.A. LAUWERIER, *Randwaardeproblemen, deel 1*, 1967. ISBN 90 6196 024 X.

MCS 3.2 H.A. LAUWERIER, *Randwaardeproblemen, deel 2*, 1968. ISBN 90 6196 036 3.

MCS 3.3 H.A. LAUWERIER, *Randwaardeproblemen, deel 3*, 1968. ISBN 90 6196 043 6.

MCS 4 H.A. LAUWERIER, *Representaties van groepen*, 1968. ISBN 90 6196 037 1.

MCS 5	J.H. VAN LINT, J.J. SEIDEL & P.C. BAAYEN, *Colloquium Discrete wiskunde*, 1968. ISBN 90 6196 044 4.
MCS 6	K.K. KOKSMA, *Cursus ALGOL 60*, 1969. ISBN 90 6196 045 2.
MCS 7.1	*Colloquium Moderne rekenmachines, deel 1*, 1969. ISBN 90 6196 046 0.
MCS 7.2	*Colloquium Moderne rekenmachines, deel 2*, 1969. ISBN 90 6196 047 9.
MCS 8	H. BAVINCK & J. GRASMAN, *Relaxatietrillingen*, 1969. ISBN 90 6196 056 8.
MCS 9.1	T.M.T. COOLEN, G.J.R. FÖRCH, E.M. DE JAGER & H.G.J. PIJLS, *Elliptische differentiaalvergelijkingen, deel 1*, 1970. ISBN 90 6196 048 7.
MCS 9.2	W.P. VAN DEN BRINK, T.M.T. COOLEN, B. DIJKHUIS, P.P.N. DE GROEN, P.J. VAN DER HOUWEN, E.M. DE JAGER, N.M. TEMME & R.J. DE VOGELAERE, *Colloquium Elliptische differentiaalvergelijkingen, deel 2*, 1970. ISBN 90 6196 049 5.
MCS 10	J. FABIUS & W.R. VAN ZWET, *Grondbegrippen van de waarschijnlijkheidsrekening*, 1970. ISBN 90 6196 057 6.
MCS 11	H. BART, M.A. KAASHOEK, H.G.J. PIJLS, W.J. DE SCHIPPER & J. DE VRIES, *Colloquium Halfalgebra's en positieve operatoren*, 1971. ISBN 90 6196 067 3.
MCS 12	T.J. DEKKER, *Numerieke algebra*, 1971. ISBN 90 6196 068 1.
MCS 13	F.E.J. KRUSEMAN ARETZ, *Programmeren voor rekenautomaten; De MC ALGOL 60 vertaler voor de EL X8*, 1971. ISBN 90 6196 069 X.
MCS 14	H. BAVINCK, W. GAUTSCHI & G.M. WILLEMS, *Colloquium Approximatietheorie*, 1971. ISBN 90 6196 070 3.
MCS 15.1	T.J. DEKKER, P.W. HEMKER & P.J. VAN DER HOUWEN, *Colloquium Stijve differentiaalvergelijkingen, deel 1*, 1972. ISBN 90 6196 078 9.
MCS 15.2	P.A. BEENTJES, K. DEKKER, H.C. HEMKER, S.P.N. VAN KAMPEN & G.M. WILLEMS, *Colloquium Stijve differentiaalvergelijkingen, deel 2*, 1973. ISBN 90 6196 079 7.
MCS 15.3	P.A. BEENTJES, K. DEKKER, P.W. HEMKER & M. VAN VELDHUIZEN, *Colloquium Stijve differentiaalvergelijkingen, deel 3*, 1975. ISBN 90 6196 118 1.
MCS 16.1	L. GEURTS, *Cursus Programmeren, deel 1: De elementen van het programmeren*, 1973. ISBN 90 6196 080 0.
MCS 16.2	L. GEURTS, *Cursus Programmeren, deel 2: De programmeertaal ALGOL 60*, 1973. ISBN 90 6196 087 8.
MCS 17.1	P.S. STOBBE, *Lineaire algebra, deel 1*, 1974. ISBN 90 6196 090 8.
MCS 17.2	P.S. STOBBE, *Lineaire algebra, deel 2*, 1974. ISBN 90 6196 091 6.
MCS 17.3	N.M. TEMME, *Lineaire algebra, deel 3*, 1976. ISBN 90 6196 123 8.
MCS 18	F. VAN DER BLIJ, H. FREUDENTHAL, J.J. DE IONGH, J.J. SEIDEL & A. VAN WIJNGAARDEN, *Een kwart eeuw wiskunde 1946-1971, Syllabus van de Vakantiecursus 1971*, 1974. ISBN 90 6196 092 4.
MCS 19	A. HORDIJK, R. POTHARST & J.Th. RUNNENBURG, *Optimaal stoppen van Markovketens*, 1974. ISBN 90 6196 093 2.

MCS 20 T.M.T. COOLEN, P.W. HEMKER, P.J. VAN DER HOUWEN & E. SLAGT, *ALGOL 60 procedures voor begin- en randwaardeproblemen*, 1976. ISBN 90 6196 094 0.

MCS 21 J.W. DE BAKKER (red.), *Colloquium Programmacorrectheid*, 1975. ISBN 90 6196 103 3.

MCS 22 R. HELMERS, F.H. RUYMGAART, M.C.A. VAN ZUYLEN & J. OOSTERHOFF, *Asymptotische methoden in de toetsingstheorie; Toepassingen van naburigheid*, 1976. ISBN 90 6196 104 1.

MCS 23.1 J.W. DE ROEVER (red.), *Colloquium Onderwerpen uit de biomathematica, deel 1*, 1976. ISBN 90 6196 105 X.

MCS 23.2 J.W. DE ROEVER (red.), *Colloquium Onderwerpen uit de biomathematica, deel 2*, 1976. ISBN 90 6196 115 7.

MCS 24.1 P.J. VAN DER HOUWEN, *Numerieke integratie van differentiaalvergelijkingen, deel 1: Eenstapsmethoden*, 1974. ISBN 90 6196 106 8.

MCS 25 *Colloquium Structuur van programmeertalen*, 1976. ISBN 90 6196 116 5.

MCS 26.1 N.M. TEMME (ed.), *Nonlinear analysis, volume 1*, 1976. ISBN 90 6196 117 3.

MCS 26.2 N.M. TEMME (ed.), *Nonlinear analysis, volume 2*, 1976. ISBN 90 6196 121 1.

MCS 27 M. BAKKER, P.W. HEMKER, P.J. VAN DER HOUWEN, S.J. POLAK & M. VAN VELDHUIZEN, *Colloquium Discretiseringsmethoden*, 1976. ISBN 90 6196 124 6.

MCS 28 O. DIEKMANN, N.M. TEMME (EDS), *Nonlinear Diffusion Problems*, 1976. ISBN 90 6196 126 2.

MCS 29.1 J.C.P. BUS (red.), *Colloquium Numerieke programmatuur, deel 1A, deel 1B*, 1976. ISBN 90 6196 128 9.

MCS 29.2 H.J.J. TE RIELE (red.), *Colloquium Numerieke programmatuur, deel 2*, 1976. ISBN 144 0.

* MCS 30 P. GROENEBOOM, R. HELMERS, J. OOSTERHOFF & R. POTHARST, *Efficiency begrippen in de statistiek*, 1977. ISBN 90 6196 149 1.

MCS 31 J.H. VAN LINT (red.), *Inleiding in de coderingstheorie*, 1976. ISBN 90 6196 136 X.

MCS 32 L. GEURTS (red.), *Colloquium Bedrijfssystemen*, 1976. ISBN 90 6196 137 8.

MCS 33 P.J. VAN DER HOUWEN, *Differentieschema's voor de berekening van waterstanden in zeeën en rivieren*, 1977. ISBN 90 6196 138 6.

MCS 34 J. HEMELRIJK, *Oriënterende cursus mathematische statistiek*, ISBN 90 6196 139 4.

MCS 35 P.J.W. TEN HAGEN (red.), *Colloquium Computer Graphics*, 1977. ISBN 90 6196 142 4.

MCS 36 J.M. AARTS, J. DE VRIES, *Colloquium Topologische Dynamische Systemen*, 1977. ISBN 90 6196 143 2.

MCS 37 J.C. van Vliet (red.), *Colloquium Capita Datastructuren*, . ISBN 90 6196 159 9.

*	MCS 38	T.H. Koornwinder (ed.), *Representations of locally compact groups with applications*, . ISBN 90 6196 161 0.
	MCS 39	O.J. Vrieze & G.L. Wanrooij, *Colloquium Stochastische spelen*, 1978. ISBN 906196 167 X.

De met een * gemerkte uitgaven moeten nog verschijnen.