0550 NL

STICHTING

MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49

AMSTERDAM

AFDELING MATHEMATISCHE STATISTIEK

S 341

Markov - programmering

door

G. de Leve

Februari 1965

Printed at the Mathematical Centre at Amsterdam,49,2nd Boerhaavestraat.
The Netherlands.

The Mathematical Centre, founded the 11th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for Pure Scientific Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

1. Inleiding

Wellicht kunt U mijn praatje van vanmiddag het beste beschouwen als een dissonant in de rij van voordrachten die wij gehad hebben en nog zullen krijgen in dit seizoen. Wij hoorden van de toepassing van O.R. bij bedrijven zoals Shell en Unilever. Hulshof heeft ons het één en ander verteld over O.R. activiteiten bij kleine bedrijven. Verder staan de staatsmijnen nog op het programma.

Misschien verwacht U van mij dat ik iets zal zeggen over de toepassing van O.R. op het Mathematisch Centrum. Helaas moet ik U dan teleurstellen, want zover ik weet is bij het oplossen van onze eigen "bedrijfsproblemen" nooit gebruik gemaakt van O.R. technieken.

De echte O.R. problemen, welke op het Mathematisch Centrum zijn opgelost, waren afkomstig van andere bedrijven of instellingen. Ik kan hierover uiteraard niet vrij spreken. Ik sprak over echte O.R. problemen en daarmee heb ik een onderscheid willen maken tussen de werkelijke problemen van onze opdrachtgevers en die welke wij zelf verzonnen hebben. Over deze laatste problemen zou ik vanmiddag met U van gedachten willen wisselen.

Ik kan mij voorstellen dat U het zinloos vindt om over geconstrueerde O.R. problemen te spreken, temeer daar er nog zoveel onpgeloste echte O.R. problemen op een antwoord wachten.

Misschien kunnen de hierna volgende namaakproblemen U van het tegengestelde overtuigen.

Het eerste probleem

Een automobilist heeft een schadeverzekering afgesloten. In de bijbehorende polis worden o.a. de volgende voorwaarden vermeld:

1) De looptijd van de verzekering is één jaar. Aan het eind van ieder jaar kan zij worden verlengd. De premie moet aan het begin van ieder premie-jaar worden voldaan.
2) De premie bedraagt ƒ 320,- , tenzij
   a) in de voorafgaande periode van één jaar geen schade is geclaimd. In dat geval bedraagt de premie ƒ 280,- , tenzij
   b) in de voorafgaande periode van twee jaar geen schade is geclaimd. In dat geval bedraagt de premie ƒ 240,- , tenzij
   c) in de voorafgaande periode van drie jaar geen schade is geclaimd. In dat geval bedraagt de premie ƒ 220,-.
3) Schaden moeten onmiddellijk worden geclaimd. Slechts het verschil tussen de schade en een vast bedrag van ƒ 80,- , het zogenaamde eigen risico, wordt door de verzekering uitbetaald.

Als verder gegeven is dat de tijsduren tussen opéénvolgende schaden en de schaden zelf in guldens onafhankelijk negatief exponentiëel verdeeld zijn met gemiddelden $\frac{1}{2}$ resp. 200, gevraagd voor elk tijdstip aan te geven welke schaden eventueel geclaimd zullen worden en welke niet.

Het is duidelijk dat de automobilist nooit een schade van minder dan ƒ 80,- zal claimen.

Het is ook duidelijk, dat hij, als nog geen schade is geclaimd dat jaar, met het oog op de premiereducties voor schade-vrij rijden geen schaden zal claimen, welke slechts een weinig hoger zijn dan het eigen risico. De vraag is nu waar precies de grens ligt tussen de schaden die wel en die niet moeten worden geclaimd. Het behoeft geen betoog dat de grenswaarden zullen afhangen van de laatste betaalde premie en van het tijdstip in het premie-jaar. In onderstaande figuur wordt de optimale strategie aangegeven. Alle schaden worden geclaimd tenzij zij corresponderen met punten in het in het achtervlak aangegeven gearceerde gebied. Wij merken nu reeds op dat niet het claimen van een schade een wilsdaad van de automobilist is, maar juist het onderdrukken van de schade aangifte.

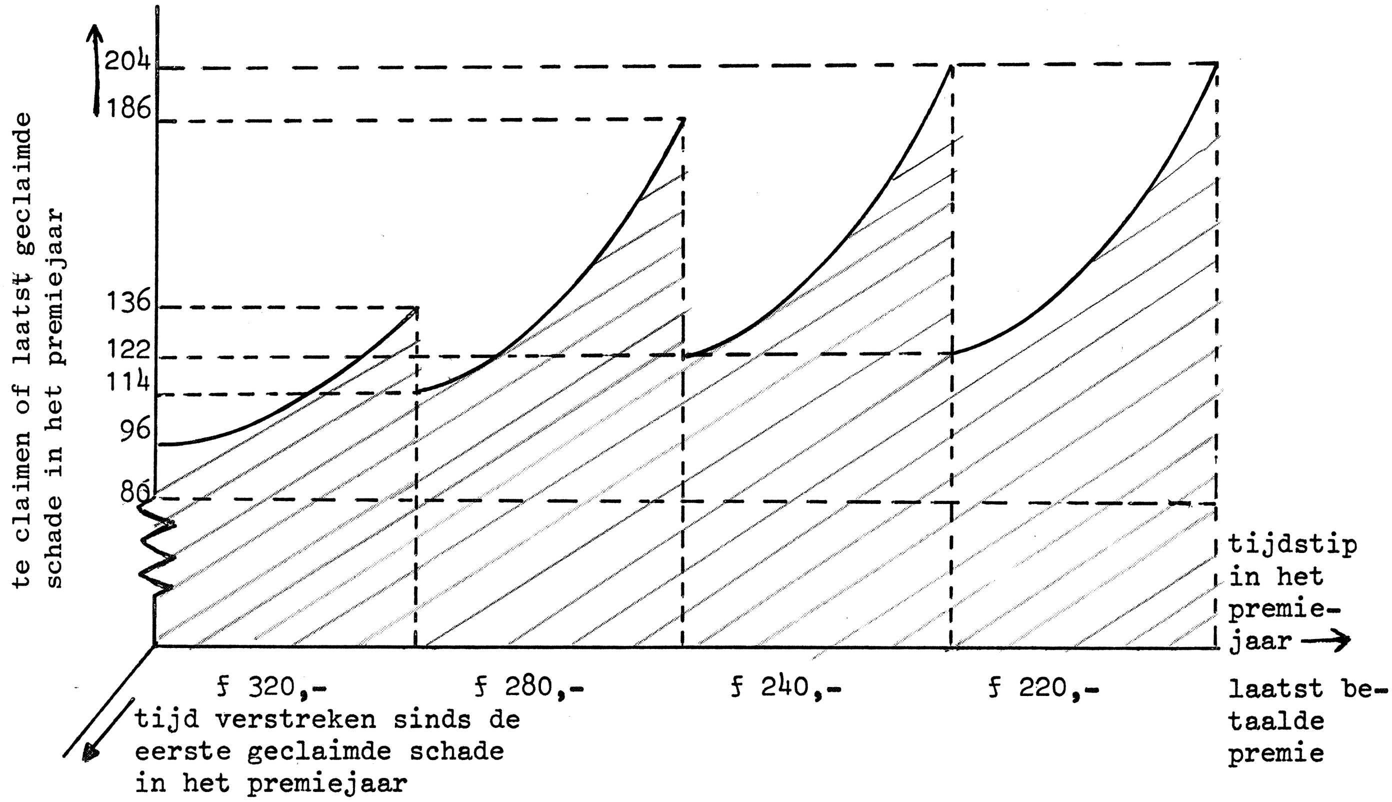


fig. 1

Toestandsruimte met aangegeven strategie

Het tweede probleem

Een fabriek kan hoogstens drie gelijke machines inschakelen bij de vervaardiging van één product. Voor het product geldt dat de gemiddelde vraag per eenheid kleiner is dan de productie van drie machines per tijdseenheid. De voorraadcapaciteit is beperkt en bedraagt M eenheden. De voorraadkosten per tijdseenheid zijn evenredig met de grootte van de voorraad. De productiekosten per eenheid worden mede bepaald door het aantal ingeschakelde machines. Het omschakelen naar hogere of lagere productiesnelheden brengt extra kosten met zich mede, die slechts afhangen van de corresponderende opéénvolging van productiesnelheden. Indien de voorraad niet toereikend is dan worden de gevraagde goederen geleverd via een zusterfabriek. De tussen beide fabrieken overeengekomen verrekeningsprijs per eenheid is gegeven. Deze verrekeningsprijs ligt hoger dan de gemiddelde kostprijs. Gevraagd wordt voor iedere productiesnelheid na te gaan voor welke voorraden men moet omschakelen en waarheen?

In onderstaande figuur wordt een strategie aangeduid. Zodra de combinatie voorraad-productie kan worden aangegeven door een punt van een dik getrokken lijn dan wordt de productiesnelheid op de aangegeven wijze veranderd.

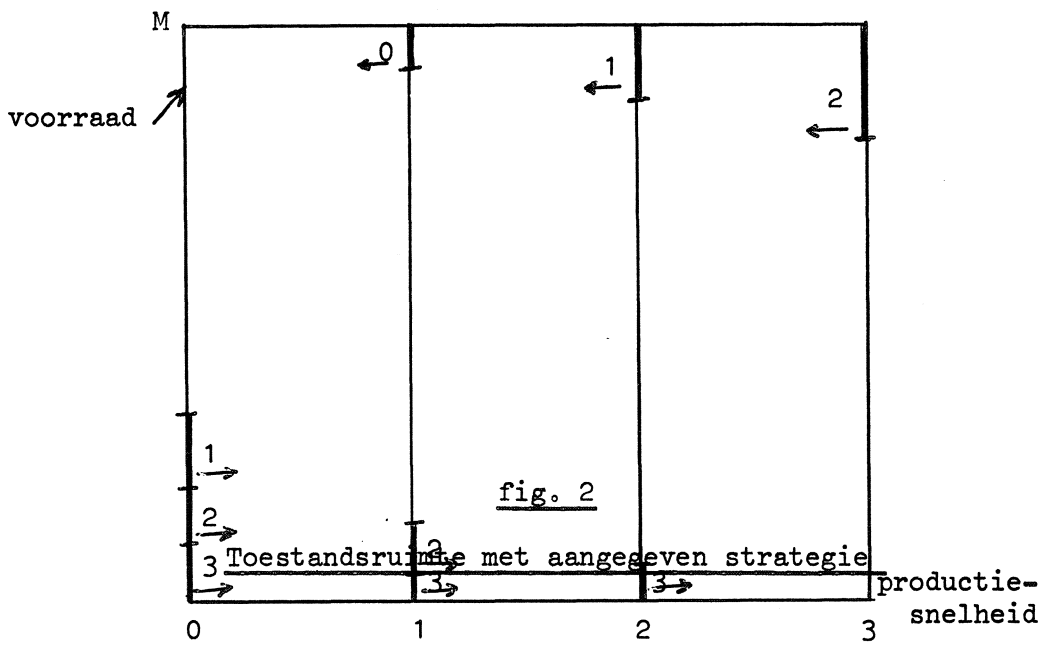


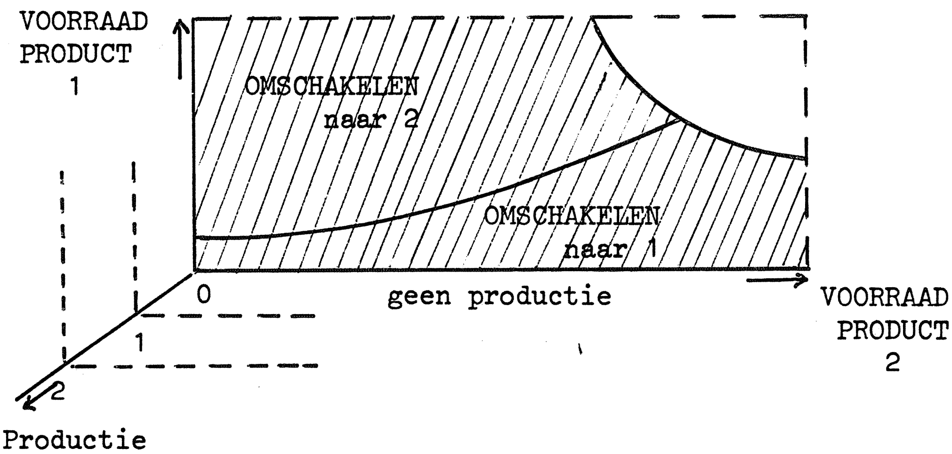
fig. 2

Toestandsruimte met aangegeven strategie

Het derde probleem

Een fabriek kan op één machine twee verschillende producten vervaardigen. In de "long run" is de productiecapaciteit voldoende om aan de gezamenlijke vraag te voldoen. De maximale voorraadcapaciteit is voor ieder product afzonderlijk gegeven. De voorraadkosten per tijdseenheid zijn evenredig met de grootte van de voorraad. De productiesnelheden en kosten per eenheid zijn gegeven. Het omschakelen van productie (eventueel afzetten van de machine) brengt extra kosten met zich mede, die slechts afhangen van de corresponderende opéénvolging van producties. Indien de voorraad niet toereikend is dan worden de gevraagde goederen geleverd via een zusterfabriek. De tussen beide fabrieken overeengekomen verrekeningsprijs per eenheid is gegeven en ligt hoger dan de gemiddelde kostprijs. Gevraagd wordt na te gaan voor welke voorraadcombinaties de productie moet worden omgeschakeld en hoe?

In onderstaande figuur (3 evenwijdige vlakken) wordt voor verschillende producties een strategie aangeduid. Zodra de voorraadcombinatie aangegeven kan worden door een punt in een gearceerd gebied dan wordt de productie op de aangegeven wijze omgeschakeld. Voor de overzichtelijkheid zijn twee van drie vlakken afzonderlijk getekend.



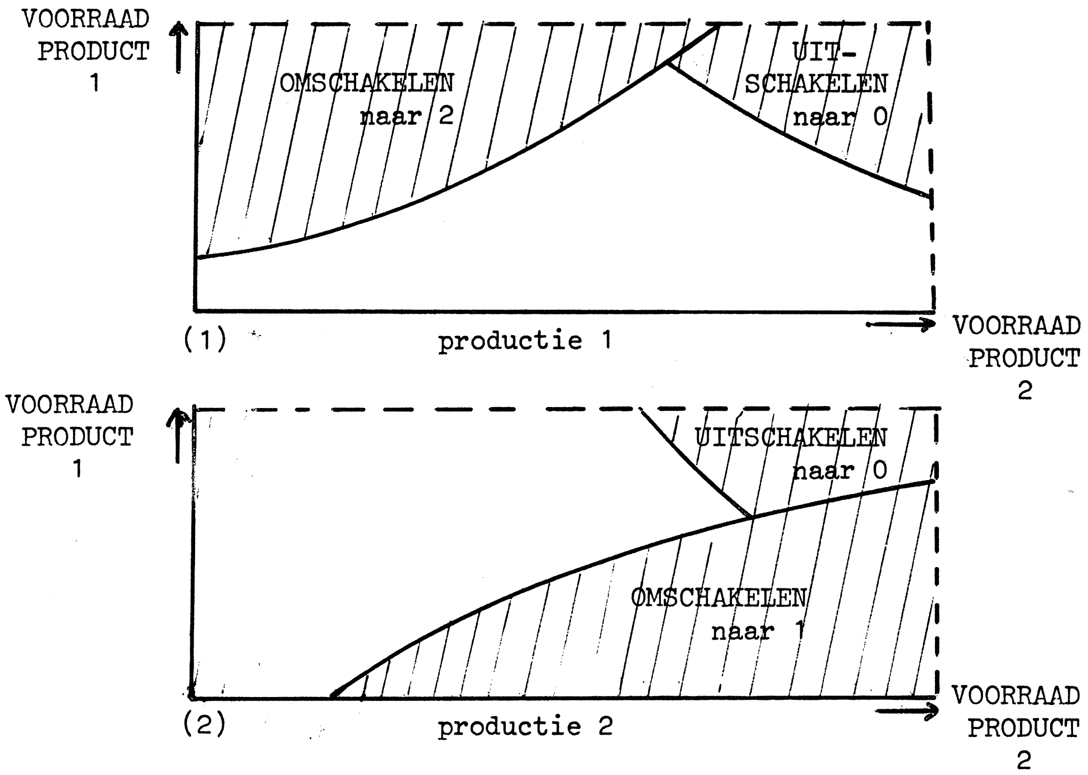


fig. 3

Toestandsruimte met aangegeven strategie

Wanneer wij deze namaakproblemen analyseren dan ontdekken wij dat zij te herleiden zijn tot één standaard probleem. Beschouwen wij hiertoe nogmaals de gestelde problemen. In alle drie de voorbeelden hadden wij een systeem. In het eerste voorbeeld was het systeem de auto en de bijbehorende verzekering, in het tweede voorbeeld de productie van één artikel, terwijl in het derde voorbeeld het systeem de productie van twee artikelen aanduidt. De toestand van het systeem kan steeds worden aangegeven met een punt in een Cartesisch coördinatenstelsel. In de lopp van de tijd wandelt het systeem door de toestandsruimte X. Het proces dat aan deze wandeling ten grondslag ligt wordt het natuurlijk proces genoemd. Verder maken wij nog gebruik van het begrip strategie. Met een strategie bedoelen wij hier:

1e) een gearceerd gebied in de toestandsruimte X dat dient om aan te geven in welke toestanden van het systeem de beslisser zal ingrijpen;

2e) een transformatie vanuit de "ingrijptoestand" naar een nieuwe toestand.

In de onderstaande figuur is voor twee "ingrijptoestanden" de bijbehorende transformatie aangegeven.

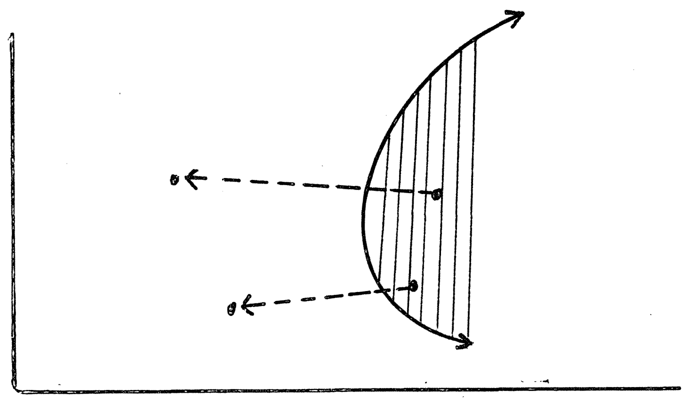


fig. 4

Het algemene geval

Wanneer wij afzien van het aantal getekende dimensies en de vorm van het gearceerde gebied dan kunnen wij de toestandsruimten met strategie (figuren 1, 2 en 3) zien als bijzondere gevallen van de toestandsruimte geschetst in fig. 4.

Men kan zeggen dat het systeem door het natuurlijk proces gedwongen door de toestandsruimte wandelt (of springt) en dat steeds als het een gearceerd gebied binnen gaat het met een boogje er weer wordt uitgeknikkerd. Het komt er dus op neer dat men moet trachten het optimale gearceerde gebied + uitwerpmechanisme te bepalen.

Om redenen van eenvoud verdient het voorkeur om aan te nemen dat op <u>ieder</u> tijdstip een beslissing wordt genomen. Een beslissing d kan men aangeven door een punt d in een niet nader te beschrijven beslissingsruimte D. Wij maken een onderscheid tussen ingrepen en z.g.n. nulbeslissingen. Door een nulbeslissing wordt het systeem "geplaatst" in een toestand waarin het zich reeds bevond. Met een ingreep correspondeert een nieuwe al of niet stochastische toestand van het systeem. Het volgt uit het karakter van menig probleem dat een aantal beslissingen niet zijn toegelaten. Wij zullen derhalve aannemen dat bij iedere toestand x in de toestandsruimte X een verzameling D(x) van beslissingen d in de beslissingsruimte D hoort.

In de volgende paragraaf zullen wij een methode bespreken welke dikwijls leidt tot een oplossing van problemen van het hierboven geschetste type.

2. De methode

In de inleiding spraken wij van een natuurlijk proces dat in de loop van de tijd de toestand van het systeem doet wijzigen. De wijze waarop de toestand van het systeem zich wijzigt wordt slechts dan volledig door het natuurlijke proces aangegeven wanneer de beslisser zich volstrekt afzijdig houdt. Volgt de beslisser echter een strategie en grijpt hij zo nu en dan in, dan ontstaat er een nieuw proces, het zgn. beslissingsproces. Het beslissingsproces is dus de resultante van het natuurlijke proces en de door de strategie voorgeschreven ingrepen (transformaties).

Wij zullen nu aannemen dat het beslissingsproces beschreven kan worden door een stationnair sterk Markovproces. Daarmee willen wij aangeven dat

a) het beslissingsproces niet deterministisch is;
b) kansen kunnen worden toegekend aan mogelijke toekomstige ontwikkelingen in de toestand van het systeem;
c) deze kansen afhangen van de huidige toestand en niet van de voorgeschiedenis;
d) deze kansen niet afhangen van het beschouwde tijdstip. Ook niet als het tijdstip zelf een stochastische grootheid is (b.v. het moment waarop de beslisser ingrijpt).

Wij zullen hierop niet nader ingaan maar volstaan met de mededeling dat een sterk Markov proces kanstheoretisch kan worden gedefiniëerd. Indien de wijze waarop de toestand van het systeem zich in de loop van de tijd ontwikkelt kan worden beschreven met behulp van een sterk Markov proces, dan kunnen zich de volgende structuren voordoen:

1) Er is een verzameling van toestanden in de toestandsruimte, waarin het systeem nooit meer terug komt als het eens buiten deze verzameling is geweest. Deze verzameling zullen wij aangeven met de naam "Verzameling van doorgangstoestanden";
2) Er zijn verzamelingen in de toestandsruimte waar het systeem nooit meer uit komt als het eens die verzameling is binnengekomen. Een

dergelijke verzameling heet een fuik als deze verzameling niet gesplitst kan worden in twee of meer deelverzamelingen met dezelfde eigenschap;

3) In een fuik bestaat een klasse van verzamelingen, die het systeem steeds in een gegeven volgorde bezoekt. Deze klasse noemt men een kringfuik.

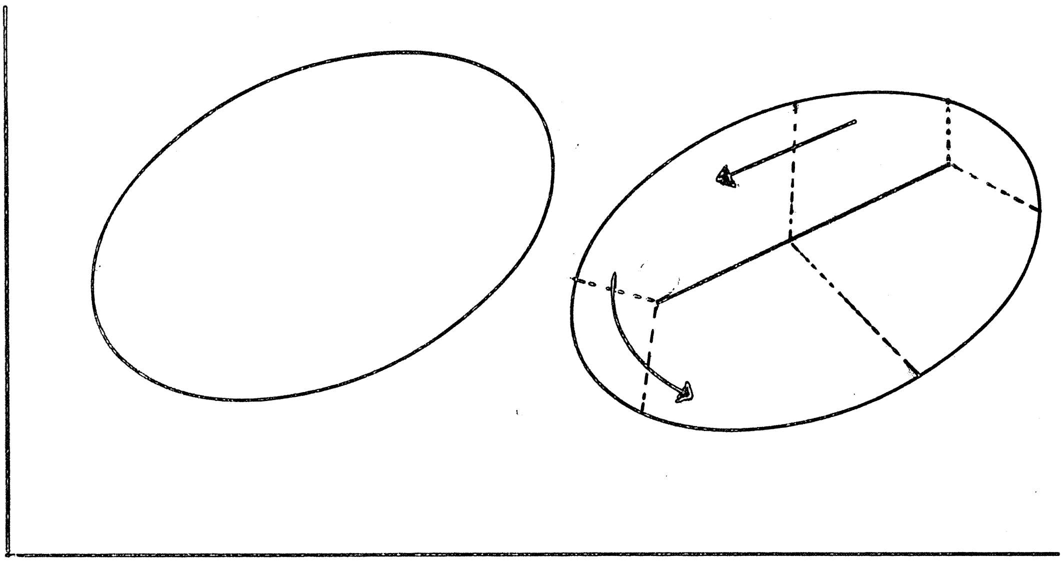


fig. 5

Fuik en Fuik + kringfuik

De hierboven geschetste indeling van de toestandsruimte in fuiken, kringfuiken en verzameling van doorgangstoestanden is niet altijd ondubbelzinnig. Maar daarover zullen wij ons thans niet bekommeren.

In het hierna volgende zullen wij ons in het bijzonder interesseren voor de toestanden $\{2\underline{I}_n ;\ n = 1, 2, \ldots\}$ waarin daadwerkelijk in het proces wordt ingegrepen. De toestanden $\{\underline{I}_n ;\ n = 1, 2, \ldots\}$ liggen dus in het gearceerde gebied, dat wij in het vervolg met A_z zullen aangeven als z de toegepaste strategie voorstelt. De kansverdeling van deze toestanden zullen wij aanduiden met $\{p^{(n)}(A; z; x);\ n = 1, 2, \ldots\}$, waarbij A een of andere verzameling van toestanden is, z de toegepaste strategie en x de begintoestand.

Verder zullen wij spreken van de limiettoestand $\underline{I}_\infty$. Deze toestand zou men kunnen zien als een zeer ver in de toekomst gelegen "ingrijptoestand". Onder zekere voorwaarden heeft deze toestand een kansverdeling, welke wij zullen aangeven met $p(A; z; x)$.

Nu geldt altijd

$$p(A; z; x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} p^{(j)}(A; z; x) \tag{1}$$

en vaak (geen kringfuiken in A_z!)

$$p(A; z; x) = \lim_{n\to\infty} p^{(n)}(A; z; x). \tag{2}$$

Deze laatste eigenschap levert de hierboven gegeven interpretatie van het begrip "limiet toestand".

Verder kan men bewijzen dat, als voor een beslissingsproces x_1 en x_2 mogelijke begintoestanden zijn uit één fuik, geldt

$$p(A; z; x_1) = p(A; z; x_2). \tag{3}$$

Wij keren nu terug tot het systeem dat een stochastische wandeling maakt door de toestandsruimte. Wij nemen aan dat tijdens de wandeling kosten worden gemaakt en dat deze kosten slechts afhangen van de door de strategie voor te schrijven ingrepen in het natuurlijk proces en de te doorlopen toestanden. Gelijk wij de toestanden van het systeem konden identificeren met een punt in de toestandsruimte, zo kunnen wij ook de mogelijke realiseringen (complete wandelingen) aangeven met een punt ω in een niet nader aan te geven ruimte Ω. Als de strategie z wordt toegepast en het systeem volgt de wandeling ω, dan geven wij de te maken kosten in een periode van de lengte T aan door de functie $k_T(\omega; z)$.

Het is duidelijk dat als $T \to \infty$ de kosten $k_T(\omega; z)$ veelal onbegrensd groot worden.

Ergo $\lim\limits_{T\to\infty} k_T(\omega; z)$ is geen handig kriterium voor het bepalen van de optimale strategie. Een ander onoverkomelijk nadeel is dat men van te voren de te volgen wandeling ω niet kent. Hoe vinden wij nu een kriterium voor de optimale strategie?

Het eerste bezwaar

$$\lim_{T\to\infty} k_T(\omega;\ z) = \infty \tag{4}$$

zou men kunnen ondervangen door de kosten te verdisconteren. Dat wil zeggen dat men op een voorgeschreven wijze aan kosten in het verre verschiet minder gewicht toekent dan aan kosten van gelijke omvang in de nabije toekomst. Voor een dergelijke handelwijze bestaat een economische rechtvaardiging. De op deze wijze verkregen kriteriumfunctie hangt dan ten nauwste samen met de numerieke waarde (welke?) van de verdisconteringsfactor.

Het tweede bezwaar was dat men de nog te volgen wandeling ω niet kent. Wij hebben vastgesteld dat wij een kwantitatief inzicht hebben (kansen) met betrekking tot de realiseerbaarheid van mogelijke wandelingen.
Op grond van dit inzicht kunnen wij een kriteriumfunctie bepalen, die de verwachting aangeeft van de totaal verdisconteerde kosten. Waardoor men ook aan het tweede bezwaar enigszins tegemoet kan komen. Wij zullen echter een andere weg bewandelen. Stel dat $k(I;\ z)$ de verwachting aangeeft van de kosten te maken in een periode tussen twee opéénvolgende ingrijptoestanden, waarbij I de toestand is op het moment van de eerste ingreep en z de toegepaste strategie. Stel dat $t(I;\ z)$ de verwachting van de lengte van die periode voorstelt.
Men kan nu onder zekere algemene veronderstellingen de volgende stelling bewijzen:

Stelling 1

Als ω een aselecte wandeling is van het systeem en als de begintoestand x_0 van ω behoort tot een fuik, dan bestaat de limiet

$$\lim_{T\to\infty} \frac{k_T(\omega;\ z)}{T} \tag{5}$$

met kans 1 en is gelijk aan

$$\frac{\int p(dI;\ z;\ x_0)\ k(I;\ z)}{\int p(dI;\ z;\ x_0)\ t(I;\ z)}\ . \tag{6}$$

Uit stelling 1 volgt dat onverschillig de gevolgde weg ω de gemiddelde kosten per tijdseenheid voor "bijna alle" wandelingen ω, die een begintoestand hebben uit een gegeven fuik, gelijk zijn. En wel gelijk aan het quotiënt van de verwachting van de kosten en van de tijdslengte elk gemeten tussen twee ingrijptoestanden in de stationnaire toestand. Voor begintoestanden behorende tot verschillende fuiken zullen in het algemeen de gemiddelde kosten per tijdseenheid verschillen.

Als de begintoestand x_0 een doorgangstoestand is dan wordt de waarde van de gemiddelde kosten per tijdseenheid bepaald door de fuik waarin het systeem zal worden gevangen. Deze fuik is echter in de begintoestand x_0 nog niet bekend. Men kan in de begintoestand x_0 wel spreken van de verwachte gemiddelde kosten per tijdseenheid. Eenvoudig kan worden nagegaan dat deze verwachting gegeven wordt door

$$\int p(dy;\ z;\ x_0)\ \frac{\int p(dx;\ z;\ y)\ k(x;\ z)}{\int p(dx;\ z;\ y)\ t(x;\ z)}\ . \tag{7}$$

Wij kiezen nu als kriterium voor de optimale strategie de functie $r(z;\ x_0)$ gegeven door

$$r(z;\ x_0) = \int p(dy;\ z;\ x_0)\ \frac{\int p(dx;\ z;\ y)\ k(x;\ z)}{\int p(dx;\ z;\ y)\ t(x;\ z)}\ . \tag{8}$$

Voor het geval x_0 in een fuik ligt gaat het rechterlid van (8) over in (6), hetgeen impliceert dat "bijna zeker" $r(z;\ x_0)$ de werkelijke gemiddelde kosten per tijdseenheid aangeeft.

Alhoewel voor begintoestanden uit een fuik geldt dat de <u>gemiddelde</u> kosten per tijdseenheid "bijna zeker" gelijk zijn, betekent dit niet dat wij geen voorkeur zouden hebben voor bepaalde toestanden binnen de fuik.

Men kan nu een functie $c(z;\ x)$ definiëren zodanig dat voor ieder tweetal begintoestanden x_1 en x_2 in één fuik de uitdrukking

$$c(z;\ x_1) - c(z;\ x_2) \tag{9}$$

de verwachting van het verschil in de totale kosten voorstelt.

Indien de toestandsruimte kan worden opgesplitst in één verzameling van doorgangstoestanden en m disjuncte fuiken en als de toestanden $\{e_j;\ j = 1, 2, \ldots, m\}$ willekeurige ingrijptoestanden zijn (in iedere fuik één) dan wordt voor iedere ingrijptoestand I_1 een functie $c(z;\ I_1)$ gegeven door de volgende relaties

$$c(z;\ I_1) = k(I_1;\ z) - r(z;\ I_1)\ t(I_1;\ z) + \int p^{(1)}(dI_2;\ z;\ x)\ c(z;\ I_2) \tag{10}$$

$$c(z;\ e_j) = 0 \qquad j = 1, 2, \ldots, m. \tag{11}$$

Men kan $c(z;\ I_1)$ interpreteren als de verwachting van de kosten die de beslisser nog uit eigen zak moet betalen, wanneer hij van de "eigenaar" van het systeem een vergoeding per tijdseenheid krijgt, die in iedere ingrijptoestand $\underline{I}_j$ opnieuw wordt vastgesteld en gelijk gekozen aan $r(z;\ I_j)$.

Ook voor niet ingrijptoestanden kan de functie $c(z;\ x)$ worden gedefiniëerd. Om rekentechnische redenen geven wij echter de voorkeur aan een iets gewijzigde definitie van de functie $c(z;\ x)$. De hierboven gegeven niet-mathematische interpretatie gaat dan weliswaar enigszins verloren, maar de rekentechnische voordelen wegen daar zeker tegenop. In verband met de nieuwe definitie van $c(z;\ x)$ beschouwen wij een verzameling A_0 met de eigenschap dat voor iedere in aanmerking komende strategie geldt:

$$A_z \supset A_0. \tag{12}$$

Voor iedere toestand x en beslissing $d \in D(x)$ bestaat er een tweetal stochastische wandelingen aangegeven met $\underline{W}^0$ en $\underline{W}^d$. Gedurende de wandeling $\underline{W}^0$ is het systeem onderworpen aan het <u>natuurlijk proces</u> met begintoestand x. De wandeling eindigt zodra het systeem een toestand uit de verzameling A_0 aanneemt. Aan het begin van de wandeling $\underline{W}^d$ wordt het systeem door de beslissing d getransformeerd in een al of niet stochastische toestand $\underline{y}$. Na deze transformatie is het systeem onderworpen aan het natuurlijk proces met begintoestand $\underline{y}$. De wandeling $\underline{W}^d$ zal eindigen zodra een toestand uit A_0 wordt aangenomen.

Wij voeren nu in de functies $k(x;\ d)$ en $t(x;\ d)$.

De $(x;\ d)$-functie $k(x;\ d)$ stelt het verschil in verwachte kosten voor gemaakt tijdens de wandelingen $\underline{W}^d$ en $\underline{W}^0$.

De $(x;\ d)$-functie $t(x;\ d)$ stelt het verschil in verwachte tijdsduur voor van de wandelingen $\underline{W}^d$ en $\underline{W}^0$.

Als d een nulbeslissing is dan geldt uiteraard

$$k(x;\ d) = 0 \tag{13}$$

$$t(x;\ d) = 0. \tag{14}$$

Wij merken op dat de functies $t(x;\ d)$ en $k(x;\ d)$ in tegenstelling tot de functies $k(x;\ z)$ en $t(x;\ z)$ niet afhangen van één bepaalde strategie z.

In fig. 6 hebben wij de wandelingen $\underline{W}^0$ en $\underline{W}^d$ schematisch aangegeven.

De beslissing d wordt nu door de strategie z in I_1 voorgeschreven. Men merke op dat de eerder gedefiniëerde functies $k(I_1;\ z)$ en $t(I_1;\ z)$ betrekking hebben op het eerste gedeelte van de wandeling $\underline{W}^d$ en wel tot de toestand $\underline{\underline{I}}_2$.

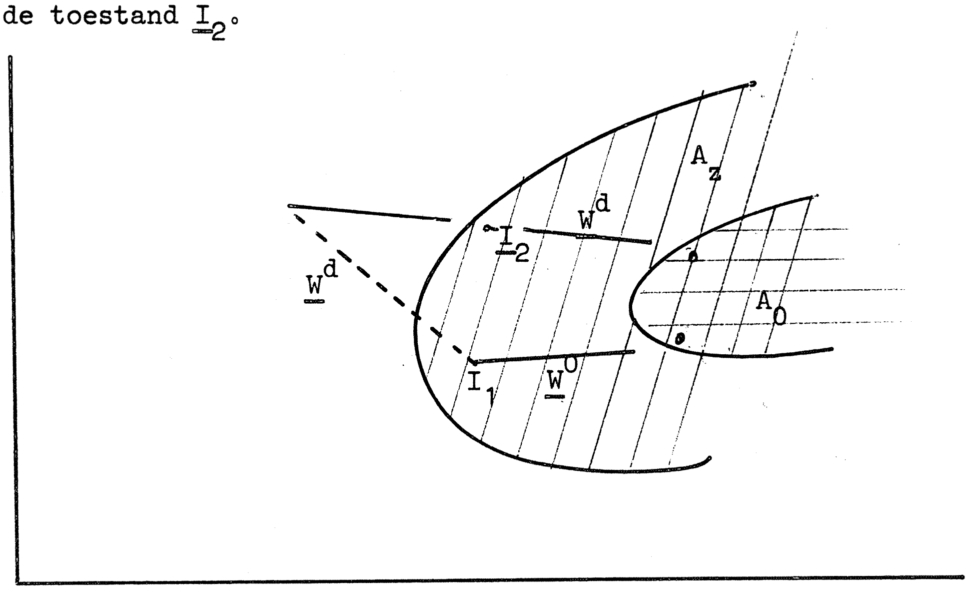


fig.6

Schematische voorstelling van $\underline{W}^0$ en $\underline{W}^d$

Als de toestand $\underline{I}_1$ de limietverdeling $p(A;\ z;\ y)$ volgt dan geldt dat ook $\underline{I}_2$ volgens $p(A;\ z;\ y)$ verdeeld is (eigenschap van de limietverdeling). Derhalve zijn de verwachte kosten en tijdsduren van de wandelingen vanuit $\underline{I}_1$ en $\underline{I}_2$ naar A_0 gelijk. Bijgevolg geldt:

$$\int p(dx;\ z;\ y)\ k(x;\ z) = \int p(dx;\ z;\ y)\ k(x;\ z(x)) \tag{15}$$

$$\int p(dx;\ z;\ y)\ t(x;\ z) = \int p(dx;\ z;\ y)\ t(x;\ z(x)), \tag{16}$$

waarbij $z(x)$ de door de strategie z voorgeschreven beslissing in de toestand x aangeeft.
Wij mogen dus de functies $k(x;\ z)$ en $t(x;\ z)$ in (6), (7) en (8) vervangen denken door $k(x;\ z(x))$ resp. $t(x;\ z(x))$.

Als $\underline{I}^*$ de eerst volgende ingrijptoestand aangeeft, dan worden de functies $r(z;\ x)$ en $c(z;\ x)$ als volgt gegeven:

$$r(z;\ x) = \mathcal{E}\{r(z;\ \underline{I}^*)/x;\ z\} \tag{17}$$

$$c(z;\ x) = k(x;\ z(x)) - r(z;\ x)\ t(x;\ z(x)) + \mathcal{E}\{c(z;\ \underline{I}^*/x;\ z\} \tag{18}$$

$$c(z;\ e_j) = 0 \qquad j = 1,\ 2,\ \ldots,\ m, \tag{19}$$

waarbij e_j $(j = 1,\ 2,\ \ldots,\ m)$ willekeurig gekozen toestanden zijn in de m fuiken (in iedere fuik één).

Men ziet dus dat $r(z;\ x)$ en $c(z;\ x)$ oplossingen zijn van een aantal functionaal vergelijkingen.

Met behulp van de functies $r(z;\ x)$ en $c(z;\ x)$ kan men nieuwe functies $r((z')z;x)$ en $c((z')z;x)$ definiëren.

De functie $r((z')z;\ x)$ geeft (de verwachting van) de gemiddelde kosten per tijdseenheid aan als x de begintoestand is en men eerst éénmaal ingrijpt volgens de strategie z' en daarna altijd de strategie z toepast.

De functie $c((z')z;\ x)$ geeft de verwachting aan van het bedrag dat de beslisser uit eigen zak moet betalen als men eerst éénmaal ingrijpt volgens strategie z' en daarna altijd volgens strategie z.

De functies $r((z')z;\ x)$ en $c((z')z;\ x)$ worden gedefiniëerd door:

$$r((z')z;\ x) \overset{\text{def}}{=} \mathcal{E}\ \{r(z;\ \underline{I}_2)/x;\ z'\} \tag{20}$$

$$c((z')z;\ x) \overset{\text{def}}{=} \mathcal{E}\{k(\underline{I}_1;\ z'(\underline{I}_1)) - r((z')z;\ \underline{I}_1)\ \cdot\ t(\underline{I}_1;\ z'(\underline{I}_1))/x;\ z'\} + \mathcal{E}\{c(z;\ \underline{I}_2)/x;\ z'\}, \tag{21}$$

waarbij $\underline{I}_1$ de toestand aangeeft waarin ingegrepen wordt volgens strategie z' en $\underline{I}_2$ de toestand waarin voor het eerst de strategie z wordt toegepast.

Wij zijn nu in staat voor een klasse van strategieën Z_0 een iteratieprocedure voor te schrijven, die onder zeer algemene voorwaarden *) een rij van strategieën oplevert met de eigenschap

$$\lim_{n\to\infty} r(z^{(n)};\ x) = \inf_{z\in Z_0}\ r(z;\ x). \tag{22}$$

Deze iteratieprocedure kunnen wij beginnen met iedere willekeurige strategie $z^{(0)} \in Z_0$.
Als de strategie $z^{(n-1)}$ wordt verkregen aan het eind van de (n-1)-de cyclus dan verloopt de n-de cyclus als volgt:

<u>1ste stap</u>

Bepaal de functies $r(z^{(n-1)};\ x)$ en $c(z^{(n-1)};\ x)$. (B.v. met de relaties (17), (18) en (19).)

<u>2de stap</u>

Bepaal de $(z;\ x)$-functies $(z \in Z_0)$, $r((z)z^{(n-1)};\ x)$ en $c((z)z^{(n-1)};\ x)$. (B.v. met de relaties (20) en (21).)

<u>3de stap</u>

Bepaal de subklasse $Z^{(n)}$ van Z_0 waarvoor geldt:

$$Z^{(n)} = \{z \mid r((z)z^{(n-1)};\ x) = \min_{z^* \in Z_0}\ r((z^*)z^{(n-1)};\ x)\}. \tag{23}$$

---

*) Voor de voorwaarden waaronder deze procedure leidt tot de optimale strategie wordt de lezer verwezen naar 3).

<u>4^{de} stap</u>

Bepaal een strategie $z \in Z^{(n)}$ z.d.d.

$$c((z)z^{(n-1)};\ x) = \min_{z^* \in Z^{(n)}} c((z^*)z^{(n-1)};\ x). \tag{24}$$

In de volgende cyclus wordt deze strategie aangegeven met $z^{(n)}$.

<u>Einde</u> van de n-de cyclus.

De n-de cyclus kan als volgt worden beschreven:
"De beslisser is verplicht na de eerstvolgende ingreep de strategie $z^{(n)}$ toe te passen. Voor de eerste ingreep moet hij een strategie z kiezen, die de door de eigenaar te betalen vergoeding minimaliseert. De vrijheid, welke hem nog is gelaten, zal hij uiteraard zo benutten dat de verwachting van de kosten uit eigen zak te betalen zo klein mogelijk wordt.

Zodra hij de goede "beginstrategie" heeft bepaald wordt deze door de "slimme" eigenaar voorgeschreven en wel met dien verstande dat de beslisser opnieuw een "beginstrategie" moet bepalen ((n+1)-ste cyclus)".

3. Technische bijzonderheden

De wijze waarop de optimale beginstrategie in de iteratieprocedure moet worden bepaald verdient een nadere toelichting. Daartoe voeren wij in de functies r(d ∘ z; x) en c(d ∘ z; x) gedefiniëerd door

$$r(d \circ z;\, x) \overset{\text{def}}{=} \mathcal{E}\{r(z;\, \underline{y})/d\} \tag{25}$$

$$c(d \circ z;\, x) \overset{\text{def}}{=} k(x;\, d) - r(d \circ z;\, x)\; t(x;\, d) + \mathcal{E}\{c(z;\, \underline{y})/d\}, \tag{26}$$

waarbij $\underline{y}$ de toestand aangeeft waarin het systeem terecht komt als in de toestand x de beslissing d wordt genomen.

De functie r(d ∘ z; x) geeft (de verwachting van) de gemiddelde kosten per tijdseenheid aan als <u>na</u> de begintoestand x de strategie z wordt toegepast maar <u>in</u> de begintoestand eerst de beslissing d moet worden genomen.

De functie c(d ∘ z ; x) geeft de verwachting aan van het bedrag dat de beslisser uit eigen zak moet betalen als hij na de begintoestand x de strategie z volgt maar in de begintoestand de beslissing d neemt.

Vervolgens voeren wij in de functies r(A ∘ z; x) en c(A ∘ z; x) gedefiniëerd door

$$r(A \circ z;\, x) \overset{\text{def}}{=} \mathcal{E}\{r(z;\, \underline{y})/x;\, A\} \tag{27}$$

$$c(A \circ z;\, x) \overset{\text{def}}{=} \mathcal{E}\{c(z;\, \underline{y})/x;\, A\}, \tag{28}$$

waarbij $\underline{y}$ nu de eerste toestand is uit een gegeven verzameling A, die na de begintoestand x wordt aangenomen.

De functie r(A ∘ z; x) geeft (de verwachting van) de gemiddelde kosten per tijdseenheid aan als na de begintoestand x de strategie z "geblokkeerd" is tot het moment waarop het systeem voor het eerst een toestand uit de verzameling A aanneemt.

De functie c(A ∘ z; x) geeft de verwachting aan van het bedrag dat de beslisser uit eigen zak moet betalen als na de begintoestand x de strategie z "geblokkeerd" is tot het moment waarop het systeem voor het eerst een toestand uit de verzameling A aanneemt.

Verder beschouwen wij de klasse K_z van alle gesloten verzamelingen $A \supset A_0$ met de eigenschap

$$\bar{A} \subset \{x / r(A \circ z;\ x) < r(z;\ x)\} \cup \{x / r(A \circ z;\ x) = r(z;\ x);$$
$$c(A \circ z;\ x) \leq c(z;\ x)\}, \qquad (29)$$

waarbij $\bar{A}$ het complement is van A.
Uit (29) volgt dat voor alle toestanden buiten A blokkering tot A wenselijk is. Men kan eenvoudig nagaan dat $A_z \in K_z$.

Wij beschouwen nu de volgende deelverzamelingen

$$D_z(x) \overset{\text{def}}{=} \{d / d \in D(x);\ r(d \circ z;\ x) = \min_{d^* \in D(x)} r(d^* \circ z;\ x)\} \qquad (30)$$

en

$$A'_z = \bigcap_{A \in K_z} A. \qquad (31)$$

De verzameling $D_z(x)$ bevat die beslissingen d welke in de toestand x (de verwachting van) de gemiddelde kosten per tijdseenheid minimaliseren.

De verzameling A'_z in de toestandsruimte is de beste "blokkeringsverzameling" voor de strategie z.

Wij beschouwen nu de volgende iteratieprocedure *):

<u>Voorbereidend gedeelte</u>

Bepaal de (x; d)-functies k(x; d) en t(x; d).

<u>Iteratief gedeelte</u>

Als $z^0 \in Z_0$ een willekeurige strategie is en als $z^{(n-1)}$ wordt verkregen aan het eind van de (n-1)-de cyclus dan verloopt de n-de cyclus als volgt:

---

*) Voor de voorwaarden waaronder deze procedure leidt tot de optimale strategie wordt de lezer verwezen naar 3).

1^{ste} stap

Bepaal de functies $r(z^{(n-1)}; x)$ en $c(z^{(n-1)}; x)$ met behulp van de relaties (17), (18) en (19).

2^{de} stap

a) Bepaal de functies $r(d \circ z^{(n-1)}; x)$ en $c(d \circ z^{(n-1)}; x)$ met behulp van de relaties (25) en (26).

b) Bepaal voor iedere toestand x de deelverzameling $D_{z^{(n-1)}}(x)$.

c) Minimaliseer de d-functie $c(d \circ z; x)$ voor iedere x onder de voorwaarde $d \in D_{z^{(n-1)}}(x)$. Voeg aan iedere x een "minimaliserende" beslissing d toe en wel zo dat om redenen van convergentie deze beslissing gelijk gekozen wordt aan $z^{(n-1)}(x)$, indien $z^{(n-1)}(x)$ een minimaliserende beslissing is.

Men beschouwt nu een strategie

$$z_1^{(n-1)}(x) = d_{z^{(n-1)}; x}, \tag{32}$$

waarbij $d_{z^{(n-1)}; x}$ de gekozen beslissing voorstelt.

3^{de} stap

Bepaal de functies $r(z_1^{(n-1)}; x)$ en $c(z_1^{(n-1)}; x)$ met behulp van de relaties (17), (18) en (19).

4^{de} stap

Bepaal de beste "blokkeringsverzameling" $A'_{z_1^{(n-1)}}$ en voer in de strategie $z^{(n)}$ gegeven door

$$z^{(n)}(x) \overset{\text{def}}{=} \begin{cases} z_1^{(n-1)}(x) \text{ als } x \in A'_{z_1^{(n-1)}} \\ \text{nulbeslissingen als } x \notin A'_{z_1^{(n-1)}} \end{cases} \tag{33}$$

Einde n-de cyclus.

Uit het voorafgaande blijkt dat de iteratieprocedure voor de optimale strategie valt en staat met de bepaling van de x-functies $r(z^{(n-1)}; x)$ en $c(z^{(n-1)}; x)$. Deze functies worden gegeven door een stelsel van functionaalvergelijkingen.

Indien deze functionaalvergelijkingen niet eenvoudig oplosbaar zijn, dan rest vaak nog de mogelijkheid om voor een gegeven strategie met behulp van een simulatie de functies $r(z^{(n-1)}; x)$ en $c(z^{(n-1)}; x)$ numeriek te bepalen. De iteratieprocedure geeft dan aan hoe men in de opéénvolgende simulaties de strategieën moet kiezen.

De wijze waarop de beste "blokkeringsverzameling" moet worden bepaald hangt sterk af van de structuur van het beschouwde beslissingsprobleem. Gewoonlijk bestaat de blokkeringsverzameling uit gebieden waarvoor geldt dat het op de rand onverschillig is of met betrekking tot de kosten men al of niet ingrijpt. Deze eigenschap levert dan de mathematische gedaante van de randen op. In het eerste probleem (§ 1) leidde deze eigenschap tot de uitspraak dat de randen van de beste "blokkeringsgebieden" horizontaal verschoven stukken waren uit een gegeven kromme (zie fig. 7).

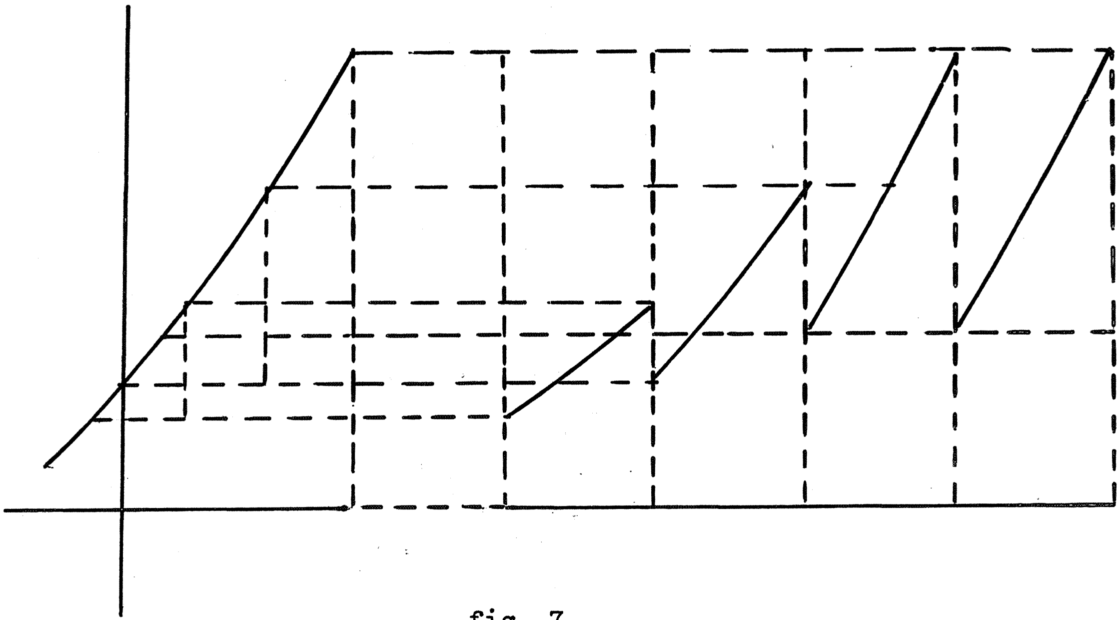


fig. 7

Constructie van de beste "blokkeringsgebieden"

in het eerste probleem

Literatuur

1) R. Bellman, Dynamic Programming, Princeton Univ. Press, 1957.

2) R.A. Howard, Dynamic Programming and Markov Processes, John Wiley, 1960.

3) G. de Leve, Generalized Markovian Decision Processes, M.C. Tract 3, 1964.