STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

S 297

Steeds weer meten en weten

door

G. de Leve

maart 1962

Steeds weer meten en weten

door

G.de Leve

§ 1 Inleiding

Het opstellen en toepassen van oplossingsmethoden voor meer-staps beslissingsproblemen vormt een wezenlijk onderdeel van de besliskunde.

Tot de klasse der meer-staps beslissingsproblemen behoren o.a. voorraad-, produktie- en vervangingsproblemen.

Zoals de naam reeds doet vermoeden wordt in ieder meerstaps beslissingsprobleem meerdere malen een beslissing genomen; hierbij kunnen de beslissingstijdstippen van te voren gegeven zijn of ze kunnen vrij gekozen worden. Het ligt in de aard van het beslissingsprobleem, dat de toestand waarin een beslissing wordt genomen van invloed is op de keuze van de beslissing. Spreekt men echter van toestand dan moet er ook iets bestaan waarop deze toestand betrekking heeft. In de wiskundige formulering spreekt men van de toestand van het systeem. Het systeem is het object van onze beschouwingen. Zo kan bij een wachttijdprobleem het systeem worden gevormd door de rij wachtenden voor een loket terwijl bij een voorraadprobleem met het systeem meestal de voorraad zal worden bedoeld.

Aangenomen wordt dat het systeem zich in verschillende toestanden kan bevinden en dat deze toestanden beschreven kunnen worden met behulp van een aantal kwantitatieve grootheden. Een toestand wordt dus gegeven door een rij van getallen, die voor deze toestand de bijbehorende waarden van de kwantitatieve grootheden aangeven.

Een dergelijke rij van getallen noemt men een toestandsvector. Indien wij een voorraad moeten beheren van drie verschillende artikelen dan is de voorraad het systeem. De toestand van het systeem, zo zullen wij aannemen, kan dan worden gegeven door drie getallen, waarvan elk de omvang van de voorraad van één der artikelen aangeeft.

Stel dat de toestanden van een systeem beschreven kunnen

worden met behulp van m kwantitatieve grootheden; dan kan iedere toestand geidentificeerd worden met een punt in een door m orthogonale coördinaatassen opgespannen ruimte of met een toestandsvector die de verbindingslijn is van dat punt met de oorsprong van het assenstelsel. Op de assen van het assenstelsel worden de eerder genoemde kwantitatieve grootheden uitgezet. Deze ruimte zullen wij de toestands-ruimte noemen.

De mathematische formulering van het probleem beperkt zich uiteraard niet tot de beschrijving van de toestand van het systeem. Wij zullen aannemen, dat ook iedere mogelijke beslissing kan worden weergegeven met behulp van een aantal kwantitatieve grootheden.

In het hierboven genoemde voorraadprobleem kan iedere beslissing worden vastgelegd door drie getallen. Deze getallen geven de bestelgrootten aan van de drie artikelen. Voor het geval men de beslissingen kan aanduiden met behulp van r kwantitatieve grootheden, correspondeert met iedere beslissing een punt in een r-dimensionale ruimte. De verbindingslijn van dat punt met de oorsprong zullen wij de beslissingsvector noemen, terwijl de ruimte zelf met beslissingsruimte wordt aangeduid.

Indien alleen de toestand van het systeem op het beslissingstijdstip bepalend is voor de keuze van de beslissing, dan wordt de oplossing van het meerstapsbeslissingsprobleem verkregen door voor ieder beslissingstijdstip bij elke mogelijke toestand van het systeem de bijbehorende optimale beslissing te bepalen. Als X de toestand aangeeft op het n^{de} beslissingstijdstip en als d de bijbehorende optimale beslissing is dan kan deze toevoeging worden gegeven door:

$$d = S_n(X) \qquad (1.1)$$

§ 2 Het dynamisch programmeringsprobleem

In een dynamisch programmeringsprobleem wordt verondersteld, dat iedere beslissing een opbrengst ten gevolge heeft. De omvang van deze opbrengst hangt af van de beslissing en de toestand waarin deze beslissing is genomen[1].
Indien men de opbrengst aangeeft met z en de toestands- en beslissingsvector met X resp. d dan geldt:

$$z = h(X,d). \qquad (2.1)$$

Stel dat uit de probleemstelling volgt, dat op N opéénvolgende tijdstippen een beslissing moet worden genomen. Dit houdt onder meer in dat N maal een opbrengst wordt verkregen. Nu zou men op ieder beslissingstijdstip steeds die beslissing kunnen kiezen, welke de bijbehorende opbrengst maximaliseert. Deze werkwijze zal echter in het algemeen niet optimaal zijn, omdat de te nemen beslissing niet alleen de opbrengst maar ook de toestand van het systeem op het volgende beslissingstijdstip zal bepalen. Om deze reden zal de maximale totale opbrengst doorgaans slechts door het volgen van een ingewikkelder beslissingspolitiek kunnen worden verkregen[1].

Wij zullen ons nu beperken tot die problemen, welke steeds voldoen aan één van de volgende eigenschappen:

1) De toestand op het k^{de} beslissingstijdstip wordt volledig bepaald door de toestand op het $(k-1)^{ste}$ beslissingstijdstip en de daarop genomen beslissing.

2) De kansverdeling van de mogelijke toestanden op het k^{de} beslissingstijdstip wordt volledig bepaald door een toestand op het $(k-1)^{ste}$ beslissingstijdstip en de daarop genomen beslissing.

Zodra in een N-staps beslissingsprobleem de eerste beslissing is genomen, gaat het N-staps beslissingsprobleem over

---

1) Er zijn problemen, waarin de opbrengst behorende bij een beslissing stochastisch is. In die gevallen beschouwt men de verwachting van de opbrengst.

in een (N-1)-staps beslissingsprobleem.

Indien x de toestand is op het eerste beslissingstijdstip dan zullen wij met $f_N(x)$ die totale opbrengst aangeven, welke wordt verkregen als steeds optimaal wordt beslist[1].

Voor het geval het N-staps beslissingsprobleem voldoet aan de eerste eigenschap dan geldt voor de toestand y op het tweede beslissingstijdstip:

$$y = T_{N-1}(x,d) \tag{2.2}$$

waarbij d de beslissing op het eerste beslissingstijdstip aangeeft.

Aangezien de maximaal te verkrijgen totale opbrengst voor de laatste (N-1) stappen gegeven wordt door $f_{N-1}(y)$ vinden wij de volgende relatie:

$$f_N(x) = \max_d \left\{h(x,d)+f_{N-1}\left[T_{N-1}(x,d)\right]\right\}. \tag{2.3}$$

Op analoge wijze vindt men:

$$f_j(x) = \max_d \left\{h(x,d)+f_{j-1}\left[T_{j-1}(x,d)\right]\right\} \tag{2.4}$$

$$(j=1,2,\ldots,N)$$

met $$f_o(x)=0. \tag{2.5}$$

Met behulp van (2.4) en (2.5) kan men achtereenvolgens de functies $f_1(x)$, $f_2(x),\ldots f_N(x)$ bepalen. Kent men eenmaal deze functies dan zijn ook de beslissingsvoorschriften:

$$d=S_j(x) \qquad (j=1,\ldots N) \tag{2.6}$$

bepaald.

Voor het geval het gestelde probleem voldoet aan de tweede eigenschap dan kan men eenvoudig nagaan dat in plaats van (2.4) geldt:

$$f_j(x)=\max_d \left\{ h(x,d)+ \mathcal{E}\left[f_{j-1}(\underline{y}\,|x,d)\right]\right\}. \tag{2.7}$$

Ook nu kan men achtereenvolgens de opbrengstfuncties $f_1(x), f_2(x),\ldots$ en $f_N(x)$ bepalen.

In een ∞-staps beslissingsprobleem zal in het algemeen voor iedere toestand x op het eerste beslissingstijdstip de maximaal te verkrijgen totale opbrengst $f_\infty(x)$ onbegrensd zijn.

Dit kan echter worden voorkomen door de directe opbrengsten behorende bij latere beslissingen te verdisconteren. Hiertoe voeren wij in de verdisconteringsfactor α, waarvoor geldt:

$$0 \leq \alpha < 1. \tag{2.8}$$

In het hierna volgende zullen wij steeds onderstellen dat de relatie (2.2) onafhankelijk is van het beslissingstijdstip.
Met andere woorden: voor problemen, welke voldoen aan de eerste eigenschap, geldt:

$$y = T(x,d). \tag{2.9}$$

Indien wij de opbrengst behorende bij de $k^{\underline{de}}$ beslissing vermenigvuldigen met de factor α^{k-1} dan gaan (2.4) resp. (2.7) over in:

$$f_j(x) = \max_d \left\{ h(x,d) + \alpha f_{j-1}\left[T(x,d)\right] \right\} \tag{2.10}$$

$$f_j(x) = \max_d \left\{ h(x,d) + \alpha \mathcal{E}\left[f_{j-1}(\underline{y}|x,d)\right] \right\}. \tag{2.11}$$

Indien de rij van functies $\{ f_j(x) \}$ (voor iedere waarde van x) convergeert naar de functie $f_\infty(x)$ dan geldt onder bepaalde voorwaarden:

$$f_\infty(x) = \max_d \left\{ h(x,d) + \alpha f_\infty\left[T(x,d)\right] \right\} \tag{2.12}$$

of

$$f_\infty(x) = \max_d \left\{ h(x,d) + \alpha \mathcal{E}\left[f_\infty(\underline{y}|x,d)\right] \right\}. \tag{2.13}$$

De optimale strategie:

$$d = S(x) \tag{2.14}$$

volgt nu uit de oplossing van een functionaalvergelijking.

§ 3 Markovian decision processes

Er bestaat een uitgebreide groep van problemen, waarin een natuurlijk proces de toestand van het systeem doet veranderen. Men denke slechts aan een slijtage-proces bij een machine of aan de invloed van de vraag op de voorraad, bij een voorraadprobleem.
Wij zullen nu veronderstellen, dat deze toestandsveranderingen kunnen worden beschreven met behulp van een stationair Markov proces.

Deze ontwikkelingen in de toestand van het systeem kan men storen door het nemen van beslissingen. In een voorraadprobleem zal het storen wellicht bestaan uit het aanvullen van de voorraad. In een wachttijdprobleem zal men misschien storen door een nieuw loket te openen, terwijl in een vervangingsprobleem tenslotte het proces wordt gestoord, door de oude machine te vervangen door een nieuwe.
Ter vereenvoudiging van de discussie zullen wij aannemen, dat het systeem zich slechts in eindig veel toestanden i kan bevinden ($i=1,\ldots,n$); verder zullen wij aannemen, dat ook nu de opbrengst $h(i,d)$ behorende bij een beslissing d, slechts afhangt van de toestand van het systeem en de genomen beslissing. Men kan eenvoudig inzien, dat in het geval van een Markov proces, een beslissingspolitiek slechts dan optimaal kan zijn, als in gelijke toestanden, ongeacht het beslissingstijdstip, gelijke beslissingen genomen worden.
Met andere woorden: voor het optimale beslissingsvoorschrift geldt dat aan iedere toestand in de toestandsruimte op ondubbelzinnige wijze een van het beslissingstijdstip onafhankelijke beslissing is toegevoegd. Zo'n beslissingsvoorschrift zullen wij een strategie noemen en aangeven met χ.
Wij hebben reeds vastgesteld dat door een beslissing de toestand op het beslissingstijdstip verandert.
Het zal duidelijk zijn, dat door de toepassing van een strategie het verloop van het oorspronkelijke natuurlijke proces ingrijpend gewijzigd zal worden.
De wetmatigheden in het verloop van het nieuwe stochastische

proces zullen afhangen van de gekozen strategie χ.

Stel dat de kans op overgang van toestand i naar toestand j in één beslissingsfase gegeven wordt door $p_{ij}(\chi)$, waarbij χ de toegepaste strategie voorstelt. Indien $p_{ij}^{(k)}(\chi)$ de kans voorstelt op overgang van toestand i naar toestand j in k beslissingsfasen dan geldt:

$$p_{ij}^{(k)}(\chi)=\sum_{l=1}^{n} p_{il}^{(k-1)}(\chi)\, p_{lj}(\chi). \tag{3.1}$$

Nu kan men bewijzen, dat de uitdrukking:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)}(\chi)=q_{ij}(\chi) \tag{3.2}$$

bestaat. Men noemt de grootheden q_{ij} de invariante kansen van het Markov proces.

Stel nu dat $h(i;\chi)$ de opbrengst is, wanneer men in de toestand i een beslissing neemt volgens χ. De verwachting van de totale opbrengst in m stappen wordt dan gegeven door:

$$\sum_{k=1}^{m} \sum_{j=1}^{n} p_{ij}^{(k)}(\chi)h(j;\chi). \tag{3.3}$$

Indien men een strategie zoekt waarvoor de opbrengst van de eerste m stappen maximaal is, dan moet voor deze strategie gelden, dat (3.3) maximaal is.
Het maximaliseren van de uitdrukking (3.3) is gelijkwaardig met het maximaliseren van:

$$\frac{1}{m} \sum_{k=1}^{m} \sum_{j=1}^{n} p_{ij}^{(k)}(\chi)h(j;\chi)=\sum_{j=1}^{n}\left[\frac{1}{m} \sum_{k=1}^{m} p_{ij}^{(k)}(\chi)\right] h(j;\chi). \tag{3.4}$$

Indien $m\to\infty$ volgt uit (3.3) en (3.4) als criterium voor de optimale strategie:

$$u(\chi;i)=\sum_{j=1}^{n} q_{ij}(\chi)h(j;\chi). \tag{3.5}$$

Met behulp van (3.5) kan men dus twee alternatieve strategieen vergelijken.
R. Howard [1] heeft een methode ontwikkeld om op een itera-

tieve wijze de optimale strategieën te bepalen.

§ 4 Generalisaties

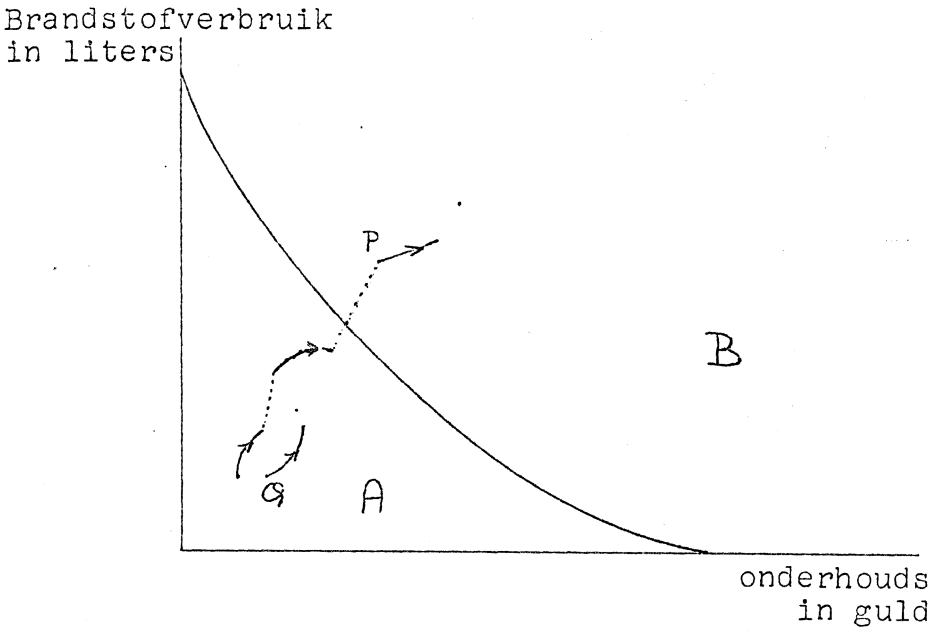
In § 3 hebben wij ons beperkt tot systemen, welke slechts een eindig aantal verschillende toestanden kunnen aannemen. Op analoge wijze kan men gelijksoortige resultaten afleiden voor meer ingewikkelder toestandsruimten. Indien het aantal mogelijke toestanden onbegrensd is, dan dient men zich er wel eerst van te overtuigen of het gestoorde Markov-proces een invariante kansverdeling bezit.

Verder kan men in plaats van te eisen, dat beslist wordt op van te voren vastgestelde tijdstippen, ook veronderstellen dat beslist mag worden op momenten, waarop de toestand van het systeem zich wijzigt. (Howard [1]). Hierdoor verkrijgt men een zekere vrijheid in de keuze van het beslissingstijdstip. Immers, in veel gevallen zullen de optimale beslissingstijdstippen met deze tijdstippen samenvallen. Indien echter één van de toestandsgrootheden van het systeem de tijd aangeeft welke verstreken is sinds de laatste beslissing (dit kan o.a. van belang zijn bij voorraad- en vervangingsproblemen), dan kan men, als men vrij wil zijn in de keuze van het beslissingstijdstip deze methoden niet gebruiken. Immers een dergelijke tijdslengte verandert continu in de tijd en dientengevolge zou men continu moeten beslissen.

Het systeem, dat in deze paragraaf verondersteld wordt onderworpen te zijn aan een Markov proces, maakt een stochastische wandeling in de toestandsruimte. De omstandigheid of wij op een tijdstip al of niet een beslissing zullen nemen wordt bepaald door de toestand van het systeem op dat tijdstip. Dit betekent dat de toestandsruimte uiteénvalt in twee verzamelingen van toestanden A en B.

Zolang het systeem toestanden doorloopt uit de verzameling A, dan is dit acceptabel. Zodra echter de toestand met een punt van B moet worden geïdentificeerd, dan zal men willen storen (beslissen). Een beslissing houdt in dat het

systeem wederom wordt overgebracht naar een toestand van A. Om de gedachten te bepalen: veronderstel dat in een vervangingsprobleem de toestand van het systeem kan worden gegeven door het weekverbruik van brandstoffen en door de onderhoudskosten in die periode. De toestandsruimte is aangegeven in fig. (4.1) met behulp van een 2-dimensionaal orthogonaal assenstelsel.



Toestandsruimte behorende bij een vervangingsprobleem

fig.4.1

In fig.4.1 zijn de verzamelingen A en B aangegeven alsmede een gedeeltelijke realisering van een stochastische wandeling (sprongen zijn gestippeld).

In het punt P neemt het systeem voor het eerst na enige tijd een toestand aan uit de verzameling B. Indien op dat moment het proces niet zou worden gestoord, dan zou, zoals is aangegeven, het systeem vanuit P verder wandelen in B.

Ten gevolge van de genomen beslissing zet het systeem zijn wandeling evenwel niet voort vanuit P maar vanuit Q .

In sommige ∞-staps beslissingsproblemen is men min of meer vrij in de keuze van het punt Q . Wanneer niet alle componenten van de toestandsvector zich wijzigen bij een beslissing, dan kan men slechts kiezen uit punten in een deelruimte van lagere dimensie. Bovendien volgen uit de probleemstellingen dikwijls nog beperkingen van een ander type.

Soms is het zelfs principiëel onmogelijk om op het beslissingstijdstip de plaats van Q te kiezen. Zo zal in een vervangingsprobleem Q de toestand aangeven van de nieuwe machine. Indien niet alle machines van een gegeven merk gelijk zijn, dan verkeert men in P nog in onzekerheid omtrent de ligging van Q . Men zal daarentegen dikwijls wel een kansverdeling kunnen geven van de ligging van het punt Q . Dit betekent, dat een beslissing in de toestand P een keuze is uit kansverdelingen van Q . Zo kan men dus in het algemeen stellen dat een beslissing geïdentificeerd kan worden met een kansverdeling al of niet geconcentreerd in een enkel punt[2)].

Ook nu voegt een strategie aan iedere toestand van de toestandsruimte een beslissing toe uit de beslissingsruimte. In het algemeen zal door toepassing van de strategie χ het natuurlijke proces in de toestandsruimte niet meer gebruikt kunnen worden om de toestandsveranderingen van het systeem te beschrijven. Een nieuw proces zal derhalve geconstrueerd moeten worden met behulp van het oude natuurlijke proces en

---

2) Ieder punt in de beslissingsruimte geeft dus een kansverdeling aan. Voor het geval dat de toestandsruimte een tweedimensionale ruimte (x_1, x_2) voorstelt en de beslissingen normale verdelingen zijn, dan kunnen alle toegelaten beslissingen worden geïdentificeerd met punten in een ruimte opgespannen door 5 orthogonale coördinaatassen. Op deze assen worden dan uitgezet de waarden $(\mu_{x_1}, \mu_{x_2}, \sigma_{x_1}, \sigma_{x_2}$ en $\rho)$.

de toegepaste strategie. Onder bepaalde voorwaarden kanmen aantonen dat het gestoorde proces wederom een Markov proces is met een invariante kansverdeling.

Evenals bij de in § 3 besproken methoden is het ook nu van belang de kansverdelingen van de toestanden op de beslissingstijdstippen te kennen. Men kan bewijzen dat als men het systeem slechts gadeslaat op de beslissingstijdstippen het gedrag van het systeem op die tijdstippen kan worden beschreven met behulp van een tijdsdiscreet Markov proces. Ook dit Markov proces hangt uiteraard af van de toegepaste strategie en het oorspronkelijke Markov proces in de toestandsruimte.

Wij zullen nu aannemen dat in elk tijdsinterval kosten worden gemaakt, die slechts zullen afhangen van de door het systeem in dat tijdsinterval doorlopen toestanden. Vervolgens zullen wij veronderstellen dat ook de toepassing van de strategie kosten met zich meebrengt, die op haar beurt worden bepaald door de in dat tijdsinterval genomen beslissingen.

Om een functie $h(x;\chi)$ te kunnen construeren, welke geheel overeenkomt met die in § 3 voeren wij een nieuwe verzameling C in, welke geheel binnen B zal komen te liggen (zie fig.4.2).

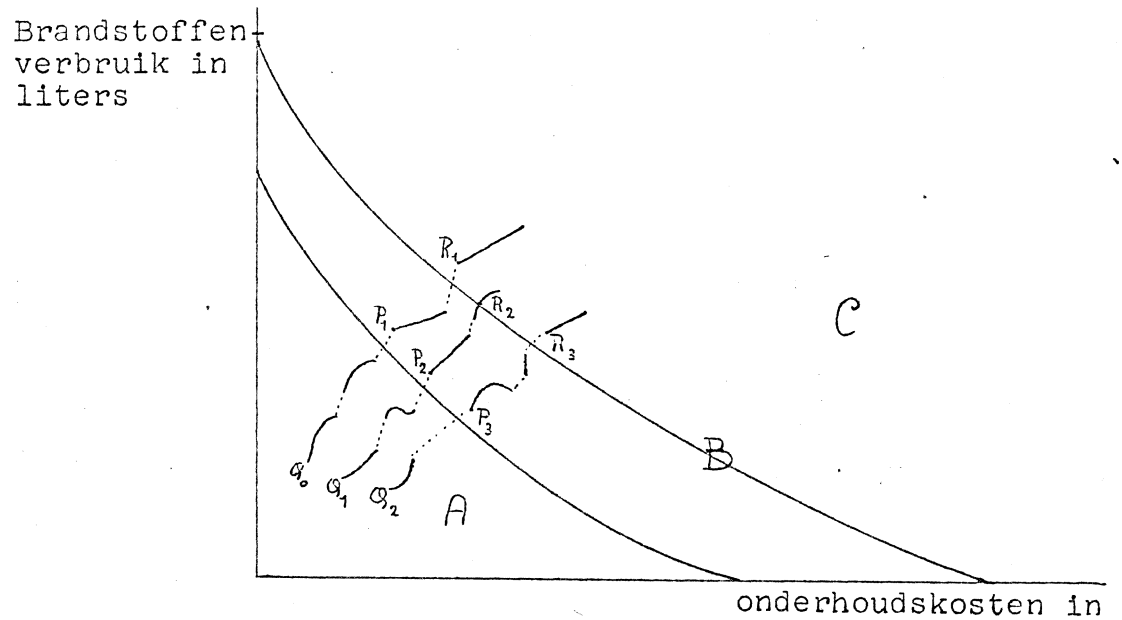


fig.4.2 Toestandsruimte behorende bij een vervangingsprobleem

Stel dat het systeem na een spanne tijds in P_1 voor het eerst een toestand uitde verzameling B aanneemt. Indien de toestand in P_1 kan worden aangegeven met X_1 dan wordt $h(X_1;\chi)$ gedefiniëerd als het verschil van de volgende twee verwachtingen (zie fig.4.2):

a) de verwachting van de kosten verbonden aan de stochastische wandeling van P_1 naar Q_1 (ten gevolge van een storing) en daarna van Q_1 via P_2 naar R_2

b) de verwachting van de kosten verbonden aan de stochastische wandeling van P_1 direct naar R_1,

waarbij R_1 en R_2 de eerste toestanden van C zijn die na P_1 worden aangenomen.

Op analoge wijze kan men een functie $t(X_1;\chi)$ definiëren; deze functie is het verschil van de volgende twee verwachtingen:

a) de verwachting van de tijdsduur van een stochastische wandeling van P_1 naar Q_1 (ten gevolge van een storing) en daarna van Q_1 via P_2 naar R_2.

b) de verwachting van de tijdsduur van een stochastische wandeling van P_1 direct naar R_1.

Indien $q(E;\chi;X_1)$ de invariante kans aangeeft op een storingstoestand in een verzameling E in B en indien X_1 de toestand voorstelt op het eerste beslissingstijdstip dan kan men bewijzen dat onder bepaalde voorwaarden het criterium voor de optimale strategie gegeven wordt door:

$$\lambda(\chi;X_1)=\frac{\int_B q(dX;\chi;X_1)h(X;\chi)}{\int_B q(dX;\chi;X_1)t(X;\chi)} \cdot \qquad (4.1)$$

Indien men de verzameling C steeds gelijk kiest aan B, dan worden de hierboven onder b) genoemde verwachtingen steeds gelijk aan nul. Bij deze keuze van C geeft de teller van het rechterlid van (4.1) voor de stationaire toestand de verwachting aan van de kosten tussen twee opéénvolgende be-

slissingen, terwijl de noemer de verwachte lengte aangeeft voor de stationaire toestand en voor de overeenkomstige periode. Het is eenvoudig in te zien, dat strategieën, welke leiden tot beslissingen op van te voren gegeven equidistante tijdstippen, een gelijke noemer hebben in (4.1). Het criterium (4.1) is dan een directe generalisatie van (3.5) voor ingewikkelder toestandsruimten.

Zoals reeds eerder is vastgesteld behoort bij iedere toegelaten strategie een verzameling B. Stel dat het mogelijk is om C zodanig te kiezen, dat deze verzameling a priori binnen de verzameling van de optimale strategie B ligt. Indien wij nu verschillende strategieën gaan vergelijken, dan kiezen wij C steeds gelijk. Uit de constructie van de functies $h(X;\chi)$ en $t(X;\chi)$ volgt dan, dat deze functies slechts afhangen van de strategie χ via de beslissing in de toestand X. Kiest men daarentegen voor iedere strategie χ de verzameling C gelijk aan de bij χ behorende verzameling B dan hangen beide functies op ingewikkelder wijze af van de gebruikte strategie

[1] R. Bellman, Dynamic Programming, Princeton University Press

[2] R.A. Howard, Dynamic Programming and Markov Processes

[3] G. de Leve, Decision Rules for adjusting Markovian processes Mathematisch Centrum, S 282 (VP 15, SP 77)