

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Rapport S 265 (C 13)

Leergang Besliskunde

Hoofdstuk XIV

Dynamische Programmering

door

G. de Leve

november 1960

1. Inleiding

Het maken van een scherp onderscheid tussen de dynamische programmering en de overige programmering is geen eenvoudige opgave.

De term Dynamische Programmering is afkomstig van Richard Bellman, die hierover in zijn boek "Dynamic Programming" schrijft¹⁾:

"The problems we treat are programming problems to use a terminology now popular. The adjective "dynamic", however, indicates that we are interested in processes in which time plays a significant role and in which the order of operations may be crucial."

Nu bestaat er echter een groot aantal zeer ingewikkelde produktieproblemen, waarin de tijd een zeer belangrijke rol speelt en die met lineaire programmering kunnen worden opgelost. Zijn deze produktieproblemen nu ook Dynamische Programmeringsproblemen?

Vervolgens zegt Bellman: "However, an essential feature of our approach will be the reinterpretation of many static processes as dynamic processes in which time can be artificially introduced."

Elk houvast is ons door deze uitspraak ontnomen! Op deze plaats kan men de d.p.²⁾ het beste omschrijven door vast te stellen, dat de d.p. een techniek is met behulp waarvan een klasse van programmeringsproblemen kan worden opgelost.

Het behoeft echter niet de enige techniek te zijn met deze eigenschap.

In de volgende paragrafen zullen wij nagaan, welke kenmerken het d.p. probleem bezit en op welke wijze het wordt opgelost.

2. Enkele typische d.p.-problemen

Veronderstel dat een koloniaal gebied wordt voorbereid op zelfbestuur, hetwelk over vijf jaar zal worden verleend. De beherende mogendheid wil in de nog resterende jaren trachten een gezonde basis te scheppen voor de onafhankelijkheid. Daartoe zijn verschillende plannen uitgewerkt, waarbij één betrekking heeft op de gezondheidszorg. In het door deskundigen samengestelde rapport kan men lezen, dat het wenselijk is drie cursussen te organiseren en wel die opleiden tot

1) R. Bellman, Dynamic Programming, Princeton University Press 1957.
2) d.p. = dynamische programmering.

- a) ziekenverzorger
- b) instructeur ziekenverzorger A
- c) instructeur ziekenverzorger B.

De cursussen duren elk één jaar en kunnen slechts in de hierboven gegeven volgorde worden doorlopen. De uit te zenden blanke instructeurs kunnen alle drie de opleidingen geven, terwijl de instructeur ziekenverzorger A inheems en kan opleiden tot ziekenverzorger en de instructeur ziekenverzorger B de bevoegdheid bezit leerlingen op te leiden zowel tot ziekenverzorger als tot instructeur ziekenverzorger A.

In de opleiding tot ziekenverzorger kan iedere instructeur maximaal vijf leerlingen instrueren.

In de opleiding tot instructeur ziekenverzorger A kan iedere instructeur maximaal drie leerlingen opleiden.

In de opleiding tot instructeur ziekenverzorger B kan elke blanke instructeur maximaal twee leerlingen opleiden.

Aangenomen wordt dat het leerlingenreservoir onbeperkt is en dat iedere leerling geschikt is om elke opleiding te volgen.

De deskundigen vragen zich in het rapport af, hoe de instructeurs in de verschillende jaren over de opleidingen moeten worden verdeeld om het maximale aantal "verplegingsjaren" te verkrijgen in de nog resterende vijf jaren van het beheer³⁾, als verder gegeven is dat voor dit doel vijf blanke instructeurs beschikbaar zijn.

Laten wij eens proberen of wij dit probleem niet in een meer wiskundige vorm kunnen gieten. Hiertoe voeren wij in de grootheden x_{ijk} . Deze symbolen geven aantallen personen aan. Stel dat de index i de status aangeeft van de betrokkenen en wel als volgt:

- $i=1$ ziekenverzorger
- $i=2$ instructeur ziekenverzorger A
- $i=3$ instructeur ziekenverzorger B
- $i=4$ blanke instructeur.

De index j geeft de functie aan en wel als volgt:

3) Een "verplegingsjaar" is een jaar, waarin één ziekenverzorger gedurende een jaar geheel ter beschikking van de verpleging staat. Men spreekt van 10 verplegingsjaren als 5 ziekenverzorger gedurende 2 jaar geheel ter beschikking van de verpleging staan.

$j=1$ niet bij een opleiding betrokken
 $j=2$ in opleiding
 $j=3$ opleider bij de cursus ziekenverzorger
 $j=4$ " " " " instructeur ziekenverzorger A
 $j=5$ " " " " instructeur ziekenverzorger B.

De index k vermeldt het betreffende jaar ($k=1, \dots, 5$).

De grootheid x_{ijk} geeft nu het aantal personen aan, dat aan het eind van het k^{de} jaar de status i heeft en het daarop volgende jaar in de functie j is tewerkgesteld. Zo geeft bijvoorbeeld x_{234} het aantal instructeur ziekenverzorgers A aan, dat in het vijfde jaar opleidt tot ziekenverzorger.

Ga zelf na dat geldt:

$$\begin{array}{llll}
 x_{13k}=0 & x_{21k}=0^4) & x_{31k}=0 & x_{41k}=0 \\
 x_{14k}=0 & x_{24k}=0 & x_{32k}=0 & x_{42k}=0 \\
 x_{15k}=0 & x_{25k}=0 & x_{35k}=0 & (k=1, \dots, 5)
 \end{array} \quad (2.1)$$

$$\begin{array}{ll}
 x_{2j1} = 0 \\
 x_{3j1} = x_{3j2} = 0
 \end{array} \quad (j=1, \dots, 5) \quad (2.2)$$

Kortheidshalve schrijven wij voortaan:

$$x_{i.k} = \sum_{j=1}^5 x_{ijk} \quad (2.3)$$

Uit de verbale formulering van het probleem volgen de relaties:
 Aantal ziekenverzorgers aan het eind van het k^{de} jaar = aantal ziekenverzorgers aan het eind van het $(k-1)^{\text{ste}}$ jaar verminderd met het aantal ziekenverzorgers, dat in het k^{de} jaar een opleiding volgt tot instructeur ziekenverzorger A en vermeerderd met het aantal inheemsen, dat een opleiding tot ziekenverzorger volgt in het k^{de} jaar.

M.a.w.

$$x_{1.k} = x_{1.k-1} - x_{12k-1} + 5 \left[x_{23k-1} + x_{33k-1} + x_{43k-1} \right] \quad (k=1, \dots, 5) \quad (2.4)$$

waarbij

 4) Bedenk dat het niet gebruiken van instructeurs altijd nadelig is. Immers zij kunnen toch altijd beter inheemsen tot ziekenverzorgers opleiden. Dit volgt uiteraard ook uit de oplossing.

$$x_{12k-1} \leq \min(x_{1.k-1}, 3[x_{34k-1} + x_{44k-1}]) \quad (2.5)$$

$$\text{en } x_{ij0} = 0 \quad (i=1,2,3). \quad (2.6)$$

De grootheden x_{430} en x_{440} , die voor $k=1$ in (4) voorkomen, zijn de aantallen blanken die aan het eind van het nulde jaar, d.w.z. aan het begin van het eerste jaar, inheemsen gaan opleiden tot ziekenverzorger resp. instructeur A. Uit de probleemstelling volgt nu direct

$$\begin{aligned} x_{430} &= 5 \\ x_{440} &= 0 \end{aligned} \quad (2.7)$$

De ongelijkheid (2.5) vertelt ons, dat het aantal ziekenverzorgers in opleiding tot instructeur ziekenverzorger A niet groter kan zijn dan het totale aantal ziekenverzorgers, terwijl het bovendien kleiner moet zijn dan de opleidingscapaciteit.

Evenzo gelden de relaties:

$$x_{2.k} = x_{2.k-1} - x_{22k-1} + x_{12k-1} \quad (k=2, \dots, 5) \quad (2.8)$$

waarbij

$$x_{12k-1} \leq \min[x_{1.k-1}, 3(x_{34k-1} + x_{44k-1})] \quad (2.5)$$

$$x_{22k-1} \leq \min(x_{2.k-1}, 2x_{45k-1}) \quad (2.9)$$

$$\text{en } x_{i.0} = 0 \quad (i=1,2,3) \quad (2.6)$$

$$x_{430} = 5, \quad x_{440} = 0 \quad (2.7)$$

$$x_{3.k} = x_{3.k-1} + x_{22k-1} \quad (k=3,4,5) \quad (2.10)$$

$$\text{waarbij } x_{22k-1} \leq \min(x_{2.k-1}, 2x_{45k-1}) \quad (2.8)$$

$$x_{4.k} = x_{4.k-1} = 5 \quad (2.11)$$

Het probleem luidt nu in zijn wiskundige formulering:

Bepaal het max. van de uitdrukking

$$z_0 = x_{111} + x_{112} + x_{113} + x_{114} \quad (2.12)$$

Stel dat wij de eerste vier jaar ons niet met de opleiding bemoeien. In het vijfde jaar echter zijn wij bereid mede te werken aan

het opvoeren van het aantal "verplegingsjaren". Op het moment, dat wij de instructeurs voor de laatste keer over de opleidingen zullen verdelen worden hun aantallen vastgelegd door de grootheden $x_{i.4}$ ($i=1, \dots, 4$). De optimale waarden van x_{ij4} bezitten dus de volgende eigenschappen:

- 1°) aan de voorwaarden (2.4) t/m (2.11) wordt voldaan
- 2°) de uitdrukking (2.12) is onder deze voorwaarden maximaal.

Het is zonder meer duidelijk, dat onze keuze zal zijn

$$\begin{aligned} x_{114} &= x_{1.4} \\ x_{1j4} &= 0 \quad (j \neq 1). \end{aligned} \quad (2.13)$$

De keuze van de overige waarden x_{ij4} ($i=2,3,4$) is van geen belang. Wij vinden dus voor het max. van z_0 :

$$z_1 = \max_{x_{ij4}} z_0 = x_{111} + x_{112} + x_{113} + x_{1.4}. \quad (2.14)$$

Dit antwoord behoeft ons niet te verbazen. Immers in het laatste jaar kan men niet beter doen dan alle ziekenverzorgers in de verpleging sturen om het grootste aantal "verplegingsjaren" te verkrijgen.

Het maximum van z_0 wordt ongeacht de waarde van $x_{i.4}$ gegeven door (2.14). Uit (2.14) volgt, dat het maximum een functie is van $x_{i.4}$.

Wij hebben nu het volgende probleem opgelost:
Gegeven een aantal ziekenverzorgers, instructeurs A en B en vijf blanke instructeurs. Gevraagd deze instructeurs zodanig over de opleiding te verdelen, dat aan het eind van het komende jaar het maximale aantal verplegingsjaren wordt verkregen.

Slechts de term x_{114} in (2.12) is afhankelijk van onze keuze. Schrijven wij nu voor het maximum van x_{114} :

$$H_1(x_{ij3}) = \max x_{114} = x_{1.4} \quad (2.15)$$

dan geldt op grond van (2.4):

$$H_1(x_{ij3}) = x_{113} + 5 \left[x_{233} + x_{333} + x_{433} \right]. \quad (2.16)$$

Tenslotte volgt uit (2.14) en (2.15):

$$z_1 = x_{111} + x_{112} + x_{113} + H_1(x_{1j3}). \quad (2.17)$$

Strekken wij onze zorg ook uit over het voorlaatste jaar, dan dienen wij de waarden x_{ij3} te bepalen en wel zodanig dat

- 1^o) aan de voorwaarden (2.4) t/m (2.11) voldaan wordt,
- 2^o) de uitdrukking (2.12) onder deze voorwaarden maximaal is.

Dit laatste kunnen wij overeenkomstig (2.14) bereiken door de uitdrukking

$$\begin{aligned} z_1 &= x_{111} + x_{112} + x_{113} + H_1(x_{1j3}) \\ &= x_{111} + x_{112} + 2x_{113} + 5 \left[x_{233} + x_{333} + x_{433} \right] \end{aligned} \quad (2.18)$$

te maximaliseren.

Indien wij z_1 als criterium kiezen, dan is het probleem waarvoor wij ons gesteld zijn van dezelfde vorm als het voorafgaande.

Uit (2.17) en (2.18) volgt dat voor de optimale keuze van x_{ij3} geldt:

$$\begin{aligned} 2x_{113} + 5 \left[x_{233} + x_{333} + x_{433} \right] &= \\ &= x_{113} + H_1(x_{1j3}) \end{aligned} \quad (2.19)$$

is maximaal onder de voorwaarden (2.4) t/m (2.11).

Ook nu kan men eenvoudig nagaan, dat voor de optimale keuze geldt:

$$\begin{aligned} x_{113} &= x_{1.3} & x_{333} &= x_{3.3} \\ x_{233} &= x_{2.3} & x_{433} &= x_{4.3} = 5. \end{aligned} \quad (2.20)$$

Wij vinden derhalve voor het max. van z_1 :

$$z_2 = \max_{x_{ij3}} z_1 = x_{111} + x_{112} + 2x_{1.3} + 5 \left[x_{2.3} + x_{3.3} + 5 \right]. \quad (2.21)$$

Schrijven wij nu:

$$H_2(x_{ij2}) = \max_{x_{ij3}} \left[x_{113} + H_1(x_{1j3}) \right] \quad (2.22)$$

dan geldt:

$$z_2 = x_{111} + x_{112} + H_2(x_{1j2}). \quad (2.21)$$

Wanneer ook de keuze in het derde jaar ons niet onverschillig laat, dan zullen wij x_{1j2} zodanig moeten kiezen, dat geldt:

1°) de waarden x_{1j2} voldoen aan de voorwaarden (2.4) t/m (2.11)

2°) de uitdrukking (2.21) is maximaal.

De term in (2.21), welke afhankelijk is van de keuze van x_{1j2} wordt gegeven door:

$$x_{112} + H_2(x_{1j2}) = x_{112} + 2x_{1.3} + 5[x_{2.3} + x_{3.3} + 5] \quad (2.22)$$

en dus volgens (2.4) t/m (2.11):

$$\begin{aligned} &= x_{112} + 2[x_{112} + 5(x_{232} + x_{432})] + 5[x_{2.2} + x_{1.2} - x_{112} + 5] = \\ &= 10x_{232} + 10x_{432} + 5x_{2.2} + 5x_{1.2} - 2x_{112} + 25. \end{aligned} \quad (2.23)$$

Na een vluchtige beschouwing van (2.23) lijkt het of de optimale keuze van x_{1j2} gegeven wordt door:

$$x_{232} = x_{2.2}$$

$$x_{432} = x_{4.2} = 5 \quad (2.24)$$

$$x_{112} = 0.$$

De keuze voldoet echter niet aan de voorwaarden (2.4) t/m (2.11).

Immers $x_{112}=0$ impliceert

$$x_{122} = x_{1.2} \quad (2.25)$$

en bijgevolg:

$$x_{1.2} \leq 3x_{442} \quad (2.26)$$

of

$$x_{442} \geq \left[\frac{x_{1.2}}{3} \right]^+ \cdot 5) \quad (2.27)$$

5) $\left[\frac{a}{b} \right]^+ =$ kleinste geheel getal groter dan of gelijk aan $\frac{a}{b}$.

$\left[\frac{a}{b} \right]^- =$ grootste gehele getal kleiner dan of gelijk aan $\frac{a}{b}$.

Op grond van (2.27) zijn de relaties

$$\begin{aligned} x_{432} &= x_{4.2} \\ x_{112} &= 0 \end{aligned} \quad (2.24)$$

strijdig.

Uit (2.22) volgt volgens (2.4) t/m (2.11) ook:

$$\begin{aligned} x_{112} + H_2(x_{1j2}) &= \\ &= 3x_{1.2} + 10x_{232} + 10x_{4.2} + 5x_{2.2} + 2x_{122} - 10x_{442} - 10x_{452} + 25. \end{aligned} \quad (2.28)$$

Voor elke waarde van x_{442} geldt:

$$x_{122} \leq 3x_{442} \quad (2.29)$$

en derhalve:

$$2x_{122} - 10x_{442} \leq 0. \quad (2.30)$$

Voor de optimale keuze geldt dus:

$$\begin{aligned} x_{442} &= 0 \\ x_{452} &= 0 \\ x_{432} &= x_{4.2} = 5 \\ x_{232} &= x_{2.2} \\ x_{112} &= x_{1.2}. \end{aligned} \quad (2.31)$$

Geven wij het max. van $x_{112} + H_2(x_{1j2})$ aan met $H_3(x_{1j1})$ dan vinden wij dus voor:

$$H_3(x_{1j1}) = \max_{x_{1j2}} [x_{112} + H_2(x_{1j2})] \quad (2.32)$$

$$\begin{aligned} H_3(x_{1j1}) &= 3x_{1.2} + 15x_{2.2} + 10x_{4.2} + 25 = \\ &= 3x_{1.1} + 150 + 12x_{121} - 15x_{441}. \end{aligned} \quad (2.33)$$

Om de optimale keuze van x_{1j1} te verkrijgen moeten wij de functie $x_{111} + H_3(x_{1j1})$ maximaliseren.

Nu geldt:

$$x_{111} + H_3(x_{1j1}) = 4x_{1.1} + 150 + 11x_{121} - 15x_{441}. \quad (2.34)$$

Als $x_{1.1} \leq 15$ dan zal de optimale keuze van x_{441} kleiner of gelijk aan $\left[\frac{x_{1.1}}{3}\right]^+$ moeten zijn. Kiest men x_{441} kleiner dan of gelijk aan $\left[\frac{x_{1.1}}{3}\right]^+$ dan volgt uit de vorm (2.34) dat de optimale keuze van x_{121} gelijk aan $3x_{441}$ is. Beschouwen wij uit (2.34) de term $11x_{121} - 15x_{441}$, die voor deze keuze van x_{121} gelijk wordt aan $18x_{441}$, dan volgt onmiddellijk dat voor de optimale keuze van x_{441} geldt, dat x_{441} gelijk is aan $\left[\frac{x_{11}}{3}\right]^+$ of $\left[\frac{x_{1.1}}{3}\right]^+$. In het laatste geval moet op grond van (2.34) x_{121} gelijk gekozen worden aan $x_{1.1}$. Wij hebben dus de volgende keuzen:

$$\begin{aligned} x_{441} &= \left[\frac{x_{1.1}}{3}\right]^+ \\ x_{121} &= 3\left[\frac{x_{1.1}}{3}\right]^+ \\ x_{111} &= x_{1.1} - 3\left[\frac{x_{1.1}}{3}\right]^+ \\ \text{en } x_{441} &= \left[\frac{x_{1.1}}{3}\right]^+ \\ x_{121} &= x_{1.1}. \end{aligned} \tag{2.35}$$

$$\tag{2.36}$$

De term $11x_{121} - 15x_{441}$ in (2.34) is voor de keuze gegeven in (2.35) gelijk aan:

$$33\left[\frac{x_{1.1}}{3}\right]^+ - 15\left[\frac{x_{1.1}}{3}\right]^+ \tag{2.37}$$

en voor de keuze gegeven in (2.36) gelijk aan:

$$11x_{1.1} - 15\left[\frac{x_{1.1}}{3}\right]^+ \tag{2.38}$$

Als $\left[\frac{x_{1.1}}{3}\right]^+ = \left[\frac{x_{1.1}}{3}\right]^+$ dan zijn beide uitdrukkingen gelijk. Indien echter $\left[\frac{x_{1.1}}{3}\right]^+ = \left[\frac{x_{1.1}}{3}\right]^+ - 1$ is dan en slechts dan de uitdrukking (2.37) groter dan de uitdrukking (2.38) als geldt:

$$\frac{x_{1.1}}{3} - \left[\frac{x_{1.1}}{3}\right]^+ < \frac{5}{11} \tag{2.39}$$

Dit laatste is slechts het geval als

$$\frac{x_{1.1}}{3} - \left[\frac{x_{1.1}}{3}\right]^+ = \frac{1}{3} \tag{2.40}$$

Men vindt dus als $x_{1.1} \leq 15$ voor $H_4(x_{1j1})$:

$$\begin{aligned} H_4(x_{1j0}) &= \max_{x_{1j1}} \left[4x_{1.1} + 150 + 11x_{121} - 15x_{441} \right] = \\ &= 4x_{1.1} + 150 + 18 \left[\frac{x_{1.1}}{3} \right]^- \quad \text{als} \quad \frac{x_{1.1}}{3} - \left[\frac{x_{1.1}}{3} \right]^- = \frac{1}{3} \\ &\quad 15x_{1.1} + 150 - 15 \left[\frac{x_{1.1}}{3} \right]^+ \quad \text{als} \quad \frac{x_{1.1}}{3} - \left[\frac{x_{1.1}}{3} \right]^- \neq \frac{1}{3}. \end{aligned} \quad (2.41)$$

Het maximale aantal verplegingsjaren wordt nu gevonden door de functie $H_4(x_{1j0})$ te maximaliseren. Men dient er echter voor te zorgen dat geldt $x_{1.1} \leq 15$. De optimale keuze van x_{1j0} wordt dan gegeven door:

$$\begin{aligned} x_{410} &= 2 \\ x_{430} &= 3. \end{aligned} \quad (2.42)$$

Het maximum bedraagt 300 verplegingsjaren.

Geldt daarentegen $x_{1.1} > 15$ dan zal op grond van (2.34) de optimale keuze van x_{121} gelijk zijn aan $3x_{441}$. Beschouwen wij de term $11x_{121} - 15x_{441}$, die voor deze keuze van x_{121} gelijk wordt aan $18x_{441}$, dan volgt onmiddellijk dat voor de optimale keuze van x_{441} geldt, dat x_{441} gelijk is aan 5.

Men vindt dus als $x_{1.1} > 15$ voor $H_4(x_{1j1})$:

$$\begin{aligned} H_4(x_{1j0}) &= \max_{x_{1j1}} \left[4x_{1.1} + 150 + 11x_{121} - 15x_{441} \right] = \\ &= 4x_{1.1} + 240. \end{aligned} \quad (2.43)$$

Het maximale aantal verplegingsjaren wordt nu gevonden door de functie $H_4(x_{1j0})$ te maximaliseren. Men dient er echter voor te zorgen dat geldt $x_{1.1} > 15$. De optimale keuze van x_{1j0} wordt dan gegeven door:

$$\begin{aligned} x_{410} &= 0 \\ x_{430} &= 5. \end{aligned} \quad (2.44)$$

Het maximum bedraagt 340 verplegingsjaren.

Vergelijken wij nu de maxima, welke voortvloeien uit de keuzen (2.42) resp. (2.44) dan zien wij onmiddellijk, dat de keuze (2.44) te verkiezen is boven (2.42).

Tenslotte geven wij voor iedere stap nogmaals de optimale keuze:

$$\begin{array}{cccccc}
 & x_{111}=10 & x_{112}=10 & x_{113}=110 & x_{114}=210 & \\
 & & & & & \\
 x_{430}=5 & x_{121}=5 & x_{122}=0 & x_{123}=0 & x_{124}=0 & \\
 & x_{431}=0 & x_{222}=0 & x_{233}=0 & & \\
 & x_{441}=5 & x_{232}=15 & x_{233}=15 & & \\
 & & x_{432}=5 & x_{433}=5 & & \\
 & & x_{442}=0 & x_{443}=0 & &
 \end{array}$$

Tot slot zullen wij nog een tweede d.p. probleem beschouwen.

Een reisbureau heeft voor een periode van 6 jaar een hotel gepacht in een wintersportcentrum. Met een plaatselijke kolenhandelaar is een contract afgesloten, waarin wordt bepaald dat de kolenhandelaar elk jaar een vaste hoeveelheid brandstof zal leveren tegen een in het contract genoemd bedrag. Verder mag het reisbureau in geval van wanprestatie het contract ééenzijdig opzeggen.

De kolenhandelaar verkoopt 3 soorten kolen en wel de volgende:

- a) superkolen
- b) kwaliteitskolen
- c) huishoudkolen.

Wij zullen de bij de verkoop per jaar te maken winst aangeven met a_i ($i=1,2,3$). De kolensoorten zijn genummerd in de hierboven aangegeven volgorde. Voor de getallen a_i geldt de ongelijkheid:

$$a_1 < a_2 < a_3 \quad (2.45)$$

De kans op opzegging van het contract na levering van de i^{de} kolensoort wordt aangeduid met p_i en is voor elk jaar gelijk. Voor de kansen p_i geldt de ongelijkheid:

$$p_1 < p_2 < p_3 \quad (2.46)$$

De kolenhandelaar, een harde zakenman, vraagt zich af welke kolensoorten hij in de opéénvolgende jaren bij het hotel zal afleveren.

Stel dat het reisbureau het hotel niet voor 6 jaar maar voor 1 jaar heeft gepacht. De kolenhandelaar zal ongetwijfeld in dat geval huishoudkolen hebben afgeleverd.

Indien wij zijn maximale winst aangeven met W_1 dan geldt

$$W_1 = \max [a_1; a_2; a_3] = a_3. \quad (2.47)$$

Bij een tweejarig contract zal de kolenhandelaar in het laatste jaar als hij mag leveren een winst maken van de grootte W_1 .

Levert hij in het eerste jaar de i^{de} kolensoort dan bedraagt de verwachting van zijn winst in deze twee jaren:

$$a_i + (1-p_i)W_1. \quad (2.48)$$

Voor de kolen welke hij in het eerste jaar het beste kan afleveren geldt, dat (2.48) maximaal is. Indien wij dit maximum aangeven met W_2 , dan vinden wij

$$W_2 = \max_i [a_i + (1-p_i)W_1]. \quad (2.49)$$

Strekt het contract zich over drie jaar uit, dan geldt voor de optimale keuze, dat

$$a_j + (1-p_j)W_2 \quad (2.50)$$

maximaal is. Het maximum wordt gegeven door

$$W_3 = \max_j [a_j + (1-p_j)W_2]. \quad (2.51)$$

Op deze wijze kunnen wij voortgaan en wij vinden tenslotte voor het maximum

$$W_6 = \max_k [a_k + (1-p_k)W_5], \quad (2.52)$$

waarbij W_5 op een analoge wijze is gedefinieerd.

Van bovenstaand resultaat zullen wij nu een toepassing beschouwen. Stel dat

$$\begin{array}{lll} a_1 = f & 435.- & p_1 = 0,2 \\ a_2 = f & 790.- & p_2 = 0,4 \\ a_3 = f & 1.050.- & p_3 = 0,6. \end{array} \quad (2.53)$$

Voor de uitdrukking (2.48) vinden wij dan

$$\begin{array}{l} a_1 + (1-p_1)W_1 = f \quad 435.- + 0,8 \times f \quad 1.050.- = f \quad 1.275.- \\ a_2 + (1-p_2)W_1 = f \quad 790.- + 0,6 \times f \quad 1.050.- = f \quad 1.420.- \\ a_3 + (1-p_3)W_1 = f \quad 1.050.- + 0,4 \times f \quad 1.050.- = f \quad 1.470.- \end{array} \quad (2.54)$$

Het maximum van (2.48) wordt dus bereikt voor $i=3$.
De kolenhandelaar zal dus in het voorlaatste jaar huishoudkolen leveren. Uit (2.54) volgt tevens $W_2 = f 1.470.-$.

Voor de uitdrukking (2.50) vinden wij nu:

$$\begin{aligned} a_1 + (1-p_1)W_2 &= f 435.- + 0,8 \times f 1.470.- = f 1.611.- \\ a_2 + (1-p_2)W_2 &= f 790.- + 0,6 \times f 1.470.- = f 1.672.- \quad (2.55) \\ a_3 + (1-p_3)W_2 &= f 1.050.- + 0,4 \times f 1.470.- = f 1.638.-. \end{aligned}$$

Het maximum van (2.50) wordt bereikt voor $i=2$.
De kolenhandelaar zal dus in het vierde jaar de zgn. kwaliteitskolen leveren. Uit (2.55) volgt voor W_3 : $W_3 = f 1.672.-$.

Voor het maximum W_4 van

$$\begin{aligned} a_1 + (1-p_1)W_3 &= f 435.- + 0,8 \times f 1.672.- = f 1.772,60 \\ a_2 + (1-p_2)W_3 &= f 790.- + 0,6 \times f 1.672.- = f 1.793,20 \quad (2.56) \\ a_3 + (1-p_3)W_3 &= f 1.050.- + 0,4 \times f 1.672.- = f 1.718,80 \end{aligned}$$

vinden wij $f 1.793,20$. Dit betekent, dat de optimale keuze in het derde jaar wordt gevormd door de kwaliteitskolen.

Voor het maximum W_5 van

$$\begin{aligned} a_1 + (1-p_1)W_4 &= f 435.- + 0,8 \times f 1.793,20 = f 1.869,56 \\ a_2 + (1-p_2)W_4 &= f 790.- + 0,6 \times f 1.793,20 = f 1.865,92 \quad (2.57) \\ a_3 + (1-p_3)W_4 &= f 1.050.- + 0,4 \times f 1.793,20 = f 1.767,38 \end{aligned}$$

vinden wij $f 1.869,56$. Dit betekent, dat de optimale keuze in het tweede jaar wordt gevormd door de superkolen.

Voor het maximum W_6 van

$$\begin{aligned} a_1 + (1-p_1)W_5 &= f 435.- + 0,8 \times f 1.869,56 = f 1.930,65 \\ a_2 + (1-p_2)W_5 &= f 790.- + 0,6 \times f 1.869,56 = f 1.911,74 \quad (2.58) \\ a_3 + (1-p_3)W_5 &= f 1.050.- + 0,4 \times f 1.869,56 = f 1.797,82 \end{aligned}$$

vinden wij $f 1.930,65$. Dit betekent, dat de optimale keuze in het eerste jaar wordt gevormd door de superkolen.

3. Het n-staps d.p. probleem

Alvorens een terugblik te werpen op de beide voorbeelden uit de vorige paragraaf zullen wij eerst een aantal begrippen omschrijven.

Het eerste begrip, dat wij zullen invoeren heet systeem. Het systeem is veelal het object van onze beschouwingen. Zo kan bij een wachttijdprobleem het systeem gevormd worden door de rij wachtenden voor het loket. Bij een voorraadprobleem zal met het systeem meestal de voorraad worden bedoeld. In het eerste voorbeeld uit de vorige paragraaf duidt men met het systeem "het verplegen van zieken" aan, terwijl in het laatste voorbeeld met het systeem "de verkoop van kolen" wordt bedoeld.

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

Een dergelijke rij van getallen noemt men een toestandsvector. De toestand aan het eind van het i^{de} jaar van het systeem "het verplegen van zieken" wordt gegeven door

- a) het aantal ziekenverzorgers $x_{1.i}$
- b) het aantal instructeur ziekenverzorgers A $x_{2.i}$
- c) het aantal " " " B $x_{3.i}$
- d) " " blanken $x_{4.i}$.

De mogelijke toestanden van het systeem "kolenleverantie" (2^{de} voorbeeld) zijn:

- 1°) contract is geldig voor het komende jaar
- 2°) contract is niet meer geldig voor het komende jaar.

Deze beide toestanden kunnen worden beschreven door één kwantitatieve grootheid, die voor de beide toestanden resp. de waarde één of nul aanneemt.

Stel dat de toestanden van een systeem beschreven kunnen worden met behulp van m kwantitatieve grootheden, dan kan iedere toestand geïdentificeerd worden met een punt in een door m orthogonale coördinaatassen opgespannen ruimte of met de toestandsvector, die de verbinding van dat punt met de oorsprong van het assenstelsel is. Op de assen van het assenstelsel worden de eerdergenoemde kwantitatieve grootheden uitgezet. Deze ruimte zullen wij de toestandsruimte noemen.

In elk van de voorbeelden uit de vorige paragraaf moet op een eindig aantal discrete tijdstippen een beslissing worden genomen.

Indien in een d.p. probleem op n discrete tijdstippen een beslissing genomen dient te worden, dan spreekt men van een n -steps d.p. probleem. Zodra de eerste beslissing genomen is gaat het n -steps probleem over in een $(n-1)$ -steps d.p. probleem. We zullen aannemen, dat ook iedere mogelijke beslissing vastgelegd kan worden met behulp van een aantal kwantitatieve grootheden. In het eerste voorbeeld wordt de op het k^{de} tijdstip gemaakte beslissing gegeven door x_{ijk} , terwijl in het tweede probleem de verschillende beslissingen kunnen worden beschreven met één kwantitatieve grootheid, welke slechts de waarden 1, 2 en 3 behoeft te doorlopen.

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.

Bij een zorgvuldige beschouwing van de problemen uit de vorige paragraaf valt het ons op, dat de toestand van het systeem op het volgende tijdstip van beslissen sterk wordt beïnvloed door de toestand op het huidige tijdstip van beslissen en door de beslissing zelf en echter niet door de toestanden en beslissingen op voorgaande tijdstippen. In het eerste voorbeeld wordt de toestand van het systeem aan het eind van het komende jaar volledig vastgelegd door de huidige toestand en de op dat tijdstip te nemen beslissing (zie (2.4) t/m (2.11)).

In het tweede voorbeeld daarentegen bestaat ook na de beslissing onzekerheid omtrent de toestand van het systeem op het volgende beslissingstijdstip. De kansen op realisering van de verschillende mogelijke toestanden worden echter wel bepaald door de huidige toestand en de te nemen beslissing.

Verder kunnen wij constateren, dat ten gevolge van een beslissing in elk van de beide voorbeelden sprake is van een opbrengst.¹⁾ De omvang van deze opbrengst hangt af van de genomen beslissing en de toestand waarin deze beslissing is genomen.

1) Er zijn d.p. problemen, waarbij de opbrengst stochastisch is. In die gevallen beschouwt men de verwachting van de opbrengst.

Een reeks van beslissingen zullen wij optimaal noemen als de totale opbrengst maximaal is.²⁾

Verder blijkt dat wij voor de bepaling van de optimale beslissing op één tijdstip niet kunnen volstaan door alleen maar de opbrengst te beschouwen, welke een direct gevolg is van de beslissing. Immers de beslissing bepaalt mede de toestand op het volgende beslissingstijdstip en dientengevolge de toekomstige opbrengsten. Om het effect van een beslissing op toekomstige opbrengsten te kunnen nagaan, dient men, omdat de opbrengsten hierdoor mede worden bepaald, de toekomstige beslissingen te kennen. Aan deze moeilijkheid wordt op de volgende wijze het hoofd geboden. Men begint nl. bij een n-staps d.p. probleem niet met de bepaling van de eerste optimale beslissing maar met de laatste. Hiervoor dient men de toestand te kennen op het bijbehorende tijdstip. Aangezien men die toestand niet kent lost men het probleem op voor alle mogelijke toestanden. De maximale opbrengst is derhalve een functie van de toestand op het laatste beslissingstijdstip. Deze toestanden zelf (eerste voorbeeld) of de kansverdeling behorende bij deze toestanden (tweede voorbeeld) worden volledig bepaald door de toestand van het systeem en de beslissing op het voorgaande moment van beslissen. De maximale opbrengst kan dus beschouwd worden als een functie alleen van die grootheden. Indien men een nieuwe opbrengstfunctie invoert, die gelijk is aan de oude vermeerderd met de maximum opbrengst van de volgende beslissing, dan kan deze nieuwe opbrengstfunctie gemaximaliseerd worden, waardoor de optimale beslissing voor het op één na laatste moment van beslissen wordt verkregen.

Ook nu geldt dat de toestand van het systeem op het beslissings-tijdstip niet bekend is zodat het maximum een functie is van deze toestand.

Resumerend kunnen wij vaststellen, dat wij eerst 1-staps problemen oplossen voor alle mogelijke uitgangstoestanden van het systeem en daarna met behulp van deze oplossingen 2-staps problemen oplossen voor alle mogelijke begintoestanden.

Indien men op deze wijze voortgaat, zo blijkt uit beide vraagstukken, dan wordt de oplossing van het oorspronkelijk gestelde vraagstuk verkregen.

2) Er zijn d.p. problemen, waarbij niet de totale opbrengst maximaal gekozen moet worden maar een gewogen som van opbrengsten. Indien de "opbrengst" een verlies is dan wordt uiteraard geminimaliseerd.

Bij het oplossen van een n-staps d.p. probleem wordt het oorspronkelijke probleem dus vervangen door een reeks van in eenvoud afnemende problemen, welke worden opgelost voor alle mogelijke begintoestanden.

Indien wij de toestand van het systeem en de beslissing op het m^{de} beslissingstijdstip resp. aangeven met x_m en d_m en als wij verder schrijven voor de bijbehorende opbrengst $h(x_m, d_m)$ en de toekomstige³⁾ opbrengsten $H_{n-m}(x_m, d_m)$ bij optimale keuze van d_{m+1}, d_{m+2} enz. dan geldt voor een n-staps d.p. probleem:

$$H_{n-m}(x_m, d_m) = \max_{d_{m+1}} \left[h(x_{m+1}, d_{m+1}) + H_{n-m-1}(x_{m+1}, d_{m+1}) \right] \quad (m \leq n-1) \quad (3.1)$$

waarbij

$$H_0(x_n, d_n) = 0. \quad (3.2)$$

Uit deze beschouwing volgt tevens, dat bij een n-staps d.p. probleem de eerste optimale beslissing d_1 wordt bepaald door de toestand x_1 . Men zou ook kunnen zeggen dat door de optimale keuze aan een toestandsvector in de toestandruimte een beslissingsvector in de beslissingsruimte wordt toegevoegd. Deze toevoeging kan men ook als volgt uitdrukken:

$$d_1 = S_n(x_1). \quad (3.3)$$

Bellman heeft door gebruik te maken van het abstracte begrip Functieruimte de zgn. strategieruimte ingevoerd. Ieder punt uit deze ruimte is een functie met behulp waarvan aan iedere toestandsvector een beslissingsvector kan worden toegevoegd. Ook de functie $S_n(x)$ in (3.3) is een punt uit deze ruimte en wordt een strategie-functie of ook wel kortweg strategie genoemd.

4. Het ∞ -staps d.p. probleem

In de vorige paragrafen hebben wij ons beperkt tot d.p. problemen, waarin slechts een eindig aantal beslissingen wordt genomen. In deze paragraaf echter zullen wij ons bezig houden met situaties, waarin het aantal te nemen beslissingen onbegrensd is. Het is zonder meer duidelijk, dat de oplossing van deze problemen niet verkregen wordt door eerst de beslissing van de "laatste" stap te bepalen.

3) evt. gewogen

Stel dat de maximale opbrengst bij een n-staps d.p. probleem gegeven wordt door de functie $f_n(x)$, waarbij x de toestandsvector aanduidt op het eerste tijdstip van beslissen.

Uit de definitie van $H_1(x,d)$ volgt:

$$f_n(x) = \max_d [h(x,d) + H_{n-1}(x,d)] . \quad (4.1)$$

Men kan nu de volgende gevallen onderscheiden: 1)

- a) Uit de probleemstelling volgt, dat de toestand van het systeem op het volgende beslissingstijdstip volledig bepaald wordt door de toestand op het huidige moment van beslissen en de beslissing zelf.
- b) Uit de probleemstelling volgt, dat de kansverdeling van de toestandsvector behorende bij het volgende beslissingstijdstip volledig bepaald wordt door de toestand op het huidige moment van beslissen en de beslissing zelf.

Onafhankelijk hiervan kunnen zich nog de volgende situaties voordoen 2)

- c) Uit de probleemstelling volgt, dat bij de i^{de} beslissing met de volledige opbrengst te verkrijgen uit toekomstige beslissingen rekening wordt gehouden.
- d) Uit de probleemstelling volgt, dat op het i^{de} beslistijdstip slechts rekening wordt gehouden met een fractie van de opbrengsten behorende bij toekomstige beslissingen en wel zo dat de bijdrage van de j^{de} beslissing ($j \geq i$) tot de totale opbrengst gelijk is aan:

$$\propto j^{-1} h(x_j, d_j) . \quad (4.2)$$

Schrijft men nu voor de toestandsvector en beslissingsvector behorend bij het i^{de} beslissingstijdstip, x_i resp. d_i , dan vinden wij voor de opbrengstfunctie $f_n(x)$:

- 1) De situatie, welke onder a) wordt beschreven, kan beschouwd worden als een speciaal geval van die vermeld onder b).
- 2) De situatie, welke onder c) wordt beschreven, kan beschouwd worden als een speciaal geval van die vermeld onder d).

$$\text{I} \quad f_n(x_1) = \max_{d_1} \left\{ h(x_1, d_1) + f_{n-1}[\underline{x}_2(x_1, d_1)] \right\} \quad (4.3)$$

als zowel a) als c) van toepassing is.

$$\text{II} \quad f_n(x_1) = \max_{d_1} \left\{ h(x_1, d_1) + \alpha f_{n-1}[\underline{x}_2(x_1, d_1)] \right\} \quad (4.4)$$

als zowel a) als d) van toepassing is.

$$\text{III} \quad f_n(x_1) = \max_{d_1} \left[h(x_1, d_1) + E \left\{ f_{n-1}(\underline{x}_2) / x_1; d_1 \right\} \right] \quad (4.5)$$

als zowel b) als c) van toepassing is.

$$\text{IV} \quad f_n(x_1) = \max_{d_1} \left[h(x_1, d_1) + \alpha E \left\{ f_{n-1}(\underline{x}_2) / x_1; d_1 \right\} \right] \quad (4.6)$$

als zowel b) als d) van toepassing is.

Beschouwen wij nu de rij van functies $f_n(x)$ dan zullen, als deze rij voor iedere x convergeert tot een functie $f(x)$, de relaties (4.3) t/m (4.6) overgaan in: ³⁾

$$\text{I} \quad f(x_1) = \max_{d_1} \left\{ h(x_1, d_1) + f[\underline{x}_2(x_1, d_1)] \right\} \quad (4.7)$$

$$\text{II} \quad f(x_1) = \max_{d_1} \left\{ h(x_1, d_1) + \alpha f[\underline{x}_2(x_1, d_1)] \right\} \quad (4.8)$$

$$\text{III} \quad f(x_1) = \max_{d_1} \left[h(x_1, d_1) + E \left\{ f(\underline{x}_2) / x_1; d_1 \right\} \right] \quad (4.9)$$

$$\text{IV} \quad f(x_1) = \max_{d_1} \left[h(x_1, d_1) + \alpha E \left\{ f(\underline{x}_2) / x_1; d_1 \right\} \right]. \quad (4.10)$$

Indien de opbrengstfuncties $f_n(x)$ behorende bij de overeenkomende n -steps d.p. problemen convergeren tot de functie $f(x)$, dan wordt de oplossing van het ∞ -steps d.p. probleem verkregen uit de oplossing van de corresponderende functionaalvergelijking [(4.7) t/m (4.10)].

Het onderzoek naar de ondubbelzinnigheid en de existentie van de oplossingen van deze functionaalvergelijkingen vormt een wezenlijk deel van de dynamische programmering. Aan dit belangrijke onderdeel van de theorie der dynamische programmering zullen wij echter geen aandacht besteden. ⁴⁾

1^{ste} iteratie methode. Indien de rij van opbrengstfuncties $f_n(x)$ van de overeenkomende n -steps d.p. problemen convergeert voor iedere x tot de opbrengstfunctie $f(x)$ van het ∞ -steps d.p. probleem, dan kan deze functie uiteraard iteratief worden bepaald. Heeft men

3) De relaties (4.7) t/m (4.9) zijn bijzondere gevallen van (4.10).

4) Wij verwijzen hiervoor naar R. Bellman: Dynamic Programming blz. 116 e.v.

de functie $f(x)$ gevonden dan kan men met behulp van één der relaties (4.7) t/m (4.10) voor iedere toestandsvector x de optimale beslissing d bepalen.

2^{de} iteratie methode. Indien voor iedere strategiefunctie $S(x)$ eenvoudig de bijbehorende opbrengst $f(x)$ kan worden bepaald, dan kan men dikwijls ook op een andere iteratieve wijze de functionaalvergelijkingen (4.7) t/m (4.10) oplossen. Men begint dan met een willekeurige strategiefunctie $S^{(0)}(x)$, waarvoor de bijbehorende opbrengst $f^{(0)}(x)$ wordt bepaald. Voor de strategiefunctie $d=S^{(1)}(x)$ geldt nu:

$$I \quad h(x_1, d_1) + f^{(0)}[x_2(x_1, d_1)] \quad (4.11)$$

$$II \quad h(x_1, d_1) + \alpha f^{(0)}[x_2(x_1, d_1)] \quad (4.12)$$

$$III \quad h(x_1, d_1) + E \{ f^{(0)}(\underline{x}_2)/x_1, d_1 \} \quad (4.13)$$

$$IV \quad h(x_1, d_1) + \alpha E \{ f^{(0)}(\underline{x}_2)/x_1, d_1 \} \quad (4.14)$$

is maximaal. De bijbehorende opbrengstfunctie $f^{(1)}(x)$ kan met behulp van $S^{(1)}(x)$ worden bepaald. Indien men op deze wijze voortgaat maximaliseert de strategiefunctie $d_1=S^{(i)}(x)$ de uitdrukkingen

$$I \quad h(x_1, d_1) + f^{(i-1)}[x_2(x_1, d_1)] \quad (4.15)$$

$$II \quad h(x_1, d_1) + \alpha f^{(i-1)}[x_2(x_1, d_1)] \quad (4.16)$$

$$III \quad h(x_1, d_1) + E \{ f^{(i-1)}(\underline{x}_2)/x_1, d_1 \} \quad (4.17)$$

$$IV \quad h(x_1, d_1) + \alpha E \{ f^{(i-1)}(\underline{x}_2)/x_1, d_1 \}. \quad (4.18)$$

Uit de constructie verkrijgen wij de rijen van functies $S^{(i)}(x)$ en $f^{(i)}(x)$ ($i=1, \dots$). Indien de rij $f^{(i)}(x)$ convergeert en men weet dat de oplossing van de functionaalvergelijkingen (4.7) t/m (4.10) bestaat en ondubbelzinnig is, dan kan men ook op deze wijze de optimale strategie bepalen. Deze oplossingsmethode wordt ook wel eens aangegeven met de naam Approximatie in de strategieruimte.

Indien echter bij een ∞ -steps d.p. probleem de totale opbrengst gezien vanuit de toestand op het eerste beslissingstijdstip voor meer dan één reeks van beslissingen onbegrensd groot wordt, dan kan men de gewenste oplossing niet met behulp van bovenstaande functionaalvergelijkingen verkrijgen. Men zal in zo'n situatie trachten een nieuw criterium op te stellen om alsnog objectief de resultaten van deze verschillende reeksen van beslissingen te kunnen onderscheiden.

Soms is het zinvol te beschouwen de rij van functies ⁵⁾

$$g_n(x) = \frac{f_n(x)}{n} \quad (4.19)$$

Indien de rij van opbrengstfuncties $f_n(x)$ van de overeenkomende n-staps d.p. problemen niet convergeert, maar wel de rij van strategiefuncties $s_n(x)$ [zie (3.3)] dan kan men ondanks de divergentie van de rij $f_n(x)$ toch de optimale strategie

$$S(x) = \lim_{n \rightarrow \infty} S_n(x) \quad (4.20)$$

bepalen. Veelal wenst men slechts de optimale strategie te kennen en is dus hiermede het probleem opgelost. In dergelijke situaties komt het ook veel voor dat de limiet

$$\lim_{n \rightarrow \infty} \frac{f_n(x)}{n} = \lim_{n \rightarrow \infty} g_n(x) = g(x) \quad (4.21)$$

bestaat.

De functies $g_n(x)$ stellen de gemiddelde winst per beslissing voor bij een n-stap d.p. probleem. Bij menig probleem convergeert deze rij van functies tot een functie $g(x)$.

In par. 5 zullen wij voor een klasse van d.p. problemen een methode bespreken, waarmee men de functionaalvergelijkingen op een meer directe wijze kan oplossen. Deze methode is slechts te gebruiken als de functionaalvergelijkingen van een bepaalde structuur zijn.

Een uitgebreide groep van ∞ -staps d.p. problemen kan worden opgelost met behulp van de theorie der Markov ketens. Bij deze d.p. problemen geldt dat de toestandsvectoren, behorende bij de door het systeem te doorlopen toestanden, voor een vaste strategiefunctie $S(x)$ een Markov-keten vormen. De verwachting van de opbrengst voor een beslissing in de limiettoestand van het Markovproces is dan het criterium voor de beste strategie. Van deze d.p. problemen hebben wij reeds een drietal opgelost in hoofdstuk IX (waarschijnlijkhedsrekenen). De oplossingsmethode wordt in dat hoofdstuk uitvoerig besproken, zodat op deze plaats slechts met een verwijzing kan worden volstaan.

5) Wij gebruiken in dit hoofdstuk vaak de termen soms, veelal, etc. zonder precies aan te geven onder welke voorwaarden de uitspraak waar is. Ook hiervoor geldt 4).

5. Een speciale klasse van ∞ -staps d.p. problemen

In deze paragraaf zullen wij d.p. problemen bespreken, die ons door hun speciale vorm in staat stellen zonder iteraties de gewenste resultaten te behalen.

Als prototype van deze klasse beschouwen wij het volgende probleem:

Een hartstochtelijk visser heeft in een café horen spreken van een boer, die twee vennetjes bezit met zeer mooie en grote vissen en waarin niet gevisst mag worden. Volgens zeer nauwkeurige schattingen zouden er in het vennetje V_1 X vissen zijn, terwijl in het vennetje V_2 volgens deze berekeningen Y vissen moeten rondzwemmen. Alle sprekers in het café waren van mening, dat de kans om betrappt te worden bij vennetje V_1 gelijk is aan p_1 , terwijl de overeenkomstige kans voor V_2 gegeven wordt door p_2 . De visser, die door deze mededelingen nieuwsgierig is geworden besluit eens te gaan kijken. Op grond van zijn uitgebreide ervaring meent hij na een vluchtige studie van de ligging en de vorm van de vennetjes te mogen vaststellen dat de fracties van het totale aantal aanwezige vissen, die hij per keer aan de haak kan slaan steeds gegeven wordt door r_1 resp. r_2 . Thuis gekomen overweegt hij, of hij als vreemdeling in deze omgeving geacht wordt het visverbod te kennen.

Een ∞ -staps d.p. probleem ontstaat, wanneer de visser besluit op zijn vrije middagen te gaan vissen totdat hij betrappt wordt en zich bovendien op elk van die middagen afvraagt in welk van de twee vennetjes hij zijn hengel zal uitwerpen.

In dit ∞ -staps d.p. probleem is het systeem "het vissen in de beide vennen". De toestand van het systeem wordt gegeven door de nog resterende aantallen vissen in V_1 en V_2 . De toestandsruimte wordt dus opgespannen door twee onderling loodrechte coördinaatassen, waarop wordt aangegeven hoeveel vissen nog in de respectievelijke vennetjes aanwezig zijn.

Stel dat de visser slechts éénmaal gaat vissen, dan wordt de verwachte vangst gegeven door

$$f_1(x,y) = \max \{ (1-p_1)r_1X; (1-p_2)r_2Y \}. \quad (5.1)$$

Uit deze relatie volgt onmiddellijk:

$$1^{\circ}) \quad f_1(\alpha X, \alpha Y) = \alpha f_1(X, Y) \quad (5.2)$$

$$2^{\circ}) \quad f_1(X, Y) \text{ is continu in } X \text{ en } Y.$$

Indien hij tweemaal gaat vissen dan wordt de verwachting van de totale opbrengst gegeven door:

$$f_2(X, Y) = \max \left\{ (1-p_1)r_1X + (1-p_1)f_1[(1-r_1)X, Y] ; \right. \\ \left. (1-p_2)r_2Y + (1-p_2)f_1[X; (1-r_2)Y] \right\} . \quad (5.3)$$

Uit (5.3) volgt dat ook voor $f_2(X, Y)$ geldt:

$$1^{\circ}) \quad f_2(\alpha X, \alpha Y) = \alpha f_2(X, Y) \quad (5.4)$$

$$2^{\circ}) \quad f_2(X, Y) \text{ is continu in } X \text{ en } Y.$$

Indien hij n maal gaat vissen dan wordt de verwachting van de totale vangst gegeven door:

$$f_n(X, Y) = \max \left\{ (1-p_1)r_1X + (1-p_1)f_{n-1}[(1-r_1)X, Y] ; \right. \\ \left. (1-p_2)r_2Y + (1-p_2)f_{n-1}[X; (1-r_2)Y] \right\} . \quad (5.5)$$

Met behulp van de methode der volledige inductie kan men voor elk geheel getal $n > 0$ bewijzen:

$$1^{\circ}) \quad f_n(\alpha X, \alpha Y) = \alpha f_n(X, Y) \quad (5.6)$$

$$2^{\circ}) \quad f_n(X, Y) \text{ is continu in } X \text{ en } Y.$$

Men kan bewijzen dat de limiet $f_n(X, Y)$ voor $n \rightarrow \infty$ voor iedere waarde van X en Y bestaat en dat de relatie (5.5) overgaat in ¹⁾

$$f(X, Y) = \max \left\{ (1-p_1)r_1X + (1-p_1)f[(1-r_1)X, Y] ; \right. \\ \left. (1-p_2)r_2Y + (1-p_2)f[X; (1-r_2)Y] \right\} . \quad (5.7)$$

Deze functionaalvergelijking heeft één oplossing.

Ook voor de functie $f(X, Y)$ geldt:

$$1^{\circ}) \quad f(\alpha X, \alpha Y) = \alpha f(X, Y) \quad (5.8)$$

$$2^{\circ}) \quad f(X, Y) \text{ is continu in } X \text{ en } Y.$$

Geeft men de beslissing om in V_1 te gaan vissen aan met D_1 . De alternatieve beslissing wordt dan met D_2 aangeduid. De keuze

1) Voor het bewijs van deze uitspraken wordt verwezen naar R.Bellman, Dynamic Programming blz. 64 e.v.

tussen beide beslissingen wordt bepaald door de toestand van het systeem op het beslissingstijdstip. Voor ieder punt zou men na kunnen gaan welke beslissing optimaal is. Uit (5.8) volgt dat deze keuze ook optimaal is voor alle punten op de verbindingslijn met de oorsprong. Uit (5.8) volgt bovendien, dat als in een punt D_1 te verkiezen is boven D_2 dit dan ook geldt voor een klein genoeg gekozen omgeving van het punt. Wanneer men deze feiten combineert dan moet de toestandsruimte te verdelen zijn in sectoren, waarvan de vorm aangegeven is in fig. 5.1.

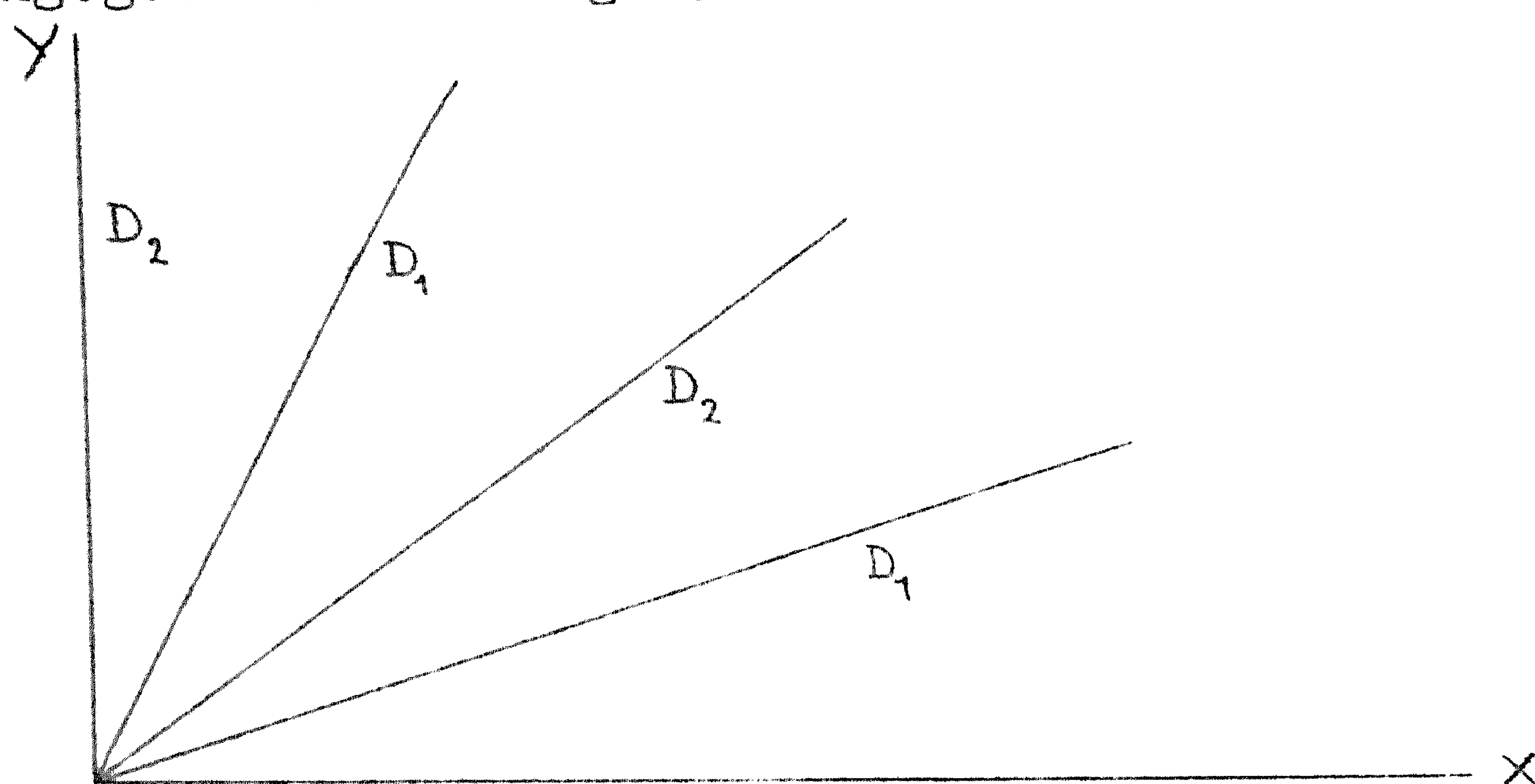


fig. 5.1

Een "op het eerste gezicht" mogelijke verdeling van de toestandsruimte

In ieder van de sectoren is steeds dezelfde beslissing optimaal.

Laten wij eens nagaan voor welke punten in de toestandsruimte het onverschillig is of men eerst de beslissing D_1 , daarna D_2 neemt en vervolgens optimaal verder gaat dan wel dat men begint met D_2 en daarna D_1 alvorens optimaal verder te beslissen. De verwachte opbrengst zullen wij aangeven met $f_{D_1 D_2}(X, Y)$ resp. $f_{D_2 D_1}(X, Y)$. Men kan eenvoudig nagaan, dat deze functies gegeven worden door:

$$f_{D_1 D_2}(X, Y) = (1-p_1)r_1X + (1-p_1)(1-p_2)r_2Y + f[(1-r_1)X, (1-r_2)Y] \quad (5.9)$$

$$f_{D_2 D_1}(X, Y) = (1-p_2)r_2Y + (1-p_2)(1-p_1)r_1X + f[(1-r_1)X, (1-r_2)Y]. \quad (5.10)$$

Voor bovenstaande toestandspunten geldt:

$$f_{D_1 D_2}(X, Y) = f_{D_2 D_1}(X, Y) \quad (5.11)$$

$$\text{of } (1-p_1)r_1X + (1-p_1)(1-p_2)r_2Y = (1-p_2)r_2Y + (1-p_1)(1-p_2)r_1X \quad (5.12)$$

$$\text{of } Y = \frac{(1-p_1)p_2r_1}{(1-p_2)p_1r_2} X. \quad (5.13)$$

Uit deze relaties volgt, dat de gezochte punten liggen op een rechte $\mathcal{L}$ door de oorsprong en welke is aangegeven in fig. 5.2.

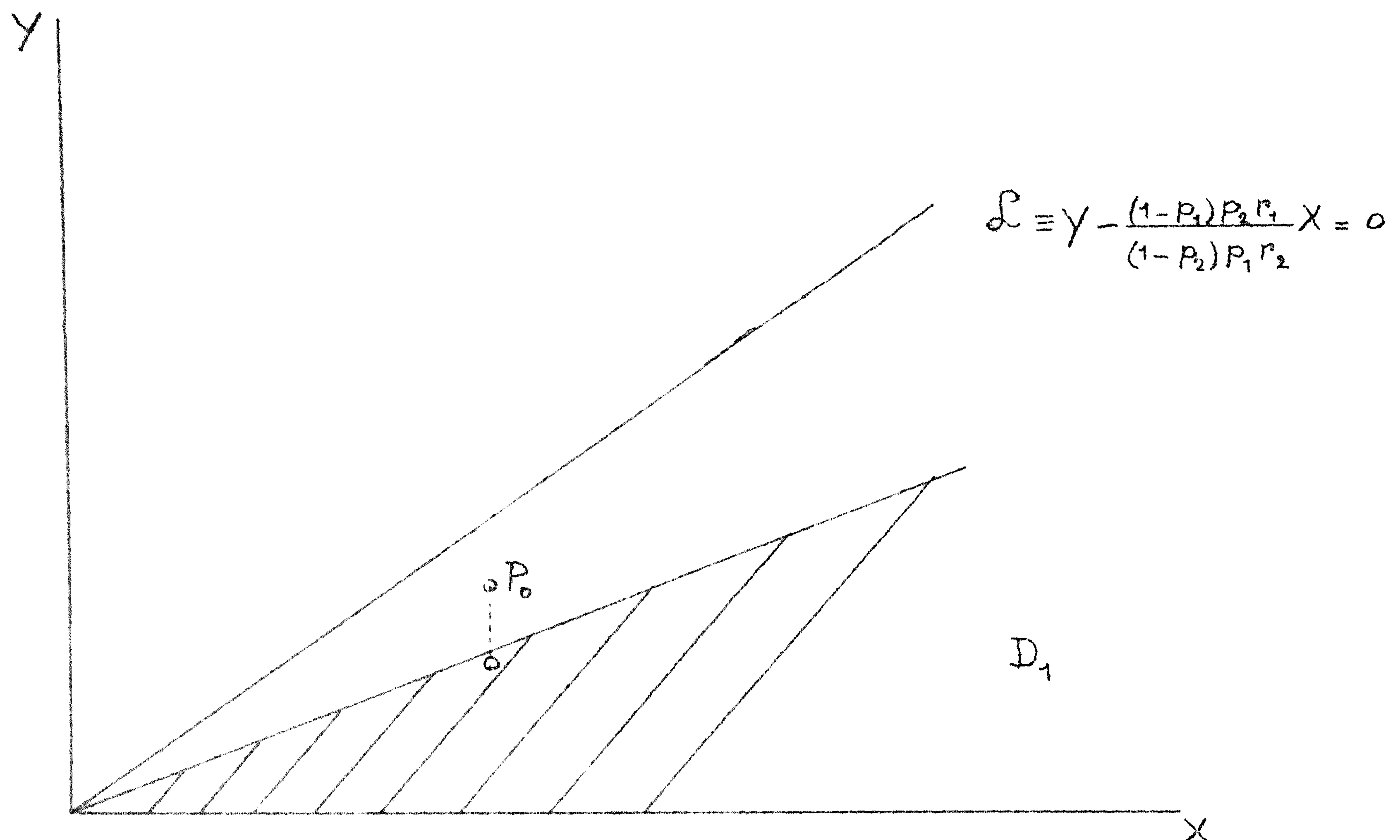


fig. 5.2

Verder blijkt uit (5.9) en (5.10) dat voor alle punten boven $\mathcal{L}$ geldt:

$$f_{D_2D_1}(X,Y) > f_{D_1D_2}(X,Y) \quad (5.14)$$

terwijl voor de punten onder $\mathcal{L}$ het omgekeerde waar is. Uit de continuïteitseigenschappen van $f(X,Y)$ volgt, dat er een sector grenzend aan de X as bestaat, waarin de beslissing D_1 te verkiezen is boven D_2 . Stel nu dat er een punt P_0 bestaat beneden de lijn $\mathcal{L}$ met coördinaten (X_0, Y_0) waarvoor geldt dat D_2 de optimale beslissing is. Stel vervolgens dat P_0 zodanig gelegen is dat het punt $[X_0, (1-r_2)Y_0]$, voorstellende de toestand op het volgende beslissings-tijdstip ligt in de bovengenoemde sector. Uit deze eigenschappen van P_0 volgt:

$$f_{D_2 D_1}(X_0, Y_0) = f(X_0, Y_0) \quad (5.15)$$

$$\text{en } f_{D_2 D_1}(X_0, Y_0) < f_{D_1 D_2}(X_0, Y_0). \quad (5.16)$$

Deze beide relaties zijn echter strijdig, zodat er geen P_0 bestaat met bovengenoemde eigenschappen. Bijgevolg zijn er geen punten beneden de $\mathcal{L}$ lijn, waarvoor de beslissing D_2 optimaal is. Op dezelfde wijze kan men aantonen, dat voor de punten boven $\mathcal{L}$ de beslissing D_2 optimaal is.

De optimale beslissing wordt dus gegeven door de volgende beslisregel ²⁾

$$\text{Ga vissen in } V_1 \text{ als } Y < \frac{(1-p_1)p_2 r_1}{(1-p_2)p_1 r_2} X. \quad (5.17)$$

$$\text{Ga vissen in } V_2 \text{ als } Y > \frac{(1-p_1)p_2 r_1}{(1-p_2)p_1 r_2} X.$$

Wij hebben tot dusver aangenomen dat de fracties r_1 en r_2 constant zijn. Zonder enig bezwaar mogen deze ook trekkingen zijn uit een verdeling. In (5.17) moet men dan r_1 vervangen denken door $E r_1$ en r_2 door $E r_2$.

Verder is het mogelijk, als geldt

$$(1-r_1)^a = (1-r_2)^b \quad (5.18)$$

waarbij a en b gehele getallen zijn, de functie $f(X, Y)$ te bepalen.

De grootheden r_1 en r_2 zijn nu constanten.

Kiezen wij ter illustratie $a=2$ en $b=1$ en beschouwen wij wederom de lijn $\mathcal{L}$ van alle punten, waarvoor het onverschillig is of men eerst D_1 kiest dan wel D_2 . De lijnen $\mathcal{L}_1$ worden gevormd door die punten, waarvoor geldt dat men volgens de beslisregel 1 maal D_1 moet kiezen en de keuze voor de $(i+1)^{\text{ste}}$ keer zowel D_1 als D_2 mag

2) De hierboven geschetste werkwijze kan gevolgd worden bij die problemen, welke de volgende functionaalvergelijkingen voortbrengen:

$$f(X, Y) = \max \left\{ \int_0^1 \{ zX + f[(1-z)X, Y] \} dG(z), \int_0^1 \{ wY + f(X, (1-w)Y) \} dH(w) \right\}.$$

De integralen in deze functionaalvergelijking zijn Stieltjes integralen.

zijn. De lijnen M_j worden gevormd door die punten, waarvoor geldt dat men volgens de beslisregel j maal D_2 moet kiezen en de keuze voor de $(j+1)^{ste}$ keer zowel D_1 als D_2 mag zijn.

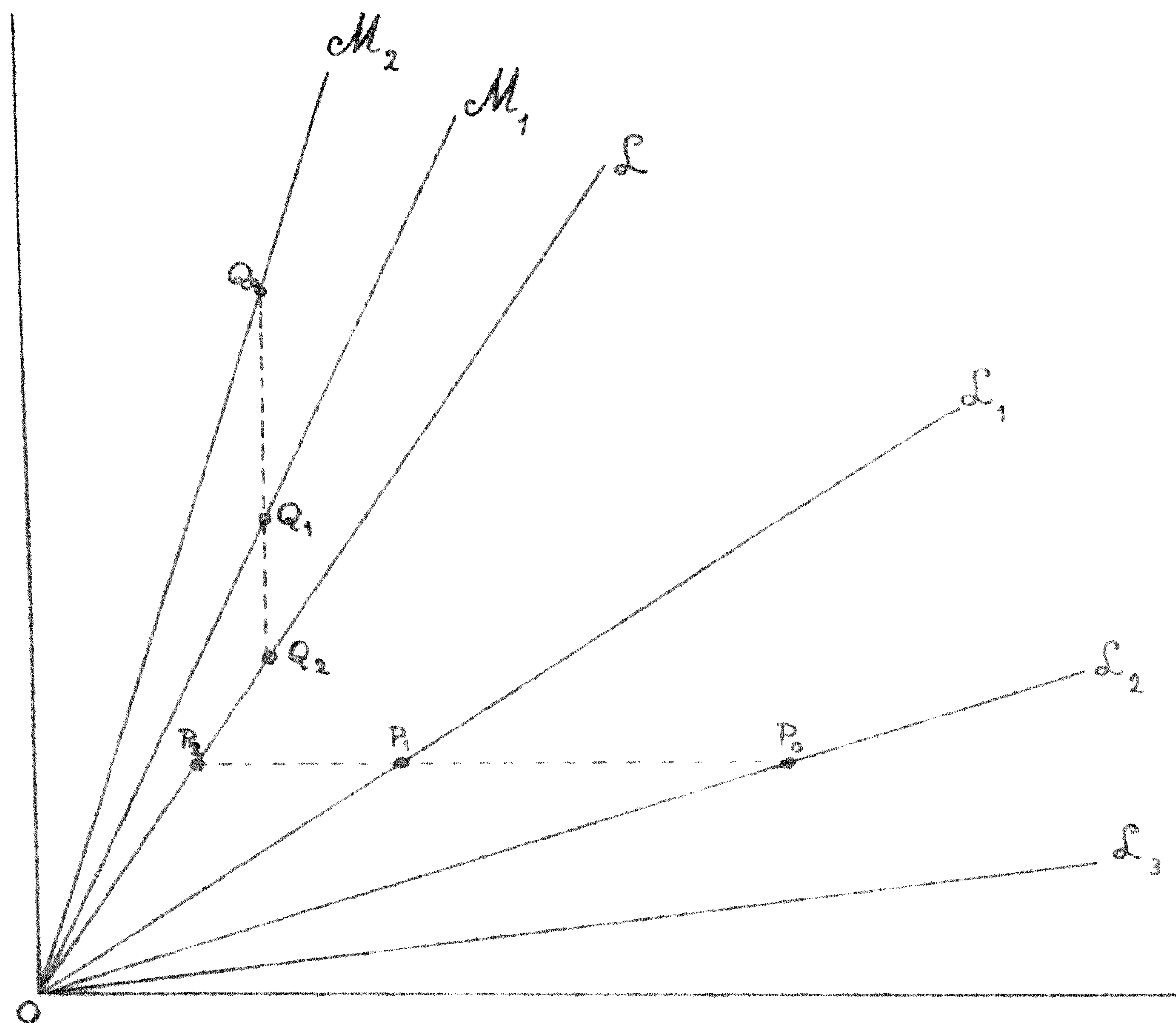


fig. 5.3

In fig. 5.3 hebben wij twee begintoestanden getekend, nl. P_0 en Q_0 . De toestanden op de volgende beslissingstijdstippen worden gegeven door P_1 en P_2 resp. door Q_1 en Q_2 .

Men kan ook zeggen, dat de lijn L_1 door een beslissing getransformeerd wordt in de lijn L_{1-1} enz.

Men kan eenvoudig nagaan dat de lijnen L_1 en M_j resp. gegeven worden door:

$$L_1 = Y - \frac{(1-p_1)p_2r_1(1-r_1)^1}{(1-p_2)p_1r_2} X = 0$$

$$M_j = Y - \frac{(1-p_1)p_2r_1}{(1-p_2)p_1r_2(1-r_2)^j} X = 0$$

Voor alle punten in de sector $L_1 O L_2$ geldt, dat het optimaal is om tweemaal een D_1 beslissing te nemen en daarna een D_2 beslis-

sing. Wij vinden dus voor een punt (X,Y) in deze sector:

$$f(X,Y) = (1-p_1)r_1X + (1-p_1)^2r_1^2X + (1-p_1)^2(1-p_2)r_2Y + \\ (1-p_1)^2(1-p_2)f[(1-r_1)^2X, (1-r_2)Y] . \quad (5.21)$$

Uit (5.21) en (5.18) volgt met $a=2$ en $b=1$:

$$f(X,Y) = (1-p_1)r_1X + (1-p_1)^2r_1^2X + (1-p_1)^2(1-p_2)r_2Y + \\ (1-p_1)^2(1-p_2)(1-r_1)^2f(X,Y) \quad (5.22)$$

of

$$f(X,Y) = \frac{(1-p_1)r_1[1+(1-p_1)r_1]X + (1-p_1)^2(1-p_2)r_2Y}{1-(1-p_1)^2(1-p_2)(1-r_1)^2} . \quad (5.23)$$

De gedaante van de functie $f(X,Y)$ in de sector $\mathcal{L}_2 O \mathcal{L}_3$ vindt men als volgt. Stel (X,Y) is een punt in deze sector, dan is $[(1-r_1)X, Y]$ een punt uit $\mathcal{L}_1 O \mathcal{L}_2$. Voor de functie $f(X,Y)$ geldt dus:

$$f(X,Y) = (1-p_1)r_1X + (1-p_1)f[(1-r_1)X, Y] = \\ = (1-p_1)r_1X + (1-p_1) \frac{(1-p_1)r_1[1+(1-p_1)r_1](1-r_1)X + (1-p_1)^2(1-p_2)r_2Y}{1-(1-p_1)^2(1-p_2)(1-r_1)^2} .$$

De gedaante van de functie $f(X,Y)$ in de sector $\mathcal{L} O \mathcal{L}_1$ vindt men als volgt. Stel (X,Y) is een punt in deze sector, dan is $[\frac{X}{1-r_1}, Y]$ een punt uit $\mathcal{L}_1 O \mathcal{L}_2$. Nu geldt

$$f(\frac{X}{1-r_1}, Y) = (1-p_1)r_1 \frac{X}{1-r_1} + (1-p_1)f(X,Y) \quad (5.25)$$

of

$$f(X,Y) = \frac{f(\frac{X}{1-r_1}, Y) - (1-p_1)r_1 \frac{X}{1-r_1}}{1-p_1} = \\ = \frac{(1-p_1)r_1[1+(1-p_1)r_1] \frac{X}{1-r_1} + (1-p_1)^2(1-p_2)r_2Y}{(1-p_1)[1-(1-p_1)^2(1-p_2)(1-r_1)^2]} - \frac{(1-p_1)r_1 \frac{X}{1-r_1}}{(1-p_1)} = \\ = \frac{r_1[1+(1-p_1)r_1] \frac{X}{1-r_1} + (1-p_1)(1-p_2)r_2Y}{1-(1-p_1)^2(1-p_2)(1-r_1)^2} - r_1 \frac{X}{1-r_1} . \quad (5.26)$$

Op de hierboven geschetste werkwijze kan men voor alle sectoren de functie $f(X,Y)$ bepalen.

Aangezien de sectorgrenzen gegeven worden door $\mathcal{L}_i$ en $\mathcal{M}_j$ kan men voor ieder punt (X,Y) de sector bepalen. Indien men de sector kent beschikt men ook over de gedaante van de functie $f(X,Y)$ in die sector en kan de functiewaarde in (X,Y) worden berekend.