

DE FORMULERING VAN EEN MARKOV BESLISSINGS-
PROBLEEM ALS PROCES NETWERK

J. van der Wal,^{*}
J. Wessels ^{*}

Samenvatting

De laatste jaren zijn en worden er zeer efficiënte methoden ontwikkeld voor de oplossing van primaire LP problemen met een netwerk structuur. Het kan daarom aantrekkelijk zijn om, indien er voor een probleem verschillende LP-formuleringen mogelijk zijn, te kiezen voor een formulering waarbij de onderliggende structuur die van een netwerk is. Voorbeeld van zo een probleem is, zoals in deze notitie wordt aangetoond, het Markov beslissingsprobleem, waarvoor naast de bekende LP-formulering een speciale formulering is te geven in de vorm van een 'maximal flow' probleem in een proces netwerk.

*Technische Hogeschool Eindhoven
Onderafd. Wiskunde & Informatica
Postbus 513
5600 MB Eindhoven

1. Inleiding

Voor verschillende typen van gestructureerde LP problemen is het mogelijk om de simplexmethode op een zeer efficiënte wijze uit te voeren door gebruik te maken van de modelstructuur. Een belangrijk voorbeeld daarvan zijn de proces netwerk problemen. Formeel is een proces netwerk een geheel van knopen die verbonden zijn via takken waarlangs 'stroom' vervoerd kan worden. In Koene [8] worden drie soorten knopen onderscheiden, te weten: raffinage-, meng- en transportknopen. In een raffinageknoop moet de uitgaande stroom in een gegeven verhouding verdeeld worden over de uitgaande takken, in een mengknoop worden de binnenkomende stromen volgens een vaste verhouding samengevoegd. De knopen zonder dergelijke verhoudingsvoorschriften zijn de transportknopen. Als in het netwerk de 'wet van behoud van stroom' geldt dan spreken we van een zuiver netwerk. Alstijdens het vervoer de stroom kan toenemen (bijv. rentewinst bij geldstromen) of afnemen (bijv. lekkage) dan spreken we van een gegeneraliseerd proces netwerk (vgl. [8]).

Voor proces netwerken zijn door velen speciale methoden ontwikkeld, o.a. Ford en Fulkerson [2]. De belangrijkste ontwikkeling van de laatste jaren is echter dat de overtuiging is gegroeid dat de meest efficiënte benadering gaat via het primaire simplex algoritme met een goed gebruik van de speciale probleemstructuur, met name van de vorm van de basis. Deze ontwikkeling is gestart met een artikel van Glover, Karney en Klingman [4]. Sindsdien hebben de speciale primaire simplex methoden veel aandacht gekregen, zie bijv. Glover en Klingman [5] en Koene [8]. Als belangrijke redenen hiervoor worden genoemd dat deze methoden (vgl. [8]):

- (i) voor zuivere netwerken tot 200 keer en voor gegeneraliseerde netwerken tot 50 keer zo snel zijn als het pakket APEX III
- (ii) veel minder geheugenruimte vragen
- (iii) door de speciale basisstructuur kunnen werken met de oorspronkelijke gegevens hetgeen afrondfouten voorkomt of beperkt.

Het zou dus verstandig kunnen zijn om een LP probleem te herformuleren als proces netwerk. En aangezien voor zowel het verdisconteerde als het gemiddelde opbrengst Markov beslissingsprobleem (MBP) LP-formuleringen te geven zijn (zie d'Epenoux [1], De Ghellinck [3] en Manne [9]) lijkt het

de moeite waard om na te gaan of en hoe een MBP te formuleren is als proces netwerk.

De opzet van deze notitie is verder als volgt.

In sectie 2 wordt allereerst de gewone LP-formulering van de MBP's gegeven. Vervolgens worden in sectie 3 proces netwerk formuleringen afgeleid waarbij summier wordt ingegaan op een aantal aspecten die een rol zouden kunnen spelen bij de beantwoording van de vraag of het uit rekentechnisch oogpunt aantrekkelijk is te kiezen voor proces netwerk formulering.

In sectie 4 wordt tenslotte nog enige aandacht besteed aan de speciaal gestructureerde MBP's zoals voorraad- en productieplanningsproblemen.

2. De gewone LP-formuleringen voor een MBP

Het Markov beslissingsprobleem zoals we dat hier beschouwen wordt gekarakteriseerd door de volgende grootheden

I = de toestandruimte, $I = \{1, 2, \dots, N\}$

K = de beslissingsruimte, $K = \{1, 2, \dots, K_0\}$

p_{ij}^k = de kans om als in toestand i beslissing k genomen wordt een overgang naar toestand j te maken

r_i^k = de directe opbrengst als in toestand i beslissing k wordt genomen.

Voor het verdisconteerde MBP met verdisconteringsfactor β , $0 \leq \beta < 1$, luidt de standaard LP-formulering, zoals gegeven door d'Epenoux [1] met x_i^k het totale verwachte verdisconteerde aantal malen dat het proces in toestand i zit en beslissing k wordt genomen:

$$(1) \quad \begin{cases} \max \sum_{i,k} x_i^k r_i^k \\ \text{onder de voorwaarden} \\ \sum_k x_i^k = \frac{1}{N} + \beta \sum_{j,l} x_j^l p_{ji}^l, & i \in I \\ x_i^k \geq 0, & i \in I, k \in K. \end{cases}$$

Oplossing van het probleem (1) levert een optimale stationaire strategie en de maximale totale verwachte verdisconteerde opbrengst voor de begin-toestandsverdeling $(\frac{1}{N}, \dots, \frac{1}{N})$.

De gewone LP-formulering voor het irreducibele MBP met als criterium de gemiddelde opbrengst per tijdseenheid is (zie De Ghellinck [3] en Manne [9])

$$(2) \quad \begin{cases} \max \sum_{i,k} x_i^k r_i^k \\ \text{onder de voorwaarden} \\ \sum_k x_i^k = \sum_{j,l} x_{jp}^l p_{ji}^l, \quad j \in I \\ \sum_{i,k} x_i^k = 1 \\ x_i^k \geq 0, \quad i \in I, k \in K \end{cases}$$

(met nu x_i^k het gemiddelde aantal keren per tijdseenheid dat het proces in i zit en beslissing k wordt genomen).

Het oplossen van (2) levert een gemiddeld optimale strategie alsmede de maximale gemiddelde opbrengst.

Zowel in het verdisconteerde als in het gemiddelde geval is er een grote overeenkomst tussen het oplossen van het LP probleem met de simplex methode en de policy iteration methode van Howard [7]. In feite zijn ze, zoals bekend, identiek als bij het oplossen van het LP probleem meerdere basisvariabelen tegelijk vervangen worden (zie bijv. Mine en Osaki [10] of Wessels en Van Nunen [11]). Elke iteratieslag op zich wordt daardoor duurder maar het aantal benodigde slagen is in het algemeen zeer gering!

Meer recent zijn ook voor het gemiddelde MBP met meerdere ketens goede LP-formuleringen gegeven door Hordijk en Kallenberg [6].

3. De formulering van een MBP als proces netwerk

De formuleringen van het verdisconteerde en het gemiddelde MBP als proces netwerk zijn vrij direct af te leiden uit de gewone LP-formuleringen (1) en (2).

Splits daartoe de gebeurtenissen in een periode in het MBP in twee delen; te weten de keuze van de beslissing en de stochastische overgang. Dit leidt tot een proces netwerk met een verzameling knopen die gesplitst kan worden in de verzamelingen $V := \{(i) \mid i \in I\}$ en $W := \{(i,k) \mid i \in I, k \in K\}$. De verzameling V correspondeert dus met de toestand van het proces vlak voor de beslissing genomen wordt en de verzameling W met de toestand vlak daarna. De knopen in V zijn transportknopen, die in W raffinageknopen. Vanuit elke knoop in V gaan er K_0 takken naar W en in elke knoop in W komt er precies 1 tak uit V aan. Een tak van V naar W representeert het nemen van een beslissing, een tak van W naar V een overgang.

We gebruiken verder de volgende notaties. De stroom van (i) naar (i,k) noteren we als $x_{i,ik}$ en de stroom van (i,k) naar j als $x_{ik,j}$. In het verdisconteerde [gemiddelde opbrengst] MBP is $x_{i,ik}$ dus het verdisconteerde [gemiddelde] aantal keren dat toestand i bezocht en beslissing k genomen wordt, en is $x_{ik,j}$ het verdisconteerde [gemiddelde] aantal keren dat i wordt bezocht, beslissing k wordt genomen en dit resulteert in een overgang naar j . Verder is r_i^k de opbrengst per eenheid stroom van (i) naar (i,k) .

De proces netwerk LP formulering van het verdisconteerde MBP wordt nu

$$(3) \quad \left\{ \begin{array}{ll} \max & \sum_{i,k} x_{i,ik} r_i^k \\ \text{onder de voorwaarden} & \\ \beta \sum_{j \in I} x_{j,i} - \sum_k x_{i,ik} = -\frac{1}{N} & , \quad i \in I \\ x_{i,ik} - \sum_j x_{ik,j} = 0 & , \quad i \in I, k \in K \\ \frac{p_{ij}^k}{p_{ijik}^k} x_{ik,j_{ik}} - x_{ik,j} = 0 & , \quad i \in I, k \in K, j \neq j_{ik} \text{ en } j_{ik} \\ & \text{zodat } p_{ijik}^k > 0 \text{ als vaste referentie knoop.} \\ \text{alle variabelen} & \geq 0 \end{array} \right.$$

Voor het gemiddelde opbrengst irreducible MBP luidt de proces netwerk formulering:

$$(4) \quad \left\{ \begin{array}{l} \max \sum_{i,k} x_{i,ik} r_i^k \\ \text{onder de voorwaarden} \\ \sum_{j,l} x_{jl,i} - \sum_k x_{i,ik} = 0 \quad , \quad i \in I \\ x_{i,ik} - \sum_j x_{ik,j} = 0 \quad , \quad i \in I, k \in K \\ \frac{p_{ij}^k}{p_{ijik}} x_{ik,jik} - x_{ik,j} = 0 \quad , \quad i \in I, k \in K, j \neq j_{ik} \\ \sum_{i,k} x_{i,ik} = 1 \\ \text{alle variabelen} \geq 0 \end{array} \right.$$

De extra restrictie $\sum x_{i,ik} = 1$ maakt het probleem tot een proces netwerk met additionele lineaire beperkingen zoals beschreven in Koene [8, Hoofdstuk 6].

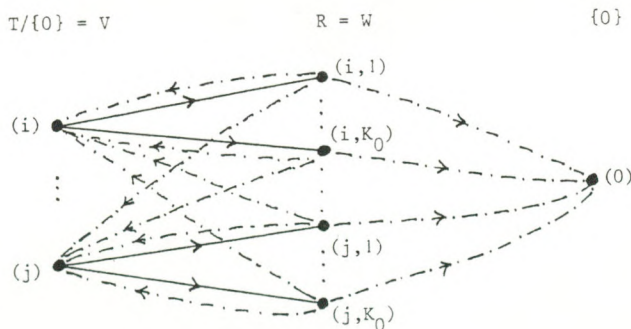
Nu is (3), de proces netwerk formulering voor het verdisconteerde MBP, een gegeneraliseerd netwerk geworden doordat de verdisconteringsfactor β in de restricties opduikt. Het probleem is echter ook als zuiver netwerk te formuleren als we de verdisconteringsfactor representeren als overgang naar een extra in te voeren absorberende toestand 0. Van de stroom die knoop (i,k) binnenkomt gaat dan een fractie βp_{ij}^k naar knoop (j) , $j \in I$, en de rest, een fractie $1 - \beta$, gaat naar knoop 0.

Dit leidt tot de volgende formulering van het verdisconteerde MBP als zuiver netwerk:

$$\begin{aligned}
 (5) \quad & \left\{ \begin{array}{l} \max \sum_{i,k} x_{i,ik} r_i^k \\ \text{onder de voorwaarden} \\ \sum_{j,l} x_{j,l,i} - \sum_k x_{i,ik} = -\frac{1}{N} \quad , \quad i \in I \\ x_{i,ik} - \sum_{j \in I} x_{ik,j} - x_{ik,0} = 0 \quad , \quad i \in I, k \in K \\ \frac{\beta p_{ij}^k}{1-\beta} x_{ik,0} - x_{ik,j} = 0 \quad , \quad i \in I, k \in K, j \in I \\ \text{alle variabelen} \geq 0. \end{array} \right.
 \end{aligned}$$

De omvang van de problemen (3) - (5) is veel groter dan die van de problemen (1) en (2). De problemen (1) en (2) hebben NK_0 variabelen en N (onafhankelijke) restricties. De proces netwerk formuleringen (3) - (5) hebben (na eliminatie van de raffinage restricties, dus met alleen de representanten) nog $N(N+2)K_0$ variabelen en $N(K_0+1)$ restricties. Voor $N = 100$ en $K_0 = 10$ neemt de omvang dus al toe van 10^3 variabelen en 10^2 restricties tot 10^5 variabelen en 10^3 restricties.

De graaf behorend bij de proces netwerk representatie van een MBP is in het algemeen erg regelmatig. Beschouweens het probleem (5) met de extra transportknoop (0). Splitsing van de knopen in de verzameling T van transportknopen en de verzameling R van raffinageknopen toont de volgende structuur.



Vanuit elke transportknoop behalve (0), zeg (i), vertrekken K_0 transporttakken naar de raffinageknopen (i,k), en vanuit elke raffinageknoop in R vertrekken er precies $K_0 + 1$ takken, dus 1 naar elke knoop in T.

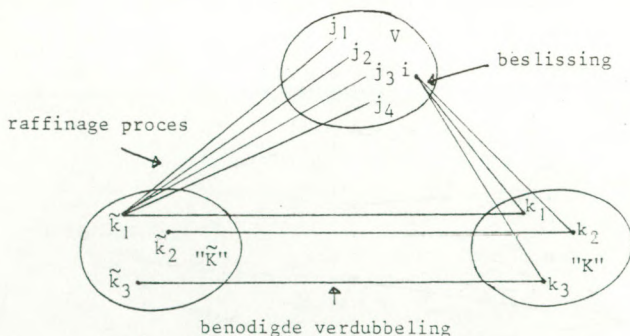
Het lijkt dat een basisoplossing bestaat (kan bestaan) uit alle raffinageprocessen alsmede per transportknoop precies 1 tak. Het veranderen van de basis betekent dan de vervanging in een transportknoop van een uitgaande tak door een andere, dus de verbetering van een beslissing in een toestand.

4. Voorraad- en productieplanningsproblemen en vervangingsproblemen

In de voorafgaande netwerk formuleringen is verondersteld dat alle knopen (i,k) in wezen verschillend zijn. Echter, bij voorraad- en produktieplanningsproblemen en vervangingsproblemen karakteriseert vaak de beslissing k alleen al de toestand van het proces direct na de beslissing. Bijv. k is het voorraadniveau direct na de bijbestelling of k is de kwaliteit van de machine direct na de onderhouds- of vervangingsbeslissing.

In deze gevallen lijkt het mogelijk de proces netwerk formulering sterk te vereenvoudigen. Het ligt voor de hand om nu alleen nog de knopen (i), $i \in I$ en (k), $k \in K$ te beschouwen. Koene [8] staat zo'n formulering echter niet toe omdat de raffinageknopen (k) slechts één ingaande tak mogen hebben. Om dit probleem te ondervangen kunnen we de beslissingsknopen verdubbelen.

De graaf krijgt daarmee de volgende structuur:



Het aantal knopen bedraagt nu $N + 2K_0$ in plaats van $N + NK_0$ in de oorspronkelijke formulering. De structuur van de graaf is wel iets minder mooi geworden.

5. Conclusie

Het is zonder meer mogelijk om voor een MBP naast de gewone LP-formulering ook een netwerk LP-formulering te geven.

De vraag of zo'n netwerk formulering voordelen biedt, meer speciaal of de mogelijkheid om speciale netwerk technieken te gebruiken opweegt tegen het vaak sterk toenemen van de omvang van het probleem, is nog geheel niet beantwoord. Bij de beantwoording van deze vraag zal er wel onderscheid gemaakt moeten worden tussen het MBP waarvoor alleen een algemene formulering als (3), (4) of (5) mogelijk is en het probleem met extra structuur in de beslissingen als beschreven in sectie 4.

Referenties

- [1] d'Epenoux, F., Sur un problème de production et de stockage dans l'aléatoire, *Revue Française de Recherche Opérationnelle*, 14 (1960), 3-16.
- [2] Ford, L.R. en D.R. Fulkerson, *Flows in networks*, Princeton University Press, Princeton, 1962.
- [3] de Ghellinck, G.T., Les problèmes de décisions séquentielles, *Cahiers du Centre d'Etudes de Recherche Opérationnelle*, 2 (1960), 161-179.
- [4] Glover, F., D. Karney en D. Klingman, Implementation and computational comparisons of primal, dual and primal-dual computer codes for minimum cost network flow problems, *Networks* 4 (1974), 191-212.
- [5] Glover, F. and D. Klingman, The simplex SON algorithm for LP/ embedded network problems, *Mathematical Programming Study* 15 (1981), 148-176.
- [6] Hordijk, A. en L.C.M. Kallenberg, Linear programming and Markov decision chains, *Management Science* 25 (1979), 352-362.
- [7] Howard, R.A., *Dynamic programming and Markov processes*, Wiley, New York, 1960.
- [8] Koene, J., Minimal cost flow in processing networks, a primal approach, MC-tract, Mathematisch Centrum, Amsterdam (te verschijnen).
- [9] Manne, A.S., Linear programming and sequential decisions, *Management Science* 6 (1960), 259-267.
- [10] Mine, H. en S. Osaki, *Markovian decision processes*, American Elsevier, New York, 1970.
- [11] Wessels, J. en J.A.E.E. van Nunen, Discounted semi-Markov decision processes: linear programming and policy iteration, *Statistica Neerlandica* 29 (1975), 1-7.