Suksessieve approximatiemetoden voor Markov beslissingsproblemen met verdiskontering

TECHNISCHE HOGESCHOOL EINDHOVEN

Onderafdeling der Wiskunde

GROEP STATISTIEK EN OPERATIONS RESEARCH

Memorandum COSOR 73-02

Suksessieve approximatiemetoden voor Markov beslissingsproblemen met verdiskontering

door

E.A. van Doorn

Eindhoven, april 1973

§ 1. Inleiding

Dit verslag behelst suksessieve approximatiemetoden voor de oplossing van een Markov-beslissingsprobleem met verdiskonteerde kosten, eindige toestandsruimte en eindige beslissingsruimte, zoals dat bijvoorbeeld beschreven staat in [1] § 1.

Er zal een kader gepresenteerd worden waarbinnen de tot dusver bekende oplosmetoden van voornoemde aard, namelijk de metode van MacQueen (zie [1]) en de waardemetode (zie [2]), als speciale gevallen figureren.

Hoewel in dit kader de waarde van de beide bekende metoden gerelativeerd wordt, is vooralsnog niet een verwerping van een van beide metoden ten gunste van een andere het resultaat; echter, mede dankzij een door Blackwell gesuggereerde notatie en enkele door Blackwell bewezen stellingen (zie [3]), zullen de beide metoden, geplaatst binnen het kader, aan doorzichtigheid in de bewijsvoering en helderheid in de notatie winnen.

§ 2. Notatie en fundamentele stellingen

Zij $X = \{1,\ldots,N\}$ de toestandsruimte en K de beslissingsruimte.
In het vervolg zal alleen sprake zijn van deterministische Markov-strategieën, kortweg strategieën genoemd, die gedefinieerd worden door een rij $(r_1,r_2,\ldots,r_n,\ldots)$ waarin $r_i \in R$: de verzameling afbeeldingen van X in K; $\pi = (r_1,r_2,\ldots,r_n,\ldots)$ betekent, dat strategie π, indien het systeem op tijdstip t zich in toestand i bevindt, de beslissing $r_t(i)$ voorschrijft.
Indien $\forall_{t\in\mathbb{N}} [r_t = r]$ spreken we van een stationaire strategie; de stationaire strategie $(r,r,\ldots,r,\ldots)$ wordt genoteerd als $r^{(\infty)}$.
u_π is de N-vektor met komponenten $u_\pi(i)$: de verwachte met α $(0 < \alpha < 1)$ verdiskonteerde opbrengst over oneindig lange tijd bij strategie π en begintoestand i; u_r is per definitie $u_{r^{(\infty)}}$. $u_\pi^{(n)}$ is de N-vektor van verwachte verdiskonteerde opbrengsten over de eerste n perioden bij strategie $\pi = (r_1,r_2,\ldots,r_n,\ldots)$ en eindwaarden nul; merk op dat $u_\pi^{(n)}$ bepaald wordt door $r_1,r_2,\ldots,r_n$. Men ziet dat $u_\pi^{(\infty)} = u_\pi$.
Zij u_* de N-vektor met komponenten $u_*(i) := \sup\limits_\pi u_\pi(i)$. Een resultaat van Blackwell ([3], theorem 7(b)) is, dat er een, voor elke begintoestand optimale, stationaire strategie bestaat, te noteren als $r_*^{(\infty)}$; er geldt dus $u_* = u_{r_*}$. In onze verdere beschouwingen zullen we ons beperken tot stationaire strategieën.
De $N \times N$-matrix van overgangskansen $\{P_{r(i)}(i,j)\}_{i,j}$ behorende bij een $r \in R$, wordt genoteerd als P_r. De N-vektor van onmiddellijke opbrengsten bij een $r \in R$: $\{q_{r(i)}(i)\}_i$ wordt genoteerd als q_r en de i-de komponent als $q_r(i)$, welke voor alle i en r groter dan nul verondersteld wordt.
Met vektoren worden steeds kolomvektoren bedoeld; vektorongelijkheid wordt als volgt gedefinieerd:

$$\forall_{u\in\mathbb{R}^N,\, v\in\mathbb{R}^N} [u > v :\Longleftrightarrow \forall_i\ [u(i) \geq v(i)]\ \&\ \exists_i\ [u(i) > v(i)]]$$

met de voor de hand liggende uitbreidingen $\geq$, $<$ en $\leq$. We maken in schrijfwijze geen onderscheid tussen vektoren en skalairen.

Het zal nuttig blijken te zijn operatoren T_r $(r \in R)$ en U te introduceren, die als volgt gedefinieerd zijn:

(1.1) Definitie.

(i) $\forall_{f \in \mathbb{R}^N} [T_r f := q_r + \alpha P_r f]$

$T_r^0 f := f;\ T_r^n f := T_r(T_r^{n-1} f).$

(ii) $\forall_{f \in \mathbb{R}^N} [Uf := \max_{r \in R} T_r f]$

$U^0 f := f;\ U^n f := U(U^{n-1} f).$

Opmerking. Eenvoudig is na te gaan dat (ii) een korrekte definitie is in die zin dat er een $r \in R$ bestaat waarvoor $T_r f$ komponentsgewijs maximaal over R is.

Zij $\mathbb{R}^N$ genormeerd met de norm $\| \ \|_\infty$, d.w.z. als $f \in \mathbb{R}^N$ dan $\|f\|_\infty := \sup_i |f(i)|$.

(1.2) Stelling.

(i) Zij $f,g \in \mathbb{R}^N$, $f \geq g$, $r \in R$, dan $Uf \geq Ug$ en $T_r f \geq T_r g$.

(ii) Zij c een N-vektor met identieke komponenten, $f \in \mathbb{R}^N$, dan geldt:

$$U(f + c) = Uf + \alpha c$$

en

$$T_r(f + c) = T_r f + \alpha c\ .$$

(iii) U en T_r ($r \in R$) zijn kontrakties met modulus α.

Bewijs. Zie Blackwell [3], theorems 3, 4 en 5. □

Opmerking. Wegens (iii) en de vaste-punt-stelling van Banach hebben U en T_r unieke vaste punten.

In het vervolg zal van de volgende door Blackwell bewezen stelling gebruik worden gemaakt:

(1.3) Stelling.

(i) $u_* = Uu_* = \lim_{n\to\infty} U^n f \qquad \forall_{f\in\mathbb{R}^N}$.

(ii) $u_r = T_r u_r = \lim_{n\to\infty} T_r^n f \qquad \forall_{f\in\mathbb{R}^N}$.

Bewijs. Zie Blackwell [3], theorems 6(a) en 7(a). □

(1.4) Stelling. $\forall_{k\in K}\ [T_k f \le f] \Rightarrow \forall_{r\in R}\ [u_r \le f]$

waarin T_k de operator uit definitie (1.1) is behorende bij die $r \in R$, waarvoor $\forall_{i\in X}\ [r(i) = k]$.

Bewijs. Zie Blackwell [3], theorem 6(d). □

§ 3. De A(λ)-algoritme

In deze paragraaf zullen de algoritmen gepresenteerd worden welke een generalisatie zijn van de algoritmen van MacQueen [1] en Kersten [2]; de relatie met deze beide algoritmen zal in § 4 uiteengezet worden.
Hier en in de rest van dit verslag worden, waar dit geen misverstand kan verwekken, de indices "r_n" van bepaalde grootheden vervangen door indices "n"; zo schrijven we T_n, P_n, q_n en u_n in plaats van T_{r_n}, P_{r_n}, q_{r_n} en u_{r_n}.
λ is in het vervolg een getal uit $\mathbb{N} \cup \{0\}$.

(3.1) algoritme A(λ)

de MQ-variant:

stap 1: initialiseer $n = 0$, $v_0 = 0$; kies $\varepsilon > 0$;

stap 2: $n := n+1$;

bepaal r_n z.d.d. $T_n v_{n-1} = U v_{n-1}$;

bepaal $z_n = T_n v_{n-1}$;

stap 3: bepaal $K'_n = \min\limits_i \, (z_n - v_{n-1})(i)$;

$$K''_n = \max_i \, (z_n - v_{n-1})(i);$$

$$v'_n = v_{n-1} + \frac{1}{1-\alpha} K'_n \quad ;$$

$$v''_n = v_{n-1} + \frac{1}{1-\alpha} K''_n \quad ;$$

als $\frac{1}{1-\alpha}(K''_n - K'_n) \le \varepsilon \min\limits_i v'_n(i)$: ga na stap 5;

stap 4: bepaal $v_n = T_n^{\lambda} z_n$; ga naar stap 2;

stap 5: stop; een schatting voor r_* is r_n

een schatting voor u_n is $\frac{1}{2}(v''_n + v'_n)$

de WM-variant

In deze variant zijn de stappen 1, 2 en 4 gelijk aan die uit de MQ-variant.

stap 3: als $n \geq 2$ bepaal $M'_{n-1} = \min_i (v_{n-1} - T^{\lambda}_{n-1} v_{n-2})(i)$;

$$M''_{n-1} = \max_i (v_{n-1} - T^{\lambda}_{n-1} v_{n-2})(i);$$

$$w'_{n-1} = v_{n-1} + \frac{\alpha}{1-\alpha} M'_{n-1} \quad ;$$

$$w''_{n-1} = v_{n-1} + \frac{\alpha}{1-\alpha} M''_{n-1} \quad ;$$

als

$$\frac{1}{1-\alpha} \{\max_i (z_n - T_{n-1} v_{n-1})(i) + \frac{\alpha^2}{1-\alpha} (M''_{n-1} - M'_{n-1})\} \leq \varepsilon \min_i w'_{n-1}(i)$$

ga naar stap 5

stap 5: stop: een schatting voor r_* is r_{n-1}

een schatting voor u_{n-1} is $\frac{1}{2}(w'_{n-1} + w''_{n-1})$

Opmerking. De beide varianten verschillen dus uitsluitend in stopkriterium, nl. in de stappen 3 en 5.

(3.2) Stelling.

I(i) $u_* \geq u_n \geq v_n \geq z_n \geq v_{n-1}$

(ii) $z_n \geq u_*^{(n)}$

II(i) $v'_n \leq u_* \leq v''_n$

(ii) v'_n is monotoon niet-dalend, v''_n is monotoon niet-stijgend

(iii) $v'_n \to u_*$, $v''_n \to u_*$ $(n \to \infty)$

(iv) $v'_n \leq u_n$

III(i) $v'_n \leq w'_n \leq u_n \leq w''_n \leq v''_n$

(ii) w'_n is monotoon niet-dalend

(iii) $u_* - u_{n-1} \leq \frac{1}{1-\alpha} \{\max_i (z_n - T_{n-1} v_{n-1})(i) + \frac{\alpha^2}{1-\alpha} (M''_{n-1} - M'_{n-1})\}$.

Opmerking 1. A priori kunnen we aan een stopkriterium de volgende eisen stellen:

(a) er moet binnen eindig veel stappen aan voldaan worden;

(b) indien aan het stopkriterium is voldaan moeten we kunnen beschikken over

(1) een schatting r voor r_* die zodanig is dat $u - u_r \leq \varepsilon u_*$;

(2) een schatting u voor u_r die zodanig is dat $|u_r - u| \leq \varepsilon u_r$.

Wat de MQ-variant betreft wordt aan eis (a) voldaan dankzij II(iii), aan eis (b)(1) dankzij II(i) en II(iv) en aan eis (b)(2) dankzij II(i) en II(iv); wat de WM-variant betreft wordt aan eis (a) voldaan dankzij III(i) (indirekt dankzij II(iii)), aan eis (b)(1) dankzij III(iii), aan eis (b)(2) dankzij III(i) (tenminste als $\alpha \geq \frac{1}{3}$).

Toelichting op de laatste bewering:
aan het stopkriterium is voldaan $\Rightarrow$

$$\Rightarrow \left(\frac{\alpha}{1-\alpha}\right)^2 (M''_{n-1} - M'_{n-1}) \leq \varepsilon \min_i w'_{n-1}(i)$$

$$\Rightarrow \frac{\alpha}{1-\alpha} (w''_{n-1} - w'_{n-1}) \leq \varepsilon \min_i w'_{n-1}(i)$$

$$\Rightarrow w''_{n-1} - w'_{n-1} \leq \varepsilon \frac{1-\alpha}{\alpha} \min_i u'_{n-1}(i) \qquad \text{(III(i))}$$

$$\Rightarrow u_{n-1} - \tfrac{1}{2}(w'_{n-1} + w''_{n-1}) \leq \varepsilon u_{n-1} \qquad \text{als } \alpha \geq \frac{1}{3}.$$

Opmerking 2. II(i) is een direkt gevolg van II(ii) en II(iii); een alternatief bewijs voor II(i) zal worden gegeven omdat dit een interessante toepassing van stelling (1.4) inhoudt, en bovendien ten behoeve van een later te gebruiken tussenresultaat.

Voor het bewijs van stelling (3.2) (en ook in het verdere verloop van dit verslag) hebben we het volgende lemma nodig:

(3.3) Lemma.

(i) $\alpha^\lambda K'_n \leq M'_n$ & $M''_n \leq \alpha^\lambda K''_n$;

(ii) $\alpha^{\lambda+1} M'_n \leq M'_{n+1}$;

(iii) $\alpha M'_n \leq K'_{n+1}$ en als $r_n = r_{n+1}$ dan $K''_{n+1} \leq \alpha M''_n$.

Bewijs. $\lambda \geq 1$.

ad (i) $\quad T_n^\lambda z_n - T_n^\lambda v_{n-1} = \alpha^\lambda P_n^\lambda (z_n - v_{n-1})$

$$\alpha^\lambda K_n' \leq \alpha^\lambda P_n^\lambda (z_n - v_{n-1}) \leq \alpha^\lambda K_n''$$

dus

$$\begin{cases} \alpha^\lambda K_n' \leq \min_i \, (T_n^\lambda z_n - T_n^\lambda v_{n-1})(i) = M_n' \\ M_n'' = \max_i \, (T_n^\lambda z_n - T_n^\lambda v_{n-1})(i) \leq \alpha^\lambda K_n'' \, . \end{cases}$$

ad (ii) $\quad z_n = T_n v_{n-1} \geq T_{n-1} v_{n-1} \geq T_{n-1}(T_{n-1}^{\lambda-1} z_{n-1} + M_{n-1}') = v_{n-1} + \alpha M_{n-1}'$

dus

$$v_n = T_n^\lambda z_n \geq T_n^\lambda (v_{n-1} + \alpha M_{n-1}') = T_n^\lambda v_{n-1} + \alpha^{\lambda+1} M_{n-1}'$$

dus

$$M_n' = \min_i \, (v_n - T_n^\lambda v_{n-1})(i) \geq \alpha^{\lambda+1} M_{n-1}' \, .$$

ad (iii) $\quad z_{n+1} \geq v_n + \alpha M_n' \quad$ (zie ad (ii))

dus

$$K_{n+1}' = \min_i \, (z_{n+1} - v_n) \geq \alpha M_n'$$

$$r_{n+1} = r_n \implies z_{n+1} = T_{n+1} v_n = T_n v_n \leq T_n(T_n^{\lambda-1} v_{n-1} + M_n'') = v_n + \alpha M_n''$$

dus

$$K_{n+1}'' \leq \alpha M_n'' \, .$$

Als $\lambda = 0$ geldt $K_n' = M_n'$ en $K_n'' = M_n''$; eenvoudig is in te zien dat de relaties geldig blijven. □

Bewijs van stelling (3.2).

ad I(i):

- $u_* \geq u_n$ triviaal;
- $z_1 = T_1 0 = \max_k q_k(.) > 0 = v_0$;

  met volledige induktie: stel $z_{\ell-1} \geq v_{\ell-2}$, dan

$z_\ell = T_\ell v_{\ell-1} \geq T_{\ell-1} v_{\ell-1} = T_{\ell-1}(T^\lambda_{\ell-1} z_{\ell-1}) \geq T^\lambda_{\ell-1}(T_{\ell-1} v_{\ell-2}) = T^\lambda_{\ell-1} z_{\ell-1} = v_{\ell-1}$

dus $z_n \geq v_{n-1}$;

- $v_n = T^\lambda_n z_n \geq T^\lambda_n v_{n-1} = T^{\lambda-1}_n z_n \geq \ldots \geq z_n$, dus $v_n \geq z_n$;

- $v_n \geq z_n \Rightarrow T^\lambda_n v_n \geq v_n \Rightarrow \forall_{\ell \in \mathbb{N}} T^{\lambda\ell}_n v_n \geq v_n$, dus ook $\lim_{\ell\to\infty} T^{\lambda\ell}_n v_n = u_n \geq v_n$.

ad I(ii):

- $z_1 = T_1 0 = \max_k q_k(.) \geq u_*^{(1)}$;

met volledige induktie: stel $z_{\ell-1} \geq u_*^{(\ell-1)}$, dan

$z_\ell = T_\ell v_{\ell-1} \geq T_* v_{\ell-1} \geq T_* z_{\ell-1} \geq T_* u_*^{(\ell-1)} = u_*^{(\ell)}$, dus $z_n \geq u_*^{(n)}$;

ad II(i):

- $T_k v''_n - v''_n = T_k(v_{n-1} + \frac{1}{1-\alpha} K''_n) - v_{n-1} - \frac{1}{1-\alpha} K''_n = T_k v_{n-1} - v_{n-1} - K''_n \leq$
$\leq T_n v_{n-1} - v_{n-1} - K''_n = z_n - v_{n-1} - K''_n \leq 0$ voor alle $k \in K$; volgens (1.4) geldt dus $u_* \leq v''_n$.

- $v'_n - T_n v'_n = v_{n-1} + \frac{1}{1-\alpha} K'_n - T_n(v_{n-1} + \frac{1}{1-\alpha} K'_n) = v_{n-1} - T_n v_{n-1} + K'_n =$
$= v_{n-1} - z_n + K'_n \leq 0$; omdat $T_n v'_n \leq U v'_n$ geldt dus $v'_n \leq U v'_n$ en ook $\forall_{\ell \in \mathbb{N}} v'_n \leq U^\ell v'_n$; volgens (1.3)(i) geldt dus $\lim_{\ell\to\infty} U^\ell v'_n = u_* \geq v'_n$.

ad II(ii):

$z_n \geq v_{n-1} + K'_n \Rightarrow T^\lambda_n z_n \geq T^\lambda_n v_{n-1} + \alpha^\lambda K'_n$,

$$\text{dus} \quad v_n \geq T^{\lambda-1}_n z_n + \alpha^\lambda K'_n \geq T^{\lambda-1}_n (v_{n-1} + K'_n) + \alpha^\lambda K'_n =$$

$$= T^{\lambda-2}_n z_n + (\alpha^{\lambda-1} + \alpha^\lambda) K'_n \geq \ldots \geq v_{n-1} + (1 + \alpha + \ldots + \alpha^\lambda) K'_n =$$

$$= v_{n-1} + \frac{1-\alpha^{\lambda+1}}{1-\alpha} K'_n$$

$$\text{dus} \quad v'_{n+1} = v_n + \frac{1}{1-\alpha} K'_{n+1} = v_n + \frac{1}{1-\alpha} \min_i (T_{n+1} v_n - v_n)(i) \geq$$

$$\geq v_{n-1} + \frac{1-\alpha^{\lambda+1}}{1-\alpha} K_n' + \frac{1}{1-\alpha} \min_i \, (T_n(T_n^{\lambda-1} z_n + \alpha^\lambda K_n') - v_n)(i) =$$

$$= v_{n-1} + \frac{1}{1-\alpha} K_n' = v_n'$$

dus $v_{n+1}' \geq v_n'$ en analoog $v_{n+1}'' \leq v_n''$.

ad II(iii):

$$v_n' = v_{n-1} + \frac{1}{1-\alpha} K_n' = v_{n-1} + \frac{1}{1-\alpha} \min_i \, (z_n - v_{n-1})(i)$$

$$\to u_* + \frac{1}{1-\alpha} \min_i \, (u_* - u_*)(i) = u_* \qquad (n \to \infty)$$

dus $v_n' \to u_*$ $(n \to \infty)$ en analoog $v_n'' \to u_*$ $(n \to \infty)$.

ad II(iv):

$v_n' - T_n v_n' \leq 0$ (zie ad (ii))

dus $\forall_{\ell \in \mathbb{N}} \, v_n' \leq T_n^\ell \, v_n'$

dus $\lim_{\ell \to \infty} T_n^\ell \, v_n' = u_n \geq v_n'$.

ad III(i):

$- \; z_n = T_n v_{n-1} \Rightarrow q_n = z_n - \alpha P_n v_{n-1}$

$$u_n = T_n u_n = \sum_{i=0}^{\infty} \alpha^i P_n^i \, q_n = \sum_{i=0}^{\infty} \alpha^i P_n^i (z_n - \alpha P_n v_{n-1}) =$$

$$= z_n + \sum_{i=1}^{\infty} \alpha^i P_n^i (z_n - v_{n-1}) =$$

$$= T_n^\lambda \, z_n + \sum_{i=\lambda}^{\infty} \alpha^i P_n^i (T_n z_n - T_n v_{n-1}) \qquad \forall_{\lambda \in \mathbb{N} \cup \{o\}}$$

$$= T_n^\lambda \, z_n + \sum_{i=1}^{\infty} \alpha^i P_n^i (T_n^\lambda \, z_n - T_n^\lambda \, v_{n-1}) =$$

$$= v_n + \sum_{i=1}^{\infty} \alpha^i P_n^i (v_n - T_n^\lambda \, v_{n-1})$$

dus $w'_n = v_n + \frac{\alpha}{1-\alpha} M'_n \le u_n \le v_n + \frac{\alpha}{1-\alpha} M''_n = w''_n$;

$$- w'_n = v_n + \frac{\alpha}{1-\alpha} M'_n =$$

$$= v_n + \frac{\alpha}{1-\alpha} (M'_n - \alpha^{\lambda} K'_n) + \frac{\alpha^{\lambda+1}}{1-\alpha} K'_n \ge$$

$$\ge v_n + \frac{\alpha^{\lambda+1}}{1-\alpha} K'_n \qquad \text{(zie lemma (3.2)(i))}$$

$$\ge v_{n-1} + \frac{1}{1-\alpha} K'_n = v'_n \qquad \text{(zie bewijs II(ii));}$$

analoog $w''_n \le v''_n$.

ad III(ii):

$$w'_{n+1} = v_{n+1} + \frac{\alpha}{1-\alpha} M'_{n+1} \ge$$

$$\ge v_{n+1} + \frac{\alpha^{\lambda+2}}{1-\alpha} M'_n \qquad \text{volgens (3.3)(ii)}$$

$$\ge T^{\lambda}_{n+1} v_n + \alpha^{\lambda+1} M'_n + \frac{\alpha^{\lambda+2}}{1-\alpha} M'_n \qquad \text{volgens ad (3.3)(ii)}$$

$$= T^{\lambda-1}_{n+1} z_{n+1} + \frac{\alpha^{\lambda+1}}{1-\alpha} M'_n \ge$$

$$\ge T^{\lambda-1}_{n+1}(v_n + \alpha M'_n) + \frac{\alpha^{\lambda+1}}{1-\alpha} M'_n \ge \ldots \ge$$

$$\ge z_{n+1} + \frac{\alpha^2}{1-\alpha} M'_n \ge v_n + \frac{\alpha}{1-\alpha} M'_n = w'_n .$$

ad III(iii):

Volgens III(i): $0 \le u_{n-1} - w'_{n-1} \le w''_{n-1} - w'_{n-1} = \frac{\alpha}{1-\alpha} (M''_{n-1} - M'_{n-1})$;

bovendien: $u_* - u_{n-1} = T_* u_* - T_{n-1} u_{n-1} = T_* u_* - T_* u_{n-1} + T_* u_{n-1} - T_{n-1} u_{n-1} =$
$= \alpha P_*(u_* - u_{n-1}) + T_* u_{n-1} - T_{n-1} u_{n-1}$;

dus $(I - \alpha P_*)(u_* - u_{n-1}) = T_* u_{n-1} - T_{n-1} u_{n-1}$

ofwel $u_* - u_{n-1} = \sum_{i=0}^{\infty} \alpha^i P_*^i\{T_* u_{n-1} - T_{n-1}u_{n-1}\} \le$

$$\le \sum_{i=0}^{\infty} \alpha^i P_*^i\{T_*(w'_{n-1} + \frac{\alpha}{1-\alpha}(M''_{n-1} - M'_{n-1})) - T_{n-1}w'_{n-1}\}$$

$$= \sum_{i=0}^{\infty} \alpha^i P_*^i\{T_* w'_{n-1} - T_{n-1}w'_{n-1} + \frac{\alpha^2}{1-\alpha}(M''_{n-1} - M'_{n-1})\} =$$

$$= \sum_{i=0}^{\infty} \alpha^i P_*^i\{T_* v_{n-1} - T_{n-1}v_{n-1} + \frac{\alpha^2}{1-\alpha}(M''_{n-1} - M'_{n-1})\} \le$$

$$\le \sum_{i=0}^{\infty} \alpha^i P_*^i\{\max_i (z_n - T_{n-1}v_{n-1})(i) + \frac{\alpha^2}{1-\alpha}(M''_{n-1} - M'_{n-1})\} =$$

$$= \frac{1}{1-\alpha}\{\max_i (z_n - T_{n-1}v_{n-1})(i) + \frac{\alpha^2}{1-\alpha}(M''_{n-1} - M'_{n-1})\} . \qquad \square$$

De navolgende lemma's zijn een eerste stap in het vergelijkende onderzoek van de beide varianten van de A(λ)-algoritme; de lemma's (3.4) t/m (3.7) geven met name enige relaties tussen de stopkriteria van de beide varianten.

Zij in het vervolg $\alpha_{\mu;\varepsilon}$ de reëelwaardige oplossing van de vergelijking $\frac{\alpha^\mu}{1-\alpha} = 1 + \varepsilon$ (met $0 \le \alpha \le 1$).

(3.4) Lemma. Voor de variabelen uit de A(λ)-algoritme geldt:

$$\forall_{\alpha \le \alpha_{\lambda+2;0}} \left[\frac{1}{1-\alpha}(K''_n - K'_n) \le \varepsilon \min_i v'_n(i)\right.$$

$$\left.\Rightarrow \frac{\alpha^2}{(1-\alpha)^2}(M''_n - M'_n) \le \varepsilon \min_i w'_n(i)\right] .$$

Bewijs. $\frac{1}{1-\alpha}(K''_n - K'_n) \le \varepsilon \min_i v'_n(i)$

$$\Rightarrow \frac{\alpha^2}{(1-\alpha)^2}(M''_n - M'_n) \le \frac{\alpha^{2+\lambda}}{1-\alpha}\varepsilon \min_i v'_n(i) \qquad \text{volgens (3.3)(i).}$$

Nu geldt $\forall_{\alpha \le \alpha_{\lambda+2;0}}$:

$$\frac{\alpha^{2+\lambda}}{1-\alpha}\min_i v'_n(i) - \min_i w'_n(i) \le$$

$$\leq \frac{\alpha^{2+\lambda}}{1-\alpha} \min_i v'_n(i) - \min_i v'_n(i) \qquad \text{volgens (3.2)III(i)}$$

$$\leq \left(\frac{\alpha^{2+\lambda}}{1-\alpha} - 1\right) \min_i v'_n(i) \leq 0$$

dus

$$\frac{\alpha^2}{(1-\alpha)^2} (M''_n - M'_n) \leq \varepsilon \min_i w'_n(i) \; . \qquad \square$$

Opmerking. Enkele waarden van $\alpha_{\lambda+2;0}$ zijn:

$\alpha_{2;0} = 0{,}62$
$\alpha_{3;0} = 0{,}68$
$\alpha_{4;0} = 0{,}72$
$\alpha_{5;0} = 0{,}75$.

Beschouwing van de stappen 3 en 5 van de A(λ)-algoritme maakt duidelijk dat dit lemma impliceert dat, indien tijdens de n-de iteratie aan het stopkriterium van de MQ-variant is voldaan én $r_{n+1} = r_n$, in de (n+1)-de stap aan het stopkriterium van de WM-variant is voldaan; dit alles binnen de gegeven grenzen voor α. Aan het slot van deze paragraaf wordt deze uitspraak nader geadstrueerd.

(3.5) Lemma. Voor de variabelen uit de A(λ)-algoritme geldt:

$$\forall_{\alpha \geq \alpha_{1;\varepsilon}} \left[r_n = r_{n-1} \;\&\; \frac{\alpha^2}{(1-\alpha)^2} (M''_{n-1} - M'_{n-1}) \leq \varepsilon \min_i w'_{n-1}(i) \right.$$

$$\left. \Rightarrow \frac{1}{1-\alpha} (K''_n - K'_n) \leq \varepsilon \min_i v'_n(i) \right] .$$

Bewijs. $w'_{n-1} - v'_n \leq v''_n - v'_n$

volgens (3.2)III(i) en (ii)

$$v''_n - v'_n = \frac{1}{1-\alpha} (K''_n - K'_n) \leq \frac{\alpha}{1-\alpha} (M''_{n-1} - M'_{n-1}) \leq \varepsilon \frac{1-\alpha}{\alpha} \min_i w'_{n-1}(i)$$

volgens (3.3)(iii) en het gegeven, dus

$$w'_{n-1} - v'_n \leq \varepsilon \frac{1-\alpha}{\alpha} \min_i w'_{n-1}(i)$$

dus

$$\min_i w'_{n-1}(i) \le \frac{1}{1 - \frac{\varepsilon(1-\alpha)}{\alpha}} \min_i v'_n(i)$$

dus

$$\frac{1}{1-\alpha}(K''_n - K'_n) \le \varepsilon \frac{1-\alpha}{\alpha} \frac{1}{1 - \frac{\varepsilon(1-\alpha)}{\alpha}} \min v'_n(i) =$$

$$= \frac{1}{\frac{\alpha}{1-\alpha} - \varepsilon} \varepsilon \min v'_n(i)$$

dus

$$\forall_{\alpha \ge \alpha_{1;\varepsilon}} \frac{1}{1-\alpha}(K''_n - K'_n) \le \varepsilon \min_i v'_n(i) \ . \qquad \square$$

Opmerking. Enkele waarden van $\alpha_{1;\varepsilon}$ zijn:

$\alpha_{1;0} = 0,50$
$\alpha_{1;0,1} = 0,52$.

De bewering welke formeel in dit lemma is uitgedrukt luidt: voor zekere α's geldt dat, indien tijdens de n-de iteratie van de A(λ)-algoritme aan het stopkriterium van de WM-variant is voldaan met de bijzonderheid dat $r_n = r_{n-1}$, tevens tijdens de n-de iteratie aan het stopkriterium van de MQ-variant zal zijn voldaan.

(3.6) Lemma. Voor de variabelen uit de A(0)-algoritme geldt:

$$\forall_{\alpha \ge \alpha_{2;\varepsilon}} [\frac{\alpha^2}{(1-\alpha)^2}(M''_n - M'_n) \le \varepsilon \min_i w'_n(i)$$

$$\Rightarrow \frac{1}{1-\alpha}(K''_n - K'_n) \le \varepsilon \min_i v'_n(i)] \ .$$

Bewijs. $w'_n - v'_n \le v''_n - v'_n$ volgens (3.3)(i)

$$= \frac{1}{1-\alpha}(K''_n - K'_n) = \frac{1}{1-\alpha}(M''_n - M'_n) \le \frac{1-\alpha}{\alpha^2} \varepsilon \min_i w'_n(i)$$

dus

$$\min_i w'_n(i) \le \frac{1}{1 - \frac{\varepsilon(1-\alpha)}{\alpha^2}} \min_i v'_n(i)$$

$$\therefore \quad \frac{1}{1-\alpha}(K_n'' - K_n') \le \frac{1-\alpha}{\alpha^2}\,\frac{1}{1-\frac{\varepsilon(1-\alpha)}{\alpha^2}}\,\varepsilon \min_i v_n'(i) =$$

$$= \frac{1}{\frac{\alpha^2}{1-\alpha}-\varepsilon}\,\varepsilon \min_i v_n'(i)$$

dus

$$\forall_{\alpha\ge\alpha_{2;\varepsilon}} \; \frac{1}{1-\alpha}(K_n'' - K_n') \le \varepsilon \min_i v_n'(i) \,. \qquad \square$$

Opmerking. Enkele waarden van $\alpha_{2;\varepsilon}$ zijn:

$\alpha_{2;0} \approx 0{,}63$
$\alpha_{2;0{,}1} \approx 0{,}63$.

De betekenis van dit lemma is dat het stopkriterium van de MQ-variant beter is dan dat van de WM-variant voor $\lambda = 0$ en α ruwweg groter dan $\frac{2}{3}$; immers uit het lemma blijkt dat, indien tijdens de (n+1)-de iteratie nog maar ten dele aan het stopkriterium van de WM-variant is voldaan, tijdens de n-de iteratie al aan het stopkriterium van de MQ-variant voldaan is.

Vooralsnog is het niet mogelijk gebleken zoveel meer teoretisch materiaal te verzamelen dan het voorafgaande, dat een ondubbelzinnige uitspraak kan worden gedaan over de relatieve efficiëntie van de MQ- resp. de WM-variant van de A(λ)-algoritme. Praktijkervaring (zie De Leeuw [4]) kan ons echter een indikatie dienaangaande geven. Welnu, uit de praktijk blijkt dat in de A(λ)-algoritme gedurende een groot aantal iteraties, vóór dat aan het stopkriterium van de WM-variant is voldaan, de strategie niet meer verandert, hetgeen betekent dat de term: $\max_i (z_n - T_{n-1}v_{n-1})$ in al die iteraties nul is.
Dit overwegende en bovendien het feit dat bij vaste λ de WM-variant van A(λ) per iteratie ongeveer een gelijke hoeveelheid rekenwerk vereist als de MQ-variant, lijkt op grond van lemma (3.5) een voorzichtige konklusie, dat voor verdiskonteringsfaktoren dichtbij 1 de MQ-variant van de A(λ)-algoritme sneller is dan de WM-variant, hier op zijn plaats.
Dankzij lemma (3.6) kunnen we de term "voorzichtig" in bovenstaande konklusie t.a.v. de A(0)-algoritme gevoeglijk achterwege laten. Het belang van lemma (3.6) kan echter pas ten volle blijken indien meer bekend is over de rol van de parameter λ in beide varianten van de A(λ)-algoritme, een nog braakliggend terrein voor onderzoek.

Lemma (3.4) maakt op grond van bovengenoemde opmerkingen plausibel dat voor kleine waarden van de verdiskonteringsfaktor de WM-variant sneller is dan de MQ-variant.

§ 4. Relatie van de A(λ)-algoritme tot bekende methoden

In het vervolg zal blijken dat de WM-variant van de A(1)-algoritme ekwivalent is met de waardemetode, zoals die door Kersten in [2] beschreven is, en de MQ-variant van de A(0)-algoritme met de metode van MacQueen, beschreven in [1].
Het onderstaande lemma stelt ons in staat de eerste van deze ekwivalenties te bewijzen.

(4.1) Lemma. Zij $\lambda \in \mathbb{N} \cup \{o\}$. Laten $\{v_0,v_1,\dots\}$ en $\{\hat{v}_0,\hat{v}_1,\dots\}$ verzamelingen vektoren uit $\mathbb{R}^n$ zijn waarvoor geldt:

$$n = 1,2,\dots \quad \begin{cases} v_0 = 0 \\ v_n = T_{r_n}^{\lambda+1} v_{n-1} + c_n \\ \text{met } r_n \text{ z.d.d. } T_{r_n} v_{n-1} = Uv_{n-1} \\ \text{en } c_n \text{ een N-vektor met identieke komponenten welke een} \\ \text{funktie zijn van } r_1,c_1,v_1,\dots,v_{n-1},r_n; \end{cases}$$

$$n = 1,2,\dots \quad \begin{cases} \hat{v}_0 = 0 \\ \hat{v}_n = T_{\hat{r}_n}^{\lambda+1} \hat{v}_{n-1} \\ \text{met } \hat{r}_n \text{ z.d.d. } T_{\hat{r}_n} \hat{v}_{n-1} = U\hat{v}_{n-1}; \end{cases}$$

dan

(i) $r_n = \hat{r}_n$

(ii) $v_n = \hat{v}_n + d_n$ waarin $\begin{cases} d_0 = 0 \\ d_n = \alpha^{\lambda+1} d_{n-1} + c_n. \end{cases}$

N.B. Neem bij meer mogelijkheden voor de keuze van $r_n(i)$ (of $\hat{r}_n(i)$) het element uit K met de laagste index dat tot de mogelijkheden behoort; hierbij is een indicering van K verondersteld.

Bewijs. Met volledige induktie:

$$v_0 = \hat{v}_0 = 0$$

$$\therefore \quad Uv_0 = U\hat{v}_0 = U0\ ;$$

wegens het eenduidige keuzekriterium geldt dus

$$r_1 = \hat{r}_1$$

$$v_1 = T_{r_1}^{\lambda+1} v_0 + c_1 = T_{\hat{r}_1}^{\lambda+1} \hat{v}_0 + c_1 = \hat{v}_1 + d_1 \;.$$

Stel

$$\left.\begin{array}{l} r_\ell = \hat{r}_\ell \\ v_\ell = \hat{v}_\ell + d_\ell \end{array}\right\} \; \ell = 1,2,\ldots,n-1 \;,$$

dan:

$$\forall_{r\in R} \; T_r v_{n-1} = T_r \hat{v}_{n-1} + \alpha d_{n-1} \;;$$

Omdat d_{n-1} konstant is en het keuzekriterium eenduidig, worden $\max\limits_r (T_r \hat{v}_{n-1} + \alpha d_{n-1})$ en $\max\limits_r (T_r \hat{v}_{n-1})$ door dezelfde r gerealiseerd, dus

$$r_n = \hat{r}_n$$

$$v_n = T_{r_n}^{\lambda+1} v_{n-1} + c_n = T_{\hat{r}_n}^{\lambda+1} (\hat{v}_{n-1} + d_{n-1}) + c_n =$$

$$= T_{\hat{r}_n}^{+1} \hat{v}_{n-1} + \alpha^{\lambda+1} d_{n-1} + c_n = \hat{v}_n + d_n \;;$$

hiermee is het gestelde voor alle n bewezen. □

(4.2) Stelling. De waardenmetode van Kersten [2] is ekwivalent met de WM-variant van de A(1)-algoritme.

Bewijs. We volstaan hier met een schets van het bewijs:
Gebruik makend van een door lemma (4.1) gesuggereerde transformatie van de grootheden v^ℓ uit [2] (v^ℓ vervult dan de rol van v_n uit lemma (4.1) en $\frac{\beta}{1-\beta} m_0^\ell$ correspondeert met c_n) ontstaat een algoritme die nog slechts in notatie afwijkt van de WM-variant van de A(1)-algoritme. □

(4.3) Stelling. De metode van MacQueen [1] is ekwivalent met de MQ-variant van de A(0)-algoritme.

Bewijs. Ook hier volstaan we met een schets:
MacQueen [1] werkt met relatieve v_n-waarden t.o.v. een bepaald element uit X. Een geschikte transformatie van de grootheden uit MacQueen [1] resulteert in een algoritme welke alleen nog in notatie afwijkt van de MQ-variant van de A(0)-algoritme. □

Literatuur

[1] J. MacQueen: A modified Dynamic Programming Method for Markovian Decision Problems.
J. Math. An. Appl. 14 (1966), 38-43.

[2] T.A.G.M. Kersten: Enkele Narkov beslissingsproblemen.
Afstudeerverslag T.H.E. 1972.

[3] D. Blackwell: Discounted Dynamic Programming.
Ann. Math. Stat. 36 (1965), 226-234.

[4] A.A.C.M. de Leeuw: Een onderzoek naar verschillende aspecten van de waardemethode.
Stageverslag T.H.E. 1973.