KYBERNETIKA ČÍSLO 2, ROČNÍK 5/1969

O metodě postupných aproximací pro nalezení optimálního řízení markovského řetězce

KAREL SLADKÝ

V práci je ukázáno, že při hledání optimálního řízení markovského řetězce metodou postupných aproximací odvozenou v práci [2] lze snadno v každém kroku iteračního postupu konstruovat dolní a horní odhady optimálního průměrného zisku, které monotonně konvergují k hodnotě optimálního průměrného zisku. Rovněž průměrný zisk příslušný řízení získanému v každém kroku iteračního postupu je omezen odpovídajícím dolním a horním odhadem.

ÚVOD

V práci [1] Howard odvodil na základě metod dynamického programování iterační postup pro nalezení optimálního řízení markovského řetězce s ohodnocenými přechody mezi jednotlivými stavy pro případ, že celkový počet přechodů roste nade všechny meze a kritériem optimality je buď průměrný zisk připadající na jeden přechod nebo celkový diskontovaný zisk (t.j. součet hodnot získaných při jednotlivých přechodech, jestliže hodnota získaná při k-tém přechodu je vynásobena α^k (pro $k = 1, 2, \ldots, \infty$), kde $\alpha \in \langle 0; 1)$ je tak zvaný diskontní faktor (anglicky „discount factor"). Howardova iterační metoda, pomocí které obdržíme hledané optimální řízení po několika krocích, spočívá v ohodnocení použitého řízení pomocí váhových koeficientů jednotlivých stavů a v dalším zlepšování použitého řízení ve všech stavech na základě známých váhových koeficientů. Nevýhodou této metody je, že ohodnocení použitého řízení je po numerické stránce dosti náročné, neboť je třeba řešit soustavu n rovnic o n neznámých (kde n je počet stavů vyšetřovaného markovského řetězce).

Pro případ, že kritériem optimality vyšetřované úlohy je průměrný zisk připadající na jeden přechod, White (viz [2]) ukázal, že za jistých podmínek je možno najít optimální řízení metodou postupných aproximací. Tato metoda se vyznačuje tím, že v každém kroku iteračního postupu je prováděno zlepšování vypočítávaných váhových koeficientů a průměrného zisku, které je kombinováno se zlepšováním použitého řízení. Uvedená metoda se doporučuje zejména pro ruční výpočet a pro případ, kdy řešíme úlohu o velkém počtu stavů.

Pro kritérium celkového diskontovaného zisku metoda postupných aproximací byla podrobně diskutována v [3] a [4]. V [4] bylo rovněž ukázáno, jak lze generovat v každém kroku iteračního postupu dolní a horní odhady optimalizovaných veličin. Posloupnosti dolních a horních odhadů potom konvergují monotonně k řešení vyšetřované úlohy.

V této práci je ukázáno, že pro případ, kdy kritériem optimality je průměrný zisk připadající na jeden přechod, lze velmi snadno v každém kroku metody postupných aproximací konstruovat dolní a horní odhady optimálního průměrného zisku, které monotonně konvergují k hodnotě optimálního průměrného zisku. Rovněž průměrný zisk příslušný řízení nalezenému v každém iteračním kroku je omezen odpovídajícím dolním a horním odhadem. Odhady optimálního průměrného zisku mohou být zejména užitečné při ručním výpočtu.

FORMULACE PROBLÉMU A JEHO ŘEŠENÍ

Hledání optimálního řízení markovského řetězce můžeme přesně formulovat takto:

Mějme markovský řetězec s $n + 1$ stavy. V i-tém stavu lze použít jedné z s_i zadaných strategií. Při použití strategie k v i-tém stavu přejdeme s pravděpodobností p_{ij}^k z i-tého do j-tého stavu. Použití strategie k v i-tém stavu přinese zisk c_{ik}. Soubor použitých strategií v jednotlivých stavech nazveme řízením markovského řetězce.

Hledáme takové řízení, pro které průměrný zisk připadající na jeden přechod bude co nejvyšší, jestliže celkový počet přechodů roste nade všechny meze.

V práci [1] Howard odvodil, že při optimálním řízení ergodického markovského řetězce musí být splněn systém rovnic

$$v_i + g = \max_{k=1,2,\ldots,s_i} \left[c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k v_j\right] \tag{1}$$

$$(\text{pro } i = 1, 2, \ldots, n, n + 1)\,,$$

když klademe jedno $v_i = 0$ (v našem případě $v_{n+1} = 0$). v_i je váhový koeficient i-tého stavu, g je hodnota průměrného zisku.

V práci [2] navrhl White pro řešení systému rovnic (1) tento postup:

Za předpokladu, že existuje takový stav markovského řetězce (který označíme indexem $n + 1$), celé číslo $u > 0$ a jisté $\varepsilon > 0$ tak, že z libovolného stavu při libovolném přípustném řízení se po u přechodech s pravděpodobností větší nebo rovnou ε dostaneme do stavu $n + 1$, potom pomocí níže uvedeného algoritmu 1 obdržíme řešení systému rovnic (1). (Je patrné, že za těchto podmínek vyšetřovaný markovský řetězec bude vždy ergodický).

Algoritmus 1 (White). 1. Vyjdeme z libovolných hodnot váhových koeficientů v_i (pro $i = 1, 2, \ldots, n$) a hodnoty průměrného zisku g. Klademe $v_{n+1} = 0$.

2. Na základě dříve zjištěných hodnot váhových koeficientů v_i určíme zlepšené hodnoty váhových koeficientů v_i a průměrného zisku g a zlepšené řízení podle vztahů:

$$V_i(m+1) = \max_{k=1,2,\ldots,s_i} \left[c_{ik} + \sum_{k=1}^{n} p_{ij}^k \cdot v_j(m)\right], \tag{2}$$

kde

$$V_i(m) = v_i(m) + g'(m) \quad \text{pro} \quad i = 1, 2, \ldots, n$$

a

$$g(m) = g'(m+1) = \max_{k=1,2,\ldots,s_{n+1}} \left[c_{n+1,k} + \sum_{j=1}^{n} p_{n+1,j}^k \cdot v_j(m)\right].$$

m značí krok iteračního postupu.

Bod 2 opakujeme až je dosaženo uspokojivé přesnosti hledaného řešení.

Důkaz konvergence algoritmu 1 za výše uvedených předpokladů lze nalézt v práci [2].

Dále dokážeme větu, která umožňuje generaci dolních a horních odhadů optimálního průměrného zisku, které monotonně konvergují k hodnotě optimálního průměrného zisku. Tyto odhady mohou být generovány nepatrnou úpravou výše uvedeného algoritmu 1 při jednotlivých krocích iteračního postupu.

Věta 1. *Jestliže výchozí hodnoty váhových koeficientů při iteračním postupu podle algoritmu 1 jsou nulové, potom lze v každém kroku iteračního postupu zjistit dolní (resp. horní) odhad optimálního průměrného zisku g_{dol} (resp. g_{hor}) podle rovnice:*

$$g_{\text{dol}}(m) = \min_{i=1,2,\ldots,n+1} \left\{ \max_{k=1,2,\ldots,s_i} \left[c_{ik} + \sum_{j=1}^{n} p_{ij}^k \cdot v_j(m)\right] - v_i(m)\right\} \tag{3}$$

resp.

$$g_{\text{hor}}(m) = \max_{i=1,2,\ldots,n+1} \left\{ \max_{k=1,2,\ldots,s_i} \left[c_{ik} + \sum_{j=1}^{n} p_{ij}^k \cdot v_j(m)\right] - v_i(m)\right\}. \tag{3*}$$

Posloupnost hodnot $g_{\text{dol}}(m)$ (resp. $g_{\text{hor}}(m)$) konverguje monotonně k hodnotě optimálního průměrného zisku.

Poznámka. Konstrukce dolních a horních odhadů optimálního průměrného zisku není nikterak náročná, neboť výrazy $V_i(m+1) = \max_k \left[c_{ik} + \sum_{j=1}^{n} p_{ij}^k \cdot v_j(m)\right]$ jsou vypočítávány již v algoritmu 1. Z úsporných důvodů budeme používat též zkráceného zápisu $\max_i$, $\max_k$.

Větu 1 dokážeme pomocí několika lemmat.

Lemma 1. *Jestliže výchozí hodnoty $v_i(1)$ v iteračním postupu podle algoritmu 1 jsou nulové, potom váhové koeficienty $v_i(m)$ jsou pro $m \geqq 1$ typu:*

$$v_i(m) = A_i(m) - A_{n+1}(m) \quad (i = 1, 2, \ldots, n, n+1), \tag{4}$$

kde

$$A_i(m) = \max_{k=1,2,\ldots,s_i} \left[c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k A_j(m-1)\right] \quad pro \quad m > 1\,;$$

a

$$A_i(1) = 0\,.$$

Důkaz provedeme úplnou indukcí. Nejprve však využitím vztahu $v_{n+1}(m) = 0$ a vyloučením hodnoty $g(m)$ přepíšeme soustavu rovnic (2) na tvar:

$$v_i(m+1) = \max_{k=1,2,\ldots,s_i} \left[c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \,.\, v_j(m)\right] - \tag{2*}$$

$$- \max_{k=1,2,\ldots,s_{n+1}} \left[c_{n+1,k} + \sum_{j=1}^{n+1} p_{n+1,j}^k \,.\, v_j(m)\right].$$

Pro $m = 1$ tvrzení lemmatu zřejmě platí.

Předpokládejme dále platnost lemmatu pro jisté $m \geqq 1$. Dosazením vztahu (4) do vztahu (2*) dostáváme pro $v_i(m+1)$:

$$v_i(m+1) = \max_k \left[c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k (A_j(m) - A_{n+1}(m))\right] -$$

$$- \max_k \left[c_{n+1,k} + \sum_{j=1}^{n+1} p_{n+1,j}^k (A_j(m) - A_{n+1}(m))\right].$$

Protože $\sum_{j=1}^{n+1} p_{ij}^k = 1$, $\sum_{j=1}^{n+1} p_{n+1,j}^k = 1$ a pro libovolná B_j, B platí $\max_j (B_j - B) = \max_j B_j - B$ dostáváme:

$$v_i(m+1) = \max_k \left[c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \,.\, A_j(m)\right] - \max_k \left[c_{n+1,k} + \sum_{j=1}^{n+1} p_{n+1,j}^k \,.\, A_j(m)\right].$$

Tento vztah lze rovněž psát ve tvaru :

$$v_i(m+1) = A_i(m+1) - A_{n+1}(m+1)\,,$$

kde

$$A_i(m+1) = \max_k \left[c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \,.\, A_j(m)\right] \quad (i = 1, 2, \ldots, n, n+1)\,.$$

Lemma 2. *Jestliže použijeme značení z lemmatu* 1, *potom výrazy* (3) (*resp.* (3*)) *pro dolní* (*resp. horní*) *odhad optimálního průměrného zisku lze psát ve tvaru*:

$$g_{\mathrm{dol}}(m) = \min_{i=1,2,\ldots,n+1} \{ \max_{k=1,2,\ldots,s_i} [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k A_j(m)] - A_i(m)\}$$

resp.

$$g_{\mathrm{hor}}(m) = \max_{i=1,2,\ldots,n+1} \{ \max_{k=1,2,\ldots,s_i} [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k A_j(m)] - A_i(m)\} .$$

Důkaz. Protože $v_{n+1}(m) = 0$ lze výraz (3) psát ve tvaru:

$$g_{\mathrm{dol}}(m) = \min_{i=1,2,\ldots,n+1} \{ \max_{k=1,2,\ldots,s_i} [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \, . \, v_j(m)] - v_i(m)\} .$$

Dosazením výrazů zjištěných v lemmatu 1 pro $v_i(m)$ obdržíme:

$$g_{\mathrm{dol}}(m) = \min_i \{\max_k [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k(A_j(m) - A_{n+1}(m))] - (A_i(m) - A_{n+1}(m))\} .$$

Protože $\sum_{j=1}^{n+1} p_{ij}^k = 1$ a pro libovolná B_j, B platí $\max_j (B_j - B) = \max_j B_j - B$ po úpravě dostáváme:

$$g_{\mathrm{dol}}(m) = \min_i \{\max_k [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \, . \, A_j(m)] - A_i(m)\} .$$

Zcela analogický postup lze provést i pro $g_{\mathrm{hor}}(m)$.

Lemma 3. *Jestliže* $p_{ij} \geqq 0, \sum_{j=1}^{n+1} p_{ij} = 1$ *potom pro libovolná* a_j, b_j *platí:*

1. $$\max_i a_i \geqq \max_i \sum_{j=1}^{n+1} p_{ij} \, . \, a_j \, , \quad (\text{resp. } \min_i a_i \leqq \min_i \sum_{j=1}^{n+1} p_{ij} \, . \, a_j) \, .$$

2. $$\max_i (a_i - b_i) \gtreqless \max_i a_i - \max_i b_i \, .$$

3. $$\min_i (a_i - b_i) \lesseqgtr \max_i a_i - \max_i b_i \, .$$

Důkaz. Vztahy 1. a 2. jsou triviální. Vztah 3. dokážeme využitím vztahu 2. Platí totiž:

$$- \min_i (a_i - b_i) = \max_i (b_i - a_i) \gtreqless \max_i b_i - \max_i a_i \Rightarrow$$

$$\Rightarrow \min_i (a_i - b_i) \lesseqgtr \max_i a_i - \max_i b_i \, .$$

 Nyní můžeme přistoupit k důkazu věty 1.

Důkaz věty 1. Jestliže zvolíme takové řízení, že strategie k_i použitá v i-tém stavu splňuje vztah $c_{i,k_i} = \max_k [c_{ik}]$, potom průměrný zisk při tomto řízení bude zřejmě splňovat vztah $g \geqq \min_i \{\max_k [c_{ik}]\} = g_{\text{dol}}(1)$. Průměrný zisk kteréhokoliv přípustného řízení je zřejmě omezen vztahem $g \leqq \max_i \{\max_k [c_{ik}]\} = g_{\text{hor}}(1)$.

Jak je dokázáno v [2], iterační postup podle algoritmu 1 konverguje k řešení soustavy rovnic (1), a proto musí být splněno

$$\lim_{m\to\infty} g_{\text{dol}}(m) = g_{\text{opt}} = \lim_{m\to\infty} g_{\text{hor}}(m)\,,$$

kde g_{opt} značí průměrný zisk při optimálním řízení. Stačí proto dokázat pouze monotonnost konvergence dolních a horních odhadů optimálního průměrného zisku, tj. prokázat platnost vztahů:

$$\text{(5)} \qquad \min_{i=1,2,\ldots,n+1} \{ \max_{k=1,2,\ldots,s_i} [c_{jk} + \sum_{j=1}^{n+1} p_{ij}^k \,.\, A_j(m)] - A_i(m)\} \leqq$$

$$\leqq \min_{i=1,2,\ldots,n+1} \{ \max_{k=1,2,\ldots,s_i} [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \,.\, A_j(m+1)] - A_i(m+1)\}\,,$$

$$\text{(5*)} \qquad \max_{i=1,2,\ldots,n+1} \{ \max_{k=1,2,\ldots,s_i} [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \,.\, A_i(m)] - A_i(m)\} \geqq$$

$$\geqq \max_{i=1,2,\ldots,n+1} \{ \max_{k=1,2,\ldots,s_i} [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \,.\, A_j(m+1)] - A_i(m+1)\}\,.$$

Dokážeme nyní platnost vztahu (5). Využitím lemmatu 3 lze psát:

$$\min_i \{\max_k [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \,.\, A_j(m)] - A_i(m)\} \leqq$$

$$\leqq \min_i \{\min_k \sum_{j=1}^{n+1} p_{ij}^k \,.\, \{\max_{k'} [c_{jk'} + \sum_{l=1}^{n+1} p_{jl}^{k'} \,.\, A_l(m)] - A_j(m)\}\} =$$

$$= \min_i \{\min_k [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \,.\, \{\max_{k'} [c_{jk'} + \sum_{l=1}^{n+1} p_{jl}^{k'} \,.\, A_l(m)]\} - [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \,.\, A_j(m)]]\} \leqq$$

$$\leqq \min_i \{\max_k [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \,.\, \{\max_{k'} [c_{jk'} + \sum_{l=1}^{n+1} p_{jl}^{k'} \,.\, A_l(m)]\}] - \max_k [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \,.\, A_j(m)]\} =$$

$$= \min_i \{\max_k [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \,.\, A_j(m+1)] - A_i(m+1)\}\,.$$

Podobným způsobem lze dokázat platnost vztahu (5*). Platí:

$$\max_i \{\max_k [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \, . \, A_j(m)] - A_i(m)\} \geqq$$

$$\geqq \max_i \{\max_k \sum_{j=1}^{n+1} p_{ij}^k \, . \, \{\max_{k'} [c_{jk'} + \sum_{l=1}^{n+1} p_{jl}^{k'} \, . \, A_l(m)] - A_j(m)\}\} =$$

$$= \max_i \{\max_k [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \, . \, \{\max_{k'} [c_{jk'} + \sum_{l=1}^{n+1} p_{jl}^{k'} \, . \, A_l(m)]\} - [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \, . \, A_j(m)]]\} \geqq$$

$$\geqq \max_i \{\max_k [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \, . \, \{\max_{k'} [c_{jk'} + \sum_{l=1}^{n+1} p_{jl}^{k'} \, . \, A_l(m)]\}] -$$

$$- \max_k [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \, . \, A_j(m)]\} =$$

$$= \max_i \{\max_k [c_{ik} + \sum_{j=1}^{n+1} p_{ij}^k \, . \, A_j(m + 1)] - A_i(m + 1)\} \, .$$

Odhad průměrného zisku, který přísluší řízení nalezenému v každém kroku algoritmu 1 umožňuje věta 2.

Věta 2. *Jestliže jsou splněny předpoklady věty* 1, *potom průměrný zisk příslušející řízení nalezenému v m_0-tém kroku je rovněž omezen hodnotami $g_{\mathrm{dol}}(m_0)$* a *$g_{\mathrm{hor}}(m_0)$, které jsou dané vztahy* (3) *a* (3*).

Důkaz. Průměrný zisk příslušný řízení nalezenému v m_0-tém kroku lze nalézt tak, že v dalších krocích iteračního postupu podle algoritmu 1 bude používat stejného řízení jako v m_0-tém kroku (proto budeme vynechávat index k).

Vztah (5) (resp. (5*)) lze potom pro $m \geqq m_0$ přepsat následovně:

(6)
$$\min_{i=1,2,\ldots,n+1} [c_i + \sum_{j=1}^{n+1} p_{ij} \, . \, A_j(m) - A_i(m)] \leqq$$

$$\leqq \min_{i=1,2,\ldots,n+1} [c_i + \sum_{j=1}^{n+1} p_{ij} \, . \, A_j(m + 1) - A_i(m + 1)]$$

resp.

(6*)
$$\max_{i=1,2,\ldots,n+1} [c_i + \sum_{j=1}^{n+1} p_{ij} \, . \, A_j(m) - A_i(m)] \geqq$$

$$\geqq \max_{i=1,2,\ldots,n+1} [c_i + \sum_{j=1}^{n+1} p_{ij} \, . \, A_j(m + 1) - A_i(m + 1)] \, ,$$

kde

$$A_i(m + 1) = c_i + \sum_{j=1}^{n+1} p_{ij} \, . \, A_j(m) \, .$$

Tabulka 1. Řešení příkladu 1

Krok	1		2		3		4		5		6	
Stav i	(S_1)	$v_i^{(S_1)}$	(S_2)	$v_i^{(S_2)}$	(S_3)	$v_i^{(S_3)}$	(S_4)	$v_i^{(S_4)}$	(S_5)	$v_i^{(S_5)}$	(S_6)	$v_i^{(S_6)}$
1	1	0,00	2	−1,00	2	0,13	2	1,24	2	1,17	2	1,18
2	1	0,00	2	8,00	2	12,19	2	13,76	2	13,81	2	13,83
3	1	0,00	1	0,00	2	0,00	2	0,00	2	0,00	2	0,00
Průměrný zisk	8,00		9,75		11,91		13,30		13,33		13,34	
Dolní odhad	7,00		9,75		11,91		13,23		13,33		13,34	
Horní odhad	16,00		13,94		13,48		13,36		13,35		13,34	

V jednotlivých krocích iteračního postupu jsou podle vztahu (2) algoritmu 1 určeny tvar řízení (S_j), hodnoty váhových koeficientů $v_i^{(S_j)}$ a průměrného zisku. Klademe $v_3 = 0$. Dolní a horní odhady optimálního průměrného zisku jsou vypočteny podle vztahů (3) a (3*).

Protože vztahy (6) a (6*) jsou speciálními případy vztahů (5) a (5*) pro případ, že $s_1 = s_2 = \ldots = s_{n+1} = 1$, plyne jejich platnost z platnosti dříve dokázaných vztahů (5) a (5*). Pro ilustraci popsané metody spočteme příklad, který je uveden v Howardově monografii [1] na str. 44.

Příklad 1. Hledáme optimální řízení markovského řetězce se 3 stavy. Ve stavu 1 můžeme použít jedné ze 3 strategií zadaných hodnotami:

$$\begin{array}{llll} p_{11}^1 = 0{,}25\,; & p_{12}^1 = 0{,}25\,; & p_{13}^1 = 0{,}5\,; & c_{11} = 7{,}0\,; \\ p_{11}^2 = 0{,}125\,; & p_{12}^2 = 0{,}75\,; & p_{13}^2 = 0{,}125\,; & c_{12} = 4{,}0\,; \\ p_{11}^3 = 0{,}75\,; & p_{12}^3 = 0{,}0625\,; & p_{13}^3 = 0{,}1875\,; & c_{13} = 4{,}5\,. \end{array}$$

Ve stavu 2 máme k dispozici 2 strategie:

$$\begin{array}{llll} p_{21}^1 = 0{,}5\,; & p_{22}^1 = 0{,}0\,; & p_{23}^1 = 0{,}5\,; & c_{21} = 16{,}0\,; \\ p_{21}^2 = 0{,}0625\,; & p_{22}^2 = 0{,}875\,; & p_{23}^2 = 0{,}0625\,; & c_{22} = 15{,}0\,. \end{array}$$

Ve stavu 3 můžeme použít jedné ze 3 strategií: Ve stavu 3 můžeme použít jedné ze 3 strategií:

$$\begin{array}{llll} p_{31}^1 = 0{,}5\,; & p_{32}^1 = 0{,}25\,; & p_{33}^1 = 0{,}25\,; & c_{31} = 8{,}0\,; \\ p_{31}^2 = 0{,}0625\,; & p_{32}^2 = 0{,}75\,; & p_{33}^2 = 0{,}1875\,; & c_{32} = 2{,}75\,; \\ p_{31}^3 = 0{,}25\,; & p_{32}^3 = 0{,}125\,; & p_{33}^3 = 0{,}625\,; & c_{33} = 4{,}25\,. \end{array}$$

Výsledky obdržené řešením příkladu 1 podle algoritmu 1 jsou uvedeny v tabulce 1. V každém kroku iteračního postupu je prováděn dolní a horní odhad optimálního průměrného zisku. Výpočet byl proveden na počítači LGP-21, program byl sestaven v jazyce ACT-V.

(Došlo dne 31. srpna 1968.)

LITERATURA

[1] R. A. Howard: Dynamic Programming and Markov Processes. Technology Press and Wiley Press, New York 1960. Ruský překlad Sovetskoe radio 1965.

[2] D. J. White: Dynamic Programming Markov Chains and the Method of Successive Approximations. Journal of Mathematical Analysis and Applications *6* (1963), 373—376.

[3] F. d'Épenoux: A Probabilistic Production and Inventory Model. Management Science *10* (1963), 98—108.

[4] J. Mac Queen: A Modified Dynamic Programming Method for a Markovian Decision Problems. Journal of Mathematical Analysis and Applications *14* (1966), 38—43.

SUMMARY

On Successive Approximations Method for Finding Optimal Control of a Markov Chain

Karel Sladký

White [2] uses some sort of successive approximations method to determine optimal control for a Markov chain with returns for the average return criterion of optimality. This paper describes how to produce upper and lower bounds of the optimal average return at each iteration of the algorithm proposed in [2]. These upper and lower bounds converge monotonely to the optimal average return. Also the policy determined at each stage achieves an average return which is lower than the corresponding upper bound and at least as good as the corresponding lower bound. Similar results for the case with discounting were obtained in [4].

Ing. Karel Sladký, Ústav teorie informace a automatizace ČSAV, Vyšehradská 49, Praha 2.