KYBERNETIKA ČÍSLO 6, ROČNÍK 5/1969

Lagrangeovy násobitelé a optimalizace nepřetržitých Markovových řetězců

PETR MANDL

V práci je použito metody Lagrangeových násobitelů k úloze maximalizace průměrného výnosu na jeden krok v Markovově řetězci, jehož přechodová matice je funkcí $P(u)$ parametru u.

Teorie řízených Markovových řetězců byla rozvíjena hlavně metodou dynamického programování ([1], [2], též příloha časopisu Kybernetika [3]). Podstatným pro použití této metody je předpoklad, že hodnota parametru řízení je funkcí okamžitého stavu soustavy. Přitom k řízení soustavy může být zvolena libovolná taková funkce. Není-li tento předpoklad splněn, není možno dynamického programování přímo použít. To nastává např. nelze-li některé stavy soustavy pozorováním odlišit. Tato práce se týká optimalizace průměrného výnosu v nepřetržitých řetězcích. Ačkoliv se jedná o nalezení podmíněného extrému, nezdá se, že by někdo o této úloze z hlediska klasické metody Lagrangeových násobitelů dříve pojednal.* Práce dává interpretaci některých výsledků dynamického programování pomocí Lagrangeových násobitelů. Nepoužívá však zmíněného předpokladu, vyžadovaného metodou dynamického programování. Je ukázáno, že platí globální nutná podmínka optimálnosti, připustíme-li smíšené strategie.

Každému $u \in U \subset R^s$ budiž přiřazena stochastická matice $P(u) = \|p_{ij}(u)\|_{i,j=1}^{r}$, kterou nelze přerovnáváním řádků a sloupců uvést na tvar

$$\begin{pmatrix} P_1, & 0, & 0 \\ 0, & P_2, & 0 \\ R_1, & R_2, & R_3 \end{pmatrix},$$

kde $P_1 \neq 0$, $P_2 \neq 0$ jsou čtvercové matice. Tento předpoklad znamená, že matice není totálně rozložitelná neboli, v terminologii Markovových řetězců, definuje

* Když byla tato práce v tisku, vyšel článek P. J. Schweitzera (Perturbation theory and undiscounted Markov renewal programming. Operations Research *17* (1969), 716—727), který obsahuje vyjádření gradientu odpovídající formuli (10).

pouze jednu izolovanou rekurentní třídu. Označme $P(u)^n = \|p_{ij}^{(n)}(u)\|_{i,j=1}^r$. Limity

$$\lim_{N\to\infty} N^{-1} \sum_{n=1}^{N} p_{ij}^{(n)}(u), \quad j = 1, \ldots, r, \tag{1}$$

nezávisejí na i a představují jediné řešení soustavy rovnic

$$xP(u) = x, \quad x = (x^1, \ldots, x^r), \tag{2}$$

s podmínkou $\sum_j x^j = 1$, tj.

$$x \,.\, e = 1, \quad e = (1, \ldots, 1). \tag{3}$$

Tečka značí skalární součin.

Chápejme nyní $P(u)$ jako matici pravděpodobností přechodu Markovova řetězce a předpokládejme, že s výskytem řetězce ve stavu j je spojen výnos velikosti $\varrho^j(u)$. Průměrný výnos na jeden krok $\Theta(u)$ je potom roven

$$\Theta(u) = \lim_{N\to\infty} N^{-1} \sum_{n=1}^{N} \sum_j p_{ij}^{(n)}(u)\, \varrho^j(u) = \sum_j x^j \varrho^j(u),$$

kde x je řešení (2) splňující (3). Úloha nalezení takového $u \in U$, pro něž je průměrný výnos maximální, spočívá tedy v nalezení maxima funkce

$$x \,.\, \varrho(u) \tag{4}$$

za podmínek (2), (3). Ve (4) je $\varrho(u) = (\varrho^1(u), \ldots, \varrho^r(u))$. u představuje parametr řízení.

Použijeme metody Lagrangeových násobitelů. Zpočátku budeme předpokládat, že $P(u)$, $\varrho(u)$ jsou definovány na nějaké otevřené množině $G \supset U$ a mají spojité parciální derivace v G. Zaveďme vektor Lagrangeových násobitelů $\lambda = (\lambda^1, \ldots, \lambda^r)$ odpovídajících podmínkám (2) a násobitel μ pro podmínku (3). Utvořme Lagrangeovu funkci

$$L = x \,.\, \varrho(u) + \lambda \,.\, (xP(u) - x) + \mu(1 - x \,.\, e).$$

Podmínka stacionárnosti L vzhledem k proměnným $x^1, \ldots, x^r$ je

$$0 = \frac{\partial L}{\partial x^j} = \varrho^j(u) + \sum_k p_{jk}(u)\, \lambda^k - \lambda^j - \mu = 0, \quad j = 1, \ldots, r. \tag{5}$$

Nechť x splňuje (2), (3). Vynásobením (5) x^j a sečtením pro $j = 1, \ldots, r$ dostáváme

$$0 = \sum_j x^j \frac{\partial L}{\partial x^j} = x \,.\, \varrho(u) + \lambda \,.\, (xP(u) - x) - \mu,$$

tj.

$$\mu = x \,.\, \varrho(u) = \Theta(u) \,.$$

λ je tedy řešením soustavy

$$\Theta e + \lambda = \varrho(u) + \lambda P(u)' \,. \tag{6}$$

Čárka značí transponovanou matici (vektor).

Lemma 1. *Soustava* (6) *jakožto soustava rovnic pro neznámé* Θ, $(\lambda^1, \ldots, \lambda^r)$ *má řešení pro každé* $u \in U$. *Přitom* Θ *je určeno jednoznačně,* λ *až na aditivní konstantu.*

Důkaz. Vyhovují-li λ a Θ soustavě (6), potom jí vyhovují i $\lambda + ke$, Θ. Můžeme proto voliti $\lambda^r = 0$. Tím dostáváme soustavu rovnic pro neznámé $\Theta, \lambda^1, \ldots, \lambda^{r-1}$ s determinantem

$$\begin{vmatrix} 1, & 1 - p_{11}(u), & \ldots, & -p_{1r-1}(u) \\ 1, & -p_{21}(u), & \ldots, & -p_{2r-1}(u) \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ 1, & -p_{r-1,1}(u), & \ldots, & 1 - p_{r-1,r-1}(u) \\ 1, & -p_{r,1}(u) & \ldots, & -p_{r,r-1}(u) \end{vmatrix} = (-1)^{r+1} \sum_j C_{ij}[E - P(u)] \,, \quad i = 1, \ldots, r \,. \tag{7}$$

$C_{ij}[E - P(u)]$ označuje algebraický doplněk prvku v i-tém řádku a j-tém sloupci v determinantu $|E - P(u)|$. O platnosti (7) pro libovolné i se přesvědčíme, přičteme-li k $(i+1)$-vému sloupci determinantu ostatní sloupce kromě prvého a rozvineme-li determinant podle prvního sloupce. $C_{ij}[E - P(u)]$, $j = 1, \ldots, r$, představuje pro některé i (ve skutečnosti pro všechna i) nenulové řešení soustavy (2). Toto řešení je úměrné limitám (1). Je tedy zejména $\sum_j C_{ij}[E - P(u)] \neq 0$, a proto determinant v (7) je nenulový. Vidíme, že $\Theta, \lambda^1, \ldots, \lambda^{r-1}$ jsou určeny jednoznačně. □

$\delta u \in R^s$ nazveme *přípustnou variací* $u \in U$, když

$$\tilde{u} = u + \varepsilon \delta u + o(\varepsilon) \in U \tag{8}$$

pro $\varepsilon > 0$, $\varepsilon \to 0$.

Lemma 2. *Platí-li* (8), *potom*

$$\Theta(\tilde{u}) = \Theta(u) + \varepsilon \delta \Theta(u) + o(\varepsilon) \,, \tag{9}$$

kde je

$$\delta \Theta(u) = \delta u \,.\, \nabla\{x \,.\, [\varrho(u) + \lambda P(u)']\} \,. \tag{10}$$

V (10) ∇ značí gradient vzhledem k proměnným u^j ve funkcích $\varrho(u)$ a $P(u)$. x je řešením (2), (3), λ řešením (6).

Důkaz. Nechť platí (8). Ve stručném zápisu budeme veličiny odpovídající hodnotě $\tilde{u}$ odlišovat vlnovkou. Máme

$$\tilde{\Theta} - \Theta = \tilde{x} \,.\, (\tilde{\varrho} - \Theta e) = \tilde{x} \,.\, (\tilde{\varrho} + \lambda - \varrho - \lambda P') = \tilde{x} \,.\, (\tilde{\varrho} - \varrho) + \tilde{x}(\tilde{P} - P)\,\lambda' \,.$$

Jelikož $\tilde{x} \to x$ pro $\varepsilon \to 0$, jak se přesvědčíme např. vyjádřením $\tilde{x}$ pomocí doplňků k libovolnému sloupci determinantu $|E - P(u)|$, dostáváme

$$\tilde{\Theta} - \Theta = x \,.\, (\tilde{\varrho} - \varrho) + x(\tilde{P} - P)\,\lambda' + o(\varepsilon) =$$

$$= \varepsilon \delta u \,.\, \nabla\{x \,.\, [\varrho(u) + \lambda P(u)']\} + o(\varepsilon) \,.$$

Tím je (9) a (10) ověřeno. □

Poznámka. Vztahy (9), (10) je možno stručně zapsat jako

$$\nabla\Theta(u) = \nabla\{x \,.\, [\varrho(u) + \lambda P(u)']\} \,.$$

Je však třeba přitom mít na mysli, že $\Theta(u)$ pro $u \in G - U$ nemusí být definováno, takže gradient na levé straně nemusí mít smysl.

$\hat{u} \in U$ nazveme *optimální hodnotou parametru řízení*, když

$$\Theta(\hat{u}) = \max_{u \in U} \Theta(u) \,.$$

Z maximality $\Theta(\hat{u})$ a z (9) vyplývá, že pro každou přípustnou variaci $\delta\hat{u}$ parametru $\hat{u}$ musí být $\delta\Theta(\hat{u}) \leqq 0$, tj.

$$\delta\hat{u} \,.\, \nabla\{\hat{x} \,.\, [\varrho(\hat{u}) + \hat{\lambda} P(\hat{u})']\} \leqq 0 \,. \tag{11}$$

Je-li $\hat{u}$ vnitřním bodem množiny U, takže všechny variace jsou přípustné, musí být

$$\nabla\{\hat{x} \,.\, [\varrho(\hat{u}) + \hat{\lambda} P(\hat{u})']\} = 0 \,.$$

(11) představuje *lokální kritérium optimálnosti.*

K odvození *globálního kritéria* utvořme pro $u \in U$ ze sloupcového vektoru $\varrho(u)'$ a matice $P(u)$ matici

$$R(u) = (\varrho(u)', P(u))$$

a předpokládejme, že množina

$$R = \{R(u) : u \in U\}$$

je konvexní množinou matic. Tj. pro libovolné a, $0 \leqq a \leqq 1$, spolu s maticemi $R(u_1)$, $R(u_2)$ obsahuje také matici $aR(u_1) + (1 - a)\,R(u_2)$.

Kdyby R nebyla konvexní, bylo by třeba uvažovati její konvexní obal

$$\overline{R} = \{\sum_{i=1}^{n} \gamma_i R(u_i) : u_i \in U \,, \quad \gamma_i \geqq 0,\ i = 1, \ldots, n, \sum_{i=1}^{n} \gamma_i = 1,\ n = 1, 2, \ldots\} \,.$$

$\sum_{i=1}^{n} \gamma_i P(u_i)$ je maticí pravděpodobností přechodu Markovova řetězce, ve kterém v každém kroku volíme parametr u náhodně, hodnotu u_i s pravděpodobností γ_i. Vektor $\sum_{i=1}^{n} \gamma_i \varrho(u_i)$ udává očekávaný výnos. Konvexnost množiny R lze tedy docílit náhodnou volbou parametru řízení neboli používáním *smíšených strategií*. Přitom podmínka, aby přechodová matice nebyla totálně rozložitelná zůstává zřejmě zachována.

O derivatelnosti $\varrho(u)$ a $P(u)$ podle u nejsou v dalším činěny žádné předpoklady.

Věta. *Tvoří-li R konvexní množinu a je-li $\hat{u}$ optimální hodnotou parametru řízení, potom platí*

$$\hat{x} \,.\, [\varrho(\hat{u}) + \hat{\lambda} P(\hat{u})'] = \max_{u \in U} \hat{x} \,.\, [\varrho(u) + \hat{\lambda} P(u)'] \,. \tag{12}$$

Důkaz. Mějme spolu s optimální hodnotou $\hat{u}$ libovolné $u \in U$. Uvažujme problém maximalizace vzhledem k $a \in [0, 1]$ veličiny

$$x \,.\, [a\varrho(u) + (1 - a)\, \varrho(\hat{u})] \,,$$

kde

$$x[aP(u) + (1 - a)\, P(\hat{u})] = x \,, \quad x \,.\, e = 1 \,.$$

Dle předpokladu ke každému a existuje u_a tak, že je

$$a\varrho(u) + (1 - a)\, \varrho(\hat{u}) = \varrho(u_a) \,, \quad aP(u) + (1 - a)\, P(\hat{u}) = P(u_a) \,.$$

Odtud

$$x \,.\, [a\varrho(u) + (1 - a)\, \varrho(\hat{u})] = \Theta(u_a) \,.$$

Snadno se vidí, že platí

$$\Theta(\hat{u}) = \Theta(u_0) = \max_{0 \leqq a \leqq 1} \Theta(u_a) \,.$$

Tedy $a = 0$ je optimální. Jelikož jsou splněny předpoklady derivovatelnosti, použité při odvození (11). dostáváme

$$0 \geqq \frac{\mathrm{d}}{\mathrm{d}a} \Theta(u_a)\Big|_{a=0} = \frac{\mathrm{d}}{\mathrm{d}a} \{\hat{x} \,.\, [a\varrho(u) + (1 - a)\, \varrho(\hat{u}) +$$

$$+ \hat{\lambda}(aP(u) + (1 - a)\, P(\hat{u}))']\}\Big|_{a=0} = \hat{x} \,.\, [\varrho(u) + \hat{\lambda} P(u)'] - \hat{x} \,.\, [\varrho(\hat{u}) + \hat{\lambda} P(\hat{u})'] \,.$$

Tím je (12) dokázáno. □

Bellmanova rovnice. Budiž $U = Z_1 \times Z_2 \times \ldots \times Z_r$, $u = (z^1, \ldots, z^r)$, $P(u) = \|p_{ij}(z^i)\|_{i,j=1}^{r}$. To znamená, že každý řádek matice $P(u)$ je funkcí nezávislého parametru, což je předpoklad metody dynamického programování. Rovněž tak, nechť $\varrho^i(u) = \varrho^i(z^i)$, $i = 1, \ldots, r$. Aby platilo (12) musí být zřejmě

$$\varrho^i(\hat{z}^i) + \sum_k p_{ik}(\hat{z}^i)\, \hat{\lambda}^k = \max_{z^i \in Z_i} [\varrho^i(z^i) + \sum_k p_{ik}(z^i)\, \hat{\lambda}^k] \,, \tag{13}$$

pro ta i, pro která je $\hat{x}^i > 0$. Je-li $P(\hat{u})$ nerozložitelná matice, je $\hat{x}^i > 0$ pro všechna i. Potom platí (13) pro $i = 1, \ldots, r$ a to je s ohledem na (6) vztah odvozený v [1].

Diskontovaný výnos. Vektor $y = (y^1, \ldots, y^r)$ určený vztahem

$$y = \sum_{n=0}^{\infty} \beta^n P(u)^n \varrho(u)' ,$$

kde je $0 < \beta < 1$, udává očekávaný diskontovaný výnos při počátečním stavu řetězce $i = 1, \ldots, r$. Je jediným řešením soustavy rovnic

$$y = \varrho(u) + \beta y P(u)' . \tag{14}$$

Budiž $q = (q^1, \ldots, q^r)$ dané rozložení pravděpodobností počátečního stavu řetězce. Zabývejme se maximalizací očekávaného diskontovaného výnosu, tj. veličiny $\psi(u) = q \,.\, y$, kde y splňuje (14). Z podmínky stacionárnosti Lagrangeovy funkce

$$L = q \,.\, y + \lambda \,.\, (\varrho(u) + \beta y P(u)' - y)$$

dostáváme pro Lagrangeovy násobitele rovnice

$$\lambda - \beta \lambda P(u) = q .$$

Obdobně jako v předchozím se odvodí

$$\nabla\psi(u) = \nabla\lambda \,.\, [\varrho(u) + \beta y P(u)']$$

a pro optimální hodnotu $\hat{u} \in U$ vztah

$$\hat{\lambda} \,.\, [\varrho(\hat{u}) + \beta \hat{y} P(\hat{u})'] = \max_{u \in U} \hat{\lambda} \,.\, [\varrho(u) + \beta \hat{y} P(u)'] ,$$

tvoří-li R konvexní množinu.

(Došlo dne 24. března 1969.)

LITERATURA

[1] R. Bellman: A Markovian Decision Process. J. of Math. and Mech. 6 (1957), 679—684.
[2] R. A. Howard: Dynamic Programming and Markov Processes. New York 1960.
[3] P. Mandl: Řízené Markovovy řetězce. Kybernetika 5 (1969), příloha.

Lagrange Multipliers and Optimization of Non-terminating Markov Chains

PETR MANDL

The paper is devoted to the mean reward per one step in a Markov chain, the transition matrix $P(u)$ of which depends on a parameter $u \in U \subset R^s$. The parameter is to be chosen so as to maximize the mean reward $\Theta(u)$. Employing the method of the Lagrange multipliers a formula for the gradient of Θ with respect to u is obtained. Hence a local condition for optimality is derived which is strengthened to a global one under the assumption that the transition matrices together with the corresponding reward vectors form a convex set. The paper gives an interpretation of the quantities employed in dynamic programming approach in terms of Lagrange multipliers. The same method can be followed in the case of the expected discounted reward and in the case of continuous time parameter.

Dr Petr Mandl, DrSc., Ústav teorie informace a automatizace ČSAV, Vyšehradská 49, Praha 2.