経営科学（日本オペレーションズ・リサーチ学会邦文機関誌） **第14巻 第1号**（1970年7月）

線形計画とマルコフ決定過程†

尾 崎 俊 治*

1. 序　　論

最近，オペレーションズ・リサーチ（ＯＲ）の各分野において逐次決定過程がよく議論されているが，本論文ではとくに系がマルコフ連鎖あるいはセミ・マルコフ過程によって支配されるような逐次決定過程，すなわち，マルコフ決定過程 (Markovian Decision Process, 略して MDP) あるいはセミ・マルコフ決定過程 (Semi-Markovian Decision Process, 略して SMDP) について議論する．

MDP は1957年 Bellman［2］によって初めて議論された．彼はダイナミック・プログラミング (Dynamic Programming, 略して DP) にマルコフ連鎖を適用した．Howard の有名な著書［10］の出現以来，MDP は多くの人々により研究されるようになった．MDPはＯＲにおける決定問題として生起したものであるが，このモデルの持つ一般的性格のために，多くの問題，たとえば，在庫管理［5］[17]，取替問題 [22]，品質管理 [16]，抜取検査 [23]，Reject Allowance [15] [24]，最適制御問題［1］［9］，信頼性理論［7］［8］[14]，その他で応用されており，今後も理論的発展とともにさらに多くの分野に適用されるものと思われる．

MDP を解くアルゴリズムは大別して2つにわけられる．1つは Howard [10] の政策反復 (Policy-Iteration, 略して PI) アルゴリズムであり，もう1つは Manne [17]，D'Epenoux［5］等による線形計画 (Linear Programming, 略して LP) アルゴリズムである．PI アルゴリズムは DP のいわゆる「政策空間における逐次近似」を用いた方法であり，Blackwell［3］および Veinott [25] その他の人々により，厳密に論ぜられている．一方，Manne 等は同じ問題を LP で定式化している．LP は電子計算機のプログラムが利用できるという利点を持っている．

この論文は MDP および SMDP を直接に定式化した LP 問題を主問題として考えたとき，Howard 型の PI アルゴリズムは本質的にはこの LP 問題と等価であることを示す．

まず，Howard の意味での完全エルゴード過程 (Completely Ergodic Process) の場合について詳細に等価性を示し，これらの2つのアルゴリズムの比較を，Howard [10] のタクシー問題および自動車取替問題について詳細に述べ，さらにそれらの改良を試みる．割引率を考慮した過程，

† 1968年2月24日受理．

* 京都大学大学院，現在広島大学工学部．

および終点のある過程についても等価性の存在を示す．さらに，SMDPについても同じようにして等価性が成立することを示す．

2. マルコフ決定過程

整数 $i=1,2,\cdots\cdots,m$ によって表わされる有限の状態の集合 S よりなる系を考える．単位期間ごとに，例えば1日毎に，それらのうちの1つの状態を観測し，しかるのちに1つの決定を下さなければならないとする．各状態 i における決定は整数 $i=1,2,\cdots\cdots,K_i$ で表わされる有限個の集合 $\boldsymbol{K}_i$ より1つ選ばれる．状態 i $(i\epsilon S)$ で決定 k $(k\epsilon\boldsymbol{K}_i)$ を下すことにより，つぎの2つのことが起こる．

（i） 利得 r_i^k を受けとる．

（ii） つぎの期間では，系は推移確率 $p_{ij}^k(j\epsilon S)$ によって支配される．

ここで，r_i^k, p_{ij}^k は時刻 $n(n=1,2,\cdots\cdots)$ と独立であり，r_i^k はすべて有限とする．また，明らかに，

$$(2.1)\qquad \sum_{j\epsilon S} p_{ij}^k=1,\quad p_{ij}^k\geqq 0,\quad i\epsilon S,\ j\epsilon S,\ k\epsilon\boldsymbol{K}_i$$

となる．さらに，初期分布

$$(2.2)\qquad \boldsymbol{a}=(a_1,a_2,\cdots\cdots,a_m)$$

を与えれば，この系は決定される．すなわち，この系は利得を持った非定常マルコフ連鎖となる．以上述べた規則にしたがって，逐次決定してゆくことにより得られる有限期間，または無限期間の総期待利得，あるいは単位期間当りの平均期待利得を最大にする政策およびその値を求める問題がMDPである．さらに，割引率を導入することもできる．以下，この論文では無限期間の場合のみを取り扱う．

マルコフ連鎖の分類[1]によっていくつかの問題が起こる．まず，決定がなんであっても，考えているマルコフ連鎖が常にエルゴード的になる場合を考える．Howard [10] はこの系を完全エルゴード過程と呼んでいるが，ここではエルゴード・マルコフ連鎖と呼ぼう．同様に，決定がなんであっても，同じ吸収状態と同じ過度状態が定まる場合には，吸収マルコフ連鎖と呼ぼう．

つぎに，政策（Policy）を定義しよう．状態空間を S とし，政策空間を $F=\boldsymbol{K}_1\times\boldsymbol{K}_2\times\cdots\cdots\times\boldsymbol{K}_m$（$\boldsymbol{K}_i$ の直積空間）としよう．任意の決定を $f\epsilon F$ で表わすとする．そのとき，時刻 $n(n=1,2,\cdots\cdots)$ における決定を f_n とする．政策は $\pi=(f_1,f_2,\cdots\cdots,f_n,\cdots\cdots)$ で表わされる．すなわち，政策とは各時刻における決定の列である．f_n が時刻 n と独立のとき，定常政策と呼び，$\pi=f^\infty$ で表わす．任意の $f\epsilon F$ に対し，$r(f)$ は決定 f を行なったときの利得 r_j^k の m 次列ベクトルであり，$Q(f)$ は推移確率 p_{ij}^k を要素とする $m\times m$ 行列である．また，$Q_n(\pi)=Q(f_1)Q(f_2)\cdots Q(f_n)$ $(n=1,2,\cdots\cdots)$ であり，とくに $Q_0(\pi)=I$（単位行列）と定義する．

2・1 エルゴード・マルコフ連鎖

エルゴード・マルコフ連鎖に対しては，総期待利得は一般に発散するので，そのかわりに単位

1) 有限マルコフ連鎖の分類については，Kemeny and Snell [13] に従うとする．

時間当りの平均期待利得の下極限，すなわち，

$$(2.3)\qquad G_1(\pi)=\lim_{n\to\infty}\inf\frac{1}{n+1}\sum_{i=0}^{n}\boldsymbol{a}Q_i(\pi)r(f_{i+1})$$

を考える．この $G_1(\pi)$ を最大にする政策 π およびその値を求めるのがわれわれの目的である．さて，この系に対して，Derman［6］は混合非定常政策，すなわち各時刻での決定を確率的にとる場合に，つぎの定理が成り立つことを示した．

［定理 2.1］ エルゴード・マルコフ連鎖においては最適な定常純粋政策が存在する．

証明は DP の関数方程式を用いてなされている．この定理により，最適平均期待利得は，

$$(2.4)\qquad g_1=\max_{f\in F}\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}\boldsymbol{a}[Q(f)]^i r(f)$$

により与えられる．式 (2.4) より，この問題はつぎの LP 問題に定式化されることが知られている（文献 [6] 参照）．

$$(2.5)\qquad \mathrm{Max}\sum_{j\in S}\sum_{k\in K_j}r_j^k x_j^k$$

subject to

$$(2.6)\qquad x_j^k\geqq 0\ ,\quad j\in S,\ k\in\boldsymbol{K}_i$$

$$(2.7)\qquad \sum_{k\in K_j}x_j^k-\sum_{i\in S}\sum_{k\in K_j}p_{ji}^k x_i^k=0,\quad j\in S$$

$$(2.8)\qquad \sum_{j\in S}\sum_{k\in K_j}x_j^k=1$$

ここで，最適決定は，

$$(2.9)\qquad d_j^k=x_j^k/\sum_{k\in K_j}x_j^k\ ,\quad j\in S,\ k\in\boldsymbol{K}_j$$

で与えられる．すなわち，d_j^k は状態 j で決定 k を選ぶ確率である．この LP 問題について，つぎの定理が成立する．

［定理 2.2］ LP 問題 (2.5)～(2.8) の最適解の中には各 $j\in S$ に対し，ただ1つの $x_j^k>0$ で他の x_j^k は 0 となるものが存在する．

証明は LP の基底解の性質を用いて簡単に行なわれている（Wolfe and Dantzig [26] 参照）．この定理と式 (2.9) より，$d_j^k=0$ あるいは 1 となる．すなわち，純粋政策が最適となるから，定理 2.1 とも一致する．

さて，LP問題 (2.5)～(2.8) の双対問題を考えよう．m個の制限式 (2.7) のうち，1個はマルコフ連鎖の性質より冗長であるから，$j=m$に関する制限式を除いて，双対問題を考える．双対変数を $v_1, v_2, \cdots\cdots v_m$ とすれば，双対問題は，

$$(2.10)\qquad \mathrm{Min}\ v_m$$

subject to

$$(2.11)\qquad v_m+v_i\geqq r_i^k+\sum_{j=1}^{m-1}p_{ij}^k v_j$$

$$i=1,2,\cdots\cdots,m-1,\quad k\in\boldsymbol{K}_i$$

$$(2.12)\qquad v_i：\text{符号制限なし},\quad i\in S$$

	x_1^1	……	$x_1^{K_1}$	x_2^1	……	$x_2^{K_2}$	……	……	x_m^1	……	$x_m^{km}\geqq(0)$	
v_1	$1-p_{11}^1$	……	$1-p_{11}^{K_1}$	$-p_{21}^1$	……	$-p_{21}^{K_2}$	……	……	$-p_{m1}^1$	……	$-p_{m1}^{Km}$	$=0$
v_2	$-p_{12}^1$	……	$-p_{12}^{K_1}$	$1-p_{22}^1$	……	$1-p_{22}^{K_2}$	……	……	$-p_{m2}^1$	……	$-p_{m2}^{Km}$	$=0$
	⋮		⋮	⋮		⋮			⋮		⋮	⋮
v_{m-1}	$-p_{1,m-1}^1$	……	$-p_{1,m-1}^{K_1}$	$-p_{2,m-1}^1$	……	$-p_{2,m-1}^{K_2}$	……	……	$-p_{m,m-1}^1$	……	$-p_{m,m-1}^{Km}$	$=0$
g_1	1	……	1	1	……	1	……	……	1	……	1	$=1$
	VII		VII	VII		VII			VII		VII	
	r_1^1	……	$r_1^{K_1}$	r_2^1		$r_2^{K_2}$	……	……	r_m^1	……	r_m^{Km}	

図 2.1 エルゴード MDP に対する Tucker 図表

となる．定理 2.1 より，最適解は存在し，それは g_1 で与えられるから，双対定理を用いて $v_m=g_1$ とおけば，つぎの LP 問題

(2.13) Min g_1

subject to

(2.14) $g_1+v_i\geqq r_i^k+\sum_{j=1}^{m-1}p_{ji}^k v_j$

$i=1,2,\cdots\cdots,m-1,\ k\in\boldsymbol{K}_i$

(2.15) v_i, g_1：符号制限なし，$i=1,2,\cdots\cdots,m-1$

となる．これらの主および双対問題の関係を理解するために，Tucker 図表を図 2.1 に示す．

つぎに，ＬＰ問題 (2.5)～(2.8) を解くアルゴリズムを考える．このＬＰ問題は等号制限であるから，普通は 2 段階法あるいはその他の複合法 (Composite Algorithm) を用いなければならない．しかし，定理 2.2 より基底に入る変数は各 $j\in S$ に対しただ 1 つであり，また図 2.1 より明らかなように基底に入る変数は必ず j 番目の制限式に入る（すなわち，j 番目の基底になる）．そこで一挙に基底解を得るために，通常の単体判定基準を各 $j\in S$ に関して求める．たとえば，

(2.16) $-r_j^B=\min_{k\in Kj}[-r_j^k]\ ,\quad j\in S$

を用いれば，基底に入る変数が決まる．そこで基底を表わす添字 B を付ける．さて，得られた基底行列[2] を

(2.17) $$\boldsymbol{B}=\begin{bmatrix} 1-p_{11}^B & -p_{21}^B & \cdots\cdots & -p_{m1}^B \\ -p_{12}^B & 1-p_{22}^B & \cdots\cdots & -p_{m2}^B \\ \vdots & \vdots & & \vdots \\ -p_{1,m-1}^B & -p_{2,m-1}^B & \cdots\cdots & -p_{m,m-1}^B \\ 1 & 1 & \cdots\cdots & 1 \end{bmatrix}$$

とする．ただし，p_{ij}^B は基底に対応する推移確率である．さらに，目標関数に関する行をも付け加えた拡張された基底行列を

2) ＬＰの術語については文献 [19] 参照．

$$(2.18)\qquad \overline{\boldsymbol{B}}=\begin{bmatrix} 1 & -r_1^B & \cdots\cdots & -p_m^B \\ \cdots & \cdots\cdots & & \cdots\cdots \\ 0 & 1-p_{11}^B & \cdots\cdots & -p_{m1}^B \\ \vdots & \vdots & & \\ 0 & 1 & \cdots\cdots & 1 \end{bmatrix}=\begin{bmatrix} 1 & -\boldsymbol{r}^B \\ \cdots & \cdots\cdots \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix}$$

で表わす．$\overline{\boldsymbol{B}}$の逆行列（あるいは$\boldsymbol{B}$の逆行列）の存在することはエルゴード・マルコフ連鎖の性質より簡単に示すことができる．式 (2.18) の逆行列を

$$(2.19)\qquad \overline{\boldsymbol{B}}^{-1}=\begin{bmatrix} 1 & \boldsymbol{B}_0 \\ \cdots & \cdots\cdots \\ \boldsymbol{0} & \boldsymbol{B}^{-1} \end{bmatrix}$$

とすれば，$\overline{\boldsymbol{B}}\cdot\overline{\boldsymbol{B}}^{-1}=I$（単位行列）であるから，

$$\boldsymbol{B}_0=\boldsymbol{r}^B\boldsymbol{B}^{-1}$$

を得る．あるいは，

$$(2.20)\qquad {}^T\boldsymbol{B}^T\boldsymbol{B}_0={}^T\boldsymbol{r}^B$$

とも書ける．ただし，添字Tは行列の転置を表わす．ここで，m次行ベクトル $\boldsymbol{B}_0$ は式 (2.19) の定義よりこのＬＰ問題の単体乗数となっている．また，それは同時に双対変数でもあるから，$\boldsymbol{B}_0=[v_1, v_2, \cdots\cdots, v_{m-1}, g_1]$ となる．式 (2.21) を要素ごとに書けば，

$$(2.21)\qquad g_1+v_i=r_i^B+\sum_{j=1}^{m=1}p_{ij}^Bv_j\ ,\quad i\epsilon S$$

となる[3] これは，Howard の PI アルゴリズムでいえば，VDO (Value Determination Operation) に相当し，この主問題でいえば，単体乗数の満たす式に相当する．

さて，この単体乗数を用いて，つぎのステップの単体判定基準を作れば

$$(2.22)\qquad \Delta_i^k=-r_i^k\sum_{j=1}^{m-1}p_{ij}^kv_j+g_1+v_i\ ,\quad i\epsilon S,\ k\epsilon\boldsymbol{K}_i$$

となる（図 2.1 参照）．明らかに，基底変数に対しては，

$$(2.23)\qquad \Delta_i^B=-r_i^B-\sum_{j=1}^{m-1}p_{ij}^Bv_j+g_1+v_i=0\ ,\quad i\epsilon S$$

となる．さらに，すべての $i\epsilon S$, $k\epsilon\boldsymbol{K}_i$ に対して，

$$\Delta_i^k=-r_i^k-\sum_{j=1}^{m-1}p_{ij}^kv_j+g_1+v_i\geqq 0$$

あるいは，式 (2.23) を用いて

$$(2.24)\qquad r_i^B+\sum_{j=1}^{m-1}p_{ij}^Bv_j\geqq r_i^k+\sum_{j=1}^{m-1}p_{ij}^kv_j$$

ならば，最適解であることを示し，そのときの最適値は g_1 によって与えられる．これはPIアルゴリズムでいえば，PIR (Policy Improvement Routine) において最適政策を得たことに相当する．一方，もし

$$\Delta_i^k=-r_i^k-\sum_{j=1}^{m-1}p_{ij}^kv_j+g_1+v_i<0$$

---

3) 式 (2.21) で $i=m$ に関しては $v_m=0$ と考える．ここで，v_m は式 (2.11) の v_m とは異なる (Howard [10, p. 35] 参照．

あるいは，式 (2.23) を用いて，

$$(2.25)\qquad r_i^B+\sum_{j=1}^{m-1}p_{ij}^B v_j<r_i^k+\sum_{j=1}^{m-1}p_{ij}^k v_j$$

なる対 $(i,\ k)$ が少なくとも1つでも存在すれば，この政策は改善可能であることを示している．これはPIRにおいて政策の改善可能な場合を示している．LPでは，普通は1個の基底変数を入れ替えてゆくが，Howard の PI アルゴリズムでは，一挙に高々m個の基底変数を入れ替えることになる．そのとき，政策の改善，すなわち g_1 が増加することは，Howard [10] (pp. 42-43) によって証明されている．しかし，Howard の PI アルゴリズムではたとえ1つの式 (2.25) を満たす対 $(i,\ k)$ が存在しているときでも，あらためて m 元連立1次方程式を解かなければならない．これは，多数回の掃出し演算が必要となることを意味し，非常に無駄なことである．この問題については 2・2 で述べる．

以上述べたように，Howard の PI アルゴリズムは本質的には逆行列型改訂単体法と同じである．しかし，この MDP の持っている性質，たとえば定理 2.2 を上手に用いているという点では，通常のＬＰの解法と較べて秀れている．そこで，以上述べた議論を用いて，このアルゴリズムの改良を試みよう．

2・2 2つのアルゴリズムの比較とそれらの改良

まず，前節で述べた LP アルゴリズムを用いて，Howard [10] のタクシー問題および自動車取替問題[4] を解いて，PI アルゴリズムと比較してみよう．最初の基底解を得るためには，いろいろな判定基準が考えられるが，ここでは式 (2.16) の判定基準を用いる．したがって，最初のステップは Howard [10] の与えた PI アルゴリズムの数値例と同じである．しかし，以後は通常の単体表を用いて計算する．LP アルゴリズムでは，1つの変数について，政策を改善してゆくので，必ずしも PI アルゴリズムとは一致しない．また，掃出しの回数からいえば，LP の m ステップが PI アルゴリズムの1回の反復に相当する．ただし，LP では1ステップごとに単体判定基準を用いるので，この点を考慮すれば，ステップ数のみでどちらが計算量が少ないか断定できない．

図2・2はタクシー問題 (Howard [10]. pp. 44-45) のゲイン g_1 の増加を2つのアルゴリズムについて比較したものである．この図で横軸は，LP ではステップ数，PI アルゴリズムでは反復数にとる．ただし，m ステップ＝1反復にとっておく．

同様に図2・3に自動車取替問題 (Howard [10], pp. 54-56) のステップ数（反復数）とゲイン g_1 との関係を示す．これらの図からわかるように，一般に LP の方がステップ数に換算すれば，早く最適解に到達する．しかし，単体判定基準は多くなる．

そこで，この MDP のアルゴリズムの改良を試みよう．通常の単体法を用いずに，式 (2.22) の単体判定基準において，各 $i\in S$ に対し，Δ_i^k の最小値が負になる対 $(i,\ k)$ については，一度に掃出しを行なう．すなわち，通常の PI アルゴリズムと同じであるが，連立方程式を解くかわり

4) 自動車取替問題は分離形 MDP [4] となるので，もっと簡単に解ける．ここでは，一般の MDP の数値例として考える．

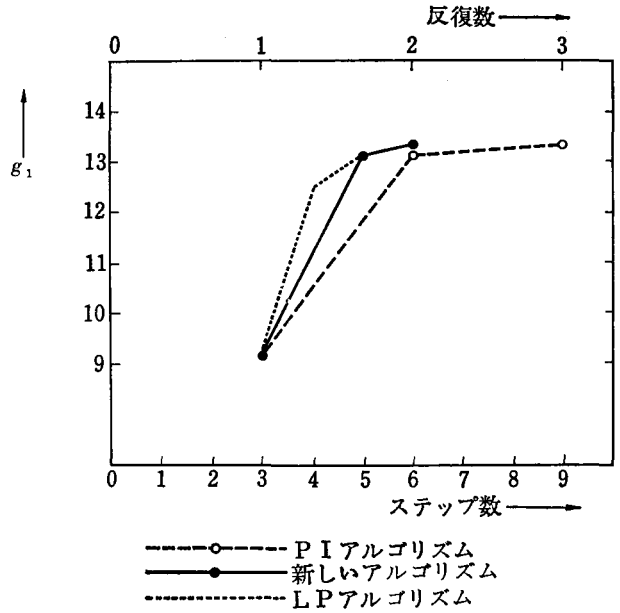

図2・2　3つのアルゴリズムの比較（タクシー問題）

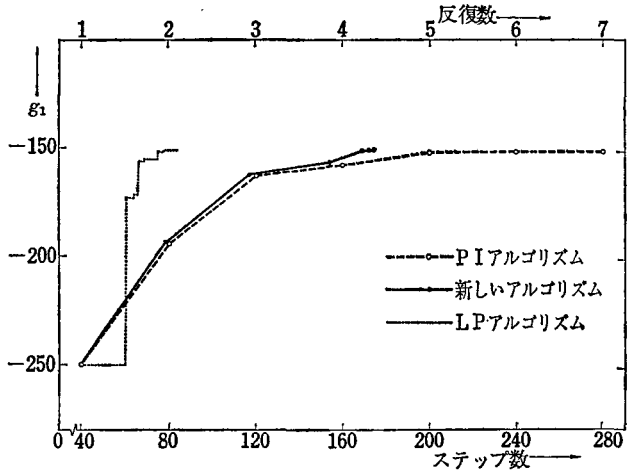

図2・3　3つのアルゴリズムの比較（自動車取替問題）

に，逆行列型改訂単体法を用いて基底を入れ替えてゆく．ただし，単体判定基準は PI アルゴリズムと同じところで使う．そのようにしてゆけば，判定基準を用いる回数は PI アルゴリズムと同じであり，ステップ数はかなり少なくなる．

図2・2のタクシー問題の場合には，この新しいアルゴリズムは掃出し演算6回で最適解に到達する．これは普通の LP による解の掃出し演算回数と同じで，単体判定基準は1回減っている．また，図2・3の自動車取替問題では，LP が84回の掃出し演算と45回の単体判定基準を必要とするのに対し，ここで述べた新しいアルゴリズムを用いると，175 回の掃出し演算と7回の単体判定基準が必要である．したがって，われわれのアルゴリズムの計算時間はかなり短くなる．一

方，PI アルゴリズムでは，40×7=280 回の掃出し演算を必要とするから，われわれのアルゴリズムは PI のそれに較べて約40%計算時間を節約できる．

これらの例からわかるように，LP と PI アルゴリズムを併用する新しいアルゴリズムは，非常に有効となる．

2・3 割引率を持ったマルコフ決定過程

2・3 では割引率を持った MDP を考える．割引率 $\beta(0\leqq\beta<1)$ を導入しよう．すなわち，ある時刻で1単位の利得は n 期間後には β^n となる．この過程に対しては総期待利得が収束するので，総期待利得を最大にする政策およびその値を求めるのが目的である．初期分布 (2.1) より出発したときの総期待利得は，

$$G_2(\pi)=\sum_{n=0}^{\infty}\boldsymbol{a}\beta^n Q_n(\pi)r(f_{n+1})$$

となる．したがって，すべての政策 π に対して，

$$g_2=G_2(\pi^*)\geqq G_2(\pi)$$

となる最適政策 π^* および g_2 を求めることである．この過程に対しても，つぎの有用な定理がある．

[**定理 2.3**] 定常な最適政策が存在する．

証明は Blackwell [3] によってなされた．この定理より，割引率を持った MDP はつぎの LP 問題となる [5]．

$$\text{(2.26)}\qquad \text{Max}\sum_{j\in S}\sum_{k\in K_j}r_j^k x_j^k$$

subject to

$$\text{(2.27)}\qquad \sum_{j\in S}\sum_{k\in K_j}(\delta_{jl}-\beta p_{jl}^k)x_j^k=a_l,\quad l\in S$$

$$\text{(2.28)}\qquad x_j^k\geqq 0\ ,\quad j\in S,\ k\in \boldsymbol{K}_j$$

この LP 問題に対しても定理 2.2 と同じ定理が成り立つ．証明は前とほぼ同じである．この事実より，状態 j で決定 k を選ぶ確率 d_j^k は0または1となる．すなわち，純粋政策が最適であることを示している．

LP 問題 (2.26)～(2.28) の双対問題は双対変数を $v_1, v_2, \cdots, v_m$ として

$$\text{(2.29)}\qquad \text{Min}\sum_{i\in S}a_i v_i$$

subject to

$$\text{(2.30)}\qquad v_i\geqq r_i^k+\beta\sum_{j\in S}p_{ij}^k v_j\ ,\quad i\in S,\ k\in \boldsymbol{K}_i$$

$$\text{(2.31)}\qquad v_i：\text{符号制限なし},\quad i\in S$$

となる．この LP 問題は PI アルゴリズムより直ちに得られる．ただし，目標関数は初期分布の加重平均になっているが，後に示すように，初期分布とは独立になるので，任意の v_i を最大にすると考えればよい．これらの2つの LP 問題を図表に示せば，図 2・4 になる．

さて，LP 問題 (2.26)～(2.28) を主問題と考えて，この LP 問題を解こう．まず，つぎの定理

	x_1^1	⋯⋯	$x_1^{K_1}$	x_2^1	⋯⋯	$x_2^{K_2}$	⋯⋯	⋯⋯	x_m^1	⋯⋯	$x_m^{K_m}$ $(\geqq 0)$	
v_1	$1-\beta p_{11}^1$	⋯⋯	$1-\beta p_{11}^{K_1}$	$-\beta p_{21}^1$	⋯⋯	$-\beta p_{21}^{K_2}$	⋯⋯	⋯⋯	$-\beta p_{m1}^1$	⋯⋯	$-\beta p_{m1}^{K_m}$	$=a_1$
v_2	$-\beta p_{12}^1$	⋯⋯	$-\beta p_{12}^{K_1}$	$1-\beta p_{22}^1$	⋯⋯	$1-\beta p_{22}^{K_2}$	⋯⋯	⋯⋯	$-\beta p_{m2}^1$	⋯⋯	$-\beta p_{m2}^{K_m}$	$=a_2$
⋮	⋮		⋮	⋮		⋮			⋮		⋮	⋮
⋮	⋮		⋮	⋮		⋮			⋮		⋮	⋮
v_m	$-\beta p_{1m}^1$	⋯⋯	$-\beta p_{1m}^{K_1}$	$-\beta p_{2m}^1$	⋯⋯	$-\beta p_{2m}^{K_2}$	⋯⋯	⋯⋯	$1-\beta p_{mm}^1$	⋯⋯	$1-\beta p_{mm}^{K_m}$	$=a_m$
	VII		VII	VII		VII			VII		VII	
	r_1^1	⋯⋯	$r_1^{K_1}$	r_2^1	⋯⋯	$r_2^{K_2}$	⋯⋯	⋯⋯	r_m^1	⋯⋯	$r_m^{K_m}$	

図 2・4　割引率を持った MDP に対する Tucker 図表

が成り立つ.

[定理 2.4] LP 問題 (2.26)～(2.28) において，最適の基底解の組は初期分布 $\boldsymbol{a}$ と独立である．すなわち，最適政策は初期分布と独立である.

[証明] 定理 2.2 より x_j^k のうち基底に入る変数については各 $j\epsilon S$ に対してただ 1 つの k が定まる．また，図 2・4 からも明らかなように，その基底は j 番目の制限式に入る．もし対応する a_j が 0 ならば，目標関数は増大しないが，a_i に関する制限 $a_i\geqq 0$, $\sum_{i\epsilon S} a_i=1$ より必ず $a_i>0$ なる i が存在し，その i に基底を入れれば，右辺はすべて正となるから，一度基底解が決まれば，以後は必ず右辺は正となり，また基底の入る制限式は定まっている．すなわち，基底解の組は初期分布 $\boldsymbol{a}$ と独立である．この定理は双対変数 v_i の意味を考えれば，直ちに理解される.

そこで，通常の LP の単体判定基準を用いるとすれば，

(2.33)　$$-r_j^B=\min_{k\epsilon K_j}[-r_j^k]\ ,\quad j\epsilon S$$

を得る．対応する基底行列は，

(2.33)　$$\boldsymbol{B}={}^T[I-\beta\boldsymbol{P}^B]=[\delta_{ji}-\beta p_{ji}^B]$$

となる．ここで，$\boldsymbol{P}^B=[p_{ji}^B]$ とする．また，拡張された基底行列は，

(2.34)　$$\overline{\boldsymbol{B}}=\begin{bmatrix}1 & \vdots & -\boldsymbol{r}^B\\ \cdots & \cdots & \cdots\\ \boldsymbol{0} & \vdots & \boldsymbol{B}\end{bmatrix}$$

となる．したがって，$\overline{\boldsymbol{B}}$ の逆行列を

(2.35)　$$\overline{\boldsymbol{B}}^{-1}=\begin{bmatrix}1 & \vdots & \boldsymbol{B}_0\\ \cdots & \cdots & \cdots\\ \boldsymbol{0} & \vdots & \boldsymbol{B}^{-1}\end{bmatrix}$$

とおけば，単体乗数 $\boldsymbol{B}_0=[v_1, v_2, \cdots, v_m]$ は

(2.36)　$$\boldsymbol{B}_0=\boldsymbol{r}^B\boldsymbol{B}^{-1}$$

となる．これはまた

(2.37)　$${}^T\boldsymbol{B}\,{}^T\boldsymbol{B}_0={}^T\boldsymbol{r}^B$$

あるいは

(2.38)　$$v_i=r_i^B+\beta\sum_{j\epsilon S}p_{ij}^B v_j\ ,\quad i\epsilon S$$

とも書ける．すなわち，単体乗数（双対変数）を求めることが，PIアルゴリズムではVDOに相当し，つぎのステップの単体判定基準は，

(2.39) $$\Delta_i^k=-r_i^k-\beta\sum_{j\in S}p_{ij}^k v_j+v_i\ ,\quad i\in S,\ k\in \boldsymbol{K}_i$$

となる．したがって，基底解に対しては，

(2.40) $$\Delta_i^B=-r_i^B-\beta\sum_{j\in S}p_{ij}^B v_j+v_i=0\ ,\quad i\in S$$

となり，もしすべての $i\in S$，$k\in \boldsymbol{K}_i$ に対し

(2.41) $$\Delta_i^k=-r_i^k-\beta\sum_{j\in S}p_{ij}^k v^j+v_i\geqq 0$$

ならば，最適解を得たことになる．一方，もし

(2.42) $$\Delta_i^k=-r_i^k-\beta\sum_{j\in S}p_{ij}^k v_j+v_i<0$$

なる対 $(i,\ k)$ が1つでも存在すれば，この政策は改善可能である．これらのことはPIアルゴリズムのPIRに相当する．

これらのことから，エルゴード連鎖の場合と同様にして，判定基準はPIアルゴリズムと同じところで用い，基底の入れ替えは連立方程式を解かずに，逆行列型改訂単体法を用いる新しいアルゴリズムを使用すれば，非常に早く最適解が得られる．

2・4 終点のあるマルコフ決定過程

MDPの最後の場合として，終点のあるMDPについて議論する．すなわち，つぎの仮定を導入する．

終点仮定：決定がなんであっても，どの状態も有限期間のうちに，1つの共通の吸収状態に到達する確率が存在する．

この仮定はまたつぎのようにも言える．どのような決定を選んでも，状態は1つの共通の吸収状態と他の残りの過渡状態とに分けられる．そこで，状態1を吸収状態に，状態 $i=2,3,\cdots\cdots,m$ を過渡状態とする．ここでは，われわれは吸収されるまでの系の行動に関心がある．すなわち，吸収されるまでに得る総期待利得を最大にする政策およびその値を求めたい．総期待利得が収束することは系が状態1に確率1でもって有限期間のうちに吸収されることから明らかである．そこで，$i=2,\cdots\cdots,m$ よりなる集合を S' と表わす．まず，ある政策 π を用いたときの総期待利得は，

(2.43) $$G_3(\pi)=\sum_{n=0}^{\infty}\boldsymbol{\alpha}'Q_n'(\pi)r'(f_{n+1})$$

となる．ここで，"′" は今までの理論とは異なり，すべての状態は $i=2,\cdots\cdots,m$ の上で考えるとする．すなわち，第1行，第1列を除いた行列あるいはベクトルである．したがって，$G_3(\pi)$ を最大にする政策 π およびその値を求めることが問題である．この系に対しても，つぎの定理が成り立つ．

[定理 2.5] 定常な最適政策が存在する．

証明はBlackwell［3］の前半の割引率を考慮した場合とほぼ同様にして証明できるが，ここでは省く．あるいは，Derman［6］による別証明がある．一般のMDPにおいても定常な最適

政策が存在することは Blackwell [3] が証明しているので，その特殊な場合とも考えられる．この定理を用いると，この系に対してもつぎの LP 問題を得る [18].

(2.44) $\mathrm{Max} \sum_{j\epsilon S'} \sum_{k\epsilon K_j} r_j^k x_j^k$

subject to

(2.45) $\sum_{j\epsilon S'} \sum_{k\epsilon K_j} (\delta_{jl} - p_{jl}^k) x_j^k = a_l \ , \quad l\epsilon S'$

(2.46) $x_j^k \geqq 0 \ , \quad j\epsilon S', \ k\epsilon \boldsymbol{K}_j$

この問題の双対問題は双対変数を $v_2, \cdots\cdots, v_m$ とすれば，

(2.47) $\mathrm{Min} \sum_{i\epsilon S'} a_i v_i$

subject to

(2.48) $v_i \geqq r_i^k + \sum_{j\epsilon S'} p_{ij}^k v_j \ , \quad i\epsilon S', \ k\epsilon \boldsymbol{K}i$

(2,49) v_i；符号制限なし，$i\epsilon S'$

となる．

この場合にも，定理 2.2 と同様な定理が成り立つから，一挙に基底解を求めうる．つぎのステップの単体判定基準は，単体乗数（双対変数）を用いて，

(2.50) $\Delta_i^k = -r_i^k - \sum_{j\epsilon S'} p_{ij}^k v_j + v_i \ , \quad i\epsilon S', \ k\epsilon \boldsymbol{K}_i$

となる．とくに，基底解に対しては，

(2.51) $\Delta_i^B = -r_i^B - \sum_{j\epsilon S'} p_{ij}^B v_j + v_i = 0 \ , \quad i\epsilon S$

となる．また，すべての $i\epsilon S'$, $k\epsilon \boldsymbol{K}_i$ に対し $\Delta_i^k \geqq 0$ ならば，最適政策を得る．一方，1 つでも $\Delta_i^k < 0$ なる対 (i, k) が存在すれば，政策は改善可能である．したがって，この終点のある MDP についても 2.2 で述べた新しいアルゴリズムを適用することができる．

3. セミ・マルコフ決定過程

MDP について展開した議論を連続時間の決定過程，すなわち SMDP あるいはマルコフ再生計画 (Markov Renewal Programming) まで拡張しよう．

まず，セミ・マルコフ過程について簡単に述べる．確率過程 $\{Z_t ; t \geqq 0\}$ を考える．ここで，$Z_t = i$ は時刻 t において状態 i にあることを表わす．また，状態は 2. と同様に $i = 1, 2, \cdots\cdots, m \epsilon S$ で表わされるとする．さて，$Q_{ij}(t) = p_{ij} F_{ij}(t)$ は $[0, \infty]$ で定義された非減少関数で，

(3.1) (i) $Q_{ij}(0) = p_{ij} F_{ij}(0) = 0 \ , \quad i\epsilon S, \ j\epsilon S$

(3.2) (ii) $\sum_{j\epsilon S} Q_{ij}(\infty) = \sum_{j\epsilon S} p_{ij} F_{ij}(\infty) = \sum_{j\epsilon S} p_{ij} = 1 \ , \quad i\epsilon S$

をみたすものである．ここで，p_{ij} は状態 i から状態 j への推移確率であり，系は状態の推移のみに着目したとき，推移確率 p_{ij} にしたがう．そこで，推移確率 p_{ij} を持つマルコフ連鎖は隠れマルコフ連鎖 (Imbedded Markov Chain) と呼ばれる．一方，$F_{ij}(t)$ はつぎの状態が j であると

きの状態 i に留まる時間の分布関数である．とくに，

$$(3.3)\qquad F_{ij}(t)=\begin{cases}0, & 0\leqq t<1,\\ 1, & t\geqq 1,\end{cases}\qquad i\epsilon S,\ j\epsilon S$$

とすれば，離散的マルコフ連鎖となり，一方

$$(3.4)\qquad F_{ij}(t)=1-e^{-\lambda_{ij}t},\quad i\epsilon S,\ j\epsilon S$$

とすれば連続時間マルコフ連鎖となる．初期分布

$$(3.5)\qquad \boldsymbol{a}=(a_1,a_2,\cdots\cdots,a_m)$$

を与えれば，この過程は決定される．そのとき，この過程はセミ・マルコフ過程と呼ばれる．とくに，m次元の再生量

$$(3.6)\qquad \boldsymbol{N}(t)=[N_1(t),\ N_2(t),\cdots\cdots,N_m(t)]$$

を考えるときはマルコフ再生過程 [21] と呼ばれる．

さて，

$$(3.7)\qquad H_i(t)=\sum_{j\epsilon S}Q_{ij}(t),\quad i\epsilon S$$

を定義する．これは，状態 i におけるつぎの状態を考えない無条件の留まる時間の分布関数となる．$F_{ij}(t)$ の平均を

$$(3.8)\qquad b_{ij}=\int_0^\infty tdF_{ij}(t),\quad i\epsilon S,\ j\epsilon S$$

とすれば，無条件分布の平均は

$$(3.9)\qquad \eta_i=\int_0^\infty tdH_i(t)=\sum_{j\epsilon S}p_{ij}\int_0^\infty tdQ_{ij}(t)=\sum_{j\epsilon S}p_{ij}b_{ij},\quad i\epsilon S$$

となる．ここでは，すべての b_{ij} は有限と仮定する．そのとき，$\eta_i(i\epsilon S)$ も明らかに有限となる．また，$F_{ij}(t)$ は時刻0で確率1でもって1になるような無限推移を除いた普通の分布関数とする．

セミ・マルコフ過程の状態の分類は隠れマルコフ連鎖のそれにしたがうとする [13]．

以上の準備のもとで，SMDP を考えよう．状態 $i\epsilon S$ で $k=1,2,\cdots\cdots$，$k_i\epsilon \boldsymbol{K}_i$ の中より1つの決定 k を選ぶものとする．このとき，系は

$$(3.10)\qquad Q_{ij}^k(t)=p_{ij}^kF_{ij}^k(t),\quad j\epsilon S$$

によって支配される．また，同時に単位時間当りの利得を r_i^k とする．すなわち，単位時間状態 i に留まることにより利得 r_i^k を得る．2. と同様にして，状態空間を S とし，政策空間を $F=\boldsymbol{K}_1\times\boldsymbol{K}_2\times\cdots\cdots\times\boldsymbol{K}_m$ としよう．そのとき，任意の決定を $f\epsilon F$ で表わすとする．ここでは，政策は定常政策のみを考える．状態の分類，および割引率を導入することによって，2. と同様に3つの問題を述べてみよう．

3・1 エルゴード・セミ・マルコフ決定過程

隠れマルコフ連鎖が決定の如何にかかわらずエルゴード的であるときは，総期待利得は発散するので，単位時間当りの平均期待利得を最大にする政策およびその値を求めるのが問題である．初期分布 $\boldsymbol{a}$ から出発したときの単位時間当りの平均期待利得は初期分布と独立になる．このとき，つぎの LP 問題を得る（詳細は文献 [20] 参照）．

(3.11) $\mathrm{Max} \sum_{j\in S} \sum_{k\in K_j} \eta_j^k r_j^k y_j^k$

subject to

(3.12) $y_j^k \geqq 0$, $j\in S$, $k\in \boldsymbol{K}_j$

(3.13) $\sum_{k\in K_j} y_j^k - \sum_{i\in S} \sum_{k\in K_i} p_{ij}^k y_i^k = 0$, $j\in S$

(3.14) $\sum_{j\in S} \sum_{k\in K_j} \eta_j^k y_j^k = 1$

となる．さらに，最適政策は

(3.15) $d_j^k = y_j^k / \sum_{k\in K_j} y_j^k$, $j\in S$, $k\in \boldsymbol{K}_j$

で与えられる．ここで，d_j^k は状態 j で決定 k を選ぶ確率である．また，つぎの定理が定理 2.2 と同様にして成り立つ．

[定理 3.1] LP 問題 (3.11)～(3.14) において，最適解の中には各 $j\in S$ に対しただ 1 つの $y_j^k>0$ となり，残りは $y_j^k=0$ となるものが存在する．証明は定理 2.2 とほぼ同様であるので，省略する．

式 (3.14) の $j=m$ に関する冗長な制限式を除いて双対問題を考える．最適値を g_1^S とすれば，双対定理により双対変数を $[v_1, v_2, \cdots, v_{m-1}, g_1^S]$ とおく．そのとき，つぎの双対問題

(3.16) $\mathrm{Min}\ g_1^S$

subject to

(3.17) $v_i + \eta_i^k g_1^S \geqq \eta_i^k r_i^k + \sum_{j=1}^{m-1} p_{ij}^k v_j$, $i=1,2,\cdots,m-1$, $k\in \boldsymbol{K}_i$

(3.18) v_i, g_1^S ; 符号制限なし

を得る．これらの主および双対問題の関係を図 3·1 の Tucker 図表に示す．

したがって，定理 3.1 を用いて，適当な基底解を得るためには，例えば

(3.19) $-r_j^B = \min_{k\in K_j} [-r_j^k]$, $j\in S$

を適用すればよい．そのとき，単体乗数は，前同様にして，

(3.20) $v_i + \eta_i^B g_1^S = \eta_i^B r_i^B + \sum_{j=1}^{m-1} p_{ij}^B v_j$, $i\in S$

	y_1^1	⋯⋯	$y_1^{K_1}$	y_2^1	⋯⋯	$y_2^{K_2}$	⋯⋯	⋯⋯	y_m^1	⋯⋯	$y_m^{K_m}$ ($\geqq 0$)	
v_1	$1-p_{11}^1$	⋯⋯	$1-p_{11}^{K_1}$	$-p_{21}^1$	⋯⋯	$-p_{21}^{K_2}$	⋯⋯	⋯⋯	$-p_{m1}^1$	⋯⋯	$-p_{m1}^{K_m}$	$=0$
v_2	$-p_{12}^1$	⋯⋯	$-p_{12}^{K_1}$	$1-p_{22}^1$	⋯⋯	$1-p_{22}^{K_2}$	⋯⋯	⋯⋯	$-p_{m2}^1$	⋯⋯	$-p_{m2}^{K_m}$	$=0$
⋮	⋮		⋮	⋮		⋮			⋮		⋮	⋮
v_{m-1}	$-p_{1,m-1}^1$	⋯⋯	$-p_{1,m-1}^{K_1}$	$-p_{2,m-1}^1$	⋯⋯	$-p_{2,m-1}^{K_2}$	⋯⋯	⋯⋯	$-p_{m,m-1}^1$	⋯⋯	$-p_{m,m-1}^{K_m}$	$=0$
g_1^S	η_1^1	⋯⋯	$\eta_1^{K_1}$	η_2^1	⋯⋯	$\eta_2^{K_2}$	⋯⋯	⋯⋯	η_m^1	⋯⋯	$\eta_m^{K_m}$	$=1$
	VII		VII	VII		VII			VII		VII	
	$\eta_1^1 r_1^1$	⋯⋯	$\eta_1^{K_1} r_1^{K_1}$	$\eta_2^1 r_2^1$	⋯⋯	$\eta_2^{K_2} r_2^{K_2}$	⋯⋯	⋯⋯	$\eta_m^1 r_m^1$	⋯⋯	$\eta_m^{K_m} r_m^{K_m}$	

図 3・1　エルゴード SMDP に対する Tucker 図表

を解くことによって得られる．ただし，$v_m=0$ とする．これは，Jewell [12] あるいは Howard [11] の VDO に相当する．また，つぎのステップの単体判定基準は，単体乗数を用いて，

$$\Delta_i^k=-\eta_i^k r_i^k-\sum_{j=1}^{m-1}p_{ij}^k v_j+v_i+\eta_i^k g_1^S\ ,\quad i\epsilon S,\ k\epsilon \boldsymbol{K}_i \tag{3.21}$$

となる．とくに，基底解に対しては

$$\Delta_i^B=-\eta_i^B r_i^B-\sum_{j=1}^{m-1}p_{ij}^B v_j+v_i+\eta_i^B g_1^S=0\ ,\quad i\epsilon S \tag{3.22}$$

となる．すべての $i\epsilon S$，$k\epsilon \boldsymbol{K}_i$ について，$\Delta_i^k\geqq 0$，あるいは式 (3.22) を用いて，

$$r_i^k+\frac{1}{\eta_i^k}\left[\sum_{j=1}^{m-1}p_{ij}^k v_j-v_i\right]\leqq r_i^B+\frac{1}{\eta_i^B}\left[\sum_{j=1}^{m-1}p_{ij}^B v_j-v_i\right]\ ,\quad i\epsilon S,\ k\epsilon \boldsymbol{K}_i \tag{3.23}$$

ならば，最適解（最適政策）である．一方，もし $\Delta_i^k<0$ なる対 (i, k) が存在すれば，あるいは，式 (3.22) を用いて

$$r_i^k+\frac{1}{\eta_i^k}\left[\sum_{j=1}^{m-1}p_{ij}^k v_j-v_i\right]>r_i^B+\frac{1}{\eta_i^B}\left[\sum_{j=1}^{m-1}p_{ij}^B v_j-v_i\right] \tag{3.24}$$

ならば，政策は改善可能である．

これらの事実を用いれば，PI および LP アルゴリズムを併用した新しいアルゴリズム（2・2 参照）が直ちに適用できる．

3・2 割引率を持ったセミ・マルコフ決定過程

2・3 で用いた割引率 β のかわりに，連続時間の過程に対しては，指数型の割引率 $\alpha(\alpha>0)$ を用いる．すなわち，ある時刻で 1 単位の利得は時間 t を経たのちには，$e^{-\alpha t}$ となる．また，$[0, t]$ 間の利得 r_i は

$$\int_0^t r_i e^{-\alpha\tau}d\tau=\frac{r_i}{\alpha}[1-e^{-\alpha t}] \tag{3.25}$$

となる．この場合には総期待利得が収束するので，初期分布 $\boldsymbol{a}$ より出発したときの総期待利得を最大にする問題は，つぎの LP 問題となる．

$$\text{Max}\sum_{j\epsilon S}\sum_{k\epsilon K_j}\rho_j^k(\alpha)x_j^k \tag{3.26}$$

subject to

$$x_j^k\geqq 0\ ,\quad j\epsilon S,\ k\epsilon \boldsymbol{K}_j \tag{3.27}$$

$$\sum_{j\epsilon S}\sum_{k\epsilon K_j}(\delta_{jl}-q_{jl}^k(\alpha))x_j^k=a_l,\quad l\epsilon S \tag{3.28}$$

ここで，

$$h_i^k(\alpha)=\int_0^\infty e^{-\alpha t}dH_i^k(t)\ ,\quad i\epsilon S,\ k\epsilon \boldsymbol{K}_i \tag{3.29}$$

$$q_{ij}^k(\alpha)=\int_0^\infty e^{-\alpha t}dQ_{ij}^k(t)\ ,\quad i\epsilon S,\ j\epsilon S,\ k\epsilon \boldsymbol{K}_i \tag{3.30}$$

$$\rho_i^k(\alpha)=\frac{r_i{}^k}{\alpha}[1-h_i^k(\alpha)]\ ,\quad i\epsilon S,\ k\epsilon \boldsymbol{K}_i \tag{3.31}$$

である．また，$h_i^k(\alpha)=\sum_{j\epsilon S}q_{ij}^k(\alpha)$ でもある．

この問題の双対問題は，

(3.32) $\text{Min} \sum_{i\epsilon S} a_i v_i$

subject to

(3.33) $v_i \geqq p_i^k(\alpha) + \sum_{j\epsilon S} q_{ij}^k(\alpha) v_j \ , \quad i\epsilon S, \ k\epsilon \boldsymbol{K}_i$

(3.34) v_i；符号制限なし，$i\epsilon S$

となる．ここで，双対変数 $v_1, v_2, \cdots\cdots, v_m$ は同時に主問題の単体乗数でもある．したがって，2・3と同時に PI アルゴリズムの対応が言え，また新しいアルゴリズムも適用できる．

3・3 終点のあるセミ・マルコフ決定過程

最後に終点のある過程について考えてみよう．この場合には，隠れマルコフ連鎖に対して，2・4と同じ終点仮定が成り立つとする．

LP 問題はつぎのようになる．

(3.35) $\text{Max} \sum_{j\epsilon S'} \sum_{k\epsilon K_j} \eta_j^k r_j^k x_j^k$

subject to

(3.36) $\sum_{j\epsilon S'} \sum_{k\epsilon K_j} (\delta_{jl} - p_{jl}^k) x_j^k = a_l \ , \quad l\epsilon S'$

(3.37) $x_j^k \geqq 0 \ , \quad j\epsilon S', \ k\epsilon \boldsymbol{K}_j$

ここで，集合 $S' = \{2, 3, \cdots\cdots, m\}$ とする．したがって，MDP の場合の利得 r_i^k のかわりに $r_i^k \eta_i^k$ を考えれば，あとはすべて同じ議論ができる．

SMDP の詳細については文献 [20] を参照されたい．

4. 結　　論

以上述べたように，MDP および SMDP はいずれも LP 問題に定式化され，数理計画の立場から言えば，この PI アルゴリズムは逆行列型改訂単体法の1つの変形である．すなわち，単体乗数（双対変数）を求めて，つぎのステップの単体判定基準を作り，各 i について Δ_i^k の最小値が負となるすべての対 (i, k) について基底変数 x_i^k を入れ替えるということになる．ただし，通常の LP とは異なり，あらためて単体乗数を求めている．

われわれは，これらの問題を改訂単体法として解き，単体判定基準は PI アルゴリズムと同じ方法を用いるアルゴリズムを開発した．この新しいアルゴリズムが非常に有効であることは 2・2 に述べた通りである．

また，同時に多くの基底を入れ替えても，総期待利得あるいは平均期待利得が増加することは，Howard [10]，Blackwell [3] 等の結果から保証される．PI アルゴリズムを可能にするのは，この過程の持っている特別な性質，たとえば定理 2.2 および 3.1 と双対変数の意味で，この双対変数さえ求めれば，この系を表わす量がすべて求まるということである．政策については，MDP の場合には最適な定常政策が存在するという性質がLPで定式化する場合は非常に役に立つ．SMDP については，定常政策のみに限って議論を進めたが，非定常な場合まで拡張しても，同じ結

果が期待される.

ここでは，割引率を考慮しない場合には，エルゴード・マルコフ連鎖，吸収マルコフ連鎖に分けて議論したが，もっと一般の場合にも拡張できる.

最後に，SMDP において，すべての分布関数 $F_{ij}^k(t)$ が単位時間で退化すればすなわち，$F_{ij}^k(t)=0(0\leqq t<1)$，$F_{ij}^k(t)=1(t\geqq 1)$ ならば，すべての $\eta_i^k=1$ となり，$q_{ij}^k(\alpha)=e^{-\alpha}p_{ij}^k$ となるから，$e^{-\alpha}=\beta$ とおくことにより，MDP の場合に帰着されることを注意しておく.

謝　　辞

最後に，日頃御指導頂きます京都大学工学部三根　久教授に厚く感謝します.

参 考 文 献

[1] Åström, K. J., "Optimal Control of Markov Processes with Imcomplete state Information," *J. Math. Anal. Appl.* 10 (1965), 174-205.
[2] Bellman, R., "A Markovian Desision Process," *J. Math. Mech.*, 6 (1957), 679-684.
[3] Blackwell, D., "Discrete Dynamic Programming," *Ann. Math. Stat.*, 33 (1962), 719-726.
[4] De Ghellinck, G. T. and G. D. Eppen, "Linear Programming Solutions for Separable Markovian Decision Problems," *Management Science*, 15 (1967), 371-394.
[5] D'Epenoux, F., "A Probabilistic Production and Inventory Problem," *Management Sci.* 10(1963), 98-108.
[6] Derman, C., "On Sequential Decisions and Markov Chains," *Management Sci.*, 9 (1962), 16-24.
[7] ——, "Optimal Replacement and Maintenance under Markovian Deterioration with Probability Bounds on Failure," *Management Sci.*, 9 (1963), 478-481.
[8] ——, "On Optimal Replacement Rules When Changes of State Are Markovian," *Mathematical Optimization Techniques*, edited by R. Bellman, University of California Press, Berkeley and Los Angeles, 1963, 201-210.
[9] Eaton, J. H. and L. A. Zaden, "Optimal Pursuit Strategies in Discrete-State Probabilistic Systems," *J. Basic Engineering*, 84 (1962), 23-29.
[10] Howard, R. A., *Dynamic Programming and Markov Processes*, The M. I. T. Press, Cambridge, Massachusetts, 1960.
[11] ——, "Research in Semi-Markovian Decision Structures," *J. Opns. Res. Soc. Jap.*, 6 (1964), 163-199.
[12] Jewell, W. S., "Markov-Renewal Programming. I, II," *Opns. Res.*, 11 (1963), 938-948, 949-971.
[13] Kemeny, J. G. and J. L. Snell, *Finite Markov Chains*, D. Van Nostrand, Princeton, New Jersey, 1960.
[14] Klein, M., "Inspection-Maintenenance-Replacement Schedules under Markovian Deterioration," *Management Sci.*, 9 (1962), 25-32.
[15] ——, "Markovian Decision Models for Reject Allowance Problem," *Management Sci.*, 12 (1966), 349-358.
[16] Lave, Jr., R. E., "A Markov Decision Process for Economic Quality Control," *IEEE Trans. on System Science and Cybernetict*, **SSC**-2 (1966), 45-54.
[17] Manne, A. S., "Linear Programming and Sequential Decisions," *Managment Sci.*, 6 (1960), 259-267.
[18] 三根久，尾崎俊治，"A Relation between Linear and Dynamic Programming in Markovian Decision Problems," 日本オペレーションズ・リサーチ学会秋季研究発表会アブストラクト集，1967年11月，35-36.
[19] 小野勝章，計算を中心とした線型計画法，日科技連，1967.
[20] Osaki, S. and H. Mine, "Linear Programming Algorithms for Semi-Markovian Decision Processes," *J. Math. Anal. Appl.*, 22 (1968), 356-381.
[21] Pyke, R., "Markov Renewal Processes with Finitely Many States," *Ann. Math. Stat.*, 32 (1961),

1234-1253.

[22] Taylor, III, H. M., "Markovian Sequential Replacement Processes," *Ann. Math. Stat.*, 36 (1965), 1677-1694.

[23] White, L. S., "Markovian Decision Models for the Evaluation of a Large Class of Continuous Sampling Inspection Plans," *Ann. Math. Stat.*, 36 (1965), 1408-1420.

[24] ——, "Bayes Markovian Decision Models for a Multiperiod Reject Allowance Problems," *Opns. Res.*, 15 (1967), 857-865.

[25] Veinott, Jr., A. F., "On The Finding Optimal Policies in Discrete Dynamic Programming with No Discounting," *Ann. Math. Stat.*, 37 (1966), 1284-1294.

[26] Wolfe, P. and G. B. Dantzig, "Linear Programming in a Markov Chain," *Opns. Res.*, 10 (1962), 371-394.