Statistics, Probability and Game Theory
IMS Lecture Notes – Monograph Series (1996) Volume 30

HAMILTONIAN CYCLE PROBLEM AND SINGULARLY PERTURBED MARKOV DECISION PROCESS[1]

JERZY A. FILAR AND KE LIU[2]
University of South Australia

Abstract

In 1962 Blackwell derived the partial Laurent's series expansion of the discounted reward Markov decision process. In this paper we establish a connection between Blackwell's expansion and a famous problem in combinatorial optimization and operations research known as the Hamiltonian cycle problem. Our results are obtained via an embedding of this combinatorial optimization problem in a suitably perturbed Markov decision process. It follows that all Hamiltonian cycles of a directed graph are the minimizers of a simple function of the first two coefficients of Blackwell's expansion.

1 Introduction

In 1962, in a fundamental paper, Blackwell [1] introduced or formalized many of the techniques that have become building blocks for the subject of Markov decision processes (MDP's for short). Among various interesting results contained in that paper was the partial Laurent's series expansion of the discounted reward. This was later completed by Miller and Veinott [17] and Veinott [18] and has led to many further developments.

In this paper we establish a connection between Blackwell's expansion and a famous problem in combinatorial optimization and operations research known as the *Hamiltonian Cycle Problem* (HCP for short). Our results are obtained via an embedding of this combinatorial optimization problem in a suitably perturbed MDP. To the extent that our perturbation alters the ergodic structure of the underlying Markov chains it is, indeed, a singular perturbation in the sense of Abbad and Filar [12]. The result presented here can be viewed as continuation of the approaches introduced in Filar and Krass [8] and Chen and Filar [13]. The main difference is that in [8] and [13] the properties of only the first term of Blackwell's expansion are utilized, whereas in the present paper both the first and second terms are used to

[1] *Key words:* Markov Decision Processes, Hamiltonian Cycle, Singular perturbation, Optimal policy.

This research was supported partially by Australia Research Council Grant ♯A49532206 and by National Natural Science Foundation of China. We are indebted to Professors Eric Denardo and Ulrich Rieder for discussing this problem with us and their many helpful suggestions.

[2] On temporary leave from Institute of Applied Mathematics, Academia Sinica

derive a novel characterization of a Hamiltonian cycle (HC, for short) an arbitrary directed graph.

The paper is organized as follows: In the second section, we give preliminaries, notations and some results from [8] and [13]. In the third section, we give a statement of our main result and the idea behind its proof. In the fourth section, we give the details of the proof. In the Appendix we illustrate the main line of argument with a numerical example.

2 Preliminaries, Notation and Review

We shall now consider the Hamiltonian cycle problem. It would be impractical to supply a complete bibliography of works on this problem, instead we refer the reader to the book of Papadimitriou and Steiglitz [16].

In graph theoretic terms, the problem is to find a simple cycle of N arcs, that is, a Hamiltonian cycle in a directed graph G with N nodes and with arcs (i, j), or determine that none exist.

Consider a moving object tracing out a directed path on the graph G with its movement "controlled" by a function f mapping the nodes N into the arcs A. This function induces a "zero-one" $N \times N$ Markov matrix $P(f)$ whose positive entries correspond to the arcs "selected" by f at the respective nodes. Suppose further, that this motion continues forever, and we regard $P(f)$ as a Markov Chain, and consider its "stationary distribution", contained in its *limit Cesaro-sum matrix*:

$$P^*(f) = \lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} P^t(f), \qquad \text{where } P^0(f) = I_N. \tag{1}$$

In [8] and [13] the relationship between the ergodic class/transient state structures of such Markov Chains, and the possible cycles in the graph were studied.

In order to make the above statements precise we now formally introduce a finite state/action MDP as a four-tuple $\Gamma = \{S, A, r, p\}$ where $S = \{1, 2, \cdots, N\}$ is the set of *states*, $A = \cup A(i)$ with $A(i) = \{1, 2, \cdots, m_i\}$ denoting the set of *actions* available in state i for each $i \in S$, $r = \{r(i, a) | a \in A(i), i \in S\}$ denotes the set of possible (immediate) rewards and $p = \{p(j|i, a) \ | a \in A(i),\ i, j \in S\}$ is the set of (one-step) transition probabilities. A *stationary policy* π in Γ is a set of N probability vectors $\pi(i) = (\pi(i, 1)$, $\pi(i, 2), \cdots, \pi(i, m_i))$, where $\pi(i, k)$ denotes the probability of choosing action k in state i whenever i is visited. We denote the set of all stationary policies by $C(S)$. A *deterministic policy* f is simply a stationary policy such that a single action is selected with probability 1 in every state; and write $f(i) = k$ for $i \in S$. We denote the set of all deterministic policies by $C(D)$ and its cardinality $m := \prod_{i=1}^{N} m_i$.

2.1 Embedding of a Directed Graph in an MDP

Now consider a directed weighted graph G with the vertex set $V=\{1,\cdots,N\}$, the arc set $\mathcal{A}$ and with weights c_{ij} associated with the arcs (i,j). Let $G_1 \subseteq G$ be a directed subgraph of G with the same vertex set V, the arc set $\mathcal{A}_1 = \{(i,j) \mid \text{for every } i \in V, \text{ there is only one } j \in V \text{ such that } (i,j) \in \mathcal{A}\}$ and with weights c_{ij} when $(i,j) \in \mathcal{A}_1$. That means for any directed subgraph G_1 only one arc emanates from each vertex of G_1. The first MDP which we shall associate with G will be the process $\Gamma = \{S, A, r, p\}$, $S = \{1, 2, \cdots, N\}$= the set of vertices of G, $A(i) = \{j \in S \mid (i,j) \in \mathcal{A}\}$ for each $i \in S$ and $A = \bigcup_{i=1}^{N} A(i)$, $r = \{r(i,j) = -c_{ij} \mid j \in A(i), i \in S\}$, and $p = \{p(j \mid i,a) \mid a \in A(i), i,j \in S\}$ with $p(j \mid i,a) = \delta_{aj}$, the Kronecker delta. Also, we assume that 1 is the initial state. We shall say that a deterministic policy f in Γ is a HC in G if the subgraph G_1 with the set of arcs $\{(1,f(1)),(2,f(2)),\cdots,(N,f(N))\}$ is a Hamiltonian cycle in G. If the subgraph G_1 contains cycles of length less than N, we say that f has *sub-cycle* in G. If the subgraph G_1 contains a cycle of length k, we say that f has a $k-$sub-cycle.

The above can be illustrated on a completed graph (without self-loops) on four nodes. For instance a policy f such that $f(1) = 2$, $f(2) = 1$, $f(3) = 4$ and $f(4) = 3$ induces a subgraph $G_f = \{(1,2),(2,1),(3,4),(4,3)\}$ which contains two 2-sub-cycles. Observe that f also induces a Markov chain with the probability transition matrix

$$P(f) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

which has two ergodic classes corresponding to the sub-cycles of G_f.

Move generally, any stationary policy $\pi \in C(S)$ induces a probability transition matrix

$$P(\pi) = (p(j|i,\pi))_{i,j}^{N,N},$$

where for all $i,j \in S$

$$p(j|i,\pi) = \sum_{a\in A(i)} p(j|i,a)\pi(i,a).$$

If, for every $\pi \in C(S)$, the Markov chain given by $P(\pi)$ contains only a single ergodic class (plus, a possibly empty, set of transient states), then the MDP Γ is called *unichain*. For a variety of technical reasons unichain MDP's are simpler to analyze. We have seen from the above example that the direct embedding of G in Γ induces a multichain ergodic structure. This and some other, technical difficulties would vanish if Γ were a unichain MDP.

In view of the above in [8] and [13] the law of motion of Γ was perturbed to $p(\epsilon) := \{p(j \mid i,a)(\epsilon) \mid (i,a,j) \in S \times A(i) \times S\}$ where for any $\epsilon \in (0,1)$ we define

$$p(j \mid i,a)(\epsilon) = \begin{cases} 1 & \text{if } i = 1 \text{ and } a = j, \\ 0 & \text{if } i = 1 \text{ and } a \neq j, \\ 1 & \text{if } i > 1 \text{ and } a = j = 1, \\ \epsilon & \text{if } i > 1,\ a \neq j, \text{ and } j = 1, \\ 1-\epsilon & \text{if } i > 1,\ a = j, \text{ and } j > 1, \\ 0 & \text{if } i > 1,\ a \neq j, \text{ and } j > 1. \end{cases} \tag{2}$$

Note that 1 denotes the "home" node. For each pair of nodes i, j (not equal to 1) corresponding to a (deterministic) arc (i,j), our perturbation replaces that arc by a pair of "stochastic arcs " $(i,\ 1)$ and (i,j) (see Figure 1) with

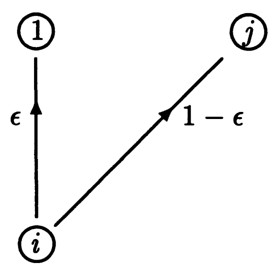

Figure 1: Perturbation of deterministic action

weights ϵ and $1-\epsilon$ respectively ($\epsilon \in (0,1)$). This stochastic perturbation has the interpretation that a decision to move along arc (i,j) results in movement along (i,j) only with probability of $(1-\epsilon)$, and with probability ϵ it results in a return to the home node 1.

Note also that the ϵ-perturbed process $\Gamma(\epsilon) = \{S, A, r, p(\epsilon)\}$ clearly "tends" to Γ as $\epsilon \to 0$. This process has the following properties, that can be found in [8].

Lemma 2.1 *(i) The MDP $\Gamma(\epsilon)$ is unichain.*
(ii) Consider the Markov Chain induced by a stationary policy π in $\Gamma(\epsilon)$ and let $S_1 \subseteq S$ be the ergodic class in that chain. Then $1 \in S_1$.

We shall now derive a useful partition of the class $C(D)$ of deterministic policies that is based on the graphs they "trace out" in G. In particular, note that with each $f \in C(D)$ we can associate a subgraph G_f of G defined by

$$\text{arc } (i,j) \in G_f \iff f(i) = j.$$

We shall also denote a simple cycle of length m and beginning at 1 by a set of arcs

$$c_m^1 = \{(i_1 = 1, i_2), (i_2, i_3), \ldots, (i_m, i_{m+1} = 1)\}\,; \quad m = 2, 3, \cdots, N \tag{3}$$

Of course, c_N^1 is a HC. If G_f contains a cycle c_m^1 we write $G_f \supset c_m^1$. Let $C_m^1 := \{f \in C(D) | G_f \supset c_m^1\}$, namely, the set of deterministic policies that trace out a simple cycle of length m, beginning at 1, for each $m = 2, 3, \ldots, N$. Of course, C_N^1 is the set of policies that correspond to HC's and any single C_m^1 can be empty, depending on the structure of the original graph G. Thus a typical policy $f \in C_3^1$ traces out a graph G_f in G that might look as follows:

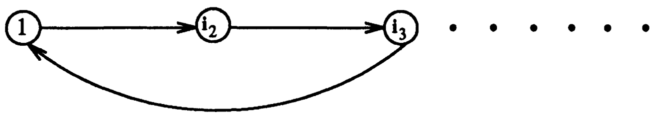

Figure 2: A cycle in C_3^1

where the dots indicate the "immaterial" remainder of G_f that corresponds to states that are transient in $P(f)$, as a result of the perturbation.

The partition of the deterministic policies that seems to be most relevant for our purposes is

$$C(D) \;=\; \left[\bigcup_{m=2}^{N} C_m^1\right] \bigcup B, \tag{4}$$

where B contains[3] all the deterministic policies that are not in any of the C_m^1's. Note that a typical policy f in B traces out a graph G_f in G that might look as follows:

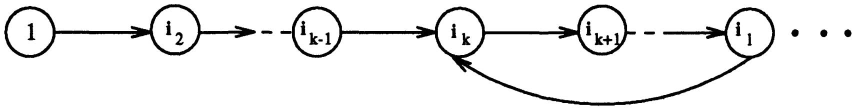

Figure 3: For the f in the B

where the dots again denote the immaterial part of G_f. However, it is important to note that for any $\epsilon > 0$, the states $1, i_2, \ldots, i_{k-1}$ are not transient in $\Gamma(\epsilon)$.

[3]It will soon be seen that the policies in B are in a certain sense "bad" or, more precisely, difficult to analyze thereby motivating the symbol B.

For any policy $\pi \in C(S)$, initial distribution α, $j \in S$ and $a \in A(j)$, define

$$x^T_{ja}(\pi) := \frac{1}{T+1}\sum_{t=0}^{T}\sum_{i=1}^{N}\alpha(i)P_\pi(S_t = j, A_t = a \mid S_0 = i). \tag{5}$$

Further, let $\mathbf{x}(\pi)$ denote the limit point of the vectors $\{\mathbf{x}^T(\pi) \mid T = 0, 1, 2, \cdots\}$, where $\mathbf{x}^T(\pi)$ is an m-dimensional vector with entries given by (5). The entries $x_{ja}(\pi)$ of $\mathbf{x}(\pi)$ can be interpreted as the *long-run expected state-action frequencies* induced by π. Similarly, the long-run expected frequencies of visits to any state $j \in S$ under π are given by

$$x_j(\pi) := \sum_{a\in A(j)} x_{ja}(\pi). \tag{6}$$

We shall denote by X the set of all frequency vectors $\mathbf{x}(\pi)$ and refer to it as the *frequency space.* Also we consider the map $\hat{T} : \pi \to \mathbf{x}(\pi)$ is defined by:

$$x_{ia}(\pi) = p^*_\pi(i)\pi(i,a), \qquad a \in A(i),\ i \in S. \tag{7}$$

In the above, $p^*_\pi(i)$ is the i-th entry of the unique invariant probability vector (stationary distribution vector) of $P(\pi)$. The transformation $\hat{T}$ has been studied by a number of authors (e.g. Derman [6], Denardo [2], Kallenberg [15]).

We define,

$$d_m(\epsilon) := 1 + \sum_{i=2}^{m}(1-\epsilon)^{i-2}, \tag{8}$$

where $m = 2, 3, \cdots, N$. The following lemmata can be found in [13] and [8].

Lemma 2.2 *Let $\epsilon \in (0,1), f \in C(D)$ and $\mathbf{x}(f)$ be its long-run frequency vector. The frequency of visits to state 1 is given by*

$$x_1(f) = \sum_{a\in A(1)} x_{1a}(f) = \begin{cases} \frac{1}{d_m(\epsilon)}, & \text{if } G_f \supset c^1_m, \\ \frac{\epsilon}{1+\epsilon}, & \text{if } f \in B. \end{cases} \tag{9}$$

Lemma 2.3 *Suppose G has a HC. Let f be a deterministic policy in $\Gamma(\epsilon)$ (and thereby also in Γ) which is a HC in G, and assume that i is the k-th node of this HC (starting at 1). Now, if $\mathbf{x}(f) = \hat{T}(f)$, then*

$$x_{ia}(f) = \begin{cases} \frac{(1-\epsilon)^{k-2}}{d_N(\epsilon)}, & \text{if } k > 1 \text{ and } a = f(i), \\ \frac{1}{d_N(\epsilon)}, & \text{if } k = 1 \text{ and } a = f(i), \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Now, consider a linear function on the frequency space X defined by

$$l(\mathbf{x}) = \sum_{a \in A(1)} x_{1a}.$$

If follows from Lemma 2.2 and Lemma 2.3 (also see Chen and Filar [13]) that $l(\mathbf{x}(f))$ can take only N possible values as f ranges over $C(D)$. In particular, if $f \in C_N^1$, then it is a HC and

$$l(\mathbf{x}(f)) = \frac{1}{d_N(\epsilon)}. \tag{11}$$

Unfortunately, a HC does not minimize $l(\mathbf{x}(f))$ over $C(D)$ because if $g \in B$ then

$$l(\mathbf{x}(g)) = \frac{\epsilon}{1+\epsilon} < \frac{1}{d_N(\epsilon)} < \frac{1}{d_m(\epsilon)} \tag{12}$$

for all $\epsilon \in (0,1)$ and $m = 2, 3, \cdots, N-1$.

In the search for a HC it would be useful to have a function that achieves its minimum at any HC, that is, whenever $f \in C_N^1$.

We shall derive such a function with the help of the Blackwell's expansion of the main part of the discounted payoff criterion, namely $[I - \beta P(f)]^{-1}$, where $\beta \in (0,1)$. In particular, Blackwell [1] has shown that

$$[I - \beta P(f)]^{-1} = \frac{1}{1-\beta} P^*(f) + \mathcal{D}(f) + O(1-\beta), \tag{13}$$

where $P^*(f)$ is as in (1), $O(1-\beta)$ tends to zero as fast as $(1-\beta)$ when $\beta \to 1^-$ and

$$\mathcal{D}(f) = [I - P(f) + P^*(f)]^{-1} - P^*(f).$$

Arguably, in deriving (11) and (12) with the help of (7) we used only the first term of (13). In the main result proved in sequel we show that by using both the first and second terms of (13) we can construct a simple functional that is minimized (over $C(D)$) only by a Hamiltonian cycle, provided that the perturbation parameter ϵ is sufficiently small.

3 Main Results

As the discussion in Section 2 indicates, if we use x_1 as the criterion, a HC (if any) lies on the second lowest level of $l(\mathbf{x})$ (see (12)). Does there exist a simple function such that a HC (if any) lies on the lowest level of that function?

Consider a new function $L(\mathbf{x}(f),\mathbf{y}(f))$ ($L(f)$ for short) defined on $C(D)$ by:

$$L(f) = L(\mathbf{x}(f),\mathbf{y}(f)) = x_1(f) + y_1(f) = \sum_{a\in A(1)} [x_{1a}(f) + y_{1a}(f)], \tag{14}$$

where $\mathbf{x}(f) \in X(\epsilon)$ and $\mathbf{y}(f)$ is:

$$\mathbf{y}(f) \;=\; [I - P(f,\epsilon) + P^*(f,\epsilon)]^{-1}\mathbf{r}(f) - P^*(f,\epsilon)\mathbf{r}(f) \tag{15}$$

and where $\mathbf{r}(f)$ is the vector $(1,0,\cdots,0)^T$. Of course, $y_1(f)$ is the first entry of vector $\mathbf{y}(f)$, and the matrix $[I - P(f,\epsilon) + P^*(f,\epsilon)]$ is known to be an invertible matrix (see Blackwell [1]). Note that $P(f)$, $P^*(f)$, $\mathbf{x}(f)$ and $\mathbf{y}(f)$ all depend on ϵ, but for notational simplicity this dependence is suppressed except in Corollary 3.1 and Lemma 3.6. Of course, $x_1(f)$ is also the expected long-run average reward, starting in state 1, under policy f.

From the observation of some numerical examples, we have conjectured that the following is a theorem.

Theorem 3.1 *Let $f^* \in C(D)$ be a HC. Then for ϵ nonnegative and sufficiently small*

$$\begin{aligned} x_1(f^*) + y_1(f^*) &= [I - P(f^*) + P^*(f^*)]^{-1}(1,1) && (16)\\ &= \min_{f\in C(D)} \{x_1(f) + y_1(f)\}, && (17) \end{aligned}$$

where $[H]^{-1}(1,1)$ is the (1,1)-entry of the inverse of matrix [H], and the dependence on ϵ is suppressed. □

In order to prove the above theorem, we shall need a number of properties of the objective function $L(f)$ over the partition of $C(D)$ given in (4). These will be stated in sequel as lemmata and corollaries with proofs of the most technical ones postponed till Section 4.

Lemma 3.1 *For the subgraph induced by $f = (2,3,\cdots,N,1)$, the matrix $P(f)$ is defined by:*

$$P(f) = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \epsilon & 0 & 1-\epsilon & 0 & \cdots & 0 & 0 \\ \epsilon & 0 & 0 & 1-\epsilon & \cdots & 0 & 0 \\ \epsilon & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \epsilon & 0 & 0 & 0 & \cdots & 0 & 1-\epsilon \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \tag{18}$$

and

$$L(f) = \frac{N+1}{2N} + O(\epsilon). \tag{19}$$

Corollary 3.1 *For any HC f, we have:*

$$L(f) = \frac{N+1}{2N} + O(\epsilon). \tag{20}$$

Proof: Suppose $f = (i_1, i_2, \cdots, i_N)$ is a HC with the transition matrix $P(f)$. We can reorder the states as $f' = (2, 3, 4, \cdots, N, 1)$ with the transition matrix $P(f')$ by elementary column and row transformations of the transition matrix $P(f)$, so that the entry $(1,1)$ has never been changed and the column and row elementary transformations only change a row's and/or column's positions in the matrix. That is

$$\begin{aligned} P(f') &= Q_1 Q_2 \cdots Q_n P(f) Q_n^{-1} \cdots Q_2^{-1} Q_1^{-1} \\ &= Q P(f) Q^{-1}, \end{aligned} \tag{21}$$

where Q_i is the row elementary transformation matrix $i = 1, 2, \cdots, n$, $Q = Q_1 Q_2 \cdots Q_n$ and $Q^{-1} = Q_n^{-1} \cdots Q_2^{-1} Q_1^{-1}$. In the same case, when $\epsilon > 0$,

$$P(f', \epsilon) = Q P(f, \epsilon) Q^{-1}. \tag{22}$$

By (1) we have:

$$\begin{aligned} P^*(f', \epsilon) &= \lim_{T \to \infty} \frac{1}{T+1} \sum_{t=0}^{T} P^t(f', \epsilon) \\ &= \lim_{T \to \infty} \frac{1}{T+1} \sum_{t=0}^{T} Q P^t(f, \epsilon) Q^{-1} \\ &= Q P^*(f, \epsilon) Q^{-1}. \end{aligned} \tag{23}$$

Because $\Gamma(\epsilon)$ is unichain, $P^*(f', \epsilon)$ and $P^*(f, \epsilon)$ are matrices with identical rows, we have:

$$\begin{aligned} P^*(f', \epsilon) &= Q_1 Q_2 \cdots Q_n P^*(f, \epsilon) Q_n^{-1} \cdots Q_2^{-1} Q_1^{-1} \\ &= P^*(f, \epsilon) Q_n^{-1} \cdots Q_2^{-1} Q_1^{-1}. \end{aligned} \tag{24}$$

By (22) and (23), we have:

$$[I - P(f', \epsilon) + P^*(f', \epsilon)] = [I - Q P(f, \epsilon) Q^{-1} + Q P^*(f, \epsilon) Q^{-1}], \tag{25}$$

and then,

$$[I - P(f,\epsilon) + P^*(f,\epsilon)] = Q^{-1}[I - P(f',\epsilon) + P^*(f',\epsilon)]Q. \tag{26}$$

Hence,

$$[I - P(f,\epsilon) + P^*(f,\epsilon)]^{-1} = Q^{-1}[I - P(f',\epsilon) + P^*(f',\epsilon)]^{-1}Q. \tag{27}$$

Since Q and Q^{-1} do not change the first entry of the matrix we have that: $L(f') = L(f)$. □

Lemma 3.2 *For the subgraph induced by $f = (2,3,\cdots,i,1,\cdots)$, the matrix $P(f)$ is of the form*

$$P(f) = \begin{bmatrix} A & 0 \\ B & C \end{bmatrix}, \tag{28}$$

where the $i \times i$ submatrix A has the form:

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ \epsilon & 0 & 1-\epsilon & \cdots & 0 & 0 \\ \epsilon & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \epsilon & 0 & 0 & \cdots & 0 & 1-\epsilon \\ 1 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{29}$$

Then we have:

$$L(f) = \frac{i+1}{2i} + O(\epsilon). \tag{30}$$

Proof: Similar to the proofs of Lemmata 3.6 and 3.1 that is supplied in Section 4. □

Corollary 3.2 *For any subgraph induced by f which has an i-sub-cycle c_i^1 (see (3)), equation (30) holds.*

Proof: Similar to the proof of Corollary 3.1. □

Lemma 3.3 *(i) If $f \in C(D)$ induces the Markov matrix $P(f)$ of the form*

$$P(f) = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \epsilon & 0 & 1-\epsilon & 0 & \cdots & 0 & 0 \\ \epsilon & 0 & 0 & 1-\epsilon & \cdots & 0 & 0 \\ \epsilon & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \epsilon & 0 & 0 & 0 & \cdots & 0 & 1-\epsilon \\ \epsilon & 1-\epsilon & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \tag{31}$$

then the stationary probability distribution is:

$$(\frac{\epsilon}{1+\epsilon}, \frac{1}{k_N}, \frac{1-\epsilon}{k_N}, \frac{(1-\epsilon)^2}{k_N}, \cdots, \frac{(1-\epsilon)^{N-2}}{k_N}), \tag{32}$$

where

$$k_N = (1+\epsilon)\sum_{i=0}^{N-2}(1-\epsilon)^i. \tag{33}$$

(ii) If the Markov matrix $P(f,\epsilon)$ is of the form

$$P(f,\epsilon) = \begin{array}{c} \begin{array}{cccccccccc} & & & & & i & & & & \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \\ \epsilon & 0 & 1-\epsilon & 0 & \cdots & 0 & \cdots & 0 & 0 \\ \epsilon & 0 & 0 & 1-\epsilon & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \epsilon & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 1-\epsilon \\ \epsilon & 0 & 0 & 0 & \cdots & 1-\epsilon & \cdots & 0 & 0 \end{bmatrix} \end{array}, \tag{34}$$

then its stationary probability distribution is:

$$\begin{aligned}(\frac{\epsilon}{1+\epsilon}, \frac{\epsilon}{1+\epsilon}, \frac{\epsilon(1-\epsilon)}{1+\epsilon}, \cdots, \frac{\epsilon(1-\epsilon)^{i-3}}{1+\epsilon},& \\ \frac{1}{N-i+1}+O(\epsilon), \cdots, \frac{1}{N-i+1}+O(\epsilon)).&\end{aligned} \tag{35}$$

Lemma 3.4 *(i) For the subgraph induced by $f = (2,3,\cdots,N,2)$, the matrix $P(f)$ is defined by (31), and*

$$L(f) = 1 + O(\epsilon). \tag{36}$$

(ii) For the subgraph induced by $f = (2,3,\cdots,N,i)$, the matrix $P(f)$ is defined by (34), and

$$L(f) = 1 + O(\epsilon).$$

Lemma 3.5 *For the subgraph induced by $f = (2,3,\cdots,N-1,2,k)$, where k can be any number from 1 to $N-1$, then we have:*

$$L(f) = 1 + O(\epsilon). \tag{37}$$

Proof: Similar to the proof of part (ii) of Lemma 3.4. □

Lemma 3.6 *For any subgraph induced by $f \in B$, we have:*

$$L(f) = 1 + O(\epsilon). \tag{38}$$

Proof: Without loss of generality suppose $f = (2, 3, \cdots, i, i+1 \cdots, j, i, \cdots)$ (we can change the order except for state 1), where $2 \le i \le j \le N$. If $j = N$ and $i = 2$, by the Lemmata 3.3 and 3.4, the result holds. If $j = N$ and $i > 2$, by the part (ii) of Lemma 3.4, the lemma is true. If $j < N$, then $f(j) = i < j < N$. So the states $j+1, j+2, \cdots, N$ are transient states in the Markov chain induced by the transition matrix $P(f, \epsilon)$. By the theory of Markov chains(see Chung [7]) the matrix $P^*(f, \epsilon)$ will be of the form:

$$P^*(f, \epsilon) = \begin{bmatrix} A & 0 \\ B & 0 \end{bmatrix}$$

where A is a $j \times j$ matrix. So the invertible matrix $H = I - P(f, \epsilon) + P^*(f, \epsilon)$ has the form:

$$H = \begin{bmatrix} A' & 0 \\ B' & C' \end{bmatrix}$$

where A' is a $j \times j$ matrix and C' is a $(N-j) \times (N-j)$ matrix. Note that

$$\begin{aligned} I &= H^{-1}H = HH^{-1} && (39) \\ &= \begin{bmatrix} A' & 0 \\ B' & C' \end{bmatrix} \begin{bmatrix} T_1 & T_2 \\ T_3 & T_4 \end{bmatrix} && (40) \\ &= \begin{bmatrix} A'T_1 & A'T_2 \\ B'T_1 + C'T_3 & B'T_2 + C'T_4 \end{bmatrix} && (41) \\ &= \begin{bmatrix} I_{j\times j} & 0 \\ 0 & I_{(N-j)\times(N-j)} \end{bmatrix}. && (42) \end{aligned}$$

And hence, H^{-1} has the form:

$$H^{-1} = \begin{bmatrix} (A')^{-1} & T_2 \\ T_3 & T_4 \end{bmatrix}. \tag{43}$$

Because we are only interested in the $(1,1)$-entry of H^{-1} it is enough to consider the submatrix A' and its inverse matrix A'^{-1}. Because the matrix A' is induced only by the restricted policy $f' = (2, 3, \cdots, j, i)$ (to the first j states and $j < N$), the result now follows just as in Lemma 3.5. □

Proof of Theorem 3.1:

By Lemma 3.1 and Corollary 3.1, for any f that is an HC we have: $L(f) = \frac{N+1}{2N} + O(\epsilon)$. By Lemma 3.6, if we consider a subgraph for $g \in B$, then we have: $L(f) = \frac{N+1}{2N} + O(\epsilon) < 1 + O(\epsilon) = L(g)$, when ϵ is small enough. Now let us investigate the cases of Lemma 3.2 and Corollary 3.2 which correspond to any $f \in C_m^1$ for $m < N$. It is easy to check that the function $w(i) = \frac{i+1}{2i}$ with positive integer variable is decreasing as i increases. That means:

$$\frac{N+1}{2N} < \frac{N}{2(N-1)} < \cdots < \frac{3}{4}.$$

Hence, Theorem 3.1 holds for ϵ sufficiently small. From the proofs presented in Section 4 it is easy to see that the results are also true when $\epsilon = 0$. □

4 The Remainder of the Proofs

Proof of Lemma 3.1: By Lemma 2.3, the stationary distribution of transition probability matrix $P(f)$, $\mathbf{p}(f)$, is:

$$\begin{aligned} \mathbf{p}(f) &= (p_1, p_2, \cdots, p_N) \\ &= (\frac{1}{d_N}, \frac{1}{d_N}, \frac{1-\epsilon}{d_N}, \frac{(1-\epsilon)^2}{d_N}, \cdots, \frac{(1-\epsilon)^{N-2}}{d_N}), \\ &= \left(\frac{1}{N} + O(\epsilon), \frac{1}{N} + O(\epsilon), \cdots, \frac{1}{N} + O(\epsilon)\right) \end{aligned} \tag{44}$$

where d_N is defined by (8). By Blackwell [1],

$$\begin{aligned} \mathbf{x}(f) &= P^*(f)\mathbf{r}(f) = (x_1, x_2, \cdots, x_N) \\ &= (\frac{1}{d_N}, \frac{1}{d_N}, \cdots, \frac{1}{d_N},) \\ &= \left(\frac{1}{N} + O(\epsilon), \frac{1}{N} + O(\epsilon), \cdots, \frac{1}{N} + O(\epsilon)\right) \end{aligned} \tag{45}$$

and $\mathbf{y}(f) = (y_1, y_2, \cdots, y_N)$ is the unique solution in the following system

$$P^*(f)\mathbf{y}(f) = 0 \tag{46}$$

and

$$\mathbf{r}(f) + P(f)\mathbf{y}(f) = \mathbf{x}(f) + \mathbf{y}(f). \tag{47}$$

From (47) we can get

$$\mathbf{y}(f) - P(f)\mathbf{y}(f) = \mathbf{r}(f) - \mathbf{x}(f). \tag{48}$$

Because $\mathbf{r}(f) = (1, 0, \cdots, 0)^T$ with the help of (44) and (45) we can rewrite (46) and (48) as

$$\sum_{i=1}^{N} p_i y_i = 0 \tag{49}$$

and

$$\begin{array}{cccccccl}
y_1 & -y_2 & & & & & = & 1 - \frac{1}{d_N} \\
-\epsilon y_1 & +y_2 & -(1-\epsilon)y_3 & & & & = & -\frac{1}{d_N} \\
-\epsilon y_1 & & +y_3 & -(1-\epsilon)y_4 & & & = & -\frac{1}{d_N} \\
\vdots & & \ddots & & \ddots & & \vdots & \quad\vdots \\
-\epsilon y_1 & & & & +y_{N-1} & -(1-\epsilon)y_N & = & -\frac{1}{d_N} \\
-y_1 & & & & & +y_N & = & -\frac{1}{d_N}.
\end{array} \tag{50}$$

Using the observations that $\frac{1}{d_N} = \frac{1}{N} + O(\epsilon)$, $\frac{1}{1-\epsilon} = 1 + O(\epsilon)$, and $\frac{1}{d_N(1-\epsilon)^i} = \frac{1}{N} + O(\epsilon)$ for $i = 1, 2, \cdots, N-2$, we can express $y_2, y_3, \cdots, y_N$ as

$$y_2 = y_1 - 1 + \frac{1}{N} + O(\epsilon) \tag{51}$$

$$y_3 = y_1 - 1 + \frac{2}{N} + O(\epsilon) \tag{52}$$

$$y_4 = y_1 - 1 + \frac{3}{N} + O(\epsilon) \tag{53}$$

$$\vdots \quad \vdots \quad \vdots$$

$$y_N = y_1 - 1 + \frac{N-1}{N} + O(\epsilon) \tag{54}$$

Substitute (51), $\cdots$, (54) into (49), then recalling that $\sum_{i=1}^{N} p_i = 1$ and $(a + O(\epsilon))(b + O(\epsilon)) = ab + O(\epsilon)$ we obtain (with the help of (44))

$$\begin{aligned}
0 &= \sum_{i=1}^{N} p_i y_i \\
&= p_1 y_1 + \sum_{i=2}^{N} p_i (y_1 - 1 + \frac{i-1}{N} + O(\epsilon)) \\
&= \sum_{i=1}^{N} p_i y_1 + \sum_{i=2}^{N} \left(\frac{1}{N} + O(\epsilon)\right)\left(-1 + \frac{i-1}{N} + O(\epsilon)\right) \\
&= y_1 + \frac{1}{N} \sum_{i=2}^{N} \left(-1 + \frac{i-1}{N}\right) + O(\epsilon) \\
&= y_1 + \frac{1}{N}\left(\frac{-N+1}{2}\right) + O(\epsilon).
\end{aligned} \tag{55}$$

Now we have

$$y_1 = \frac{N-1}{2N} + O(\epsilon). \tag{56}$$

From (44) and (56) we obtain

$$\begin{aligned} L(f) &= x_1 + y_1 \\ &= \frac{1}{N} + O(\epsilon) + \frac{N-1}{2N} + O(\epsilon) \\ &= \frac{N+1}{2N} + O(\epsilon). \end{aligned} \tag{57}$$

□

Proof of Lemma 3.3: For simplicity, we present only the proof of part (i) of Lemma 3.3. The proof of part (ii) is similar to the proof of part (i). The details can be found in Ke Liu [3].

By definition of the stationary probability distribution (that is, any row of $P^*(f)$), we can derive it as the solution of

$$\mathbf{p}^T P(f) = \mathbf{p}^T \tag{58}$$

and

$$\sum_{i=1}^{N} p_i = 1, \tag{59}$$

where $p_i \geq 0$, $i = 1, 2, \cdots, N$.

By (31), (58) and (59), we have

$$\begin{array}{cccccccc} p_1+ & p_2+ & p_3+ & \cdots & + & p_{N-1}+ & p_N & = & 1 \\ & \epsilon p_2+ & \epsilon p_3+ & \cdots & + & \epsilon p_{N-1}+ & \epsilon p_N & = & p_1 \\ p_1+ & & & & & & (1-\epsilon)p_N & = & p_2 \\ & (1-\epsilon)p_2 & & & & & & = & p_3 \\ & & (1-\epsilon)p_3 & & & & & = & p_4 \\ & & & \ddots & & & & \vdots & \vdots \\ & & & & & (1-\epsilon)p_{N-1} & & = & p_N. \end{array} \tag{60}$$

From the first two equations of (60) we have

$$p_1 = \frac{\epsilon}{1+\epsilon},$$

and from the others we have

$$\begin{array}{cccccccc} p_2+ & p_3+ & \cdots & + & p_{N-1}+ & p_N & = & \frac{1}{1+\epsilon} \\ -p_2+ & & & & & (1-\epsilon)p_N & = & \frac{-\epsilon}{1+\epsilon} \\ (1-\epsilon)p_2 & & & & & & = & p_3 \\ & (1-\epsilon)p_3 & & & & & = & p_4 \\ & & \ddots & & & & \vdots & \vdots \\ & & & & (1-\epsilon)p_{N-1} & & = & p_N. \end{array} \tag{61}$$

From (61) we have

$$\sum_{i=2}^{N} p_i = p_2 + (1-\epsilon)p_2 + (1-\epsilon)^2 p_2 + \cdots + (1-\epsilon)^{N-2} p_2 = \frac{1}{1+\epsilon}. \qquad (62)$$

So

$$\begin{aligned} p_2 &= \tfrac{1}{1+\epsilon}\tfrac{1}{\sum_{i=0}^{N-2}(1-\epsilon)^i} \\ &= \tfrac{1}{k_N}, \end{aligned} \qquad (63)$$

where k_N is defined as in (33), and

$$p_i = \frac{(1-\epsilon)^{i-2}}{k_N} \qquad i = 3, 4, \cdots, N.$$

Thus we obtained (32) and the proof is complete. □

Proof of Lemma 3.4: For simplicity, we prove only part (i) of Lemma 3.4. The proof of part (ii) is quite similar to the proof of part (i). The details can be found in Ke Liu [3].

By Lemma 3.2 the stationary distribution $\mathbf{p}(\mathrm{f})$ of the transition probability matrix $P(f)$ is:

$$\begin{aligned} \mathbf{p}(f) &= (p_1, p_2, \cdots, p_N) \\ &= (\frac{\epsilon}{1+\epsilon}, \frac{1}{k_N}, \frac{1-\epsilon}{k_N}, \cdots, \frac{(1-\epsilon)^{N-2}}{k_N}) \\ &= (\frac{\epsilon}{1+\epsilon}, \frac{1}{N-1} + O(\epsilon), \cdots, \frac{1}{N-1} + O(\epsilon)). \end{aligned}$$

Similarly to the proof of Lemma 3.1 $\mathbf{x}(f) = (\frac{\epsilon}{1+\epsilon}, \frac{\epsilon}{1+\epsilon}, \cdots, \frac{\epsilon}{1+\epsilon})$ and $\mathbf{y}(f) = (y_1, y_2, \cdots, y_N)$ is the unique solution of (46) and (47). For our case, (46) and (47) are

$$\sum_{i=1}^{N} p_i y_i = 0 \qquad (64)$$

and

$$\begin{array}{cccccccc} y_1 & -y_2 & & & & = & 1-\frac{\epsilon}{1+\epsilon} \\ -\epsilon y_1 & +y_2 & -(1-\epsilon)y_3 & & & = & -\frac{\epsilon}{1+\epsilon} \\ \vdots & \vdots & \ddots & & & \vdots & \vdots \\ -\epsilon y_1 & & & +y_{N-1} & -(1-\epsilon)y_N & = & -\frac{\epsilon}{1+\epsilon} \\ -\epsilon y_1 & -(1-\epsilon)y_2 & & & +y_N & = & -\frac{\epsilon}{1+\epsilon} \end{array} \qquad (65)$$

So we have

$$\begin{aligned} y_2 &= y_1 - 1 + \frac{\epsilon}{1+\epsilon} \\ &= y_1 - 1 + O(\epsilon) && (66) \\ y_3 &= \frac{1}{1-\epsilon}[y_2 - \epsilon y_1 + \frac{\epsilon}{1+\epsilon} + O(\epsilon)] \\ &= y_1 - 1 + O(\epsilon) && (67) \\ y_4 &= y_1 - 1 + O(\epsilon) && (68) \\ &\vdots \\ y_N &= y_1 - 1 + O(\epsilon) && (69) \end{aligned}$$

Substitute (66), (67), $\cdots$, (69) in (64) to obtain

$$\begin{aligned} 0 &= p_1 y_1 + \sum_{i=2}^{N} p_i y_i \\ &= p_1 y_1 + \left[\sum_{i=2}^{N} p_i(y_1 - 1 + O(\epsilon))\right] \\ &= \sum_{i=1}^{N} p_i y_1 + \sum_{i=2}^{N} p_i(-1 + O(\epsilon)) \\ &= y_1 - \sum_{i=2}^{N} \frac{(1-\epsilon)^{i-2}}{k_N} + O(\epsilon) \\ &= y_1 - (1 + O(\epsilon)) + O(\epsilon) && (70) \end{aligned}$$

So we have

$$y_1 = 1 + O(\epsilon). \tag{71}$$

Now we obtain

$$x_1 + y_1 = \frac{\epsilon}{1+\epsilon} + 1 + O(\epsilon) = 1 + O(\epsilon). \tag{72}$$

The proof is now complete. □

5 Conclusions

At this stage, it is difficult to comment on the significance of Theorem 3.1. Perhaps, it is only a mathematical curiosity that the first two coefficients of Blackwell's expansion (13) are related to the famous Hamiltonian Cycle

Problem. On the other hand, it is not inconceivable that the inherent difficulty of the HCP can be better understood by explaining the behavior of the fundamental matrices $H(f, \epsilon)$ as $\epsilon \to 0$. After all, the structure of the graph is contained in these matrices and the embedding in the singularly perturbed MDP allows for an analogy to be drawn between sub-cycles and ergodic classes.

We are grateful to the referee for pointing out that Theorem 3.1 is also valid when the perturbation is dropped, that is, in the case $\epsilon = 0$. The idea of using equations (46) and (47) which simplified our previous proof is also due to the referee. Arguably, the ultimate goal of this line of research is to construct an objective criterion that will lend itself to numerical minimisation schemes. If these schemes were to be carried out over the frequency space X it may be that the case $\epsilon > 0$ will be more tractable because of the correspondence between its extreme points and deterministic policies of $\Gamma(\epsilon)$.

REFERENCES

[1] D. Blackwell. Discrete Dynamic Programming. *Ann. Math. Statist.* 33:719–726, (1962).

[2] E.V. Denardo. On Linear Programming in Markov Decision Processes. *Man. Sci.* 16:281–288, (1970).

[3] Ke Liu. *Theory and Applications of Markov Decision Processes and Their Perturbations.* School of Mathematics, University of South Australia, Australia, in preparation.

[4] D. Blackwell. Discounted dynamic programming. *Ann. Math. Statist.*, 36:226–235, (1965).

[5] C. Derman and M. Klein. Surveillance of Multi-component Systems: A Stochastic Traveling Salesman Problem. *NRLQ*, 13:103–111, (1966).

[6] C. Derman. *Finite State Markovian Decision Processes.* Academic Press, New York, (1970).

[7] K. L. Chung. *Markov Chains with Stationary Transition Probabilities.* Spring-Verlag, Berlin, (1960).

[8] J. A. Filar and D. Krass. Hamiltonian Cycles and Markov Chains. *Math. Oper. Res.*, 19:223–237, (1994).

[9] D. Krass. *Contributions to the Theory and Applications of Markov Decision Processes.* The Johns Hopkins University, USA, July, (1989).

[10] M. K. GHOSH AND S. I. MARCUS. Infinite horizon controlled diffusion problems with some nonstandard criteria. *J. Math. Sys. and Contr.*, 1:45–70, (1991).

[11] D. KRASS AND J. A. FILAR AND S. SINHA. A Weighted Markov Decision Process. Dep. of Math., UMBC, Tech. Rep., Baltimore USA, (1987).

[12] M. ABBAD AND J. A. FILAR. Perturbation and Stability Theory for Markov Control Problems. *IEEE Trans. Auto. Control*, 37:1415–1420, (1992).

[13] MING CHEN AND J. A. FILAR. Hamiltonian Cycles, Quadratic Programming, and Ranking of Extreme Points. in *Global Optimization*, C. Floudas and P. Pardalos, eds. Princeton University Press, (1992).

[14] A. HORDIJK AND L. C. M. KALLENBERG. Constrained Undiscounted Stochastic Dynamic Programming. *Math. Oper. Res.*, 9:276–289, (1984).

[15] L. C. M. KALLENBERG. *Linear Programming and Finite Markovian Control Problems.* vol. 148, Mathematical Center Tracts, Amsterdam, (1983).

[16] C. PAPADIMITRIOU AND K. STEIGLITZ. Combinatorial Optimization: Algorithms and Complexity. Prentice Hall, New Jersey, (1982).

[17] B. L. MILLER AND A. F. JR. VEINOTT. Discrete dynamic programming with small interest rate. *Ann. Math. Statist.*, 40:366–370, (1969).

[18] A. F. JR. VEINOTT. Discrete Dynamic Programming with Sensitive Discount Optimality Criteria. *Ann. Math. Statist.*, 40:1635–1660, (1969).

CENTRE FOR INDUSTRIAL AND APPLIED MATHEMATICS
SCHOOL OF MATHEMATICS
UNIVERSITY OF SOUTH AUSTRALIA, THE LEVELS
ADELAIDE, SOUTH AUSTRALIA 5095