Existence of average optimal strategies in Markovian decision problems with strictly unbounded costs

Citation for published version (APA):

Document status and date:
Published: 01/01/1977

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 19. Feb. 2019
Existence of average optimal strategies in
Markovian decision problems with strictly
unbounded costs

by

J. Wijngaard

Memorandum COSOR 77-14

Eindhoven, Juni 1977

The Netherlands
Existence of average optimal strategies in Markovian decision problems with strictly unbounded costs.

J. Wijngaard

1. INTRODUCTION

The existence of average optimal strategies in Markovian decision processes is investigated by Blackwell [1] (finite state space), Derman [2], Ross [4], Hordijk [3] (countable state space), Tijms [7], Wijngaard [8] (general state space). Conditions sufficient for the existence of average optimal strategies consist in general of some recurrence conditions and some continuity and compactness conditions. The easiest recurrence conditions to work with are of course the conditions which guarantee recurrence to one point. (Derman, Ross, Hordijk). One can also use recurrence to a finite subset of the state space. The simultaneous Doeblin condition, introduced by Hordijk [3], is a strong form of this recurrence. The conditions used in [8] include the finiteness of the expected time and costs until the first visit to some subset of the state space.

In a lot of problems with a countable state space the condition of recurrence to a finite subset are not very strong. In inventory problems for instance the costs are high if the inventory level is far from zero. That means that "good" strategies have to bring back the inventory level near to zero. The unboundedness of the costs makes that the recurrence conditions are satisfied for all "good" strategies. In this paper we try to replace the recurrence conditions by an unboundedness condition. To guarantee that the set of "good" strategies is non void we state the existence of at least one strategy for which the average costs exist and are bounded on the state space.
2. PRELIMINARIES

We consider a Markovian decision process on a countable state space $V$. Let $\Omega$ be the set of all stationary strategies. The process under strategy $\alpha \in \Omega$ is described by a Markov chain $P_{\alpha}$. The one-period costs under $\alpha$ are given by the function $c_{\alpha}(.): V \to \mathbb{R}$, $c_{\alpha}(u)$ are the costs starting in $u$. We assume that $c_{\alpha}(u) \geq 0$ for all $u \in V$, $\alpha \in \Omega$. The functions $c_{\alpha}$ are assumed to be strictly unbounded, that means that there is a positive function $h(.)$ on $V$ such that $h(.)$ is unbounded from above on each infinite subset of $V$, $c_{\alpha}(u) \geq h(u)$ for all $u \in V$, $\alpha \in \Omega$.

To guarantee that the problem is interesting we assume the existence of a strategy $\alpha^* \in \Omega$ such that the average costs under $\alpha^*$, $g_{\alpha^*}$, exist and are bounded on $V$, $g_{\alpha^*}(u) = \lim_{n \to \infty} \frac{1}{n} \sum_{\ell=0}^{n-1} (P_{\alpha^*} c_{\alpha^*})(u) \leq b < \infty$.

Define $\Omega' := \{ \alpha \in \Omega | \lim_{n \to \infty} \frac{1}{n} \sum_{\ell=0}^{n-1} (P_{\alpha} c_{\alpha})(u) < b, u \in V \}$

The set $\Omega'$ contains all "good" strategies.

It is clear, by the assumptions on $c_{\alpha}$, that all strategies in $\Omega'$ have to bring back the state of the system to a finite set. This is worked out in lemma 1. First we have to give some definitions and notations.

For each $\alpha \in \Omega$, $P_{\alpha}$ is the embedded (sub-)Markov process of $P_{\alpha}$ on $A$, $Q_{\alpha}$ is the embedded (sub-)Markov process of $P_{\alpha}$ on $A$,

$Q_{\alpha}(u, v) := \sum_{n=0}^{\infty} (P_{\alpha}^n)_{A \setminus A'}(u, v)$, (where $1_{\{v\}}$)

$(T_{\alpha} f)(u) := \sum_{n=0}^{\infty} (P_{\alpha}^n f)(u)$ for each $u \in V$, nonnegative function $f$ on $V$.

Hence $(T_{\alpha} c_{\alpha})(u)$ is equal to the expected costs until the first visit to $A$, starting in $u$.

Lemma 1. For each $\alpha \in \Omega'$ the sub-Markov process $Q_{\alpha}$ is a Markov process $((P_{\alpha}^n)_{A'}(1)(u) + 0$ for $n \to \infty)$. Let $\pi_{\alpha}$ be an invariant probability of $Q_{\alpha}$ for $\alpha \in \Omega'$, then $(T_{\alpha} 1)(u) \leq \frac{2}{\pi_{\alpha}(u)}$ and $(T_{\alpha} c_{\alpha})(u) \leq \frac{2b}{\pi_{\alpha}(u)}$ for all $u \in A$. 


For the proof of this lemma we refer to [9].

Now, let $\alpha \in \Omega'$ and $u \in A$ be such that $\pi_\alpha(u) > 0$ for some invariant probability $\pi_\alpha$ of $Q_\alpha$. Let $Q_{\alpha u} := Q_{\alpha D}$ with $D := A \setminus \{u\}$. Then the sum $\sum_{i=0}^{\infty} (Q_{\alpha u}^i, T_{\alpha}^i)(v)$ is equal to the expected costs until the first visit to $\{u\}$, starting in $v$ and the average costs under strategy $\alpha$, starting in $u$, are equal to

$$g_\alpha(u) = \frac{\sum_{i=0}^{\infty} (Q_{\alpha u}^i, T_{\alpha}^i)(u)}{\sum_{i=0}^{\infty} (Q_{\alpha u}^i, T_{\alpha}^i)(u)}$$

Define for each $u \in A$ the subset $\Omega_u$ of $\Omega'$ by

$$\Omega_u := \{\alpha \in \Omega' \mid Q_\alpha has an invariant probability $\pi_\alpha with \pi_\alpha(u) > 0\}$$

Define for each subset $C \subseteq A$ the set of strategies $\Omega_C$ by $\Omega_C := \bigcup_{u \in C} \Omega_u$

Let $A_r := \{u \in A \mid \Omega_u \neq \emptyset\}$, $g_u := \inf_{\alpha \in \Omega_u} g_\alpha(u)$ for $u \in A_r$ and $g := \min_{u \in A_r} g_u$

Let $C$ be an arbitrary subset of $A$ such that $g_u = g$ for all $u \in C$ and let $C' := V \setminus C$. Consider the set of strategies $\Omega_C$. Let for each $\alpha \in \Omega_C$ the union of the maximal invariant sets of $Q_\alpha$ which have a non void intersection with $C$ be denoted by $E_{\alpha C}$. Define $\tilde{E}_{\alpha C} := \{u \in V \mid Q_\alpha(u, E_{\alpha C}) = 1\}$

Notice that $\tilde{E}_{\alpha C}$ is an invariant subset of $P_\alpha$ and that $E_{\alpha C} \subseteq \tilde{E}_{\alpha C}$.

For all $u \in \tilde{E}_{\alpha C}$ the sum $\sum_{i=0}^{\infty} Q_{\alpha C}^i T_{\alpha}^i (c_\alpha - g)(u)$ exists. This sum is equal to $+\infty$ if it is possible to reach, from $u$, points $v$ with $T_{\alpha}^i (c_\alpha - g)(v) = +\infty$ and the sum is finite if that is not possible.

Define the function $v_{\alpha C}$ on $V$ by

$$v_{\alpha C}(u) = \sum_{i=0}^{\infty} Q_{\alpha C}^i T_{\alpha}^i (c_\alpha - g)(u), u \in \tilde{E}_{\alpha C} and v_{\alpha C}(u) = +\infty, u \not\in \tilde{E}_{\alpha C}$$

Notice that for all $u \in V$

$$v_{\alpha C}(u) = T_{\alpha} (c_\alpha - g)(u) + (Q_{\alpha C} v_{\alpha C})(u) = c_\alpha(u) - g + (P_{\alpha C} v_{\alpha C})(u)$$

In the next lemma it is proved that for all $\alpha_1, \alpha_2 \in \Omega_C$ there is an $\alpha_0 \in \Omega_C$ such that $v_{\alpha_0 C}(u) \leq \min\{v_{\alpha_1 C}(u), v_{\alpha_2 C}(u)\}, u \in V$
Lemma 2. Let $C \subseteq A$ such that $g_u = g$ for all $u \in C$. Let for all $\alpha \in \Omega_C$ the sets $E_\alpha$ and $\bar{E}_\alpha$ and the functions $\nu_\alpha$ be defined as above. The strategy $\alpha_0$ is a combination of $\nu_1$ and $\nu_2$, apply $\nu_1$ on the subset of $E_\alpha \cap \bar{E}_\alpha$ where $\nu_\alpha C < \nu C$ and apply $\nu_2$ on the other points of $V$.

Then $\alpha_0 \in \Omega_C$ and $\nu C(u) \leq \min(\nu_1 C(u), \nu_2 C(u))$, $u \in V$.

Proof. Let $F$ be the subset of $V$ where $\nu_1$ is applied and $G := V \setminus F$. Let $F^*$ and $G^*$ be the intersections of $F$ and $G$ with $V \setminus C$. The sub-Markov process $P_{\alpha C}$ describes the state of the system under strategy $\nu$ until the first visit to $C$. The embedded sub-Markov processes of $P_{\alpha C}$ on $G^*$ and $P_{\alpha C}$ on $F^*$ are denoted by $S_1$ and $S_2$. The process $R := S_1 S_2$ describes the state of the system under strategy $\alpha_0$ on the moments that the set $F^*$ is entered (until the first visit to $C$).

Define $W := \{u \in V | \min(\nu_1 C(u), \nu_2 C(u)) < \infty\}$

If $\nu_1 C(u) < \infty$ the sum $\sum_{n=0}^{\infty} P^n a_1 (c-g)(u)$ converges absolute for each $D \supset C$

and if $\nu_2 C(u) < \infty$ the sum $\sum_{n=0}^{\infty} P^n a_2 (c-g)(u)$ converges absolute for each $E \supset C$

For $n = 1, 2, 3, \ldots$ the functions $\nu_0 a_1$ on $W$ are defined by

$$v_0 a_1 (u) = \sum_{n=0}^{\infty} P^n a_2 (c-g)(u) + (S_2 v_1 a_1)(u), u \in W \cap G^*$$

$$v_0 a_1 (u) = \sum_{n=0}^{\infty} P^n a_1 (c-g)(u) + (S_1 v_0 a_1)(u), u \in W \cap F^*$$

and

$$v_0 a_1 (u) = \sum_{n=0}^{\infty} P^n a_2 (c-g)(u) + (S_2 v_0 a_1)(u), u \in W \cap G^*$$

$$v_0 a_1 (u) = \sum_{n=0}^{\infty} P^n a_1 (c-g)(u) + (S_1 v_0 a_1)(u), u \in W \cap F^*$$

The value $v_a a_1 (u)$ can be interpreted as the costs until absorption in $C$ of a process, starting in $u$, with one period costs $c-a$ where the strategy $\alpha_0$ is applied until the nth entrance in the set $F^*$ and from then on the strategy $\nu C$. 
It is easy to verify that

\[(1) \quad \min(v_{a_1}^C(u), v_{a_2}^C(u)) - v_{a_0 a_1}(u) \geq \sum_{k=1}^n R_k^0(v_{a_2}^C - v_{a_1}^C)(u), \quad u \in W\]

which implies that

\[v_{a_0 a_1}(u) \leq \min(v_{a_1}^C(u), v_{a_2}^C(u)) < \infty \text{ for } u \in W\]

Define \(v(u) := \min(v_{a_1}^C(u), v_{a_2}^C(u))\)

Since \(v_{a_1}^C(u) = T_{a_1}(c_{a_1} - g)(u) + (Q_{a_1}^C v_{a_1}^C)(u)\)

and \(v_{a_2}^C(u) = T_{a_2}(c_{a_2} - g)(u) + (Q_{a_2}^C v_{a_2}^C)(u)\)

we know that \(T_{a_0}^0(c_{a_0} - g)(u) + (Q_{a_0}^C v)(u) \leq v(u)\) (see Strauch [6])

Hence \(T_{a_0}^0(c_{a_0} - g)(u) < \infty\) for \(u \in W\) and therefore also \((T_{a_0}^0 c_{a_0})(u) < \infty\) and \((T_{a_0}^0)(u) < \infty\)

We will prove now that all ergodic sets of \(Q_{a_0}\) which can be reached with positive probability from some point in \(W\) have a non void intersection with \(C\).

Assume the existence of an ergodic set \(E_0\) of \(Q_{a_0}\) with \(E_0 \cap C = \emptyset\) which can be reached with positive probability from some point \(u \in W\).

Let \(E_0^*\) be the set of all points in \(V\) which can be reached with positive probability from some point in \(E_0\). Then \(E_0^* \subset W\). The average costs under \(a_0, g_{a_0}\), exist and are constant on \(E_0^*\) and there is a function \(v_{a_0}\) on \(E_0^*\) such that \(v_{a_0} = c_{a_0} - g_{a_0} + P_{a_0} v_{a_0}\) (see [10]).

Using the definition of \(v_{a_0}\) it is possible to derive that

\[v_{a_0 a_1}(u) = v_{a_0}(u) + R^0(v_{a_1}^C - v_{a_0}^C)(u) + f_n(g_{a_0} - g)(u), \quad u \in E_0^*\]

where \(f_n(g_{a_0} - g) \to +\infty\) for \(n \to \infty\) if \(g_{a_0} > g\).
We have \( v_{a_0}(u) = T_{a_0} (c_{a_0} - g_{a_0})(u) + Q_{a_0} v_{a_0} \leq T_{a_0} (c_{a_0} - g)(u) + (Q_{a_0} c', v)(u) + (Q_{a_0} v_{a_0})(u) - (Q_{a_0} c', v)(u) \leq \)

\[ \leq v(u) + (Q_{a_0} v_{a_0})(u) - (Q_{a_0} c', v)(u) \leq \]

\[ \leq v_{a_1 C}(u) + (Q_{a_0} v_{a_0})(u) - (Q_{a_0} c', v)(u) \]

Hence \( v_{a_1 C}(u) - v_{a_0}(u) \geq (Q_{a_0} c', v)(u) - (Q_{a_0} v_{a_0})(u) \) which implies that \( R_n(v_{a_1 C} - v_{a_0})(u) \) is bounded from below in \( n \) and therefore \( g_{a_0} = g \).

From (1) it follows that \( \sum_{n=1}^{\infty} R_n(v_{a_2 C} - v_{a_1 C})(u) < \infty \) for \( u \in E^*_0 \), that means that starting in \( E^*_0 \) the state of the system stays at last in \( F \cap E^*_0 \) or in \( G \cap E^*_0 \), the process \( R \) is absorbing. This is only possible if \( E^*_0 \subset F \) or \( E^*_0 \subset G \). (see [10] for details). The strategy \( a_0 \) is therefore identical to the strategy \( a_1 \) or identical to the strategy \( a_2 \). This implies that \( E_0 \cap C \) can not be empty which gives a contradiction.

It is proved now that all ergodic sets of \( Q_{a_0} \) which can be reached with positive probability from some point in \( W \) have a non void intersection with \( C \).

Hence \( a_0 \in \Omega_C \). Since \( T_{a_0} (c_{a_0} - g)(u) < \infty \) for \( u \in W \) the function \( v_{a_0 C} \) on \( W \)

defined by \( v_{a_0 C}(u) = \sum_{n=0}^{\infty} Q_{a_0 C} T_{a_0} (c_{a_0} - g)(u) \) is finite on \( W \).

Further \( v_{a_0 a_1}(u) \to v_{a_0 C}(u) \) for \( n \to \infty \) and by (1) \( v_{a_0 C}(u) \leq \min \{ v_{a_1 C}(u), v_{a_2 C}(u) \} \).
To prove the existence of average optimal strategies we need two extra conditions.

**Condition I.** There is a positive number \( B \) such that for all \( u \in A \) the sum \( \sum_{i=0}^{\infty} (Q_{au}^i)(v) \leq B \) for all \( a \in \Omega, v \in \mathbb{E}_{au} \) (where \( \mathbb{E}_{au} \) is defined as in lemma 2 with \( C = \{u\} \)).

Notice that this condition is satisfied if there is a \( \delta > 0 \) such that for all \( a \in \Omega \) and \( u, v \in \Lambda \) the transition probability \( Q_{au}(u,v) = 0 \) or \( Q_{au}(u,v) > \delta \).

This is true for instance in problems with an inventory structure (see [9]).

Let \( \pi_a(u) \) be the unique invariant probability of \( Q_a \) with \( \pi_a(u) > 0 \).

Since \( \pi_a(u) = \frac{1}{\sum_{i=0}^{\infty} (Q_{au}^i)(u)} \) the condition I implies that \( \pi_a(u) \geq \frac{1}{B} \).

**Condition II.** Let \( B \subset V \), \( a \in \Omega \) and \( \Omega_B \) be the subset of strategies which are identical to \( a \) on \( B \). For each extended real valued function \( v \) on \( A \), bounded from below, there is an \( a_0 \in \Omega_B \) such that

\[
T_{a_0} (c_a - g) + Q_{a_0} \inf_{a \in \Omega_B} \{T_a (c_a - g) + Q_a v\}
\]

This condition guarantees the existence of an optimal stationary strategy for a total costs decision process which stops as soon as \( A \) is entered, the one-period costs are \( c_a - g \), the costs of stopping are \( v \). If \( B = \emptyset \), then \( \Omega_B = \Omega \).

See Schäl [5] for various sets of continuity and compactness conditions which guarantee that the condition II is satisfied (since \( c_a - g > 0 \) on \( V \setminus A \) it is an almost negative dynamic programming problem).

Now, let \( A_g := \{u \in A \mid g_u = g\} \)

In the next theorem it is shown that there exists a strategy which is average optimal on \( A_g \).
Theorem 3. There is a strategy $a \in \Omega$ and a set $C \supseteq A$ such that $C$ is invariant under $Q_a$ and $g_a(u) = g$ for all $u \in C$.

Proof. First the existence of a subset $C_1$ of $A$ and a strategy $a_1 \in \Omega$ will be shown, such that $g_{a_1}(u) = g$ for $u \in C_1$ and $C_1$ is invariant under $Q_{a_1}$. Then it will be shown that in case $A \setminus C_1 \neq \emptyset$ the pair $(C_1, a_1)$ can be extended to a pair $(C_2, a_2)$ with $C_2 \supseteq C_1$, $C_2 \setminus C_1 \neq \emptyset$, $C_2$ is invariant under $Q_{a_2}$, $g_{a_2}(u) = g$ for $u \in C_2$ and $a_2$ is identical to $a_1$ on $C_1$.

This extension procedure can be continued until we have a pair $(C_n, a_n)$ with $C_n \supseteq A$.

To prove the first part, let $u \in A$. Let for $a \in \Omega$, the functions $v_{a'u}$ be defined as the functions $v_{a'C}$ in lemma 2 with $C = \{u\}$.

Define $\omega(v) := \inf_{a \in \Omega} \{ v_{a'u}(v) \}$. Since $T_{a_{a_1}}(c_{a_1} - g)(v) \geq -g$ for all $v \in V$ and $v_{a'u}(v) = +\infty$ or $v_{a'u}(v) = \sum_{l=0}^{\infty} Q_{a'u}^l (c_{a_1} - g)(v)$, the condition I yields $\omega(v) \geq -g B$ for all $v \in V$. By condition II there is a strategy $a_1$ such that

\[ T_{a_{a_1}}(c_{a_1} - g)(v) + (Q_{a_1 u}, \omega)(v) \leq T_{a_{a_1}}(c_{a_1} - g)(v) + (Q_{a'u}, \omega)(v) \]

for all $v \in V$, $a \in \Omega$.

Lemma 2 implies for each $\epsilon > 0$ the existence of a strategy $a_{\epsilon} \in \Omega_{\epsilon}$ such that $v_{a_{\epsilon}'u}(v) \leq \omega(v) + \epsilon$ for all $v \in A$.

Substitution of $a_{\epsilon}$ in (1) gives

\[ T_{a_{a_1}}(c_{a_1} - g)(v) + (Q_{a_1 u}, \omega_{a_{\epsilon}})(v) \leq T_{a_{a_1}}(c_{a_1} - g)(v) + (Q_{a_{\epsilon}'u}, \omega_{a_{\epsilon}})(v) \leq \]

\[ T_{a_{\epsilon}}(c_{a_1} - g)(v) + (Q_{a_{\epsilon}'u}, \omega_{a_{\epsilon}})(v) = v_{a_{\epsilon}'u}(v) \leq \omega(v) + \epsilon \]

Hence

\[ T_{a_{a_1}}(c_{a_1} - g)(v) + (Q_{a_1 u}, \omega_{a_{\epsilon}})(v) \leq \omega(v) \]

for all $v \in A$. 

-8-
Let $C_1$ be the subset of $A$ with all points where $\omega(.)$ is finite. The point $u$ is of course an element of $C_1$ and (2) implies that $C_1$ is invariant under $Q_{\alpha_1}$.

Let $E$ be an ergodic set of $Q_{\alpha_1}$ in $C_1$. We consider two cases, $u \in E$ or $u \not\in E$. If $u \in E$, repeated use of the inequality (2) yields

$$\omega(u) \geq \sum_{k=0}^{n-1} Q_{\alpha_1}^k u, T_{\alpha_1} (c_{\alpha_1} - g)(u) + (Q_{\alpha_1}^n u, \omega)(u)$$

But $u \in E$ implies that $\alpha_1 \in \Omega_u$, $(Q_{\alpha_1}^n u, \omega)(u) = 0$ for $n \to \infty$ and $\omega(u) = v_{\alpha_1}(u)$.

Hence

$$\sum_{k=0}^{\infty} (Q_{\alpha_1}^k u, T_{\alpha_1} c_{\alpha_1})(u) - g \sum_{k=0}^{\infty} (Q_{\alpha_1}^k u, T_{\alpha_1} 1)(u) \leq 0,$$

$$\leq \sum_{k=0}^{\infty} (Q_{\alpha_1}^k u, T_{\alpha_1} c_{\alpha})(u) - g \sum_{k=0}^{\infty} (Q_{\alpha_1}^k u, T_{\alpha_1} 1)(u), \alpha \in \Omega_u.$$ 

Substitution of

$$g_{\alpha}(u) = \frac{\sum_{k=0}^{\infty} (Q_{\alpha_1}^k u, T_{\alpha_1} c_{\alpha})(u)}{\sum_{k=0}^{\infty} (Q_{\alpha_1}^k u, T_{\alpha_1} 1)(u)}, \alpha \in \Omega_u$$

yields $(g_{\alpha}(u) - g) \sum_{k=0}^{\infty} (Q_{\alpha_1}^k u, T_{\alpha_1} 1)(u) \leq (g_{\alpha}(u) - g) \sum_{k=0}^{\infty} (Q_{\alpha_1}^k u, T_{\alpha_1} 1)(u), \alpha \in \Omega_u$.

Since $(T_{\alpha} 1)(u) \leq \frac{2}{\pi(\alpha)}$ (lemma 1), $\pi_{\alpha}(u) \geq \frac{1}{B}$ and $\sum_{k=0}^{\infty} (Q_{\alpha_1}^k 1)(u) \leq B$ (condition 1) this implies $g_{\alpha}(u) = g$ and therefore $g_{\alpha_1}(v) = g$ for all $v \in E$.

If $u \not\in E$ the process $Q_{\alpha_1}u$ is identical to the process $Q_{\alpha_1}$ on $E$ and repeated use of the inequality (2) gives, for $v \in E$.

$$\omega(v) \geq \sum_{k=0}^{n-1} Q_{\alpha_1}^k T_{\alpha_1} c_{\alpha_1} - g(v) + (Q_{\alpha_1}^n \omega)(v)$$

Let $g_{\alpha_1}$ be the average costs of $\alpha_1$ on $E$. The existence of this average cost follows from the finiteness of $T_{\alpha_1} c_{\alpha_1}$ and $T_{\alpha_1} 1$ on $E$, which follow from the finiteness of $\omega$ on $E$. 


Furthermore the expression \( \sum_{\ell=0}^{n-1} Q_{a_1} T_{a_1} (c_{a_1} - g_{a_1}) (v) \) is bounded in \( n \).

Hence, by (3), \( g \geq g_{a_1} \) and therefore \( g_{a_1} = g \).

We proved that the average costs of \( a_1 \) are equal to \( g \) on all ergodic sets of \( Q_{a_1} \) in \( C_1 \). The finiteness of \( \omega \) on \( C_1 \) implies that the average costs of \( a_1 \) are equal to \( g \) on the whole set \( C_1 \).

If \( A \setminus C_1 \neq \emptyset \), choose \( u \in A \setminus C_1 \) and let \( C := C_1 \cup \{u\} \).

Let \( C^* \) be union of \( C_1 \) and the set of all points in \( V \) which can be reached with positive probability from some point in \( C_1 \) and define \( \Omega^* \) as the set of all stationary strategies which are identical to \( a_1 \) on \( C_1 \).

Let \( \Omega^*_v := \Omega^* \cap \Omega_v \), \( v \in A \) and \( \Omega^* := \Omega^* \cap \Omega_C \). Since \( \Omega^* \) dominates \( \Omega \) in the average costs sense we know that \( \inf_{a \in \Omega^*_v} g_{a} (v) = g_v = g \) for \( v \in A \).

Define \( \omega^* (v) = \inf_{a \in \Omega^*_v} \{ v \alpha_{a}(v) \} \).

By condition II there is a strategy \( a_2 \) which minimizes the expression \( T_{a_2} (c_{a_2} - g) (v) + (Q_{a_2} \omega^*) (v) \) over the set of strategies \( \Omega^* \).

As for \( a_1 \) and \( \omega \) it is possible to show that

\[
T_{a_2} (c_{a_2} - g) (v) + (Q_{a_2} \omega^*) (v) \leq \omega^* (v) \text{ for all } v \in A.
\]

and also that \( g_{a_2} (u) = g \) for all \( u \in C_2 \), where \( C_2 \) is the set of all points in \( A \) where \( \omega^*(\cdot) \) is finite.

If \( A \setminus C_2 \neq \emptyset \) we can continue in the same way. \( \square \)

The following corollary is a direct consequence of theorem 3.

Corollary 4. If it is possible to reach the set \( A \) from each point in \( V \) with finite expected costs then there exists a strategy which is average optimal on the whole state space \( V \).


