Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2012, Article ID 103698, 16 pages
doi:10.1155/2012/103698

Research Article

A Version of the Euler Equation in Discounted Markov Decision Processes

**H. Cruz-Suárez, G. Zacarías-Espinoza,
and V. Vázquez-Guevara**

*Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla,
Avenida San Claudio y Río Verde, Col. San Manuel, CU, 72570 Puebla, PUE, Mexico*

Correspondence should be addressed to H. Cruz-Suárez, hcs@fcfm.buap.mx

Received 21 July 2012; Revised 29 August 2012; Accepted 6 September 2012

Academic Editor: Vu Phat

This paper deals with Markov decision processes (MDPs) on Euclidean spaces with an infinite horizon. An approach to study this kind of MDPs is using the dynamic programming technique (DP). Then the optimal value function is characterized through the value iteration functions. The paper provides conditions that guarantee the convergence of maximizers of the value iteration functions to the optimal policy. Then, using the Euler equation and an envelope formula, the optimal solution of the optimal control problem is obtained. Finally, this theory is applied to a linear-quadratic control problem in order to find its optimal policy.

1. Introduction

This paper deals with the optimal control problem in discrete time and with an infinite horizon. This problem is presented with the help of the Markov decision processes (MDPs) theory. To describe the MDPs, it is necessary to provide a Markov control model. The components of the Markov control model are used to describe the dynamic of the system. In this way at each time t ($t = 0, 1, \ldots$) the state of the system is affected by an admissible action. This sequence of actions is called a policy. The optimal control problem consists in determining an optimal policy, which is characterized through a performance criterion. In this paper, the infinite-horizon expected total discounted reward is considered.

An approach for solving the optimal control problem is through the dynamic programming technique (DP) (see [1–4]). DP characterizes the optimal solution of the optimal control problem using a functional equation, known as the dynamic programming equation (see [1–4]). In the literature there exists conditions that guarantee the value iteration (VI) functions procedure, which is used to approximate the optimal value function of the optimal control problem. However, this technique has problems when the reward and/or dynamic have a complicated functional form (see [5, page 93]).

An alternative for solving this problem is using the Euler Equation (EE), which is well known in the applications of MDPs to economic models. This equation is established and solved in this context (in some cases empirically) (see [6–13]).

An iterative method for deterministic MDPs is presented in [14]. In this case, the EE is obtained in terms of the VI functions. Retaking this idea, this article presents an iterative method of finding the solution of EE in terms of the VI functions in stochastic MDPs.

In this paper, the Euler equation is obtained using an envelope formula (see [15–17]) under interiority conditions of the VI functions. The envelope formula characterizes the performance criterion derivative with respect to the initial state of the system. The performance criterion derivative is important in analyzing the behavior of the Markov control process. Also, in [18], a general study about the performance sensitivities in the policy space is presented. In this context, two performance sensitivity formulas are studied: one for performance derivatives at any policy in the policy space and the other one for performance differences between any two policies in the policy space.

The technique proposed in this paper is used as follows. Firstly, EE is applied to obtain the VI functions. Secondly, applying the envelope formula, the maximizers of the VI functions are obtained. Then using the maximizers convergence to the optimal policy, the optimal control problem is solved. This procedure is exemplified by a linear quadratic problem.

The paper is organized as follows: in Section 2, the theory of MDPs necessary for subsequent sections is presented. In Section 3, some conditions on the Markov Control Model are presented to ensure both the differentiability of the VI functions and the optimal value function. These conditions guarantee the validity of a version of the EE for the VI. Finally, in Section 4, a linear quadratic problem is presented to illustrate the theory.

2. Markov Decision Process

A discrete-time markov control model is a quintuple $(X, A, \{A(x) \mid x \in X\}, Q, r)$, where X is the state space, A is the action space, $A(x)$ is the set of feasible actions in the state $x \in X$, Q is a transition law and $r : \mathbb{K} \to \mathbb{R}$ is the one-step reward function (see [3]). X and A are (nonempty) Borel spaces with the Borel σ-algebras $\mathcal{B}(X)$ and $\mathcal{B}(A)$, respectively. $Q(\cdot \mid \cdot)$ is a stochastic kernel on X given $\mathbb{K}$, where $\mathbb{K} := \{(x, a) \mid x \in X, a \in A(x)\}$, and r is a measurable function.

Consider a Markov control model and, for each $t = 0, 1, \ldots$, define the space $\mathbb{H}_t$ of admissible histories up to time t as $\mathbb{H}_0 = X$, and $\mathbb{H}_t = \mathbb{K} \times \mathbb{H}_{t-1}$, for $t = 1, 2, \ldots$.

A *policy* is a sequence $\pi = \{\pi_t\}$ of stochastic kernels π_t on the action space A given $\mathbb{H}_t$. The set of policies will be denoted by Π.

Let $\mathbb{F}$ be the set of decision functions or measurable selectors, that is, the set of all measurable functions $f : X \to A$ such that $f(x) \in A(x)$ for all $x \in X$.

A sequence $\{f_t\}$ of functions $f_t \in \mathbb{F}$ is called a Markov policy. A stationary policy is a Markov policy $\pi = \{f_t\}$ such that $f_t = f$ for all $t = 0, 1, 2, \ldots$, with $f \in \mathbb{F}$, and it will be denoted by f (see [3]).

Given the initial state $x_0 = x \in X$, and any policy $\pi \in \Pi$, there is a probability measure P_x^π on the space $(\Omega, \mathcal{F})$, with $\Omega := (X \times A)^\infty$ and $\mathcal{F}$, the product σ-algebra (see [3]). The corresponding expectation operator will be denoted by E_x^π. The stochastic process $((\Omega, \mathcal{F}, P_x^\pi), \{x_t\})$ is called a discrete-time Markov decision process.

The *total expected discounted reward* is defined as

$$v(\pi, x) := E_x^{\pi}\left[\sum_{t=0}^{\infty}\alpha^t r(x_t, a_t)\right], \tag{2.1}$$

$\pi \in \Pi$ and $x \in X$, where $\alpha \in (0,1)$ is called the discount factor.

Definition 2.1. A policy $\pi^* \in \Pi$ is optimal, if for each $x \in X$,

$$v(\pi^*, x) = \sup_{\pi\in\Pi} v(\pi, x). \tag{2.2}$$

The function defined by

$$V(x) = \sup_{\pi\in\Pi} v(\pi, x), \tag{2.3}$$

$x \in X$, will be called the optimal value function.

The optimal control problem consists in determining an optimal policy.

2.1. Dynamic Programming

Definition 2.2. A measurable function $\lambda : X \to \mathbb{R}$ is said to be a solution to the optimal equation (OE) if it satisfies

$$\lambda(x) = \sup_{a\in A(x)}\left\{r(x,a) + \alpha\int_X \lambda(y)Q(dy \mid x, a)\right\}, \tag{2.4}$$

$x \in X$.

Assumption 2.3. (a) The one-step reward function r is nonpositive, upper semicontinuous (u.s.c), and sup-compact on $\mathbb{K}$. (r is a sup-compact function if the set $\{a \in A(x) \mid r(x,a) \geq \gamma\}$ is compact for every $x \in X$ and $\gamma \in \mathbb{R}$.)

(b) The transition law Q is strongly continuous.

(c) There exists a policy π such that $v(\pi, x) > -\infty$, for each $x \in X$.

Definition 2.4. The value iteration (VI) functions are defined as follows:

$$v_n(x) = \sup_{a\in A(x)}\left\{r(x,a) + \alpha\int_X v_{n-1}(y)Q(dy \mid x, a)\right\}, \tag{2.5}$$

for all $x \in X$ and $n = 1, 2, \ldots$, with $v_0(x) = 0$.

The following theorem is well-known in the literature of MDPs (see, [1–4]). The proof can be consulted in (see [3, page 46]).

Theorem 2.5. *Suppose that Assumption 2.3 holds. Then*

(a) The optimal value function V is a solution of the OE (see Definition 2.2).

(b) There exists $f \in \mathbb{F}$ such that

$$V(x) = r(x, f(x)) + \alpha \int_X V(y) Q(dy \mid x, f(x)). \tag{2.6}$$

(c) For every $x \in X$, $v_n(x) \to V(x)$, when $n \to \infty$.

Remark 2.6. Under Assumption 2.3, it is possible to demonstrate that for each $n = 1, 2, \ldots$, there exists a stationary policy $f_n \in \mathbb{F}$ such that

$$v_n(x) = r(x, f_n(x)) + \alpha \int_X v_{n-1}(y) Q(dy \mid x, f_n(x)), \tag{2.7}$$

$x \in X$ (see [3, page. 27, 28]).

3. Differentiability in MDPs

3.1. Notation and Preliminaries

Let X and Y be Euclidean spaces and consider the following notation: $C^2(X, Y)$ denotes the set of functions $l : X \to Y$ with a continuous second derivative (when $X = Y$, $C^2(X, Y)$ will be denoted by $C^2(X)$ and in some cases it will be written only as C^2). Let $\Gamma : X \times Y \to \mathbb{R}$ be a measurable function such that $\Gamma \in C^2(X \times Y, \mathbb{R})$. Γ_x, and Γ_y denote the partial derivative of Γ for x and y, respectively. The notations for the second partial derivatives of Γ are Γ_{xx}, Γ_{xy}, Γ_{yx} and Γ_{yy}.

For any set $C \subset X$, a point $x \in C$ is called an interior point of C if there exists an open set U such that $x \in U \subset C$. The interior of C is the set of all interior points of C denoted by $\operatorname{int}(C)$.

The set-value mapping Θ from X to Y is said to be

(a) nondecreasing, if $x, z \in X$ with $x < z$ then $\Theta(x) \subseteq \Theta(z)$,

(b) convex, if $x, z \in X$ and $\beta \in [0, 1]$, then $\beta a + (1-\beta)\tilde{a} \in \Theta(\beta x + (1-\beta)z)$, with $a \in \Theta(x)$ and $\tilde{a} \in \Theta(z)$.

Let $\Upsilon : \mathbb{K} \to \mathbb{R}$ be a measurable function. Define $v : X \to \mathbb{R}$ by

$$v(x) := \sup_{a \in \Theta(x)} \Upsilon(x, y), \tag{3.1}$$

$x \in X$.

The proof of the following lemma is similar to the proof of Theorem 1 in [16].

Lemma 3.1. *Suppose that*

(a) $\Upsilon \in C^2(\operatorname{int}(\mathbb{K}); \mathbb{R})$, *furthermore $\Upsilon_{yy}(x, \cdot)$ is negative definite, for every $x \in X$;*

(b) *for each $x \in X$*, argmax $\Upsilon(x, y) \subseteq \operatorname{int}(\Theta(x))$.

Then there exists a function $l : X \to A$ such that $v(x) = \Upsilon(x, l(x))$, for every $x \in X$. Moreover $l \in C^1(\mathrm{int}(X); Y)$ and $\Upsilon \in C^2(\mathrm{int}(X); \mathbb{R})$.

Remark 3.2. Observe that (a) implies that $\Lambda(x, \cdot)$ is a strictly concave function, for each $x \in X$. Then the maximizer $l : X \to A$ is unique.

The proof of the following lemma can be consulted in [19], Theorem 25.7, page 248.

Lemma 3.3. *Let $C \subset \mathbb{R}^n$ be an open and convex set. Let $g : C \to \mathbb{R}$ be a concave and differentiable function, and $\{g_n\}$ be a sequence of differentiable, concave and real-valued functions on C, such that $g_n(x) \to g(x)$, when $n \to \infty$, for all $x \in C$. Then*

$$\lim_{n \to \infty} g_n'(x) = g'(x). \tag{3.2}$$

3.2. An Envelope Formula in MDPs

Let $(X, A, \{A(x) \mid x \in X\}, Q, r)$ be a fixed Markov control model. Throughout this section it is assumed that Assumption 2.3 holds. Also, it is supposed that $X \subseteq \mathbb{R}^n$ and $A \subseteq \mathbb{R}^m$ are convex sets with nonempty interiors and X is partially ordered. It is considered that the set-valued mapping $x \to A(x)$ is nondecreasing and convex, and $A(x)$ has nonempty interior, for each $x \in X$. Also, it is assumed that the transition law Q is given by a difference equation:

$$x_{t+1} = L(F(x_t, a_t), \xi_t), \tag{3.3}$$

$t = 0, 1, \ldots$, with a given initial state $x_0 = x \in X$ fixed, where $\{\xi_t\}$ is a sequence of independent and identically distributed (iid) random variables, independent of $x_0 = x \in X$ and taking values in a Borel space $S \subset \mathbb{R}^k$. Let ξ be a generic element of the sequence $\{\xi_t\}$. The density of ξ is designated by Δ; $L : X' \times S \to X$ is a measurable function, with $X' \subset \mathbb{R}^m$, and $F : \mathbb{K} \to X'$, is a measurable function too.

Since Assumption 2.3 is assumed, then Theorem 2.5 yields. Therefore, the optimal value function (see Definition 2.1) satisfies

$$V(x) = \sup_{a \in A(x)} \{r(x, a) + \alpha E[V(L(F(x, a), \xi))]\}, \tag{3.4}$$

and the VI functions (see Definition 2.4) satisfy

$$v_n(x) = \sup_{a \in A(x)} \{r(x, a) + \alpha E[v_{n-1}(L(F(x, a), \xi))]\}, \tag{3.5}$$

for each $n = 1, 2, \ldots$, with $v_0(x) := 0$. In addition, by Theorem 2.5, there exists the optimal policy, which will be denoted by f. Furthermore, there exists the maximizer f_n of v_n for each $n = 1, 2, \ldots$ (see Remark 2.6).

Let $G : \mathbb{K} \to \mathbb{R}$ be a function defined as

$$G(x, a) := r(x, a) + \alpha H(x, a), \tag{3.6}$$

$(x, a) \in \mathbb{K}$, where

$$H(x, a) := E[V(L(F(x, a), \xi))]. \tag{3.7}$$

Define $G^n : \mathbb{K} \to \mathbb{R}$ by

$$G^n(x, a) := r(x, a) + \alpha E[v_{n-1}(L(F(x, a), \xi))], \tag{3.8}$$

for each $n = 1, 2, \ldots$, with $v_0(x) := 0$ and $(x, a) \in \mathbb{K}$.

Assumption 3.4. (a) r is a strictly concave function and $r(\cdot, a)$ is an increasing function on X for each $a \in A$ fixed;

(b) $L(\cdot, s)$ is a concave and increasing function, for each $s \in S$; F is a concave function, $F(\cdot, a)$ is an increasing function on X, for each $a \in A$.

Lemma 3.5. *Under Assumption 3.4, it results that v_n is a strictly concave function and f_n is unique, for all $n = 1, 2, \ldots$. Also, V is a strictly concave function and f is unique.*

Proof. By Assumption 3.4(a), it suffices to prove Condition C1 (see [20, Lemma 6.2]), which guarantees the result. Let $\Psi : \mathbb{K} \times S \to X$ be defined by

$$\Psi(x, a, s) := L(F(x, a), s). \tag{3.9}$$

Then for each $s \in S$, the function $\Psi(\cdot, \cdot, s)$ is concave in $\mathbb{K}$ by Assumption 3.4(b). Indeed, since F is a concave function, then

$$F(\beta(x, a) + (1-\beta)(y, b)) \geq \beta F(x, a) + (1-\beta)F(y, b), \tag{3.10}$$

$(x, a), (y, b) \in \mathbb{K}$ and $\beta \in [0, 1]$. Furthermore, it is known that $L(\cdot, s)$ is a concave and increasing function, for each $s \in S$, then

$$L(F(\beta(x, a) + (1-\beta)(y, b)), s) \geq \beta L(F(x, a), s) + (1-\beta)L(F(y, b), s). \tag{3.11}$$

From similar arguments, it can be shown that if $x < y$, then $\Psi(x, a, s) \leq \Psi(y, a, s)$, for each $s \in S$ and $a \in A(y)$. Then the result follows. □

Assumption 3.6. (a) $r \in C^2(\text{int}(\mathbb{K}); \mathbb{R})$ and $r_{aa}(x, \cdot)$ is negative definite for each $x \in X$;

(b) $F \in C^2(\text{int}(\mathbb{K}); X')$ and $F_a(x, \cdot)$ is invertible, for each $a \in A$;

(c) $L(\cdot, s) \in C^2(\text{int}(X'); X)$, for each $s \in S$. Besides, $L(\cdot, \cdot)$ has an inverse in the second variable R, such that $R(\cdot, s) \in C^2(\text{int}(X' \times X); S)$, and $|\det R_s(\cdot, s)| \in C^2(\text{int}(X'); \mathbb{R})$, for all $s \in S$, where, in this case, R_s denotes the derivative of R with respect to the second variable, and the determinant of R_s is denoted as $\det R_s$;

(d) $\Delta \in C^2(\text{int}(S); R)$ and the interchange between derivatives and integrals is valid (see Remark 3.8).

Lemma 3.7. *By Assumption 3.6 it results that $H \in C^2(\text{int}(\mathbb{K}); X')$, with H defined in* (3.7).

Proof. The proof is similar to the proof of Lemma 5 in [16]. Assumption 3.6 allows to express the stochastic kernel (see (3.3)) in the following form: for each measurable subset B of X and $(x,a) \in \mathbb{K}$,

$$Q(B \mid (x,a)) = \Pr(s \in S \mid L(F(x,a),s)) = \Pr(s \in S \mid s(s \in R(F(x,a),B)) \tag{3.12}$$

$$= \int_{R(F(x,a),B)} \Delta(u)du. \tag{3.13}$$

Then for the change of variable theorem, it results that

$$Q(B \mid (x,a)) = \int_{R(F(x,a),B)} \Delta(R(F(x,a),u))|\det R_s(F(x,a),u)|du. \tag{3.14}$$

It follows from (3.13) that H can be expressed as

$$H(x,a) = \int_{R(F(x,a),B)} V(u)\Delta(R(F(x,a),u))|\det R_s(F(x,a),u)|du. \tag{3.15}$$

Now, using Assumption 3.6, the result follows. □

Remark 3.8. In Lemma 3.7, Assumption 3.6(d) was used to guarantee the differentiability of the second order of the integral $\int K(x,a,u)du$, with respect to x or a, where

$$K(x,a,u) := V(u)\Delta(R(F(x,a),u))|\det R_s(F(x,a),u)|, \tag{3.16}$$

$(F(x,a),u) \in \operatorname{int}(X' \times X)$. This condition can be verified in practice when the derivatives of K can be bounded in the following sense: for $(F(x,a),u) \in \operatorname{int}(X' \times X)$, $|K_x(x,a,u)| \leq k_1(a,u)$,$|K_a(x,a,u)| \leq k_2(x,u)$, $|K_{xx}(x,a,u)| \leq k_3(a,u)$, $|K_{aa}(x,a,u)| \leq k_4(a,u)$, $|K_{xa}(x,a,u)| \leq k_1(a,u)$, for some functions g_i integrable with respect to u, for $i = 1,\ldots,5$ (see Remark 10 in [16]).

Assumption 3.9. (a) The optimal policy f satisfies that $f(x) \in \operatorname{int}(A(x))$, for each $x \in X$;

(b) The sequence $\{f_n\}$ of the maximizers of the VI functions satisfies that $f_n(x) \in \operatorname{int}(A(x))$, for each $x \in X$ and $n = 1,2,\ldots$.

Define W by

$$W(x,a) := \left[r_x - r_a F_a^{-1} F_x\right](x,a), \tag{3.17}$$

$(x,a) \in \mathbb{K}$.

Remark 3.10. Assumption 3.9 evidently holds; if $A(x)$ is open for every $x \in X$, then $f(x)$ $(f_n(x))$ belongs to the interior of $A(x)$, $x \in X$. Also, in some particular cases (see [8, 16]), the interiority of $f(x)$ $(f_n(x))$ is guaranteed by the mean value theorem.

Theorem 3.11. *Under Assumptions 3.4, 3.6, and 3.9(a), it results that* $f \in C^1(\mathrm{int}(X); A)$, $V \in C^2(\mathrm{int}(X); \mathbb{R})$ *and for each* $x \in \mathrm{int}(X)$,

$$V'(x) = W(x, f(x)), \tag{3.18}$$

where W *is defined in* (3.17).

Proof. Let $x \in \mathrm{int}(X)$ fixed. Note that Assumptions 3.4 and 3.6 imply that $G \in C^2(\mathrm{int}(\mathbb{K}); \mathbb{R})$ where G is defined in (3.6). Indeed, since Assumptions 3.4(a) and 3.6(a) hold, it is known that $r \in C^2(\mathrm{int}(\mathbb{K}); \mathbb{R})$ and $r_{aa}(x, \cdot)$ is negative definite. Moreover, Lemma 3.5 implies that $H(x,a) = E[V(L(F(x,a),\xi))]$ is a concave function, and by Lemma 3.7, it follows that $H \in C^2(\mathrm{int}(\mathbb{K}); \mathbb{R})$, obtaining that $H_{aa}(x, \cdot)$ is negative semidefinite (see [21, page 260]). Furthermore, by Assumption 3.9(a) and applying Lemma 3.1, it concludes that $f \in C^1(\mathrm{int}(X); A)$ and $V \in C^2(\mathrm{int}(X); \mathbb{R})$.

On the other hand, it is obtained that

$$G_a(x,a) = r_a(x,a) + \alpha E\left[V'(L(F(x,a),\xi))L'(F(x,a),\xi)\right] F_a(x,a), \tag{3.19}$$

for each $a \in \mathrm{int}(A(x))$. Then, the first order condition and the invertibility of F_a (see Assumption 3.6(b)) imply that $G_a(x, f(x)) = 0$, that is,

$$-r_a F_a^{-1}(x, f(x)) = \alpha E\left[V'(L(F(x,f(x)),\xi))L'(F(x,f(x)),\xi)\right]. \tag{3.20}$$

Moreover, since V satisfies (2.4) and $f \in \mathbb{F}$ is the optimal policy, then

$$V(x) = G(x, f(x)). \tag{3.21}$$

Using the fact that $G_a(x, f(x)) = 0$, it is possible to obtain the following envelope formula:

$$\begin{aligned} V'(x) &= G_x(x, f(x)) + G_a(x, f(x)) f'(x), \\ &= G_x(x, f(x)). \end{aligned} \tag{3.22}$$

Equivalently,

$$V'(x) = r_x(x, f(x)) + \alpha E\left[V'(L(F(x,f(x)),\xi))L'(F(x,f(x)),\xi)\right] F_x(x, f(x)). \tag{3.23}$$

Finally, substituting (3.20) in (3.23), it follows that

$$V'(x) = W(x, f(x)). \tag{3.24}$$

□

Theorem 3.12. *Under Assumptions 3.4, 3.6 and 3.9(b), it results that* $f_n \in C^1(\mathrm{int}(X); A)$, $v_n \in C^2(\mathrm{int}(X); \mathbb{R})$, *for each* $n = 1, 2, \ldots$. *Furthermore,*

$$v_n'(x) = W(x, f_n(x)), \tag{3.25}$$

for each $x \in \mathrm{int}(X)$ *and* $n = 2, 3, \ldots$.

Proof. The proof will be made by induction. Let $x \in \operatorname{int}(X)$ be fixed. Since

$$v_1(x) = \max_{a \in A(x)} G^1(x, a), \tag{3.26}$$

where G^1 is defined in (3.8) and by Assumptions 3.4(a) and 3.6(a), it follows that $G^1 \in C^2(\operatorname{int}(\mathbb{K}); \mathbb{R})$ and $G^1_{aa}(x, a)$ is negative definite. By Assumption 3.9(b), it yields that $f_1(x) \in \operatorname{int}(A(x))$, and applying Lemma 3.1, it follows that $f_1 \in C^1(\operatorname{int}(X); A)$, $v_1 \in C^2(\operatorname{int}(X); \mathbb{R})$. Straightforward computations allow to obtain that

$$v_1'(x) = G^1_x(x, f_1(x)). \tag{3.27}$$

Moreover, by Lemma 3.5 it is known that v_1 is strictly concave, then it is negative semidefinite (see [21, page 260]).

Let $n = 2$, then

$$v_2(x) = \max_{a \in A(x)} G^2(x, a), \tag{3.28}$$

where G^2 is defined in (3.8).

Since r, L, F, $v_1 \in C^2$, then $G^2 \in C^2$ too. Moreover, Lemmas 3.5 and 3.7 imply that

$$H^2(x, a) := E[v_1(L(F(x, a), \xi))] \tag{3.29}$$

is a concave function and $H^2 \in C^2$. It follows that $H^2_{aa}(x, \cdot)$ is negative semidefinite.

Consequently, G^2_{aa} is negative definite. Now, since $f_2(x) \in \operatorname{int}(A(x))$ (see Assumption 3.9(b)), applying again Lemma 3.1, it follows that $f_2 \in C^1(\operatorname{int}(X); A)$, $v_2 \in C^2(\operatorname{int}(X); \mathbb{R})$. Furthermore, the first order condition implies that $G^2_a(x, f_2(x)) = 0$. By the invertibility of F_a (see Assumption 3.6(b)), it follows that

$$-r_a F_a^{-1}(x, f_2(x)) = \alpha E\left[v_1'(L(F(x, f_2(x)), \xi)) L'(F(x, f_2(x)), \xi)\right]. \tag{3.30}$$

On the other hand,

$$v_2'(x) = r_x(x, f_2(x)) + \alpha E\left[v_1'(L(F(x, f_2(x)), \xi)) L'(F(x, f_2(x)), \xi)\right] F_x(x, f_2(x)), \tag{3.31}$$

and substituting (3.30) in (3.31), it is obtained that

$$v_2'(x) = W(x, f_2(x)), \tag{3.32}$$

where W is defined in (3.17).

Now, suppose that $v_{n-1} \in C^2(\operatorname{int}(X); \mathbb{R})$ with $n > 2$. Using arguments similar to the case $n = 2$, it is possible to demonstrate that $f_n \in C^1(\operatorname{int}(X); A)$, $v_n \in C^2(\operatorname{int}(X); \mathbb{R})$ and

$$v_n'(x) = W(x, f_n(x)). \tag{3.33}$$

□

Assumption 3.13. For each $x \in X$, the function $W(x,\cdot)$ has a continuous inverse function, denoted by w.

Theorem 3.14. *Under Assumptions 3.4, 3.6, 3.9 and 3.13, it follows that*

$$f_n(x) \longrightarrow f(x), \tag{3.34}$$

when $n \to \infty$, for each $x \in \operatorname{int}(X)$.

Proof. Let $x \in \operatorname{int}(X)$ fixed. It is known by Lemma 3.5 and Theorem 3.11 that the optimal value function V is concave and differentiable on $\operatorname{int}(X)$. In addition, it is known that for each $n \in \mathbb{N}$, v_n is a concave and differentiable function on $\operatorname{int}(X)$. Then from Lemma 3.3 it follows that

$$v_n'(x) \longrightarrow V'(x), \tag{3.35}$$

when n goes to ∞.

Now by Assumption 3.13, it concludes that for $n = 2, 3, \ldots$,

$$\begin{aligned} f_n(x) &= w(x, v_n'(x)), \\ f(x) &= w(x, V'(x)), \end{aligned} \tag{3.36}$$

where f_n is a stationary policy of v_n and f is the optimal policy. Finally, the convergence is guaranteed by the continuity of w (see Assumption 3.13). □

3.3. Euler Equation

Theorem 3.15. *Under Assumptions 3.4, 3.6, 3.9, and 3.13 it follows that*

$$\begin{aligned} v_n'(x) = {} & r_x(x, w(x, v_n'(x))) \\ & + \alpha E\left[v_{n-1}'(L(F(x, w(x, v_n'(x))), \xi))L'(F(x, w(x, v_n'(x))), \xi)\right]F_x(x, w(x, v_n'(x))), \end{aligned} \tag{3.37}$$

for each $x \in \operatorname{int}(X)$ and $n \in \mathbb{N}$, where w is the function given in Assumption 3.13.

Proof. Let $x \in \operatorname{int}(X)$ be fixed. By Lemma 3.5 and Theorem 3.12, it is known that $v_n \in C^2(\operatorname{int}(X); \mathbb{R})$ and it is a concave function. Now, from the first order condition and the invertibility of F_a (see Assumption 3.6(b)), it follows that

$$-r_a F_a^{-1}(x, f_n(x)) = \alpha E\left[v_{n-1}'(L(F(x, f_n(x)), \xi))L'(F(x, f_n(x)), \xi)\right]. \tag{3.38}$$

Since

$$v_n'(x) = W(x, f_n(x)), \tag{3.39}$$

and using the invertibility of $W(x,\cdot)$ (see Assumption 3.13), it follows that

$$f_n(x) = w\left(x, v_n'(x)\right). \tag{3.40}$$

Finally, substituting (3.40) in (3.38), (3.37) is obtained. □

Corollary 3.16. *The optimal value function satisfies*

$$\begin{aligned} V'(x) &= r_x\left(x, w\left(x, V'(x)\right)\right) \\ &\quad + \alpha E\left[V'\left(L\left(F\left(x, w\left(x, V'(x)\right)\right), \xi\right)\right) L'\left(F\left(x, w\left(x, V'(x)\right)\right), \xi\right)\right] F_x\left(x, w\left(x, V'(x)\right)\right), \end{aligned} \tag{3.41}$$

for each $x \in \operatorname{int}(X)$.

Proof. Let $x \in \operatorname{int}(X)$ be fixed. It is known that the VI functions satisfy the Euler equation (3.37), so applying Lemma 3.3, it is obtained that

$$v_n'(x) \longrightarrow V'(x), \tag{3.42}$$

when $n \to \infty$. Also from Assumption 3.13, w is a continuous function. Then, when n goes to infinite in (3.37), it follows that the optimal value function satisfies (3.41). □

4. A Linear-Quadratic Model

Consider that $\mathbb{R}^n = X = A = A(x)$, for each $x \in X$. The dynamic of the system is given by

$$x_{t+1} = Bx_t + Ca_t + \xi_t, \tag{4.1}$$

$t = 1, 2, \ldots$, with $x_0 = x \in X$ given. B and C are invertible matrices of size $n \times n$, $\{\xi_t\}$ is a sequence of iid column random vectors with values in $\mathbb{R}^n$. Let ξ be a generic element of the sequence $\{\xi_t\}$, assume that ξ has a density Δ with $\Delta \in C^2$, and $E(\xi)$ equals vector zero. Furthermore, it is assumed that if P is a symmetric negative definite matrix of size $n \times n$, then $E[\xi^T P \xi]$ is finite. In addition, it is assumed that the interchange between derivatives and integrals is valid (see Remark 3.8). A particular case of this assumption can be found in [16, page 315].

The reward function is given by

$$r(x, a) = x^T Q x + a^T R a, \tag{4.2}$$

where x^T and a^T denote the transpose of vectors x and a; Q and R are symmetric matrices of size $n \times n$, and both of them are negative definite.

Lemma 4.1. *The linear quadratic model satisfies Assumption 2.3.*

Proof. Note that

$$O_\gamma^x := \left\{ a \in A(x) \mid x^T Q x + a^T R a \geq \gamma \right\} \tag{4.3}$$

is a compact set, for each $x \in X$ and $\gamma \in \mathbb{R}$. Indeed, let $x \in X$ and $\gamma \in \mathbb{R}$. If any sequence $\{a_n\}$ of O_γ^x satisfies $x^T Q x + a_n^T R a_n \to -\infty$, then there is a contradiction. Therefore O_γ^x is a set bounded below. Moreover, since Q and R are negative definite, then O_γ^x is a set bounded above. In addition, if $\{a_n\} \subset O_\gamma^x$ so that $a_n \to a$, then by the continuity of r, it follows that $r(x, a) \geq \gamma$, implying that O_γ^x is a closed set. Therefore, the reward function r is sup-compact. Finally, note that r is a nonpositive and continuous function on $\mathbb{K}$. So Assumption 2.3(a) holds.

On the other hand, let $U \in \mathcal{B}(X)$, then

$$Q(U \mid x_t = x, a_t = a) = \Pr(x_{t+1} \in U \mid x_t = x, a_t = a) = \int I_U(Bx + Ca)\Delta(s)ds, \tag{4.4}$$

where I_U denotes the indicator function of U. Since the density Δ is continuous, it is obtained that the transition law Q is weakly continuous, that is, Assumption 2.3(b) holds.

Finally, let $h \in \mathbb{F}$ be defined as

$$h(x) = -C^{-1}Bx, \tag{4.5}$$

then the dynamic of the system is given by

$$x_t = Bx_{t-1} + Ch(x_{t-1}) + \xi_{t-1} = \xi_{t-1}, \tag{4.6}$$

for $t = 1, 2, \ldots$, with $x_0 = x \in X$.

It follows that

$$E_x^h\left[x_t^T Q x_t + h(x_t)^T R h(x_t)\right] = E\left[\xi_t^T P \xi_t\right] =: \theta, \tag{4.7}$$

where $P := (Q + (C^{-1}B)^T R(C^{-1}B))$.

Since $\{\xi_t\}$ is i.i.d, then

$$v(h, x) = \frac{\theta}{1 - \alpha} < \infty, \tag{4.8}$$

where v is given by (2.1). Therefore, Assumption 2.3(c) holds. □

Lemma 4.2. *The linear quadratic problem satisfies Assumptions 3.4, 3.6, 3.9, and 3.13.*

Proof. It is easy to obtain that r is a concave function and $r \in C^2$, implying Assumptions 3.4(a) and 3.6(a). Assumption 3.4(b) is satisfied using Condition C2 in [20]. Furthermore, observe that

$$F(x, a) = Bx_t + Ca_t, \tag{4.9}$$

$(x, a) \in \mathbb{K}$. Since C is an invertible matrix, then Assumption 3.6(b) holds.

In addition, note that

$$L(y,s) = y + s, \tag{4.10}$$

$(y,s) \in X' \times S$. Then it follows that L has an inverse R in the second variable, which is

$$R(y,u) = u - y. \tag{4.11}$$

Therefore, Assumption 3.6(c) yields. Furthermore, since $\Delta \in C^2$, it results that Assumption 3.6(d) holds. On the other hand, Assumption 3.9 is satisfied since $A(x) = \mathbb{R}^n$ for each $x \in X$. Finally, it is obtained

$$W(x,a) = 2\left(Qx - \left(\left(RC^{-1}B\right)^T\right)^{-1} a\right), \tag{4.12}$$

where W is defined in (3.17), implying that the inverse of $W(x,\cdot)$ is

$$w(x,z) = \left(\left(RC^{-1}B\right)^T\right)^{-1}\left(Qx - \frac{1}{2}z\right), \tag{4.13}$$

which is a continuous function. Therefore, Assumption 3.13 is satisfied. □

Lemma 4.3. *VI functions for the linear quadratic problem satisfy*

$$v_n'(x) = 2K_n x, \tag{4.14}$$

for each $n = 1,2,\ldots$, *where*

$$K_n = Q + \alpha B^T\left(K_{n-1} - K_{n-1}C\left(R + C^T K_{n-1} C\right)^{-1} C^T K_{n-1}\right)B, \tag{4.15}$$

with $K_1 = Q$.

Proof. Observe that the validity of Theorem 3.15 is guaranteed by Lemmas 4.1 and 4.2. Now, since Q and R are negative definite, then

$$v_1(x) = x^T Q x. \tag{4.16}$$

By Theorem 3.15, it is known that $v_2(x)$ satisfies the Euler equation (3.37), and by (4.13), it follows that

$$v_2'(x) = 2Qx + \alpha B^T E\left[v_1'\left(\left(B + C\left(\left(RC^{-1}B\right)^T\right)^{-1} Q\right)x - \frac{1}{2}C\left(\left(RC^{-1}B\right)^T\right)^{-1} v_2'(x) + \xi\right)\right]. \tag{4.17}$$

Since $v_1'(x) = 2Qx$ and $E(\xi)$ is equal to zero vector, then

$$v_2'(x) = 2\left[Q + \alpha B^T Q\left(B + C\left(\left(RC^{-1}B\right)^T\right)^{-1} Q\right)\right]x - \alpha B^T QC\left(\left(RC^{-1}B\right)^T\right)^{-1} v_2'(x), \quad (4.18)$$

and by direct calculations, it is obtained that

$$v_2'(x) = 2K_2 x, \quad (4.19)$$

where

$$K_2 = Q + \alpha B^T\left(Q - QC\left(R + C^T QC\right)^{-1} C^T Q\right)B. \quad (4.20)$$

Now, suppose that $v_n'(x) = 2K_n x$ for $n > 2$, with K_n defined in (4.15). Then, by Theorem 3.15 and (4.13), it is known that

$$v_{n+1}'(x) = 2Qx + \alpha B^T E\left[v_n'\left(\left(B + C\left(\left(RC^{-1}B\right)^T\right)^{-1} Q\right)x - \frac{1}{2}C\left(\left(RC^{-1}B\right)^T\right)^{-1} v_{n+1}'(x) + \xi\right)\right]. \quad (4.21)$$

Then

$$v_{n+1}'(x) = 2\left[Q + \alpha B^T K_n\left(B + C\left(\left(RC^{-1}B\right)^T\right)^{-1} Q\right)\right]x - \alpha B^T K_n C\left(\left(RC^{-1}B\right)^T\right)^{-1} v_{n+1}'(x), \quad (4.22)$$

and using matrix algebra, it yields that

$$v_{n+1}'(x) = 2K_{n+1}x, \quad (4.23)$$

where K_{n+1} satisfies (4.15). □

Lemma 4.4. *The optimal policy for the linear quadratic problem is*

$$f(x) = -\alpha\left(R + C^T KC\right)^{-1} C^T KBx, \quad (4.24)$$

where K satisfies

$$K = Q + \alpha B^T\left(K - KC\left(R + C^T KC\right)^{-1} C^T K\right)B. \quad (4.25)$$

Proof. Lemma 4.3 and (4.13) allow to obtain

$$f_n(x) = \left(\left(RC^{-1}B\right)^T\right)^{-1}(Q - K_n)x, \tag{4.26}$$

for each $n = 1, 2, \ldots$ and $x \in X$. Moreover, the validity of Theorem 3.14 is guaranteed by Lemma 4.2, that is, $f_n(x) \rightarrow f(x)$, implying the convergence of the sequence $\{K_n\}$ which, according to its definition in (4.15), guarantees that its limit, denoted by K, must satisfy (4.25). Finally using matrix algebra (4.24) is obtained. □

5. Conclusion

In this paper a method to solve the optimal control problem is presented. This method is based on the use of the Euler equation. The procedure proposed to solve the optimal control problem is by means of an envelope formula and the use of the convergence of the maximizers of values iteration functions to a stationary optimal policy. Future work aims to study possible error bounds for approximating the maximizers toward the optimal policy.

References

[1] D. P. Bertsekas, *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, Belmont, Tenn, USA, 1987.

[2] E. B. Dynkin and A. A. Yushkevich, *Controlled Markov Processes*, vol. 235, Springer, New York, NY, USA, 1979.

[3] O. Hernández-Lerma and J. B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, vol. 30, Springer, New York, NY, USA, 1996.

[4] O. Hernández-Lerma and J. B. Lasserre, *Further Topics on Discrete-Time Markov Control Processes*, vol. 42, Springer, New York, NY, USA, 1999.

[5] N. Stokey, R. Lucas, and E. Prescott, *Recursive Methods in Economic Dynamics*, Harvard University Press, Cambridge, UK, 1989.

[6] K. J. Arrow, "A note on uncertainty and discounting in models of economic growth," *Journal of Risk and Uncertainty*, vol. 38, no. 2, pp. 87–94, 2009.

[7] W. A. Brock and L. J. Mirman, "Optimal economic growth and uncertainty: the discounted case," *Journal of Economic Theory*, vol. 4, no. 3, pp. 479–513, 1972.

[8] H. Cruz-Suárez, R. Montes-de-Oca, and G. Zacarías, "A consumption-investment problem modelled as a discounted Markov decision process," *Kybernetika*, vol. 47, no. 6, pp. 740–760, 2011.

[9] T. Kamihigashi, "Stochastic optimal growth with bounded or unbounded utility and with bounded or unbounded shocks," *Journal of Mathematical Economics*, vol. 43, no. 3-4, pp. 477–500, 2007.

[10] I. Karatzas and W. D. Sudderth, "Two characterizations of optimality in dynamic programming," *Applied Mathematics and Optimization*, vol. 61, no. 3, pp. 421–434, 2010.

[11] A. Jaśkiewicz and A. S. Nowak, "Discounted dynamic programming with unbounded returns: application to economic models," *Journal of Mathematical Analysis and Applications*, vol. 378, no. 2, pp. 450–462, 2011.

[12] D. Levhari and T. N. Srinivasan, "Optimal savings under uncertainty," *Review of Economic Studies*, vol. 36, pp. 153–163, 1969.

[13] L. J. Mirman and I. Zilcha, "On optimal growth under uncertainty," *Journal of Economic Theory*, vol. 2, no. 3, pp. 329–339, 1975.

[14] H. Cruz-Suárez and R. Montes-de-Oca, "Discounted Markov control processes induced by deterministic systems," *Kybernetika*, vol. 42, no. 6, pp. 647–664, 2006.

[15] L. M. Benveniste and J. A. Scheinkman, "On the differentiability of the value function in dynamic models of economics," *Econometrica*, vol. 47, no. 3, pp. 727–732, 1979.

[16] H. Cruz-Suárez and R. Montes-de-Oca, “An envelope theorem and some applications to discounted Markov decision processes,” *Mathematical Methods of Operations Research*, vol. 67, no. 2, pp. 299–321, 2008.

[17] P. Milgrom and I. Segal, “Envelope theorems for arbitrary choice sets,” *Econometrica. Journal of the Econometric Society*, vol. 70, no. 2, pp. 583–601, 2002.

[18] X.-R. Cao, *Stochastic Learning and Optimization*, Springer, New York, NY, USA, 2007.

[19] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, USA, 1970.

[20] D. Cruz-Suárez, R. Montes-de-Oca, and F. Salem-Silva, “Conditions for the uniqueness of optimal policies of discounted Markov decision processes,” *Mathematical Methods of Operations Research*, vol. 60, no. 3, pp. 415–436, 2004.

[21] A. De la Fuente, *Mathematical Methods and Models for Economists*, Cambridge University Press, Cambridge, UK, 2000.