Chapter 6
Continuous-Time Controlled Jump Markov Processes on the Finite Horizon

Mrinal K. Ghosh and Subhamay Saha

6.1 Introduction

This chapter studies continuous-time Markov decision processes and continuous-time zero-sum stochastic dynamic games. In the continuous-time setup, although the infinite horizon cases have been well studied, the corresponding literature on finite horizon case is few and far between. Infinite horizon continuous-time Markov decision processes have been studied by many authors (e.g. see [5] and the references therein). In the finite horizon case, Pliska [7] has used a semi-group approach to characterise the value function and the optimal control. But his approach yields only existential results. In this chapter, we show that the value function is a smooth solution of an appropriate dynamic programming equation. Our method of proof gives algorithms for computing the value function and an optimal control.

The situation is analogous for continuous-time stochastic dynamic Markov games. In this problem as well, the infinite horizon case has been studied in the literature [6]. To our knowledge, the finite horizon case has not been studied. In this chapter, we prove that the value of the game on the finite horizon exists and is the solution of an appropriate Isaacs equation. This leads to the existence of saddle point equilibrium.

The rest of our chapter is structured as follows. In Sect. 6.2 we analyse the finite horizon continuous-time MDP. Section 6.3 deals with zero-sum stochastic dynamic games. We conclude our chapter in Sect. 6.4 with a few remarks.

M.K. Ghosh (✉) • S. Saha
Department of Mathematics, Indian Institute of Science, Bangalore 560012, India
e-mail: mkg@math.iisc.ernet.in; subhamay@math.iisc.ernet.in

D. Hernández-Hernández and A. Minjárez-Sosa (eds.), *Optimization, Control, and Applications of Stochastic Systems*, Systems & Control: Foundations & Applications,
DOI 10.1007/978-0-8176-8337-5_6, © Springer Science+Business Media, LLC 2012

6.2 Finite Horizon Continuous-Time MDP

Throughout this chapter the time horizon is T. The control model we consider is given by

$$\{X, U, (\lambda(t,x,u), t \in [0,T], x \in X, u \in U), Q(t,x,u,\mathrm{d}z), c(t,x,u)\}$$

where each element is described below.

The state space X. The state space X is the set of states of the process under observation which is assumed to be a Polish space.

The action space U. The decision-maker dynamically takes his action from the action space U. We assume that U is a compact metric space.

The instantaneous transition rate λ. $\lambda : [0,T] \times X \times U \to [0,\infty)$ is a given function satisfying the following assumption:

(A1) λ is continuous and there exists a constant M such that

$$\sup_{t,x,u} \lambda(t,x,u) \leq M.$$

The transition probability kernel Q. For a fixed $t \in [0,T], x \in X, u \in U$, $Q(t,x,u,.)$ is a probability measure on X with $Q(t,x,u,\{x\}) = 0$. Q satisfies the following:

(A2) Q is weakly continuous, i.e. if $x_n \to x$, $t_n \to t$, $u_n \to u$, then for any $f \in C_b(X)$

$$\int_X f(z) Q(t_n, x_n, u_n, \mathrm{d}z) \to \int_X f(z) Q(t,x,u,\mathrm{d}z).$$

The cost rate c. $c : [0,T] \times X \times U \to [0,\infty)$ is a given function satisfying the following assumption:

(A3) c is continuous and there exists a finite constant $\tilde{C}$ such that

$$\sup_{t,x,u} c(t,x,u) \leq \tilde{C}.$$

Next we give an informal description of the evolution of the controlled system. Suppose that the system is in state x at time $t \geq 0$ and the controller or the decision-maker takes an action $u \in U$. Then the following happens on the time interval $[t, t + \mathrm{d}t]$:

1. The decision maker has to pay an infinitesimal cost $c(t,x,u)\mathrm{d}t$, and
2. A transition from state x to a set A (not containing x) occurs with probability

$$\lambda(t,x,u)\mathrm{d}t \int_A Q(t,x,u,\mathrm{d}z) + o(\mathrm{d}t);$$

or the system remains in state x with probability

$$1-\lambda(t,x,u)\mathrm{d}t+o(\mathrm{d}t).$$

Now we describe the optimal control problem. To this end we first describe the set of admissible controls. Let

$$\mathbf{u}:[0,T]\times X\to U$$

be a measurable function. Let $\mathscr{U}$ denote the set of all such measurable functions which is the set of admissible controls. Such controls are called Markov controls. For each $\mathbf{u}\in\mathscr{U}$, it can be shown that there exists is a strong Markov process $\{X_t\}$ (see [1, 3]) having the generator

$$\mathscr{A}_t^{\mathbf{u}}f(x)=-\lambda(t,x,\mathbf{u}(t,x))f(x)+\int_X f(z)Q(t,x,\mathbf{u}(t,x),\mathrm{d}z)$$

where f is a bounded measurable function.

For each $\mathbf{u}\in\mathscr{U}$, define

$$V^{\mathbf{u}}(t,x)=\mathbb{E}_{t,x}^{\mathbf{u}}\left[\int_t^T c(s,X_s,\mathbf{u}(s,X_s))\mathrm{d}s+g(X_T)\right] \tag{6.1}$$

where $g:X\to\mathbb{R}_+$ is the terminal cost function which is assumed to be bounded, continuous and $\mathbb{E}_{t,x}^{\mathbf{u}}$ is the expectation operator under the control $\mathbf{u}$ with initial condition $X_t=x$. The aim of the controller is to minimise $V^{\mathbf{u}}$ over all $\mathbf{u}\in\mathscr{U}$. Define

$$V(t,x)=\inf_{\mathbf{u}\in\mathscr{U}}\mathbb{E}_{t,x}^{\mathbf{u}}\left[\int_t^T c(s,X_s,\mathbf{u}(s,X_s))\mathrm{d}s+g(X_T)\right]. \tag{6.2}$$

The function V is called the value function. If $\mathbf{u}^*\in\mathscr{U}$ satisfies

$$V^{\mathbf{u}^*}(t,x)=V(t,x)\quad\forall(t,x),$$

then $\mathbf{u}^*$ is called an optimal control.
The associated dynamic programming equation is

$$\begin{cases}\frac{\mathrm{d}\varphi}{\mathrm{d}t}(t,x)+\inf_{u\in U}\big[c(t,x,u)-\lambda(t,x,u)\varphi(t,x)\\ +\lambda(t,x,u)\int_X\varphi(t,z)Q(t,x,u,\mathrm{d}z)\big]=0\\ \text{on } X\times[0,T)\quad\text{and}\\ \varphi(T,x)=g(x).\end{cases} \tag{6.3}$$

The importance of (6.3) is illustrated by the following verification theorem.

Theorem 6.2.1 *If* (6.3) *has a solution* φ *in* $C_b^{1,0}([0,T]\times X)$*, then* $\varphi = V$*, the value function. Moreover, if* $\boldsymbol{u}^*$ *is such that*

$$\left[c(t,x,\boldsymbol{u}^*(t,x)) - \lambda(t,x,\boldsymbol{u}^*(t,x))\varphi(t,x) + \lambda(t,x,\boldsymbol{u}^*(t,x))\int_X \varphi(t,z)Q(t,x,\boldsymbol{u}^*(t,x),\mathrm{d}z)\right]$$
$$= \inf_{u\in U}\left[c(t,x,u) - \lambda(t,x,u)\varphi(t,x) + \lambda(t,x,u)\int_X \varphi(t,z)Q(t,x,u,\mathrm{d}z)\right], \tag{6.4}$$

then $\boldsymbol{u}^*$ *is an optimal control.*

Proof. Using Ito-Dynkin formula to the solution φ of (6.3), we obtain

$$\varphi(t,x) \le \inf_{\mathbf{u}\in\mathscr{U}} \mathbb{E}^{\mathbf{u}}_{t,x}\left[\int_t^T c(s,X_s,\mathbf{u}(s,X_s))\mathrm{d}s + g(X_T)\right].$$

For $\mathbf{u} = \mathbf{u}^*$ as in the statement of the theorem, we get the equality

$$\varphi(t,x) = \mathbb{E}^{\mathbf{u}^*}_{t,x}\left[\int_t^T c(s,X_s,\mathbf{u}^*(s,X_s))\mathrm{d}s + g(X_T)\right].$$

The existence of such a $\mathbf{u}^*$ follows by a standard measurable selection theorem [2]. □

In view of the above theorem, it suffices to show that (6.3) has a solution in $C_b^{1,0}([0,T]\times X)$.

Theorem 6.2.2 *Under* (*A*1)–(*A*3)*, the dynamic programming equation* (6.3) *has a unique solution in* $C_b^{1,0}([0,T]\times X)$*.*

Proof. Let $\varphi(t,x) = \mathrm{e}^{-\gamma t}\psi(t,x)$ for some $\gamma < \infty$. Then from (6.3) we get,

$$\begin{cases} \mathrm{e}^{-\gamma t}\frac{\mathrm{d}\psi}{\mathrm{d}t}(t,x) - \gamma\mathrm{e}^{-\gamma t}\psi(t,x) + \inf\limits_{u\in U}\big[c(t,x,u) - \lambda(t,x,u)\mathrm{e}^{-\gamma t}\psi(t,x) \\ +\lambda(t,x,u)\int_X \mathrm{e}^{-\gamma t}\psi(t,z)Q(t,x,u,\mathrm{d}z)\big] = 0 \\ \text{on} \quad X\times[0,T) \quad \text{and} \\ \psi(T,x) = \mathrm{e}^{\gamma T}g(x). \end{cases}$$

Thus (6.3) has a solution if and only if

$$\begin{cases} \frac{\mathrm{d}\psi}{\mathrm{d}t}(t,x) - \gamma\psi(t,x) + \inf\limits_{u\in U}\big[\mathrm{e}^{\gamma t}c(t,x,u) - \lambda(t,x,u)\psi(t,x) \\ +\lambda(t,x,u)\int_X \psi(t,z)Q(t,x,u,\mathrm{d}z)\big] = 0 \\ \text{on} \quad X\times[0,T) \quad \text{and} \\ \psi(T,x) = \mathrm{e}^{\gamma T}g(x) \end{cases}$$

has a solution. The above differential equation is equivalent to the following integral equation:

$$\psi(t,x) = \mathrm{e}^{\gamma t} g(x) + \mathrm{e}^{\gamma t} \int_t^T \mathrm{e}^{-\gamma s} \inf_{u \in U} \Big[\mathrm{e}^{\gamma s} c(s,x,u) - \lambda(s,x,u) \psi(s,x) \\ + \lambda(s,x,u) \int_X \psi(s,z) Q(s,x,u,\mathrm{d}z) \Big] \mathrm{d}s\,.$$

Let $C_b^{\mathrm{unif}}([0,T] \times X)$ be the space of bounded continuous functions φ on $[0,T] \times X$ with the additional property that given $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\sup_x |\varphi(t+h,x) - \varphi(t,x)| < \varepsilon \quad \text{whenever} \quad |h| < \delta.$$

Suppose $\varphi_n \in C_b^{\mathrm{unif}}([0,T] \times X)$ and $\varphi_n \to \varphi$ uniformly. Then

$$|\varphi(t+h,x) - \varphi(t,x)| \leq |\varphi(t+h,x) - \varphi_n(t+h,x)| + |\varphi_n(t+h,x) - \varphi_n(t,x)| \\ |\varphi_n(t,x) - \varphi(t,x)|\,.$$

Given $\varepsilon > 0$, there exists n_0 such that $\sup\limits_{t,x} |\varphi_{n_0}(t,x) - \varphi(t,x)| < \frac{\varepsilon}{3}$, and for this n_0, there exists $\delta > 0$ such that $\sup\limits_{x} |\varphi_{n_0}(t+h,x) - \varphi_{n_0}(t,x)| < \frac{\varepsilon}{3}$ whenever $|h| < \delta$. Putting $n = n_0$, we get from the above inequality

$$\sup_x |\varphi(t+h,x) - \varphi(t,x)| < \varepsilon \quad \text{whenever} \quad |h| < \delta\,.$$

Thus $C_b^{\mathrm{unif}}([0,T] \times X)$ is a closed subspace of $C_b([0,T] \times X)$, and hence it is a Banach space.

Now for $\varphi \in C_b^{\mathrm{unif}}([0,T] \times X)$, it follows from the assumption on Q that $\int_X \varphi(t,z)$ $Q(t,x,u,\mathrm{d}z)$ is continuous in t,x and u. Define

$$\mathscr{T} : C_b^{\mathrm{unif}}([0,T] \times X) \to C_b^{\mathrm{unif}}([0,T] \times X) \quad \text{by}$$

$$\mathscr{T}\psi(t,x) = \mathrm{e}^{\gamma t} g(x) + \mathrm{e}^{\gamma t} \int_t^T \mathrm{e}^{-\gamma s} \inf_{u \in U} \Big[\mathrm{e}^{\gamma s} c(s,x,u) - \lambda(s,x,u) \psi(s,x) \\ + \lambda(s,x,u) \int_X \psi(s,z) Q(s,x,u,\mathrm{d}z) \Big] \mathrm{d}s\,.$$

For $\psi_1, \psi_2 \in C_b^{\mathrm{unif}}([0,T] \times X)$, we have

$$\begin{aligned}|\mathscr{T}\psi_1(t,x)-\mathscr{T}\psi_2(t,x)| &\le \mathrm{e}^{\gamma t}\int_t^T \mathrm{e}^{-\gamma s}2M||\psi_1-\psi_2||\mathrm{d}s\\ &= \frac{2M}{\gamma}\mathrm{e}^{\gamma t}[\mathrm{e}^{-\gamma t}-\mathrm{e}^{-\gamma T}]||\psi_1-\psi_2||\\ &= \frac{2M}{\gamma}[1-\mathrm{e}^{-\gamma(T-t)}]||\psi_1-\psi_2||\\ &\le \frac{2M}{\gamma}||\psi_1-\psi_2||.\end{aligned}$$

Thus if we choose $\gamma = 2M+1$, then $\mathscr{T}$ is a contraction and hence has a fixed point. Let φ be the fixed point. Then $\mathrm{e}^{-(2M+1)t}\varphi$ is the unique solution of (6.3). □

6.3 Zero-Sum Stochastic Game

In this section, we consider a zero-sum stochastic game. The control model we consider here is given by

$$\{X,U,V,(\lambda(t,x,u,v),t\in[0,T],x\in X,u\in U,v\in V,Q(t,x,u,v,\mathrm{d}z),r(t,x,u,v)\}$$

where X is the state space as before; U and V are the action spaces for player I and player II, respectively; λ and Q denote the rate and transition kernel, respectively, which now depend on the additional parameter v; and r is the reward rate. The dynamics of the game is similar to that of MDP with appropriate modifications. Here player I receives a payoff from player II. The aim of player I is to maximise his payoff, and player II seeks to minimise the payoff to player I.

Now we describe the strategies of the players. In order to solve the problem, we will need to consider Markov relaxed strategies. We denote the space of strategies of player I by $\mathscr{U}$ and that of player *II* by $\mathscr{V}$ where

$$\mathscr{U} = \{\mathbf{u}\,|\,\mathbf{u}:[0,T]\times X\to\mathscr{P}(U)\text{ measurable}\},$$

$$\mathscr{V} = \{\mathbf{v}\,|\,\mathbf{v}:[0,T]\times X\to\mathscr{P}(V)\text{ measurable}\}.$$

Now corresponding to λ, Q and r, define

$$\tilde{\lambda}(t,x,\mu,\nu) = \int_V\int_U \lambda(t,x,u,v)\mu(\mathrm{d}u)\nu(\mathrm{d}v),$$

$$\tilde{Q}(t,x,\mu,\nu) = \int_V\int_U Q(t,x,u,v)\mu(\mathrm{d}u)\nu(\mathrm{d}v),$$

$$\tilde{r}(t,x,\mu,\nu) = \int_V\int_U r(t,x,u,v)\mu(\mathrm{d}u)\nu(\mathrm{d}v),$$

where $\mu \in \mathscr{P}(U)$ and $\nu \in \mathscr{P}(V)$. As in the previous section, we make the following assumptions:

(A1$'$) λ is continuous and there exists a finite constant M such that

$$\sup_{t,x,u,v} \lambda(t,x,u,v) \leq M.$$

(A2$'$) Q is weakly continuous, i.e. if $x_n \to x$, $t_n \to t$, $u_n \to u$ and $v_n \to v$, then for any $f \in C_b(X)$

$$\int_X f(z)Q(t_n,x_n,u_n,v_n,\mathrm{d}z) \to \int_X f(z)Q(t,x,u,v,\mathrm{d}z).$$

(A3$'$) r is continuous and there exists a finite constant $\tilde{C}$ such that

$$\sup_{t,x,u,v} r(t,x,u,v) \leq \tilde{C}.$$

If the players use strategies $(\mathbf{u},\mathbf{v}) \in \mathscr{U} \times \mathscr{V}$, then the expected payoff to player I is given by

$$\mathbb{E}_{t,x}^{\mathbf{u},\mathbf{v}}\left[\int_t^T \tilde{r}(s,X_s,\mathbf{u}(s,X_s),\mathbf{v}(s,X_s))\mathrm{d}s + g(X_T)\right]$$

where g is the terminal reward function which is assumed to be bounded and continuous. Now we define the upper and lower values for our game. Define

$$\overline{V}(t,x) = \inf_{\mathbf{v}\in\mathscr{V}} \sup_{\mathbf{u}\in\mathscr{U}} \mathbb{E}_{t,x}^{\mathbf{u},\mathbf{v}}\left[\int_t^T \tilde{r}(s,X_s,\mathbf{u}(s,X_s),\mathbf{v}(s,X_s))\mathrm{d}s + g(X_T)\right].$$

Also define

$$\underline{V}(t,x) = \sup_{\mathbf{u}\in\mathscr{U}} \inf_{\mathbf{v}\in\mathscr{V}} \mathbb{E}_{t,x}^{\mathbf{u},\mathbf{v}}\left[\int_t^T \tilde{r}(s,X_s,\mathbf{u}(s,X_s),\mathbf{v}(s,X_s))\mathrm{d}s + g(X_T)\right].$$

The function $\overline{V}$ is called the upper value function of the game, and $\underline{V}$ is called the lower value function of the game. In the game, player I is trying to maximise his payoff and player II is trying to minimise the payoff of player I. Thus $\underline{V}$ is the minimum payoff that player I is guaranteed to receive and $\overline{V}$ is the guaranteed greatest amount that player II can lose to player I. In general $\underline{V} \leq \overline{V}$. If $\overline{V}(t,x) = \underline{V}(t,x)$, then the game is said to have a value. A strategy $\mathbf{u}^*$ is said to be an optimal strategy for player I if

$$\mathbb{E}_{t,x}^{\mathbf{u}^*,\mathbf{v}}\left[\int_t^T \tilde{r}(s,X_s,\mathbf{u}^*(s,X_s),\mathbf{v}(s,X_s))\mathrm{d}s + g(X_T)\right] \geq \overline{V}(t,x)$$

for any $t,x,\mathbf{v}$.

Similarly, $\mathbf{v}^*$ is called an optimal policy for player II if

$$\mathbb{E}_{t,x}^{\mathbf{u},\mathbf{v}^*}\left[\int_t^T \tilde{r}(s,X_s,\mathbf{u}(s,X_s),\mathbf{v}^*(s,X_s))\mathrm{d}s+g(X_T)\right]\leq \underline{V}(t,x)$$

for any $t,x,\mathbf{u}$. Such a pair $(\mathbf{u}^*,\mathbf{v}^*)$, if it exists, is called a saddle point equilibrium. Our aim is to find the value of the game and to find optimal strategies for both the players. To this end, consider the following pair of Isaacs equations:

$$\begin{cases}\frac{\mathrm{d}\varphi}{\mathrm{d}t}(t,x)+\inf\limits_{\nu\in\mathscr{P}(V)}\sup\limits_{\mu\in\mathscr{P}(U)}\big[\tilde{r}(t,x,\mu,\nu)-\tilde{\lambda}(t,x,\mu,\nu)\varphi(t,x)\\ +\tilde{\lambda}(t,x,\mu,\nu)\int_X\varphi(t,z)\tilde{Q}(t,x,\mu,\nu,\mathrm{d}z)\big]=0\\ \text{on } X\times[0,T) \quad \text{and}\\ \varphi(T,x)=g(x).\end{cases} \tag{6.5}$$

$$\begin{cases}\frac{\mathrm{d}\psi}{\mathrm{d}t}(t,x)+\sup\limits_{\mu\in\mathscr{P}(U)}\inf\limits_{\nu\in\mathscr{P}(V)}\big[\tilde{r}(t,x,\mu,\nu)-\tilde{\lambda}(t,x,\mu,\nu)\psi(t,x)\\ +\tilde{\lambda}(t,x,\mu,\nu)\int_X\psi(t,z)\tilde{Q}(t,x,\mu,\nu,\mathrm{d}z)\big]=0\\ \text{on } X\times[0,T) \quad \text{and}\\ \varphi(T,x)=g(x).\end{cases} \tag{6.6}$$

By Fan's minimax theorem [4], we have that if $\varphi\in C_b^{1,0}([0,T]\times X)$ is a solution of (6.5), then it is also a solution of (6.6) and vice versa. The importance of Isaacs equations is illustrated by the following theorem.

Theorem 6.3.1 *Let $\varphi^*\in C_b^{1,0}([0,T]\times X)$ be a solution of (6.5) and (6.6). Then*

(i) φ^* is the value of the game.
(ii) Let $(\boldsymbol{u}^*,\boldsymbol{v}^*)\in\mathscr{U}\times\mathscr{V}$ be such that

$$\begin{aligned}&\inf_{\nu\in\mathscr{P}(V)}\Big[\tilde{r}(t,x,\boldsymbol{u}^*(t,x),\nu)-\tilde{\lambda}(t,x,\boldsymbol{u}^*(t,x),\nu)\varphi^*(t,x)+\tilde{\lambda}(t,x,\boldsymbol{u}^*(t,x),\nu)\\&\qquad\int_X\varphi^*(t,z)\tilde{Q}(t,x,\boldsymbol{u}^*(t,x),\nu,\mathrm{d}z)\Big]\\&=\sup_{\mu\in\mathscr{P}(U)}\inf_{\nu\in\mathscr{P}(V)}\Big[\tilde{r}(t,x,\mu,\nu)-\tilde{\lambda}(t,x,\mu,\nu)\psi(t,x)+\tilde{\lambda}(t,x,\mu,\nu)\\&\qquad\int_X\psi(t,z)\tilde{Q}(t,x,\mu,\nu,\mathrm{d}z)\Big]\end{aligned} \tag{6.7}$$

and

$$\begin{aligned}
&\sup_{\mu\in\mathscr{P}(U)}\Big[\tilde r(t,x,\mu,\boldsymbol{\nu}^*(t,x))-\tilde\lambda(t,x,\mu,\boldsymbol{\nu}^*(t,x))\varphi^*(t,x)+\tilde\lambda(t,x,\mu,\boldsymbol{\nu}^*(t,x))\\
&\qquad\qquad\int_X\varphi^*(t,z)\tilde Q(t,x,\mu,\boldsymbol{\nu}^*(t,x),\mathrm{d}z)\Big]\\
&=\inf_{\nu\in\mathscr{P}(V)}\sup_{\mu\in\mathscr{P}(U)}\Big[\tilde r(t,x,\mu,\nu)-\tilde\lambda(t,x,\mu,\nu)\varphi(t,x)+\tilde\lambda(t,x,\mu,\nu)\\
&\qquad\qquad\qquad\int_X\varphi(t,z)\tilde Q(t,x,\mu,\nu,\mathrm{d}z)\Big].
\end{aligned}\tag{6.8}$$

Then $\mathbf{u}^*$ *is an optimal policy for player I and* $\boldsymbol{\nu}^*$ *is an optimal policy for player II.*

Proof. Let $\mathbf{u}^*$ be as in (6.7) and $\mathbf{v}$ be any arbitrary strategy of player II. Then by Ito-Dynkin formula applied to the solution φ, we obtain

$$\begin{aligned}
\varphi^*(t,x)&\le\mathbb{E}^{\mathbf{u}^*,\mathbf{v}}_{t,x}\left[\int_t^T\tilde r(s,X_s,\mathbf{u}^*(s,X_s),\mathbf{v}(s,X_s))\mathrm{d}s+g(X_T)\right]\\
&\le\inf_{\mathbf{v}\in\mathscr{V}}\mathbb{E}^{\mathbf{u}^*,\mathbf{v}}_{t,x}\left[\int_t^T\tilde r(s,X_s,\mathbf{u}^*(s,X_s),\mathbf{v}(s,X_s))\mathrm{d}s+g(X_T)\right]\\
&\le\underline{V}(t,x).
\end{aligned}$$

Now let $\mathbf{v}^*$ be as in (6.8) and let $\mathbf{u}$ be any arbitrary strategy of player I. Then again by Ito's formula we obtain

$$\begin{aligned}
\varphi^*(t,x)&\ge\mathbb{E}^{\mathbf{u},\mathbf{v}^*}_{t,x}\left[\int_t^T\tilde r(s,X_s,\mathbf{u}(s,X_s),\mathbf{v}^*(s,X_s))\mathrm{d}s+g(X_T)\right]\\
&\ge\inf_{\mathbf{v}\in\mathscr{V}}\mathbb{E}^{\mathbf{u},\mathbf{v}^*}_{t,x}\left[\int_t^T\tilde r(s,X_s,\mathbf{u}(s,X_s),\mathbf{v}^*(s,X_s))\mathrm{d}s+g(X_T)\right]\\
&\ge\overline{V}(t,x).
\end{aligned}$$

From the above two inequalities, it follows that

$$\varphi^*(t,x)=\overline{V}(t,x)=\underline{V}(t,x).$$

Hence φ^* is the value of the game. Moreover it follows that $(\mathbf{u}^*,\mathbf{v}^*)$ is a saddle point equilibrium. □

Now our aim is to find a solution of (6.5) (and hence of (6.6)) in $C_b^{1,0}([0,T]\times X)$. Our next theorem asserts the existence of such a solution.

Theorem 6.3.2 *Under* $(A1')$–$(A3')$, *equation* (6.5) *has a unique solution in* $C_b^{1,0}([0,T]\times X)$.

Proof. Let $\varphi(t,x)=\mathrm{e}^{-\gamma t}\psi(t,x)$ for some $\gamma<\infty$. Substituting in (6.5), we get

$$\begin{cases} \mathrm{e}^{-\gamma t}\frac{\mathrm{d}\psi}{\mathrm{d}t}(t,x)-\gamma\mathrm{e}^{-\gamma t}\psi(t,x)+\inf\limits_{\nu\in\mathscr{P}(V)}\sup\limits_{\mu\in\mathscr{P}(U)}\big[\tilde{r}(t,x,\mu,\nu)-\tilde{\lambda}(t,x,\mu,\nu)\mathrm{e}^{-\gamma t}\psi(t,x)\\ +\tilde{\lambda}(t,x,\mu,\nu)\int_X\mathrm{e}^{-\gamma t}\psi(t,z)\tilde{Q}(t,x,\mu,\nu,\mathrm{d}z)\big]=0\\ \text{on}\quad X\times[0,T)\quad\text{and}\\ \psi(T,x)=\mathrm{e}^{\gamma T}g(x). \end{cases}$$

Thus (6.5) has a solution if and only if

$$\begin{cases} \frac{\mathrm{d}\psi}{\mathrm{d}t}(t,x)-\gamma\psi(t,x)+\inf\limits_{\nu\in\mathscr{P}(V)}\sup\limits_{\mu\in\mathscr{P}(U)}\big[\mathrm{e}^{\gamma t}\tilde{r}(t,x,\mu,\nu)-\tilde{\lambda}(t,x,\mu,\nu)\psi(t,x)\\ +\tilde{\lambda}(t,x,\mu,\nu)\int_X\psi(t,z)\tilde{Q}(t,x,\mu,\nu,\mathrm{d}z)\big]=0\\ \text{on}\quad X\times[0,T)\quad\text{and}\\ \psi(T,x)=\mathrm{e}^{\gamma T}g(x) \end{cases}$$

has a solution. The above differential equation is equivalent to the following integral equation:

$$\psi(t,x)=\mathrm{e}^{\gamma t}g(x)+\mathrm{e}^{\gamma t}\int_t^T\mathrm{e}^{-\gamma s}\inf_{\nu\in\mathscr{P}(V)}\sup_{\mu\in\mathscr{P}(U)}\Big[\mathrm{e}^{\gamma s}\tilde{r}(s,x,\mu,\nu)-\tilde{\lambda}(s,x,\mu,\nu)\psi(s,x)$$
$$+\tilde{\lambda}(s,x,\mu,\nu)\int_X\psi(s,z)\tilde{Q}(s,x,\mu,,\nu,\mathrm{d}z)\Big]\mathrm{d}s.$$

Let $C_b^{\mathrm{unif}}([0,T]\times X)$ be the same space as defined in the previous section. Define

$$\mathscr{T}:C_b^{\mathrm{unif}}([0,T]\times X)\to C_b^{\mathrm{unif}}([0,T]\times X)\quad\text{by}$$

$$\mathscr{T}\psi(t,x)=\mathrm{e}^{\gamma t}g(x)+\mathrm{e}^{\gamma t}\int_t^T\mathrm{e}^{-\gamma s}\inf_{\nu\in\mathscr{P}(V)}\sup_{\mu\in\mathscr{P}(U)}\Big[\mathrm{e}^{\gamma s}\tilde{r}(s,x,\mu,\nu)$$
$$-\tilde{\lambda}(s,x,\mu,\nu)\psi(s,x)+\tilde{\lambda}(s,x,\mu,\nu)\int_X\psi(s,z)\tilde{Q}(s,x,\mu,\nu,\mathrm{d}z)\Big]\mathrm{d}s.$$

For $\psi_1,\psi_2\in C_b^{\mathrm{unif}}([0,T]\times X)$, we have

$$
\begin{aligned}
|\mathscr{T}\psi_1(t,x)-\mathscr{T}\psi_2(t,x)| &\leq \mathrm{e}^{\gamma t}\int_t^T \mathrm{e}^{-\gamma s}2M||\psi_1-\psi_2||\mathrm{d}s \\
&= \frac{2M}{\gamma}\mathrm{e}^{\gamma t}[\mathrm{e}^{-\gamma t}-\mathrm{e}^{-\gamma T}]||\psi_1-\psi_2|| \\
&= \frac{2M}{\gamma}[1-\mathrm{e}^{-\gamma(T-t)}]||\psi_1-\psi_2|| \\
&\leq \frac{2M}{\gamma}||\psi_1-\psi_2||.
\end{aligned}
$$

Thus if we choose $\gamma = 2M+1$, then $\mathscr{T}$ is a contraction and hence has a fixed point. Let φ be the fixed point. Then $\mathrm{e}^{-(2M+1)t}\varphi$ is the unique solution of (6.5). □

6.4 Conclusion

In this chapter we have established smooth solutions of dynamic programming equations for continuous-time controlled Markov chains on the finite horizon. This has led to the existence of an optimal Markov strategy for continuous-time MDP and saddle point equilibrium in Markov strategies for zero-sum games. We have used the boundedness condition on the cost function c for simplicity. For continuous-time MDP, if c is unbounded above, then we can show that $V(t,x)$ is the minimal nonnegative solution of (6.3) by approximating the cost function c by $c\wedge n$ for a positive integer n and then letting $n\to\infty$. If c is unbounded on both sides and it satisfies a suitable growth condition, then again we can prove the existence of unique solutions of dynamic programming equations in $C^{1,0}([0,T]\times X)$ with appropriate weighted norm; see [5] and [6] for analogous results.

References

1. A. Arapostathis, V. S. Borkar and M. K. Ghosh, *Ergodic Control of Diffusion Processes*, Cambridge University Press, 2011.
2. V. E. Benes, *Existence of optimal strategies based on specified information for a class of stochastic decision problems*, SIAM J. Control 8 (1970), 179–188.
3. M. H. A. Davis, *Markov Models and Optimization*, Chapman and Hall, 1993.
4. K. Fan, *Fixed-point and minimax theorems in locally convex topological linear spaces*, Proc. of the Natl. Academy of Sciences of the United States of America 38 (1952), 121–126.
5. X. Guo and O. Hernández-Lerma, *Continuous-Time Markov Decision Processes. Theory and Applications*, Springer-Verlag, 2009.
6. X. Guo and O. Hernández-Lerma, *Zero-sum games for continuous-time jump Markov processes in Polish spaces: discounted payoffs*, Adv. in Appl. Probab. 39 (2007), 645–668.
7. S. R. Pliska, *Controlled jump processes*, Stochastic Processes Appl. 3 (1975), 259–282.