RISK-SENSITIVE STOCHASTIC CONTROL AND GAMES

Subrata Golui
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
GUWAHATI-781039, INDIA
March, 2023

RISK-SENSITIVE STOCHASTIC CONTROL AND GAMES

A thesis submitted
in the partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

by

Subrata Golui
(Roll Number: 186123018)

to the
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
March, 2023

DECLARATION

It is certified that the work contained in this thesis entitled **"RISK-SENSITIVE STOCHASTIC CONTROL AND GAMES"** has done by me, under the supervision of **Dr. Chandan Pal**, Assistant Professor, Department of Mathematics, Indian Institute of Technology Guwahati for the award of the degree of Doctor of Philosophy and this work has not been submitted elsewhere for a degree.

March, 2023

Subrata Golui
Roll No. 186123018
Department of Mathematics
Indian Institute of Technology Guwahati

CERTIFICATE

It is certified that the work contained in this thesis entitled **“RISK-SENSITIVE STOCHASTIC CONTROL AND GAMES”** by **Subrata Golui,** a student of Department of Mathematics, Indian Institute of Technology Guwahati, for the award of the degree of Doctor of Philosophy has been carried out under my supervision and this work has not been submitted elsewhere for a degree.

`March, 2023`

Dr. Chandan Pal
Department of Mathematics
Indian Institute of Technology Guwahati

I dedicate this thesis to my "parents"

Gobordhan Golui and Radha Golui

and my "thesis advisor"

Dr. Chandan Pal

for their eternal love, affection, sacrifice, & hard work.

ACKNOWLEDGEMENT

With the divine grace of almighty God, the eventual destination of my last five years' academic journey has come; I'm on the way to my thesis submission. As a single flower can't shape a garland, likewise, my thesis wouldn't have been completed without the moral support and live involvement of my parents, teachers, friends, relatives, and well-wishers. I delightedly remember all of these lovely people by admiring their help and company.

Foremost, I express my deep sense of gratitude towards my mentor and path-finder Dr. Chandan Pal. I want to thank him for being the backbone of my academic enhancement with his erudite suggestions, constant encouragement, healthy criticism, generous support, and counsel during my whole Ph.D. life. A long experience of his research life, his diligence, and punctuality towards his work schedule was really a source of inspiration to me. I must acknowledge that my writing skill, speaking skills, personality, and mathematical thinking have been extremely profited by working under an intellectual thinker like him. But apart from academic matters, he is one of the most fantastic human beings I have ever met in my life. He was not only my research guide but my friend and philosopher and that is what makes him so special. I shall be grateful forever for his love, affection, and care toward me. I am also thankful to madam, Sanhita Pal (wife of Dr. Chandan Pal) to give me friendly suggestions and caring. I think of their family as my family member.

Next, I take the opportunity to express my gratefulness to Dr. Subhamay Saha. His spirit toward mathematics inspires me a lot. He helped me always whenever I needed it. My special thanks to him. From the very beginning of my Ph.D. journey, Dr. Chandan Pal (my guide) and Dr. Subhamay Saha are tolerating me. I will be thankful to them forever. Also, I express my gratefulness to the members of my doctoral committee, Dr. Ayon Ganguly, Prof. N. Selvaraju, and Prof. Sriparna Bandopadhyay, for investing their valuable time throughout the progress of my research. I'm highly indebted to them because of their comments, advice, and encouragement to enhance my views and thinking on solving problems. Also, I confess my heartfelt thankfulness to all the other faculties in our department for their immense help and support in every bit of my learning phase from the beginning. In addition, I would like to give a special acknowledgment to the Indian Institute of Technology Guwahati for providing academic, financial, and medical support. Apart from this, I extend my thanks to the departmental all staff to make my journey smooth by assisting with various departmental and technical issues. I would like to take the pleasure of thanking, Dr, Somnath Pradhan, Prof. Mrinal Kanti Ghosh, Dr. Manikandan Rangaswamy, Prof. Sankar K. De, Prof. Murari Mitra, Prof. Binayak S. Choudhury, Prof. Sumita Mallik, Mrittyunjay sir, Suchinta sir, and Mohim sir for their continuous support, inspiration, and helpful suggestions.

Academic life can't be trouble-free without the help of seniors, and I am fortunate enough to have some amiable seniors to interact with. I must thank Avijit da, Mrityunjoy da, Buddha dev da, Kousik da, Nilay da, and Dibakar da for indulging me in every situation. My special appreciation goes to my six neighbors in the Lohit hostel: Niladri, Gaurav, Arup, Sayantan, Santanu, and Arnab with whom I spent the maximum time at the campus. Apart from this, I want to convey my thankfulness to Suroj, Shekhar, Sunit, Aniruddha, Sandip, Amit, Shilpi, Bivakar, Kannan, Rupchand, Sangita, Priyam, Moumita, Pranay, Suprabhat for sharing some quality time with me throughout the journey.

This space is too limited to convey my gratitude towards all my amazing friends, but the golden memories which we have enjoyed in the last five years will be cherished in my heart forever.

Last but not least, I am head over heels indebted to my family members and all relatives for their affection, support, and everything for my life. Their selfless sacrificial life always pushes me to achieve my goal. It is my parents' dream to see me with a doctorate degree. I dedicate this thesis to my parents, Shri. Gobordhan Golui and Smt. Radha Golui and my thesis supervisor, Dr. Chandan Pal under whose careful protection I have been able to enjoy my life.

"May be candle be lightened forever. The joy is not of light alone, but of the presence of those who played the role behind the curtain."

March, 2023

With regards
Subrata Golui

"I attribute my success to this: I never gave or took any excuse"

— **Florence Nightingale**

"Tell me and I forget. Teach me and I remember. Involve me and I learn"

— **Benjamin Franklin**

"Mathematics is the most beautiful and most powerful creation of the human spirit"

— **Stefan Banach**

ABSTRACT

This thesis considers risk-sensitive stochastic control and game problems on countable/Borel state space for discrete/continuous-time Markov decision processes (MDPs) under certain Lyapunov conditions. Here, infinite-horizon control/game problems are analyzed with various cost criteria. The controllers can take action in discrete/continuous-time from their admissible strategies.

In the single-player setup, we investigate infinite-horizon discounted cost criterion for continuous-time pure jump controlled Markov processes. The controller tries to minimize his/her payoff through a Markov decision process and finds an optimal risk-sensitive control in the class of Markov control.

When there is more than one controller, the stochastic control problem is referred to as a stochastic game problem, we study zero/nonzero-sum game problems. For zero-sum game problems, we consider infinite-horizon discounted/ergodic cost criteria for discrete/continuous-time Markov decision processes (MDPs) on countable/Borel state space. Here, player 1 is a maximizing player and player 2 is a minimizing player. So, player 1 tries to maximize his/her reward while player 2 always tries to minimize his/her costs. For these problems, a saddle-point equilibrium point is achieved. To study nonzero-sum game problems, we consider infinite-horizon ergodic cost criteria for continuous-time Markov decision processes (CTMDPs) on a countable state space. Here, each player tries to minimize his/her ergodic payoff criterion. We establish the existence of a Nash-equilibrium in the class of stationary strategies for non-zero sum ergodic game problems. For each model, suitable real-life examples are provided to support the models.

First, we investigate risk-sensitive continuous-time discounted control problem for pure jump Markov processes on general Borel state space. The transition and the cost rates are possibly unbounded. We establish the existence and uniqueness of the solution to the Hamilton-Jacobi-Bellman (HJB) equation under certain Lyapunov conditions. Also, we provide proof of the existence of optimal risk-sensitive control in the class of Markov control and completely characterize the optimal control. Moreover, we consider an illustrative example to support our results and assumptions.

After that, a continuous-time risk-sensitive zero-sum stochastic game for controlled Markov decision processes with discounted cost criterion on countable state space is analyzed. Here, the transition and cost rates are possibly unbounded. Under a Foster-Lyapunov condition, we prove the existence of the value of the game and saddle-point equilibrium in the class of admissible strategies by studying the corresponding Hamilton-Jacobi-Isaacs (HJI) equation. Also, an illustrative example is used to support our results.

Next, we consider risk-sensitive zero-sum stochastic games for controlled continuous-time Markov decision processes on a general state space with discounted cost criteria. The transition and cost rates are allowed to be unbounded. Under a stability assumption, we prove the existence of the value of the game and saddle-point equilibrium in the class of Markov strategies and give a characterization in terms of the corresponding Hamilton-Jacobi-Isaacs (HJI) equation. Moreover, we illustrate our results and assumptions by example.

After that, we analyze risk-sensitive zero-sum stochastic games for controlled discrete-time Markov decision processes with ergodic cost criteria on countable/compact

state space and Borel action spaces. For countable state space case, the payoff function is nonnegative and possibly unbounded and it is a real-valued and bounded function for compact state space case. Under a certain Lyapunov type stability assumption on the dynamic, we establish the existence of the value and a saddle-point equilibrium, for countable state space case. But for compact state space case, we establish these results without any Lyapunov type stability assumptions. Using the stochastic representation of the principal eigenfunction of the associated optimality equation, we completely characterize all possible saddle-point strategies in the class of stationary Markov strategies. Also, we present and analyze an illustrative example.

Subsequently, a nonzero-sum stochastic game for continuous-time Markov decision processes on a denumerable state space with risk-sensitive ergodic cost criterion is considered. We allow the transition rates and cost rates to be unbounded. Under a certain stability assumption, we show the existence of a solution of the corresponding system of coupled HJB equations which leads to the existence of a Nash equilibrium in stationary strategies. We establish this using an approach involving principal eigenvalues associated with the HJB equations. Furthermore, we completely characterize Nash equilibrium in the space of stationary Markov strategies by exploiting appropriate stochastic representation of principal eigenfunctions. Also, a controlled population system is considered to illustrate our results.

Finally, we investigate risk-sensitive continuous-time stochastic zero-sum games for controlled Markov decision processes with ergodic cost criteria. Here, the transition and the cost rates may be unbounded. Under a Lyapunov stability condition, we provide proof of the existence of the value of the game and a saddle-point equilibrium in the class of all stationary strategies. This is accomplished by establishing the existence of a principal eigenpair for the corresponding Hamilton-Jacobi-Isaacs (HJI) equation. This, in turn, is established by using a nonlinear version of Kreĭn-Rutman theorem. We also give a characterization of the saddle-point equilibrium in terms of the corresponding HJI equation. Lastly, we use a controlled population system to illustrate our results.

Contents

CHAPTER 1

Introduction

In this thesis, we investigate risk-sensitive stochastic control and stochastic game problems. Optimal control problems are single-player, single-objective dynamic decision problems. Optimal control problems can be roughly divided into two parts: one is a deterministic optimal control problem and the other is a stochastic optimal control problem. Here we are interested in the latter one, i.e., the stochastic optimal control problem. A stochastic control problem is an optimization problem, where controllers try to control the evolution of a stochastic process by taking some actions. These actions either result in a cost or yield a reward. The aim of the controller is to maximize, in case of reward, or minimize, in case of cost, which is accumulated either over a finite or infinite time horizon. Stochastic control plays an important role in many real-life applications such as in telecommunication, queueing systems, epidemiology, finance, etc. The risk-sensitive stochastic control problem is a special type of stochastic control problem. Stochastic optimal control problems for controlled Markov processes can be classified in two ways. (1) Discrete-time; (2) Continuous-time.

A discrete-time Markov decision process (DTMDP) is described by a family of controlled transition kernels. The evolution of the process is as follows: at each time epoch the controller chooses an action from a suitable action set. For his/her action, he/she has to pay a certain cost. Then at the next time epoch, the state moves to a new state, the distribution of which is given by the controlled stochastic kernel. Then again, the same procedure repeats. This process either goes for a finite time period, in which case it is known as a finite horizon problem, or goes for an infinite time period, in which case it is known as an infinite horizon problem. In infinite-horizon problems, based on cost criteria, this can be further classified into discounted cost problems and ergodic cost problems. In the control problem, the basic aim is to find optimal controls, that is a control that will optimize the total expected cost and characterize all possible optimal controls via Hamilton-Jacobi-Bellman (HJB) equation (see, Chapter 2, Section 3).

In the continuous-time setup, one of the most widely studied controlled processes is the continuous-time Markov decision process (CTMDP). In order to describe a CTMDP model we need to specify its controlled transition kernels and controlled waiting time distributions. Once the process jumps to a new state the controller chooses an action. Now the process remains in the present state for a random amount of time which is exponentially distributed. The parameter of the exponential distribution is determined by the controller through his action. Because of his/her action, the controller has to pay a cost continuously. The cost rate depends on the state and also the action. After spending a random amount of time in the present state the process jumps to a new state according to the controlled transition kernel. Here again, there are finite and infinite-horizon problems. The basic problem is the same, to characterize the value function (see, Chapter 2, Section 3) and to find optimal controls (see, Chapter 2, Section 5).

In stochastic control problems, a controller wishes to optimize only the expected value of the total random payoff. But minimizing just the first moment of a random variable may not always be good from an optimization point of view and in this approach, controllers ignore the risk. In order to address this concern one of the approaches that are available in the literature is a risk-sensitive approach. In this approach, one investigates an exponential of the total random payoff which takes into account the attitude of the controller with respect to risk. So, besides the above classification of a stochastic control problem, we can classify stochastic problems into two categories with respect to risk, say, risk-neutral and risk-sensitive problems. In a risk-neutral stochastic optimal problem, the cost criterion is of additive nature and so, the controller wishes to optimize the expected value of the random payoff without caring about the variance of the random payoff. Where in a risk-sensitive stochastic optimal problem, the cost criterion is of "exponential of integral" nature which takes into account expectation as well as the higher order moments of the random payoff. Consequently, the analysis of risk-sensitive control is technically more involved because of the exponential nature of the cost.

Now, following [90], we describe the risk-sensitive criterion in mathematical forms. Let Y be a random cost accrued over a finite/infinite time horizon. Let the (constant) coefficient of absolute risk aversion be given by $\theta \in \mathbb{R}$. Let U_θ be a utility function given by

$$U_\theta(x) = \begin{cases} sgn(\theta)e^{\theta x}, & \text{if } \theta \neq 0, \\ x, & \text{if } \theta = 0. \end{cases}$$

Suppose the decision maker evaluates the random cost Y via $E(U_\theta(Y))$. A certainty equivalent of Y is a number $J(\theta, Y)$ such that

$$U_\theta(J(\theta, Y)) = E(U_\theta(Y)). \tag{1.0.1}$$

Therefore for a person with the risk-sensitive factor θ, paying the random cost Y is tantamount to paying a deterministic cost $J(\theta, Y)$. It is easy to see that

$$J(\theta, Y) = \frac{1}{\theta}\log(Ee^{\theta Y}).$$

If $\theta > 0$, then $J(\theta, Y) \geq EY$. Thus the decision maker is paying a higher cost for being risk-averse. If $\theta < 0$, then the decision maker is risk-seeking. Finally, $\theta = 0$ corresponds to the risk-neutral case.

The risk of a random quantity is also associated with its variance in the literature. That is why the controller may wish to minimize both mean and variance as described above. In the risk-neutral case: the decision make minimizes $E(Y)$. In the risk-sensitive case: the decision maker seeks to minimize $J(\theta, Y) = \frac{1}{\theta}\log(E(e^{\theta Y}))$. For a small value of θ, by Taylor series expansion,

$$J(\theta, Y) \approx E(Y) + \frac{\theta}{2}Var(Y). \tag{1.0.2}$$

From the above explanation, we conclude that, in the risk-neutral criterion, the controller wants to optimize the expected value of the total payoff. But in the risk-sensitive criterion, the controller considers the expected value of the exponential of the total payoff and so it captures the effects of the higher order moments of the cost as well as its expectation. This difference makes the risk-sensitive case significantly different from its risk-neutral counterpart. As a result, the risk-sensitive criterion gives better protection from the risk.

The right-hand side of (1.0.2) is a standard utility employed in a portfolio optimization problem. However, the above may not be suitable for games [90]. Note that there are other non-linear risk-sensitive utility functions, e.g., power, logarithm, etc. But these utility functions do not lead to certainty equivalence.

We now briefly describe the game model. Game theory is the analysis of the conflict that involves multi-person decision-making. By choosing appropriate strategies, each person/player either tries to maximize his/her gain or minimize the total cost he/she has to pay. In game theory, one tries to find a strategy for all players such that each of them obtains his/her desired result. There are two types of games; one is deterministic and the other is stochastic. We are interested in stochastic game theory, which deals with systems having uncertainty. Like the one-player setup, stochastic games, based on their duration, can be classified into finite-horizon games and infinite-horizon games. In the infinite-horizon case, again, there are two basic payoff criteria, namely, discounted payoff and average payoff. Moreover, stochastic games have two sub-classes; (1) zero-sum stochastic games and (2) nonzero-sum stochastic games.

In zero-sum stochastic games, there are two players with opposite interests. As the name suggests, the sum of the cost functions of the players is identically equal to zero, i.e., what one player gains are lost to the other player. As we have discussed, in zero-sum games there is a single performance criterion: one player tries to maximize it and the other player tries to minimize it and the total aggregate sum of gains of the players is zero. Here, the value of the game is defined to be the maximum possible gain of maximizing player or the minimum possible loss of the minimizing player when they use their best strategies. This pair of best strategies are known to be a saddle-point equilibrium which will be discussed later elaborately. However, in a nonzero-sum game, there are two or more players and each of them tries to minimize (or maximize) his/her performance criterion by choosing appropriate strategies as a response to the other players' strategies. It is clear from the name that in a nonzero-sum game the sum of all individual performance criteria may not be zero. Here, an equilibrium is a set of strategies such that unilateral deviation from the strategy corresponding to any player by that player is disadvantageous to him/her. This leads to the Nash equilibrium which will be discussed later elaborately.

Risk-sensitive game is a generalization of a classical stochastic game in which the degree of risk aversion or risk tolerance of the players in the games is explicitly parameterized in the cost criterion and influences the outcome of the game directly.

In a zero-sum game, as defined above, two players are strictly competitive and for the risk-neutral zero-sum game case, the sum of the expected random cost and the expected random payoff is zero. Thus, in this case, we can define saddle-point equilibrium. However, this is not going to be the case for any non-linear utility function, including the one we are addressing here. Thus risk-sensitive zero-sum games have to be studied via Nash equilibria [90]; we discuss this in more detail in Chapter 7. Such (risk-sensitive) zero-sum games are called nonstandard zero-sum games. In the thesis, we use such nonstandard zero-sum games with risk-sensitive exponential utility.

Here, the risk-sensitive stochastic zero-sum game has primarily been formulated from the viewpoint of the minimizer player (suppose player 2) who is risk-averse. The maximizer is a virtual player (suppose player 1) who is antagonistic to the minimizing player. Such games have applications in queueing systems where each player treats the rest of the players as a superplayer antagonistic to him/her. Moreover, such a model is relevant in worst-case scenarios, e.g., in financial applications when a risk-averse investor is trying to maximize his long-term portfolio gain against the market which, by default, is the minimizer in this case. We refer to [1] for a zero-sum stochastic game in a flow control problem in discrete-time, and [44], [46] for analogous problems in continuous-time. Such a game is also applied in [11] for a temporal capital asset pricing model (CAPM) where each investor treats the rest of the investors as a superplayer antagonistic to him/her. But, to handle many realistic situations, it is also necessary to study games that involve many players and which are not zero-sum.

In sum up, in this thesis, we investigate risk-sensitive stochastic single and multi-controller problems for controlled Markov processes with infinite-horizon cost criterion on discrete/continuous-time. Here, we consider discounted as well as ergodic cost criteria.

In a finite-horizon problem, a well-defined ending time or ending condition can be specified which clearly defines the end of the problem while in an infinite-horizon simulation, there is no well-defined ending time or condition. In the case of infinite-horizon problems, we are almost always interested in reporting on the long-run performance of the system. Thus, the desired starting conditions of interest are sometime in the far-off future. In addition, because we want long-run performance there is no natural ending point of interest. Moreover, another significance is that there are situations where problems with infinite time horizons arise in a natural way, e.g. when the random lifetime of an investor is considered. However more important is the fact that Markov decision models with finite but large horizon can be approximated by models with infinite time horizon. The latter one is often simpler to solve and admits mostly a (time) stationary optimal policy. Some examples of infinite-horizon situations are telecommunications, hospital emergencies, etc. The infinite-horizon discounted cost (DC) and average cost (AC) evaluation criterion can be seen as two opposite extremes in the spectrum of possible criteria that can be considered, in the sense that the first one captures primarily the performance of the process at the present and near future, and the second captures the performance at the distant future. When decisions are made frequently, so that the discount rate is very close to 1, or when performance criterion cannot easily be described in economic terms, the decision maker may prefer to compare policies on the basis of their average expected reward instead of their expected total discounted reward, see [100]. The ergodic problem for controlled Markov processes refers to the problem of minimizing a time average cost over an infinite time horizon. Hence the cost over any finite initial time segment does not affect the ergodic cost. This makes the analysis of the ergodic problem analytically more difficult. In continuous-time ergodic game problems, we investigate nonzero-sum as well as zero-sum game problems. In the control problem, we consider discounted cost criterion for continuous-time pure jump processes, where we find the optimal strategy of the corresponding Hamilton-Jacobi-Bellman (HJB) equation. We study zero-sum games with infinite-horizon, discounted, and ergodic cost criteria, where we investigate the existence of a saddle-point equilibrium of the corresponding Hamilton-Jacobi-Isaacs (HJI) equations for discrete/continuous-time Markov decision processes. And, for the nonzero-sum game, we deal with infinite-horizon ergodic cost evaluation criterion for continuous-time Markov decision processes (CTMDPs) where we find the existence of a Nash-equilibrium of the corresponding couple of optimality (HJB) equations.

1.1 Literature survey

The prime intent of this thesis is to formulate and analyze the risk-sensitive stochastic control and stochastic games with various cost criteria. A short literature survey on the recent development of this topic is addressed below.

The study of optimal control theory started dating back to the early 17th century, see [105]. This class of problems received tremendous impetus from the Second World War due to a wide range of applications.

In stochastic optimal control, the Markov chain is a very important and fundamental model for stochastic dynamic systems and is widely used for modeling control problems that arise naturally in many real-life problems, for example in queueing models, epidemiology models, birth-death models, etc. In [17], [59], [102], [103] the authors studied Markov decision processes (MDPs) and showed the importance of MDPs in real-life applications.

The study of discrete-time Markov decision processes is more than 50 years old. We refer to [70], for one of the earlier works. After that, the investigation of discrete-time MDPs has been done by various authors in various setups. Finite horizon problems and infinite-horizon problems with countable state space and compact action spaces are well understood. Several authors have also analyzed discrete-time MDPs on general state space problems, see the books by Bertsekas and Shreve [17], Hernandez-Lerma [68] and Ross [103], and references therein for details. See, [8], for a detailed survey of results on average cost problems for discrete-time MDP. In the literature, continuous time MDP is also a well-studied object, see [59]. If the transition rates are bounded, the most widely used approach in solving CTMDP is to convert it into an equivalent discrete-time MDP and then use the theory of discrete-time MDP. This technique is known as the uniformization technique, for details see [17], [70], [102], and the references therein. For the case of unbounded transition rates, the analysis can be done by a suitable generalization of this approach. Several authors have studied the case of unbounded transition rates and unbounded cost under various hypotheses, see [55], [59], [60], [119] and the references therein.

All the references described till now are on risk-neutral control. Risk-sensitive control problems are studied extensively in the literature due to their applications in queueing systems and mathematical finance, especially in portfolio optimization, large deviation theory, and its connection to robust control problems and stochastic dynamic games, for more details see [14], [15], [18], [43], [73, P. 125], [114, part-II] and the references therein.

Also, see, [21], [30], [32], [47], [53], [62], [81], [82], [94], [108], [120]. Earlier works on risk-sensitive control goes back to Howard and Matheson [71] and Jacobson [73]. In [66], the authors study risk-sensitive control of discrete-time Markov chains with countable state space. Here the authors solve the problem by converting the original problem to an equivalent stochastic dynamic game. Risk-sensitive control of discrete-time Markov chains with general state space has been investigated in [30], [74]. In [24], the authors use a direct approach of multiplicative Poisson equation to study ergodic cost risk-sensitive problems for discrete-time Markov chains. For important contributions to the risk-sensitive control of DTMDP on a general state space, see [30], [32]. In the literature, risk-sensitive control problems for CTMDPs are an important class of stochastic optimal control problems and have been widely studied under different sets of conditions. Finite horizon risk-sensitive CTMDPs for countable state space were studied in [47], [53], [108] and for infinite-horizon risk-sensitive CTMDPs for countable/general state space, we refer to [47], [54], [62], [81], [82], [94], [120]. In [47], [82], the transition and cost rates are assumed to be bounded while [53], [62] considered unbounded cost and transition rates. Continuous-time infinite-horizon risk-sensitive discounted control problem is studied in [62] and the papers [21], [81], [82] studied infinite-horizon risk-sensitive CTMDPs for ergodic cost criterion while in [47], [94], the authors considered discounted as well as ergodic cost criterion. Infinite horizon risk-sensitive CTMDP for piecewise deterministic Markov decision processes has been studied in [54]. Risk-sensitive control problem has also been studied for continuous-time diffusions, see [20], [37], [114] and the references therein.

Besides the above discussion about ergodic problem, we now focus more on infinite-horizon ergodic cost criterion. As already discussed, the study of ergodic control problem is interesting and useful when we deal with infinite-horizon problem. The analysis of stochastic systems with the risk-sensitive ergodic criterion can be traced back to the seminal papers by Jacobson in [73] and Howard and Matheson in [71]. The literature on risk-sensitive MDP under ergodic cost criterion is quite extensive, e.g., [2], [21], [24], [25], [26], [30], [31], [47], [52], [67], [71], [74], [81], [82], [94], [111], [114]. The risk-sensitive ergodic cost stochastic optimal control problems for CTMDPs are first considered in [47]. In this respect, we mention some interesting works, [10], [78], [79] studying multiplicative ergodic theorem for geometrically stable Markov processes. In [79, p. 77, sec. 2.4], the authors made a strong connection between ergodic theory and Perron-Frobenius eigenvalue theory. For the classical approach to studying risk-sensitive ergodic control problems based on equivalent game formulation, one can see [36]. In [25], the authors investigated risk-sensitive ergodic cost criterion for discrete-time MDP with bounded cost using a simultaneous Doeblin condition on a countable state space. Also, see [2], [24], and the references therein for multiplicative ergodic theory. These papers used the eigenvalue approach to study risk-sensitive ergodic control problems. The authors in [47], [94] used the results of [78], [79] to study their risk-sensitive ergodic control problems. Also, in the context of controlled diffusions, the eigenvalue approach is used in [5], [6], [7], [19] to study the risk-sensitive ergodic control problems. In [21], the authors studied risk-sensitive discrete/continuous-time ergodic control problems for controlled Markov processes with countable state space . They established the existence of a principal eigenpair of the associated ergodic HJB equation. For this, they first studied the corresponding Dirichlet eigenvalue problems on finite sets and then pass to the limit by increasing the finite sets to countable state space. In [2], authors used a novel technique to provide a variational formula for infinite-horizon risk-sensitive reward on a compact state and action spaces. They build a nonlinear version of Kreĭn-Rutman theorem to study the corresponding ergodic HJB equation which leads to the existence of optimal ergodic control. In the literature, the Kreĭn-Rutman theorem has been studied extensively, see [2], [4], [80], [84], [86], [91], [92], [93], [116] and the references therein. In the pioneering works of Perron [98] and Frobenius [38], it was proved that the spectral radius of a nonnegative square matrix is an eigenvalue with a nonnegative eigenvector. In [80], Kreĭn-Rutman extended the results of Perron and Frobenius's theory to a positive compact linear operator, which is the celebrated Kreĭn-Rutman theorem. For Kreĭn-Rutman theorem of a linear/nonlinear operator on ordered Banach space (under a different set of conditions), see [2], [4], [86], [91], [92], [93], [116] and the references therein.

Most of the above discussions are on countable state space. But, we see that there are many real-life situations where the state space may be uncountable, for example, the chemical reaction model, Gaussian model, etc. One can see [65], [100], and references therein for real-life examples. Although risk-sensitive control of CTMDPs on a countable state space has been studied extensively, the corresponding literature in the context of risk-sensitive control of CTMDPs on a general state space is rather limited. Some exceptions are [49], [54], [65], [94]. In [62], the authors analyzed a continuous-time discounted control problem with unbounded transition and cost rates (nonnegative cost) on countable state space. In Chapter 2, we have extended the results of [62] for general state space. To the best of our knowledge, this is the first work that deals with infinite-horizon discounted risk-sensitive control problems for CTMDPs on a general state space with unbounded cost and transition rates and the controls can be admissible controls.

In the 1930s, Von Neumann and Morgenstern extended the results of single control problems to multiple decision-makers through their pioneering work, see [106]. But, Shapely is one of the persons who first introduced the idea of stochastic games in his paper [104] and has been studied extensively in the literature due to its immense applications; see [12], [16], [45], [48], [57], [110], [112], [118] and the references therein. Such games for MDP have been widely studied in the literature. For a survey on zero-sum games in discrete-time, see [107]. We refer [23], [34], [39], [56], [57], [58], [97], [112], for stochastic games with risk-neutral criterion. Semi-Markov games have also been investigated widely in the literature. Important works include [83] and many others.

There has been a lot of work on both zero-sum and non-zero sum games with risk-sensitive utility. Discrete-time zero-sum risk-sensitive stochastic games have been considered by several authors, see [12], [16], [41] and references therein. Infinite-horizon zero-sum risk-sensitive stochastic games on a countable state space have been studied in [12]. The paper [16] extended the results of [12] for general state space. Both papers first treat the discounted cost criterion and then study the ergodic cost criterion using vanishing discount asymptotics. Both in the papers [12] and [16], the authors considered bounded cost function. In [16], the ergodic cost criterion is analyzed under a local minorization property and a Lyapunov condition. Nonzero-sum games for risk-sensitive discrete-time Markov decision processes have been studied by [13], [77], [113].

Continuous-time risk-sensitive zero-sum stochastic games on finite-horizon are studied in [48], [109], [117], while infinite horizon zero-sum games are studied in [51], [95]. In [48], [117], unbounded costs and transition rates are considered while [109] considers unbounded transition but bounded cost on Borel state space. The discounted risk-sensitive zero-sum game for CTMDPs is studied in [51] with unbounded cost and unbounded transition rates. Let us also mention the recent work of [45], [95], which study the infinite-horizon risk-sensitive stochastic zero-sum game for CTMDPs with bounded cost and transition rates. But this boundedness requirement restricts our domain of application since in many real-life situations, we see that the reward/cost and transition rates are unbounded as for example in queueing control and population processes. As already said that [16], studied the ergodic cost criterion under a local minorization property and a Lyapunov condition. The analogous results in continuous-time setup are carried out in [45]. In this respect, we mention that the authors in the paper [22] studied risk-sensitive zero-sum ergodic game problems for controlled diffusion processes in $\mathbb{R}^d$. Using the eigenvalue approach, they have completely characterized all possible saddle-point equilibrium in the space of stationary Markov strategies.

One can see [40], [110], and the references therein for continuous-time risk-sensitive stochastic nonzero-sum game. In [110], finite-horizon cost criterion was studied, while [40] considered infinite-horizon ergodic cost criterion.

Besides the above discussion, this thesis studies more risk-sensitive stochastic games with unbounded transition/cost rates for MDP. In Chapter 3, we have extended the control problem in article [62] to a two-person infinite-horizon discounted-cost risk-sensitive zero-sum stochastic games for CTMDPs. In Chapter 4, we have generalized the work of Chapter 3 (that is a zero-sum stochastic game problem on countable space) to a zero-sum stochastic game problem on general state space. Also, this is an extension of [95] from bounded rates to unbounded rates.

In the article [21], authors studied risk-sensitive discrete/continuous-time ergodic control problems for controlled Markov processes with countable state space. But the corresponding literature in the context of ergodic risk-sensitive stochastic games for MDP with unbounded cost is very limited. One exception is [113]. We have extended the work of [21] to a discrete-time zero-sum game in Chapter 5. Also, in this chapter, we have analyzed the same problem (zero-sum game problem) for bounded costs on compact state space by getting inspired by the work of [2], where the authors analyzed an infinite-horizon risk-sensitive ergodic control problem on compact state and action space. Moreover, this chapter (Chapter 5) can be considered as an extension of the results of the manuscript [12] to the case with unbounded cost. Then, we study nonzero-sum risk-sensitive continuous-time stochastic games with ergodic costs for possibly unbounded transition and cost rates on countable state space in Chapter 6 by getting inspired from [21]. At last, we study a continuous-time zero-sum game problem for ergodic cost criterion on countable state space with possibly unbounded transition and cost rates in Chapter 7 along the line on the control problem [21]. Also, the results of this chapter can be seen as an extension of [45], where the authors studied a continuous-time zero-sum game problem with bounded transition and cost rates on a countable space.

1.2 Structure of the Thesis

This thesis is designed with eight chapters. **Chapter 1** addresses the general introduction of the risk-sensitive stochastic control and game problems and presents the historical background on the development of these problems. The objective and motivation for the thesis and some preliminaries are also provided in this chapter. The rest of the thesis is organized as follows:

In **Chapter 2**, we study risk-sensitive discounted control problems for continuous-time pure jump Markov processes taking values in general state space. The transition rates of underlying continuous-time jump Markov processes and the cost rates are allowed to be unbounded. Under certain Lyapunov conditions, we establish the existence and uniqueness of the solution to the Hamilton-Jacobi-Bellman (HJB) equation. Also, we prove the existence of optimal risk-sensitive control in the class of Markov control and completely characterize the optimal control.

Chapter 3 considers a two-person zero-sum stochastic game for controlled continuous-time Markov decision processes with risk-sensitive discounted cost criterion on countable state space. The transition and cost rates are possibly unbounded. Here, we prove the existence of the value of the game and saddle-point equilibrium in the class of admissible (feedback) strategies under a Foster-Lyapunov condition. We achieve our results by studying the corresponding Hamilton-Jacobi-Isaacs (HJI) equation.

Chapter 4 studies zero-sum stochastic games for controlled continuous-time Markov decision processes on a general state space with risk-sensitive discounted cost criterion. The transition and cost rates are possibly unbounded. Under a stability assumption, we prove the existence of a saddle-point equilibrium in the class of Markov strategies and give a characterization in terms of the corresponding Hamilton-Jacobi-Isaacs (HJI) equation. Also, we illustrate our results and assumptions by an example.

The main target of **Chapter 5** is to study zero-sum stochastic games for controlled discrete-time Markov processes with risk-sensitive ergodic cost criterion with countable/compact state space and Borel action spaces. The payoff function is nonnegative and possibly unbounded for countable state space case and for compact state space case, it is a real-valued and bounded function. For countable state space case, under a certain Lyapunov type stability assumption on the dynamics we establish the existence of the value and a saddle-point equilibrium. For compact state space case, we establish these results without any Lyapunov type stability assumptions. Using the stochastic representation of the principal eigenfunction of the associated optimality equation, we completely characterize all possible saddle-point strategies in the class of stationary Markov strategies. Also, we present and analyze an illustrative example.

The main focus of the work given in **Chapter 6** is to analyze a nonzero-sum stochastic game for continuous-time Markov decision processes on a denumerable state space with risk-sensitive ergodic cost criterion. Transition rates and cost rates are allowed to be unbounded. Under a Lyapunov type stability assumption, we show that the corresponding system of coupled HJB equations admits a solution that leads to the existence of a Nash equilibrium in stationary strategies. We establish this using an approach involving principal eigenvalues associated with the HJB equations. Furthermore, exploiting appropriate stochastic representation of principal eigenfunctions, we completely characterize Nash equilibrium in the space of stationary Markov strategies.

In **Chapter 7**, we consider a continuous-time zero-sum stochastic game with controlled Markov decision processes and with risk-sensitive ergodic cost criterion. Here the transition and the cost rates may be unbounded. We prove the existence of the value of the game and a saddle-point equilibrium in the class of all stationary strategies under a Lyapunov stability condition. This is accomplished by establishing the existence of a principal eigenpair for the corresponding Hamilton-Jacobi-Isaacs (HJI) equation. This, in turn, is established by using a nonlinear version of Kreĭn-Rutman theorem. We then obtain a characterization of the saddle-point equilibrium in terms of the corresponding HJI equation. Finally, we use a controlled population system to illustrate our results.

Finally, **Chapter 8** concludes the thesis with a summary of these works by highlighting the contribution made in the field of risk-sensitive stochastic control and game problems. Several directions for the expansion of the ideas given in this thesis are also addressed here, which can be considered as the scope of further development of risk-sensitive stochastic control and games.

CHAPTER 2

Risk-sensitive discounted cost criterion for continuous-time Markov decision processes on a general state space

2.1 Introduction

In this chapter, we study the risk-sensitive discounted cost criterion for continuous-time Markov decision processes (CTMDPs) with Borel state space. In the literature, risk-sensitive control problems for CTMDPs are an important class of stochastic optimal control problems and have been widely studied under different sets of conditions. Finite horizon risk-sensitive CTMDPs for countable state space were studied in [47], [53], [108], and for infinite horizon risk-sensitive CTMDPs we refer to [47], [54], [62], [81], [82]. For important contributions to the risk-sensitive control of discrete-time MDP on a general state space, see [30], [32]. Although risk-sensitive controls of CTMDPs on a countable state space have been studied extensively, the corresponding literature in the context of risk-sensitive control of CTMDPs on a general state space is rather limited. Some exceptions are [54], [65], [94].

In the paper [94], the authors studied risk-sensitive control of pure jump processes in a general state space. They considered bounded transition and cost rates and established an optimal risk-sensitive control in terms of the HJB equation in the class of all Markov controls. This boundedness requirement, however, imposes some restrictions in applications, for instance in queueing control and population processes, where the transition and reward/cost rates are usually unbounded. Also, there are many real-life situations where the state space may be uncountable, for example, the chemical reaction model, Gaussian model, etc. One can see [65], [100], and references therein for real-life examples. In [65], the author considered the finite-horizon risk-sensitive control problem for CTMDPs on a Borel state space with unbounded transition and cost rates and proved the existence of optimal control via the HJB equation.

In this chapter, we have extended [94] from bounded transition and cost rates to unbounded rates. In [62], the authors analyzed a continuous-time risk-sensitive discounted control problem with unbounded transition and cost rates (nonnegative cost) on countable state space. This chapter is an extension of the results of [62] for general state space. To the best of our knowledge, this is the first work that deals with infinite-horizon discounted risk-sensitive control for CTMDPs on a general state space with unbounded cost and transition rates and the controls can be admissible.

The main objective of this chapter is to prove the existence of the solution of the HJB equation and the characterization completely of the optimal risk-sensitive control in the class of Markov controls. We first consider bounded transition and cost rates and establish the existence of a solution to the corresponding HJB equation by Banach's fixed point theorem as in [94]. Then we will relax the bounded hypothesis and extend this result to unbounded transition and cost rates. We characterize the value function via the HJB equation. Also, we prove the existence of an optimal control in the class of Markov control and the HJB characterization of the optimal risk-sensitive control and prove its complete characterization. In Corollary 2.5.1, we prove that if the cost and transition rates are bounded, then an optimal control exists for our model.

The rest of this chapter is structured as follows. Section 2.2 deals with the description of the problem, required notations, some Assumptions, and preliminary results. In Section 2.3, we give a continuity-compactness Assumption and prove the stochastic representation of the solution of the HJB equation (2.3.1). In Section 2.4, we truncate our transition and cost rates and prove the existence of the unique solution to the HJB equation. A complete characterization of optimal control is proven in Section 2.5. In Section 2.6, we illustrate our theory and assumptions with an illustrative example. The content of this chapter is based on the published article [49].

2.2 Continuous-time discounted cost MDP

The model of CTMDP is a five-tuple that consists of the following elements:

$$\mathbb{M} := \{S, A, (A(x) \subset A, x \in S), c(x,a), q(\cdot|x,a)\},$$

- a Borel space S, called the state space, whose elements are referred to as states of the system and the corresponding Borel σ-algebra is $\mathcal{B}(S)$. (Throughout the whole chapter we consider that for any Borel space X, the corresponding Borel σ-algebra is $\mathcal{B}(X)$.)

- A is the action set, which is assumed to be Borel space with the Borel σ-algebra $\mathcal{B}(A)$.
- for each $x \in S$, $A(x) \in \mathcal{B}(A)$ denotes the set of admissible actions for state x. Let $K := \{(x,a) | x \in S, a \in A(x)\}$, which is a Borel subset of $S \times A$.
- the measurable function $c : K \to \mathbb{R}_+$ denotes the cost rate function. We require the cost function $c(x,a)$ to measure (or evaluate) the utility of taking action a at state x.
- given any $(x,a) \in K$, the transition rate $q(\cdot|x,a)$ is a Borel measurable signed kernel on S given K. That is, $q(\cdot|x,a)$ satisfies countable additivity; $q(D|x,a) \geq 0$ where $(x,a) \in K$, $x \notin D$ and $D \in \mathcal{B}(S)$. Moreover, we assume that $q(\cdot|x,a)$ satisfies the following conservative and stable conditions: for any $x \in S$,

  $$q(S|x,a) \equiv 0 \quad \text{and}$$
  $$q^*(x) := \sup_{a \in A(x)} q_x(a) < \infty,$$

  where $q_x(a) := -q(\{x\}|x,a) \geq 0$. We need the transition rates to specify the random dynamic evolution of the system.

Next, we give an informal description of the evolution of the CTMDPs as follows. The controller observes continuously the current state of the system. When the system is in state $x \in S$ at time $t \geq 0$, he/she chooses action $a_t \in A(x)$ according to some control. As a consequence of this, the following happens:

- the controller incurs an immediate cost at rate $c(x, a_t)$; and
- after a random sojourn time (i.e., the holding time at state x), the system jumps to a set $B \in \mathcal{B}(S)$ $(x \notin B)$ of states with the transition probability $\dfrac{q(B|x,a_t)}{q_x(a_t)}$ determined by the transition rates $q(dy|x,a_t)$. The distribution function of the sojourn time is $(1 - e^{-\int_t^{t+x} q_x(a_s)ds})$. (see Proposition B.8 in [59, p. 205] for details).

When the state of the system transits to a new state $y \neq x$, the above procedure is repeated. Thus, the controller tries to minimize his/her costs with respect to some performance criterion $\mathscr{J}_\alpha(\cdot,\cdot,\cdot)$, which in our present case is defined by (2.2.2), below. To formalize what is described above, below we describe the construction of continuous-time Markov decision processes (CTMDPs) under possibly admissible feedback controls. To construct the underlying CTMDPs (as in [63], [75], [99], [100]) we introduce some notations: let $S_\Delta := S \cup \{\Delta\}$ (with some "isolated" state $\Delta \notin S$), $\Omega_0 := (S \times (0,\infty))^\infty$, $\Omega_k := (S \times (0,\infty))^k \times S \times (\{\infty\} \times \{\Delta\})^\infty$ for $k \geq 1$ and $\Omega := \cup_{k=0}^{\infty} \Omega_k$. Let $\mathscr{F}$ be the

Borel σ-algebra on Ω. Then we obtain the measurable space $(\Omega, \mathscr{F})$. For some $k \geq 1$, and sample $\omega := (x_0', \theta_1, x_1', \cdots, \theta_k, x_k', \cdots) \in \Omega$, define

$$T_0(\omega) := 0,\ T_k(\omega) := T_{k-1}(\omega) + \theta_k,\ T_\infty(\omega) := \lim_{k\to\infty} T_k(\omega).$$

Using $\{T_k\}$, we define the state process $\{\xi_t\}_{t\geq 0}$ as

$$\xi_t(\omega) := \sum_{k\geq 0} I_{\{T_k \leq t < T_{k+1}\}} x_k' + I_{\{t\geq T_\infty\}}\Delta,\ \text{for } t \geq 0 \text{ (with } T_0 := 0). \tag{2.2.1}$$

Here, I_E denotes the indicator function of a set E, and we use the convention that $0 + z =: z$ and $0z =: 0$ for all $z \in S_\Delta$. Obviously, $\xi_t(\omega)$ is right-continuous on $[0, \infty)$. We denote $\xi_{t-}(\omega) := \liminf_{s\to t-} \xi_s(\omega)$. From eq. (2.2.1), we see that $T_k(\omega)$ $(k \geq 1)$ denotes the k-th jump moment of $\{\xi_t, t \geq 0\}$, x_{k-1}' is the state of the process on $[T_{k-1}(\omega), T_k(\omega))$, $\theta_k = T_k(\omega) - T_{k-1}(\omega)$ plays the role of sojourn time at state x_{k-1}', and the sample path $\{\xi_t(\omega), t \geq 0\}$ has at most denumerable states $x_k'(k = 0, 1, \cdots)$. The process after T_∞ is regarded to be absorbed in the state Δ. Thus, let $q(\cdot|\Delta, a_\Delta) :\equiv 0$, $A_\Delta := A \cup \{a_\Delta\}$, $A(\Delta) := \{a_\Delta\}$, $c(\Delta, a) :\equiv 0$ for all $a \in A_\Delta$, where a_Δ is an isolated point.
To precisely define the criterion, we need to introduce the concept of control as in [61], [63], [72], [76], [115]. Take the right-continuous σ-algebras $\{\mathscr{F}_t\}_{t\geq 0}$ with $\mathscr{F}_t := \sigma(\{T_k \leq s, \xi_{T_k} \in S\} : 0 \leq s \leq t, k \geq 0)$. For all $t \geq 0$, $\mathscr{F}_{s-} =: \bigvee_{0\leq t<s} \mathscr{F}_t$. Now define a σ-algebra $\mathscr{P} := \sigma(D \times \{0\}, C \times (s, \infty) : D \in \mathscr{F}_0, C \in \mathscr{F}_{s-})$, which denotes the σ-algebra of predictable sets on $\Omega \times [0, \infty)$ related to $\{\mathscr{F}_t\}_{t\geq 0}$. To complete the specification of a stochastic optimal control problem, we need, of course, to introduce an optimality criterion. This requires defining the class of controls as below.

Definition 2.2.1. *An admissible strategy/policy, denoted by* $\pi := \{\pi(t)\}_{t\geq 0}$, *is a measurable map* $\pi(t)(\omega) := \pi(\omega, t)$ *from* $(\Omega \times [0, \infty), \mathscr{P})$ *onto* $(A_\Delta, \mathcal{B}(A_\Delta))$ *satisfying* $\pi(t)(\omega) \in A(\xi_{t-}(\omega))$ *for all* $\omega \in \Omega$ *and* $t \geq 0$.

The set of all admissible policies/controls is denoted by Π_{Ad}. *A control* $\pi \in \Pi_{Ad}$, *is called a Markov if* $\pi(t)(\omega) = \pi(\xi_{t-}(\omega), t)$ *i.e.,* $\pi(\omega, t) = \pi(\xi_{t-}(\omega), t)$ *for every* $\omega \in \Omega$ *and* $t \geq 0$, *where* $\xi_{t-}(\omega) := \lim_{s\uparrow t} \xi_s(\omega)$. *For notational simplicity, we would not write* ω *anywhere throughout the rest of this chapter. We denote by* Π_M *the family of all Markov controls.*

For any compact metric space Y, let $\mathcal{P}(Y)$ denote the space of probability measures on Y with Prohorov topology. Under Assumption $(A1)$ below, for any initial state $x \in S$ and any control $\pi \in \Pi_{Ad}$, Theorem 4.27 in [76] yields the existence of a unique probability measure denoted by P_x^π on $(\Omega, \mathscr{F})$. Let E_x^π be the expectation operator with respect to P_x^π. Fix any discounted factor $\alpha > 0$. For any $\pi \in \Pi_{Ad}$ and $x \in S$, the risk-sensitive discounted criterion is defined as

$$\mathscr{J}_\alpha(\theta, x, \pi) := \frac{1}{\theta} \log\left\{ E_x^\pi \left[\exp\left(\theta \int_0^\infty e^{-\alpha t} c(\xi_t, \pi(t)) dt \right) \right] \right\}, \tag{2.2.2}$$

provided that the integral is well defined, where $\{\xi_t\}_{t\geq 0}$ is the Markov process corresponding to $\pi := \{\pi(t)\}_{t\geq 0} \in \Pi_{Ad}$ and $\theta \in (0,1]$ denotes a risk-sensitive parameter and the limiting case of $\theta \to 0$ is the risk-neutral case. For each $x \in S$, let

$$\mathscr{J}_\alpha^*(\theta, x) = \inf_{\pi \in \Pi_{Ad}} \mathscr{J}_\alpha(\theta, x, \pi).$$

A control $\pi^* \in \Pi_{Ad}$ is said to be optimal if $\mathscr{J}_\alpha(\theta, x, \pi^*) = \mathscr{J}_\alpha^*(\theta, x)$ for all $x \in S$. The objective of this chapter is to provide conditions for the existence of optimal control and introduce a HJB characterization of such control. Since the logarithm is an increasing function, instead of studying $\mathscr{J}_\alpha(\theta, x, \pi)$, we will consider $\tilde{J}_\alpha(\theta, x, \pi)$ on $[0,1] \times S \times \Pi_{Ad}$ defined by

$$\tilde{J}_\alpha(\theta, x, \pi) := E_x^\pi \left[\exp\left(\theta \int_0^\infty e^{-\alpha t} c(\xi_t, \pi(t)) dt \right) \right]. \tag{2.2.3}$$

Obviously, $\tilde{J}_\alpha(\theta, x, \pi) \geq 1$ for $(\theta, x) \in [0,1] \times S$ and $\pi \in \Pi_{Ad}$, and we have π^* is optimal if and only if $\inf_{\pi \in \Pi_{Ad}} \tilde{J}_\alpha(\theta, x, \pi) = \tilde{J}_\alpha(\theta, x, \pi^*) =: \tilde{J}_\alpha^*(\theta, x)$ $\forall x \in S$.

Since the transition rates $q(dy|x,a)$ and costs rates $c(x,a)$ are allowed to be unbounded, we next give conditions for the non-explosion of $\{\xi_t, t \geq 0\}$ and finiteness of $\mathscr{J}_\alpha(\theta, x, \pi)$, which had been widely used in CTMDPs; see, for instance, [59], [61], [62], [63], [101] and references therein.

(A1) There exists a real-valued Borel measurable function $V \geq 1$ on S and constants $\rho_0 > 0$, $M_0 > 0$, $L_0 \geq 0$ and $0 < \rho_1 < \min\{\alpha, \rho_0^{-1}\alpha^2\}$ such that

(i) $\int_S V(y) q(dy|x,a) \leq \rho_0 V(x) \quad \forall (x,a) \in K$;

(ii) $\sup_{a \in A(x)} q_x(a) \leq M_0 V(x) \quad \forall x \in S$;

(iii) $\sup_{a \in A(x)} c(x,a) \leq \rho_1 \log V(x) + L_0 \quad \forall x \in S$.

Remark 2.2.2.

(a) Note that, when the transition rates are bounded i.e., $\sup_{x\in S} q^*(x) < \infty$, Assumptions (A1)(i)-(ii) are satisfied by taking a suitable constant value of $V(x)$.

(b) Under Assumption (A1)(iii) the criterion (2.2.3) is well defined and finite; see Proposition 2.2.3(c) below.

Proposition 2.2.3. *Grant Assumption (A1). Then for any control $\pi \in \Pi_{Ad}$ and $(\theta, x) \in [0,1] \times S$, the following results are true:*

(a) $P^\pi_x(T_\infty = \infty) = 1$, $P^\pi_x(\xi_0 = x) = 1$, *and* $P^\pi_x(\xi_t \in S) = 1$ *for all* $t \geq 0$;

(b) $E^\pi_x[V(\xi_t)] \leq e^{\rho_0 t} V(x)$ *for all* $t \geq 0$;

(c) *We have*

$$\tilde{J}_\alpha(\theta, x, \pi) \leq \frac{\alpha^2}{\alpha^2 - \rho_0\rho_1\theta} e^{\theta L_0/\alpha} [V(x)]^{\frac{\rho_1\theta}{\alpha}} \leq \frac{\alpha^2}{\alpha^2 - \rho_0\rho_1} e^{L_0/\alpha} V(x).$$

Also, we get

$$\mathscr{J}^*_\alpha(\theta, x) \leq \log\left(\frac{\alpha^2}{\alpha^2 - \rho_0\rho_1}\right) + \frac{L_0}{\alpha} + \frac{\rho_1}{\alpha} \log V(x) \quad \forall \theta \in (0, 1], x \in S. \tag{2.2.4}$$

Proof. For parts (a) and (b), see, [61] and ([63, Theorem 3.1]).
Proof of part (c): Observe that $d(-e^{-\alpha t})$ is a probability measure on $[0, \infty)$. For any $\pi \in \Pi_{Ad}$ and $(\theta, x) \in [0, 1] \times S$, by (2.2.3) and Jensen's inequality we have

$$\begin{aligned}
\tilde{J}_\alpha(\theta, x, \pi) &= E^\pi_x\left[\exp\left(\int_0^\infty \frac{\theta}{\alpha} c(\xi_t, \pi(t)) d(-e^{-\alpha t})\right)\right] \\
&\leq E^\pi_x\left[\int_0^\infty \exp\left(\frac{\theta}{\alpha} c(\xi_t, \pi(t))\right) d(-e^{-\alpha t})\right].
\end{aligned}$$

By Assumption $(A1)$ and part (b) we obtain

$$\begin{aligned}
\tilde{J}_\alpha(\theta, x, \pi) &\leq E^\pi_x\left[\int_0^\infty \exp\left(\frac{\theta}{\alpha}(\rho_1 \log V(\xi_t) + L_0)\right) d(-e^{-\alpha t})\right] \\
&= e^{\theta L_0/\alpha}\left[\int_0^\infty E^\pi_x\left(V(\xi_t)^{\frac{\rho_1\theta}{\alpha}}\right) d(-e^{-\alpha t})\right] \\
&\leq e^{\theta L_0/\alpha}\left[\int_0^\infty (E^\pi_x[V(\xi_t)])^{\frac{\rho_1\theta}{\alpha}} d(-e^{-\alpha t})\right] \quad (\text{since} \ \ \rho_1\theta < \alpha) \\
&\leq \alpha e^{\theta L_0/\alpha} [V(x)]^{\frac{\rho_1\theta}{\alpha}} \left[\int_0^\infty \exp\left(\frac{\rho_0\rho_1\theta t}{\alpha} - \alpha t\right) dt\right] \\
&= \frac{\alpha^2}{\alpha^2 - \rho_0\rho_1\theta} e^{\theta L_0/\alpha} [V(x)]^{\frac{\rho_1\theta}{\alpha}},
\end{aligned}$$

where the last equality holds due to the fact that $\rho_0\rho_1\theta < \alpha^2$.
Next observe that $\sup_{\theta \in [0,1]} \tilde{J}^*_\alpha(\theta, x) \leq \frac{\alpha^2}{\alpha^2 - \rho_0\rho_1} e^{L_0/\alpha} V(x)$, and

$$\sup_{\theta \in (0,1]} \mathscr{J}^*_\alpha(\theta, x) = \sup_{\theta \in (0,1]} \frac{1}{\theta} \log \tilde{J}^*_\alpha(\theta, x) \leq \sup_{\theta \in (0,1]} \frac{1}{\theta}\left(\log \frac{\alpha^2}{\alpha^2 - \rho_0\rho_1\theta}\right) + \frac{L_0}{\alpha} + \frac{\rho_1}{\alpha} \log V(x).$$

Also, doing a simple and direct calculation, we achieve (2.2.4). ☐

In [47], [82], the authors used the Dynkin's formula within the class of Markov controls by using the Markov property of the state process $\{\xi_t\}_{t\geq 0}$. But this Markov property may fail to hold when we study within the class of admissible controls, and consequently, here we can't directly apply the Dynkin formula. Hence we assume the following condition so that we can apply the Dynkin's formula for a large enough class of functions, which had been widely used in CTMDPs; see, for instance, [53], [62], [65].

(*A*2) The Borel measurable function $V^2 \geq 1$ on S satisfies the following Lyapunov condition

$$\int_S q(dy|x,a)V^2(y) \leq \rho_2 V^2(x) + b_0 \quad \forall \ (x,a) \in K,$$

for some constants $0 < \rho_2 < \alpha$ and $b_0 \geq 0$. Here V is as in Assumption $(A1)$.
We now introduce some frequently used notations.

- $C_c^\infty(a,b)$ denotes the set of all infinitely differentiable functions on (a,b) with compact support.

- Let $A_{as}([0,1]\times S)$ denote the space of all functions which are real-valued and differentiable almost everywhere with respect to the first variable $\theta \in [0,1]$. Given any real-valued function $W \geq 1$ on S and a Borel set $[0,1]$, a real-valued function φ on $[0,1]\times S$ is called W bounded if $\|\varphi\|_W^\infty := \sup_{(\theta,x)\in[0,1]\times S} \frac{|\varphi(\theta,x)|}{W(x)} < \infty$. Denote $W([0,1]\times S)$ the Banach space of all W-bounded functions. When $W \equiv 1$, $B_1^\infty([0,1]\times S)$ is the space of all bounded functions on $[0,1]\times S$.
  Now define $L_W^\infty([0,1]\times S) := \{\varphi : [0,1]\times S \to \mathbb{R} : \varphi \in B_W^\infty([0,1]\times S) \cap A_{as}([0,1]\times S)\}$.

2.3 Stochastic representation of a solution to the HJB equation

In this section, we prove that if the HJB equation for the cost criterion (2.2.3) has a solution then we will give a stochastic representation of that solution. Using dynamic programming heuristics, the HJB equations for the discounted cost criterion (2.2.3) is given by

$$\begin{cases} \alpha\theta\frac{\partial\varphi_\alpha}{\partial\theta}(\theta,x) &= \inf\limits_{a\in A(x)}\left[\int_S q(dy|x,a)\varphi_\alpha(\theta,y) + \theta c(x,a)\varphi_\alpha(\theta,x)\right], \\ & 1 \leq \varphi_\alpha(\theta,x) \leq \frac{\alpha^2}{\alpha^2-\rho_0\rho_1\theta}e^{\theta L_0/\alpha}(V(x))^{\frac{\rho_1\theta}{\alpha}} \ \text{ for } \ (\theta,x)\in[0,1]\times S, \end{cases} \tag{2.3.1}$$

for each $x \in S$ and a.e. $\theta \in [0,1]$ where the upper bound of $\varphi_\alpha(\theta,x)$ is motivated by Proposition 2.2.3.

Remark 2.3.1. *To prove the existence of optimal control for bounded cost and transition rates, in [94], the authors studied the following HJB equation having a solution* $\phi_\alpha(\theta, x)$ *on* $[0,1] \times S$ *such that*

$$\begin{cases} \alpha\theta \frac{\partial \phi_\alpha}{\partial \theta}(\theta, x) & = \inf_{a \in A(x)} \left[\int_S q(dy|x,a)\phi_\alpha(\theta, y) + \theta c(x,a)\phi_\alpha(\theta, x) \right], \text{ for } (\theta, x) \in [0,1] \times S, \\ & \lim_{\theta \to 0} \phi_\alpha(\theta, x) = 1 \text{ uniformly in } x \in S. \end{cases} \tag{2.3.2}$$

From the arguments for the existence of a unique solution to the equation (2.3.2), it is necessary to have $\phi_\alpha(\theta, x)$ *converges to* 1 *uniformly in* x *as* $\theta \to 0$. *But, it is not true in general when the cost and transition rates are unbounded; for more details see Example 3.2 in [62]. In this chapter, we replace the uniform convergence condition with the above new one.*

To ensure the existence of an optimal control, in addition to Assumptions $(A1)$ and $(A2)$, we also need the following continuity and compactness conditions.

(*A*3) The following conditions hold:

(i) for each $x \in S$, the set $A(x)$ is compact;

(ii) for any fixed $x \in S$, the functions $c(x,a)$ and $q(\cdot|x,a)$ are continuous in $a \in A(x)$;

(iii) for any given $x \in S$, the function $\int_S V(y)q(dy|x,a)$ is continuous in $a \in A(x)$, where V is introduced in Assumption $(A1)$.

Remark 2.3.2. *Assumptions* $(A3)$ *(i)-(iii) are commonly used to find an optimal control for continuous-time MDP, see [53], [59], [62], [63], [65]. Also, note that if Assumption* $(A3)$ *(iii) is satisfied, then for any given* $x \in S$, *the function* $\int_S u(y)q(dy|x,a)$ *is continuous in* $a \in A(x)$ *for each function* $u \in B_V^\infty(S)$, *for details see [69, Lemma 8.3.7].*

In the next theorem, we show that if the HJB equation has a solution then its stochastic representation is equal to the value function corresponding to the cost criterion (2.2.3).

Theorem 2.3.1. *Under Assumptions* $(A1)$-$(A3)$ *suppose that the HJB equation (2.3.1) has a solution* $\varphi_\alpha \in L_V^\infty([0,1] \times S)$ *satisfying the bounds as in equation (2.3.1). Then, for all* $(\theta, x) \in [0,1] \times S$, *we have the probabilistic representation of* φ_α *as*

$$\varphi_\alpha(\theta, x) = \inf_{\pi \in \Pi_{Ad}} E_x^\pi \left[\exp\left(\theta \int_0^\infty e^{-\alpha t} c(\xi_t, \pi(t)) dt \right) \right] \tag{2.3.3}$$

i.e., $\varphi_\alpha(\theta, x) = \tilde{J}_\alpha^*(\theta, x)$ *for all* $(\theta, x) \in [0,1] \times S$.

Proof. First, we see that

$$\left[\theta c(x,a)\varphi_\alpha(\theta,x)+\int_S q(dy|x,a)\varphi_\alpha(\theta,y)\right]$$

is continuous in $a\in A(x)$ and $A(x)$ is compact. So by measurable selection theorem, [17, Proposition 7.33], there exists a measurable function $f^*:[0,1]\times S\to A$ such that

$$\begin{aligned}&\inf_{a\in A(x)}\left[\theta c(x,a)\varphi_\alpha(\theta,x)+\int_S q(dy|x,a)\varphi_\alpha(\theta,y)\right]\\&=\left[\theta c(x,f^*(\theta,x))\varphi_\alpha(\theta,x)+\int_S q(dy|x,f^*(\theta,x))\varphi_\alpha(\theta,y)\right].\end{aligned}\tag{2.3.4}$$

Let

$$\pi^*:S\times\mathbb{R}_+\to A$$

be defined by

$$\pi^*(x,t):=f^*(\theta e^{-\alpha t},x).$$

Now we observe from equation (2.3.1) that for any $x\in S$, $a\in A(x)$ and a.e. $\theta\in[0,1]$ that

$$-\alpha\theta\frac{\partial\varphi_\alpha}{\partial\theta}(\theta,x)+\left[\int_S q(dy|x,a)\varphi_\alpha(\theta,y)+\theta c(x,a)\varphi_\alpha(\theta,x)\right]\geq 0.\tag{2.3.5}$$

For any admissible control $\pi\in\Pi_{Ad}$ and $\theta\in[0,1]$, let $\{\xi_t,t\geq 0\}$ be the corresponding process, and define $\theta(t):=\theta e^{-\alpha t}$. Now for each $\omega\in\Omega$, by equation (2.3.5), we get for a.e. $s\geq 0$,

$$-\alpha\theta(s)\frac{\partial\varphi_\alpha}{\partial\theta}(\theta(s),\xi_s)+\left[\int_S q(dy|\xi_s,\pi(s))\varphi_\alpha(\theta(s),y)+\theta(s)c(\xi_s,\pi(s))\varphi_\alpha(\theta(s),\xi_s)\right]\geq 0.\tag{2.3.6}$$

Define a function $g:[0,\infty)\times S\times\Omega\to[0,\infty)$ by

$$g(t,x,\omega):=\exp\left(\int_0^t\theta(s)c(\xi_s,\pi(s))ds\right)\varphi_\alpha(\theta(t),x).$$

In order to use the extension of Dynkin's formula, (for details see [62, Theorem 3.3] or [53, Theorem 3.1]) to the function g, it suffices to verify that $E^\pi_x\left[\exp\left(\int_0^t 2e^{-\alpha s}c(\xi_s,\pi(s))ds\right)\right]<\infty$ for all $x\in S$ and $t\in(0,\infty)$. In view of Assumptions (A1) and (A2), we have

$$E^\pi_x\left[\exp\left(\int_0^t 2e^{-\alpha s}c(\xi_s,\pi(s))ds\right)\right]$$

$$
\begin{aligned}
&\leq E_x^{\pi}\left[\exp\left(\int_0^{\infty}\frac{2}{\alpha}c(\xi_s,\pi(s))d(-e^{-\alpha s})\right)\right]\\
&\leq E_x^{\pi}\left[\int_0^{\infty}\exp\left(\frac{2}{\alpha}c(\xi_s,\pi(s))\right)d(-e^{-\alpha s})\right]\\
&\leq E_x^{\pi}\left[\int_0^{\infty}\exp\left(\frac{2}{\alpha}(\rho_1\log V(\xi_s)+L_0)\right)d(-e^{-\alpha s})\right]\\
&\text{(by Assumption (A1))}\\
&= e^{2L_0/\alpha}\left[\int_0^{\infty}E_x^{\pi}\left(V(\xi_s)^{\frac{2\rho_1}{\alpha}}\right)d(-e^{-\alpha s})\right]\\
&\leq \alpha e^{2L_0/\alpha}\left(V^2(x)+\frac{b_0}{\rho_2}\right)\left[\int_0^{\infty}e^{\rho_2 s-\alpha s}ds\right]\\
&= \frac{\alpha e^{2L_0/\alpha}}{\alpha-\rho_2}\left(V^2(x)+\frac{b_0}{\rho_2}\right)<\infty, \qquad (2.3.7)
\end{aligned}
$$

where the second inequality is obtained by using Jensen's inequality.
Thus, using the extension of Dynkin formula in [53, Theorem 3.1] to the function g, we have

$$
\begin{aligned}
&E_x^{\pi}[g(t,\xi_t,\omega)]-\varphi_\alpha(\theta,x)\\
&=E_x^{\pi}\biggl\{\int_0^t\exp\left(\int_0^s\theta(v)c(\xi_v,\pi(v))dv\right)\times\biggl[-\alpha\theta(s)\frac{\partial\varphi_\alpha}{\partial\theta}(\theta(s),\xi_s)\\
&\qquad+\int_S q(dy|\xi_s,\pi(s))\varphi_\alpha(\theta(s),y)+\theta(s)c(\xi_s,\pi(s))\varphi_\alpha(\theta(s),\xi_s)\biggr]ds\biggr\}. \qquad (2.3.8)
\end{aligned}
$$

Now from (2.3.6) and (2.3.8), we have

$$
\varphi_\alpha(\theta,x)\leq E_x^{\pi}\left[\exp\left(\int_0^t\theta(s)c(\xi_s,\pi(s))ds\right)\varphi_\alpha(\theta(t),\xi_t)\right]. \qquad (2.3.9)
$$

Given any $p>1$, let $q>1$ such that $\frac{1}{p}+\frac{1}{q}=1$, by Hölder's inequality we have

$$
\begin{aligned}
\varphi_\alpha(\theta,x)&\leq E_x^{\pi}\left[\exp\left(\int_0^t\theta(s)c(\xi_s,\pi(s))ds\right)\varphi_\alpha(\theta(t),\xi_t)\right]\\
&\leq\left\{E_x^{\pi}\left[\exp\left(p\int_0^t\theta(s)c(\xi_s,\pi(s))ds\right)\right]\right\}^{1/p}\times\left\{E_x^{\pi}[\varphi_\alpha^q(\theta(t),\xi_t)]\right\}^{1/q}\\
&=:T_1(p,t)\cdot T_2(q,t). \qquad (2.3.10)
\end{aligned}
$$

For $T_2(q,t):=\{E_x^{\pi}[\varphi_\alpha^q(\theta(t),\xi_t)]\}^{1/q}$, by the upper bound of φ_α in (2.3.1), we have

$$
\varphi_\alpha(\theta(t),\xi_t)=\varphi_\alpha(\theta e^{-\alpha t},\xi_t)\leq\frac{\alpha^2}{\alpha^2-\theta e^{-\alpha t}\rho_0\rho_1}\exp\left(\frac{\theta e^{-\alpha t}L_0}{\alpha}\right)[V(\xi_t)]^{\frac{\rho_1\theta e^{-\alpha t}}{\alpha}}.
$$

If $t>\alpha^{-1}\log(\theta q\rho_1/\alpha)$ then $\theta e^{-\alpha t}q\rho_1/\alpha<1$. Applying Jensen's inequality and Proposition 2.2.3(b), we get

$$
T_2(q,t)\leq\left\{E_x^{\pi}\left[\left(\frac{\alpha^2}{\alpha^2-\theta e^{-\alpha t}\rho_0\rho_1}\right)^q\exp\left(\frac{q\theta e^{-\alpha t}L_0}{\alpha}\right)[V(\xi_t)]^{\frac{q\rho_1\theta e^{-\alpha t}}{\alpha}}\right]\right\}^{1/q}
$$

$$\begin{aligned}
&= \frac{\alpha^2}{\alpha^2 - \theta e^{-\alpha t}\rho_0\rho_1} \exp\left(\frac{\theta e^{-\alpha t} L_0}{\alpha}\right)\left[E_x^{\pi}[V^{\frac{q\rho_1\theta e^{-\alpha t}}{\alpha}}(\xi_t)]\right]^{\frac{1}{q}} \\
&\leq \frac{\alpha^2}{\alpha^2 - \theta e^{-\alpha t}\rho_0\rho_1} \exp\left(\frac{\theta e^{-\alpha t} L_0}{\alpha}\right)[E_x^{\pi}(V(\xi_t))]^{\frac{\rho_1\theta e^{-\alpha t}}{\alpha}} \\
&\leq \frac{\alpha^2}{\alpha^2 - \theta e^{-\alpha t}\rho_0\rho_1} \exp\left(\frac{\theta e^{-\alpha t}}{\alpha}(L_0 + \rho_0\rho_1 t)\right) V^{\frac{\theta e^{-\alpha t}\rho_1}{\alpha}}(x) =: T_3(t). \qquad (2.3.11)
\end{aligned}$$

Next take $t \to \infty$ and get

$$T_1(p,t) \to \left\{E_x^{\pi}\left[\exp\left(p\int_0^{\infty}\theta(s)c(\xi_s,\pi(s))ds\right)\right]\right\}^{1/p} \text{ and } T_3(t) \to 1. \qquad (2.3.12)$$

By (2.3.10), (2.3.11) and (2.3.12) we obtain

$$\varphi_\alpha(\theta,x) \leq \left\{E_x^{\pi}\left[\exp\left(p\theta\int_0^{\infty} e^{-\alpha t}c(\xi_t,\pi(t))dt\right)\right]\right\}^{1/p}.$$

Now, take the limit as $p \downarrow 1$ and get the result

$$\varphi_\alpha(\theta,x) \leq E_x^{\pi}\left[\exp\left(\theta\int_0^{\infty} e^{-\alpha t}c(\xi_t,\pi(t))dt\right)\right].$$

Since $\pi \in \Pi_{Ad}$ is an arbitrary control, we have

$$\varphi_\alpha(\theta,x) \leq \inf_{\pi\in\Pi_{Ad}} E_x^{\pi}\left[\exp\left(\theta\int_0^{\infty} e^{-\alpha t}c(\xi_t,\pi(t))dt\right)\right]. \qquad (2.3.13)$$

Using (2.3.1), (2.3.4) and (2.3.8), we can show that

$$E_x^{\pi^*}\left[\exp\left(\int_0^t \theta(s)c(\xi_s,\pi^*(\xi_{s-},s))ds\right)\varphi_\alpha(\theta(t),\xi_t)\right] = \varphi_\alpha(\theta,x). \qquad (2.3.14)$$

Now, using the lower bound of φ_α in (2.3.1) and Fatou's lemma, we obtain

$$\begin{aligned}
&\liminf_{t\to\infty} E_x^{\pi^*}\left[\exp\left(\int_0^t \theta(s)c(\xi_s,\pi^*(\xi_{s-},s))ds\right)\varphi_\alpha(\theta(t),\xi_t)\right] \\
&\geq \liminf_{t\to\infty} E_x^{\pi^*}\left[\exp\left(\int_0^t \theta(s)c(\xi_s,\pi^*(\xi_{s-},s))ds\right)\right] \\
&\geq E_x^{\pi^*}\left[\liminf_{t\to\infty}\exp\left(\int_0^t \theta(s)c(\xi_s,\pi^*(\xi_{s-},s))ds\right)\right] \\
&= \tilde{J}_\alpha(\theta,x,\pi^*). \qquad (2.3.15)
\end{aligned}$$

From (2.3.14) and (2.3.15), we have

$$\tilde{J}_\alpha(\theta,x,\pi^*) \leq \varphi_\alpha(\theta,x).$$

Thus

$$\inf_{\pi\in\Pi_{Ad}} \tilde{J}_\alpha(\theta,x,\pi) \leq \tilde{J}_\alpha(\theta,x,\pi^*) \leq \varphi_\alpha(\theta,x). \qquad (2.3.16)$$

From (2.3.13) and (2.3.16), we have (2.3.3). □

2.4 The existence of solution to the HJB equation

In this Section, we prove that the equation (2.3.1) is the HJB equation for the α discounted cost (2.2.3) and the equation (2.3.1) has a solution in $L^{\infty}_{V}([0,1]\times S)$. We now proceed to make a rigorous analysis of the above. First, we prove a lemma about the existence of a solution for the HJB equation for bounded transition and cost rates; see Lemma 2.4.1 below. Then in Theorem 2.4.1, we relax these boundedness conditions and prove the existence of a solution to the HJB eq. (2.3.1). For that, we first truncate our transition and cost rates which plays a crucial role to derive the HJB equations and find the solution. Fix any $n \geq 1$, $0 < \delta < 1$. For each $n \geq 1$, $x \in S$, $a \in A(x)$, let $A_n(x) := A(x)$, $S_n := \{x \in S | V(x) \leq n\}$, and $K_n := \{(x,a) | x \in S_n, a \in A_n(x)\}$. Moreover for each $x \in S$, $a \in A_n(x)$ define

$$q^{(n)}(dy|x,a) := \begin{cases} q(dy|x,a) & \text{if } x \in S_n, \\ 0 \ \text{ if } \ x \notin S_n \end{cases} \tag{2.4.1}$$

and

$$c_n(x,a) := \begin{cases} c(x,a) \wedge \ \min\{n, \rho_1 \ln V(x) + L_0\} & \text{if } x \in S_n, \\ 0 \ \text{ if } \ x \notin S_n. \end{cases} \tag{2.4.2}$$

Lemma 2.4.1. *Grant Assumptions (A1)-(A3). Then, there exists a unique function $\varphi^{(n,\delta)}_{\alpha}$ (depending on n, δ) in $L^{\infty}_{V}([0,1]\times S)$ for which the followings are true :*

1. *$\varphi^{(n,\delta)}_{\alpha} \in B^{\infty}_{1}([0,1]\times S)$ is a bounded solution to the following differential equations (DEs) for all $x \in S$ and a.e. $\theta \in (\delta, 1]$:*

$$\begin{cases} \alpha\theta \frac{\partial \varphi^{(n,\delta)}_{\alpha}}{\partial\theta}(\theta,x) & = \inf_{a\in A(x)} \left[\theta c_n(x,a)\varphi^{(n,\delta)}_{\alpha}(\theta,x) + \int_S q^{(n)}(dy|x,a)\varphi^{(n,\delta)}_{\alpha}(\theta,y)\right] \\ \varphi^{(n,\delta)}_{\alpha}(\delta,x) & = e^{n\delta/\alpha}. \end{cases} \tag{2.4.3}$$

2. *$\varphi^{(n,\delta)}_{\alpha}(\theta,x)$ has a stochastic representation as follows: for each $x \in S$ and a.e. $\theta \in (\delta,1]$,*

$$\varphi^{(n,\delta)}_{\alpha}(\theta,x) = \inf_{\pi\in\Pi_{Ad}} E^{\pi}_{x}\left[e^{n\delta/\alpha}\exp\left(\theta\int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi^{(n)}_t,\pi(t))dt\right)\right], \tag{2.4.4}$$

where $T_\delta(\theta) := \alpha^{-1}\log(\theta/\delta)$ and $\{\xi^{(n)}_t\}_{t\geq 0}$ is the process corresponding to the $q^{(n)}(\cdot|x,a)$.

Proof. (1) Since $S_n := \{x \in S | V(x) \leq n\}$, by Assumption $(A1)$(ii), we see that $q_x^{(n)}(a) := \displaystyle\int_{S/\{x\}} q^{(n)}(dy|x,a)$ is bounded. So we can use the Lyapunov function $V \equiv 1$ such that $\int_S q^{(n)}(dy|x,a)V(y) \leq \rho_0 V(x)$, and $\overline{q}^{(n)} := \sup_{(x,a)\in K} q_x^{(n)}(a) < \infty$. Now let us define a nonlinear operator T on $B_1^\infty([0,1]\times S)$ as follows:

$$Tu(\theta,x) = e^{\delta n/\alpha} + \frac{1}{\alpha}\int_\delta^\theta \inf_{a\in A(x)}\left[\frac{1}{s}\int_S q^{(n)}(dy|x,a)u(s,y) + c_n(x,a)u(s,x)\right]ds,$$

where $u \in B_1^\infty([0,1]\times S)$ and $(\theta,x) \in [\delta,1]\times S$. By using the Assumption $(A1)$ and the fact that c_n is bounded, we obtain

$$\begin{aligned}
&\sup_{\theta\in[\delta,1]}\sup_{x\in S}|Tu(\theta,x)| \\
&\leq e^{\delta n/\alpha} + \frac{1}{\alpha}\int_\delta^1 \sup_{a\in A(x)}\left\{\frac{1}{s}\sup_{x\in S}\left[\int_S |q^{(n)}(dy|x,a)||u(s,y)|\right] + n\sup_{x\in S}|u(s,x)|\right\}ds \\
&\leq e^{\delta n/\alpha} + \frac{\|u\|_1^\infty}{\alpha}\left\{\int_\delta^1 \sup_{a\in A(x)}\frac{1}{s}\sup_{x\in S}\left(2q_x^{(n)}(a)\right)ds + n(1-\delta)\right\} \\
&\leq e^{\delta n/\alpha} + \frac{1}{\alpha}\left[(-2)\overline{q}^{(n)}\log\delta + n(1-\delta)\right]\|u\|_1^\infty.
\end{aligned}$$

Therefore, T is a nonlinear operator from $B_1^\infty([0,1]\times S)$ to $B_1^\infty([0,1]\times S)$. For any $g_1, g_2 \in B_1^\infty([0,1]\times S)$ and $\theta \in [\delta,1]$, we have

$$\begin{aligned}
\sup_{x\in S}|Tg_1(t,x) - Tg_2(t,x)| &\leq \frac{1}{\alpha}\int_\delta^t \left(2\overline{q}^{(n)}/s + n\right)\sup_{x\in S}|g_1(s,x) - g_2(s,x)|ds \\
&\leq \frac{1}{\alpha}\left[2\overline{q}^{(n)}(\log t - \log\delta) + n(t-\delta)\right]\|g_1 - g_2\|_1^\infty. \qquad (2.4.5)
\end{aligned}$$

Now, we prove the following:

$$\sup_{x\in S}|T^l g_1(t,x) - T^l g_2(t,x)| \leq \frac{\|g_1-g_2\|_1^\infty}{\alpha^l\cdot l!}\left[2\overline{q}^{(n)}(\log t - \log\delta) + n(t-\delta)\right]^l \quad \forall\, l \geq 1. \quad (2.4.6)$$

By (2.4.5) and (2.4.6) we have

$$\begin{aligned}
&\sup_{x\in S}|T^{l+1}g_1(t,x) - T^{l+1}g_2(t,x)| \\
&\leq \frac{1}{\alpha}\int_\delta^t \left(2\overline{q}^{(n)}/s + n\right)\sup_{x\in S}|T^l g_1(s,x) - T^l g_2(s,x)|ds \\
&\leq \frac{\|g_1-g_2\|_1^\infty}{\alpha^{l+1}\cdot l!}\int_\delta^t \left(2\overline{q}^{(n)}/s + n\right)\left[2\overline{q}^{(n)}(\log s - \log\delta) + n(s-\delta)\right]^l ds \\
&= \frac{\|g_1-g_2\|_1^\infty}{\alpha^{l+1}\cdot(l+1)!}\left[2\overline{q}^{(n)}(\log t - \log\delta) + n(t-\delta)\right]^{l+1}.
\end{aligned}$$

Since $\sum_{k\geq 1}\frac{1}{\alpha^k\cdot k!}\Big[-2\overline{q}^{(n)}\log\delta+n(1-\delta)\Big]^k<\infty$, there exists some m such that $\beta:=\frac{1}{\alpha^m\cdot m!}\Big[-2\overline{q}^{(n)}\log\delta+n(1-\delta)\Big]^m<1$, which implies that $\|T^m g_1-T^m g_2\|_1^\infty\leq\beta\|g_1-g_2\|_1^\infty$. Therefore, T is a m-step contraction operator on $B_1^\infty([0,1]\times S)$. So, by Banach fixed point theorem, there exists a unique bounded function $\varphi_\alpha^{(n,\delta)}\in B_1^\infty([0,1]\times S)$ (depending on (n,δ)) such that $T\varphi_\alpha^{(n,\delta)}=\varphi_\alpha^{(n,\delta)}$; that is,

$$\varphi_\alpha^{(n,\delta)}(\theta,x)=e^{\delta n/\alpha}+\frac{1}{\alpha}\int_\delta^\theta\inf_{a\in A(x)}\Big[\frac{1}{s}\int_S q^{(n)}(dy|x,a)\varphi_\alpha^{(n,\delta)}(s,y)+c_n(x,a)\varphi_\alpha^{(n,\delta)}(s,x)\Big]ds.$$

Also note that $\varphi_\alpha^{(n,\delta)}(\delta,x)=e^{\delta n/\alpha}$. Hence by using (2.4.1), (2.4.2) and the above equation, we have $\varphi_\alpha^{(n,\delta)}\in L_V^\infty([0,1]\times S)$ and it satisfies equation (2.4.3).

(2) First we see that

$$\Big[\theta c_n(x,a)\varphi_\alpha^{(n,\delta)}(\theta,x)+\int_S q^{(n)}(dy|x,a)\varphi_\alpha^{(n,\delta)}(\theta,y)\Big]$$

is continuous in $a\in A(x)$ and $A(x)$ is compact. So by measurable selection theorem, [17, Proposition 7.33], there exists a measurable function $f^{*\delta}:[0,1]\times S\to A$ such that

$$\begin{aligned}&\inf_{a\in A(x)}\Big[\theta c_n(x,a)\varphi_\alpha^{(n,\delta)}(\theta,x)+\int_S q^{(n)}(dy|x,a)\varphi_\alpha^{(n,\delta)}(\theta,y)\Big]\\&=\Big[\theta c_n(x,f^{*\delta}(\theta,x))\varphi_\alpha^{(n,\delta)}(\theta,x)+\int_S q^{(n)}(dy|x,f^{*\delta}(\theta,x))\varphi_\alpha^{(n,\delta)}(\theta,y)\Big].\end{aligned}\tag{2.4.7}$$

Let

$$\pi^{*\delta}:S\times\mathbb{R}_+\to A$$

be defined by

$$\pi^{*\delta}(x,t):=f^{*\delta}(\theta e^{-\alpha t},x).$$

Let $\theta(t):=\theta e^{-\alpha t}$ for $t\in[0,\infty)$. Since c_n and $\varphi_\alpha^{(n,\delta,k)}$ are bounded, by Dynkin's formula we get

$$\begin{aligned}&E_x^\pi\Big[\exp\Big(\int_0^{T_\delta(\theta)}\theta(s)c_n(\xi_s^{(n)},\pi(s))ds\Big)\varphi_\alpha^{(n,\delta)}\Big(\theta(T_\delta),\xi_{T_\delta}^{(n)}\Big)\Big]-\varphi_\alpha^{(n,\delta)}(\theta,x)\\&=E_x^\pi\Big\{\int_0^{T_\delta(\theta)}\Big[-\alpha\theta(s)\frac{\partial\varphi_\alpha^{(n,\delta)}}{\partial\theta}(\theta(s),\xi_s^{(n)})+\int_S q^{(n)}(dy|\xi_s^{(n)},\pi(s))\varphi_\alpha^{(n,\delta)}(\theta(s),y)\\&\qquad+\theta(s)c_n(\xi_s^{(n)},\pi(s))\varphi_\alpha^{(n,\delta)}(\theta(s),\xi_s^{(n)})\Big]\times\exp\Big(\int_0^s\theta(v)c_n(\xi_v^{(n)},\pi(v))dv\Big)ds\Big\}.\end{aligned}\tag{2.4.8}$$

By using (2.4.3) and (2.4.8), we obtain

$$E_x^{\pi}\left[\exp\left(\int_0^{T_\delta(\theta)} \theta(s)c_n(\xi_s^{(n)},\pi(s))ds\right)\varphi_\alpha^{(n,\delta)}\left(\theta(T_\delta),\xi_{T_\delta}^{(n)}\right)\right] \geq \varphi_\alpha^{(n,\delta)}(\theta,x).$$

Since $\pi \in \Pi_{Ad}$ is an arbitrary control and $\varphi_\alpha^{(n,\delta)}(\theta(T_\delta(\theta)),\xi_{T_\delta}^{(n)}) = e^{n\delta/\alpha}$, we have

$$\varphi_\alpha^{(n,\delta)}(\theta,x) \leq \inf_{\pi\in\Pi_{Ad}} E_x^{\pi}\left[e^{n\delta/\alpha}\exp\left(\int_0^{T_\delta(\theta)} \theta(s)c_n(\xi_s^{(n)},\pi(s))ds\right)\right]. \tag{2.4.9}$$

Using equations (2.4.3), (2.4.7) and (2.4.8), we can show that

$$\varphi_\alpha^{(n,\delta)}(\theta,x) = E_x^{\pi^{*\delta}}\left[e^{n\delta/\alpha}\exp\left(\int_0^{T_\delta(\theta)} \theta(s)c_n(\xi_s^{(n)},\pi^{*\delta}(\xi_{s-}^{(n)},s))ds\right)\right].$$

Therefore

$$\varphi_\alpha^{(n,\delta)}(\theta,x) \geq \inf_{\pi\in\Pi_{Ad}} E_x^{\pi}\left[e^{n\delta/\alpha}\exp\left(\int_0^{T_\delta(\theta)} \theta(s)c_n(\xi_s^{(n)},\pi(s))ds\right)\right]. \tag{2.4.10}$$

Therefore, from (2.4.9) and (2.4.10), we obtain (2.4.4). This completes the proof. □

Theorem 2.4.1. *Grant Assumptions* (*A*1)*-*(*A*3)*. Then the HJB equation* (2.3.1) *has a unique solution* $\varphi_\alpha \in L_V^\infty([0,1]\times S)$ *satisfying* $1 \leq \varphi_\alpha(\theta,x) \leq \frac{\alpha^2 e^{\theta L_0/\alpha}}{\alpha^2-\rho_0\rho_1\theta}(V(x))^{\frac{\rho_1\theta}{\alpha}}$ *for all* $(\theta,x)\in[0,1]\times S$*.*

Proof. First note that $\varphi_\alpha^{(n,\delta)}$ is the solution to the equation (2.4.3), which depends on two parameters n, δ. We prove this theorem in two steps.
Step 1: In the first step, we construct a solution $\varphi_\alpha^{(n)}(\cdot,x)$ from $\varphi_\alpha^{(n,\delta)}(\cdot,x)$ by passing the limit as $\delta\to 0$, such that $\varphi_\alpha^{(n)}(\cdot,x)$ is an absolutely continuous function and satisfies the following DEs:

$$\begin{cases} \alpha\theta\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) & = \inf_{a\in A(x)}\left[\int_S q^{(n)}(dy|x,a)\varphi_\alpha^{(n)}(\theta,y)+\theta c_n(x,a)\varphi_\alpha^{(n)}(\theta,x)\right], \\ & \qquad x\in S,\ \text{a.e.},\ \theta\in[0,1], \\ 1\leq\varphi_\alpha^{(n)}(\theta,x) & \leq \frac{\alpha^2 e^{\theta L_0/\alpha}}{\alpha^2-\rho_0\rho_1\theta}(V(x))^{\frac{\rho_1\theta}{\alpha}} \quad \forall\ (\theta,x)\in[0,1]\times S. \end{cases} \tag{2.4.11}$$

Given $0<\delta<1$ and $1\leq n<\infty$ by (2.4.4) and $\sup_{(x,a)\in K} c_n(x,a)\leq n$, we have $\varphi_\alpha^{(n,\delta)}(\theta,x)\leq e^{2n/\alpha},\quad x\in S,\theta\in[\delta,1]$.

Next, we extend the domain of $\varphi_\alpha^{(n,\delta)}$ to $[0,1]\times S$ by

$$\overline{\varphi}_\alpha^{(n,\delta)}(\theta,x) = \begin{cases} \varphi_\alpha^{(n,\delta)}(\theta,x), & \delta\leq\theta\leq 1\ \forall x\in S \\ e^{n\delta/\alpha}, & 0\leq\theta<\delta\ \forall x\in S. \end{cases}$$

We consider the following expression, for any given $\pi\in\Pi_{Ad}$, $x\in S$, $\theta,\theta_0\in[\delta,1]$:

$$\Bigg|E_x^{\pi}\left[e^{n\delta/\alpha}\exp\left(\theta\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right]$$

$$
\begin{aligned}
&- E_x^\pi\left[e^{n\delta/\alpha}\exp\left(\theta_0\int_0^{T_\delta(\theta_0)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right]\Bigg|\\
&\leq P_1+P_2,
\end{aligned}
$$

where

$$
\begin{aligned}
P_1 :=&\Bigg|E_x^\pi\left[e^{n\delta/\alpha}\exp\left(\theta\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right]\\
&- E_x^\pi\left[e^{n\delta/\alpha}\exp\left(\theta_0\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right]\Bigg|,
\end{aligned}
$$

and

$$
\begin{aligned}
P_2 :=&\Bigg|E_x^\pi\left[e^{n\delta/\alpha}\exp\left(\theta_0\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right]\\
&- E_x^\pi\left[e^{n\delta/\alpha}\exp\left(\theta_0\int_0^{T_\delta(\theta_0)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right]\Bigg|.
\end{aligned}
$$

Consider $c\wedge d := min\{c,d\}$ and $c\vee d := max\{c,d\}$. Then for fix $n\geq 1$; we have

$$
\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt \leq n\int_0^{T_\delta(\theta)} e^{-\alpha t}dt \leq \frac{n}{\alpha}
$$

and

$$
\begin{aligned}
\int_{T_\delta(\theta\wedge\theta_0)}^{T_\delta(\theta\vee\theta_0)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt &\leq n\int_{T_\delta(\theta\wedge\theta_0)}^{T_\delta(\theta\vee\theta_0)} e^{-\alpha t}dt\\
&= \frac{n}{\alpha}[\exp(-\alpha T_\delta(\theta\wedge\theta_0)) - \exp(-\alpha T_\delta(\theta\vee\theta_0))] \leq \frac{\delta n|\theta_0-\theta|}{\alpha\theta\theta_0}.
\end{aligned}
$$

Using the above results and knowing the fact that $e^{bz}-1\leq(e^b-1)z$ for all $z\in[0,1]$ and $b>0$, we obtain

$$
\begin{aligned}
P_1 &= e^{n\delta/\alpha}E_x^\pi\left[\exp\left((\theta\wedge\theta_0)\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right.\\
&\quad\left.\times\left(\exp\left(|\theta_0-\theta|\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)-1\right)\right]\\
&\leq e^{2n/\alpha}E_x^\pi\left[\exp\left(|\theta_0-\theta|\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)-1\right]\\
&\leq e^{2n/\alpha}\left(\exp\left(\frac{n}{\alpha}|\theta_0-\theta|\right)-1\right)\\
&\leq e^{2n/\alpha}\left(e^{n/\alpha}-1\right)|\theta_0-\theta|.
\end{aligned}
$$

Similarly for P_2 we have

$$
P_2 = e^{n\delta/\alpha}E_x^\pi\left[\exp\left(\theta_0\int_0^{T_\delta(\theta\wedge\theta_0)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right.
$$

$$
\begin{aligned}
&\times\left(\exp\left(\theta_0\int_{T_\delta(\theta\wedge\theta_0)}^{T_\delta(\theta\vee\theta_0)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)-1\right)\Bigg]\\
&\leq e^{2n/\alpha}E_x^{\pi}\left[\exp\left(\theta_0\int_{T_\delta(\theta\wedge\theta_0)}^{T_\delta(\theta\vee\theta_0)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)-1\right]\\
&\leq e^{2n/\alpha}\left(\exp\left(\frac{n\delta|\theta-\theta_0|}{\alpha\theta}\right)-1\right)\\
&\leq e^{2n/\alpha}\left(e^{n/\alpha}-1\right)|\theta_0-\theta|.
\end{aligned}
$$

Hence for all $(\theta,x)\in[0,1]\times S$, we have

$$
|\overline{\varphi}_\alpha^{(n,\delta)}(\theta_0,x)-\overline{\varphi}_\alpha^{(n,\delta)}(\theta,x)|\leq 2e^{2n/\alpha}(e^{n/\alpha}-1)|\theta-\theta_0|. \tag{2.4.12}
$$

Now we want to show that $\overline{\varphi}_\alpha^{(n,\delta)}$ is decreasing as $\delta\to 0$ for any (θ,x). For a fixed $\alpha>0$ and $\varepsilon>0$ small enough, consider $\overline{\varphi}_\alpha^{(n,\delta+\varepsilon)}(\theta,x)-\overline{\varphi}_\alpha^{(n,\delta)}(\theta,x)$ and assume that $h_\delta:=e^{\frac{n\delta}{\alpha}}$. By measurable selection theorem we get the minimizer $\pi^{*(\delta+\varepsilon)}$ like in equation (2.4.7), corresponding to $\overline{\varphi}_\alpha^{(n,\delta+\varepsilon)}$ such that the followings cases hold.

Case 1. If $\delta+\varepsilon<\theta$ then

$$
\begin{aligned}
&\overline{\varphi}_\alpha^{(n,\delta+\varepsilon)}(\theta,x)-\overline{\varphi}_\alpha^{(n,\delta)}(\theta,x)\\
&=E_x^{\pi^{*(\delta+\varepsilon)}}\left[h_{\delta+\varepsilon}\exp\left(\theta\int_0^{T_{\delta+\varepsilon}} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^{*(\delta+\varepsilon)}(\xi_{t-}^{(n)},t))dt\right)\right]\\
&\quad-\inf_{\pi\in\Pi_{Ad}}E_x^{\pi}\left[h_\delta\exp\left(\theta\int_0^{T_\delta} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right]\\
&\geq h_\delta E_x^{\pi^{*(\delta+\varepsilon)}}\Bigg[\exp\left(\theta\int_0^{T_{\delta+\varepsilon}} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^{*(\delta+\varepsilon)}(\xi_{t-}^{(n)},t))dt\right)\\
&\quad\times\left\{h_\varepsilon-\exp\left(\theta\int_{T_{\delta+\varepsilon}}^{T_\delta} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^{*(\delta+\varepsilon)}(\xi_{t-}^{(n)},t))dt\right)\right\}\Bigg]\\
&\geq h_\delta E_x^{\pi^{*(\delta+\varepsilon)}}\left[\exp\left(\theta\int_0^{T_{\delta+\varepsilon}} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^{*(\delta+\varepsilon)}(\xi_{t-}^{(n)},t))dt\right)\left\{h_\varepsilon-\exp\left(\theta\int_{T_{\delta+\varepsilon}}^{T_\delta} e^{-\alpha t}ndt\right)\right\}\right]\\
&=h_\delta E_x^{\pi^{*(\delta+\varepsilon)}}\Bigg[\exp\left(\theta\int_0^{T_{\delta+\varepsilon}} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^{*(\delta+\varepsilon)}(\xi_{t-}^{(n)},t))dt\right)\\
&\quad\times\left\{h_\varepsilon-\exp\left(\frac{n\theta(e^{-\alpha T_{\delta+\varepsilon}}-e^{-\alpha T_\delta})}{\alpha}\right)\right\}\Bigg]\\
&=0.
\end{aligned}
$$

Case 2. $\delta<\theta\leq\delta+\varepsilon$

$$
\begin{aligned}
&\overline{\varphi}_\alpha^{(n,\delta+\varepsilon)}(\theta,x)-\overline{\varphi}_\alpha^{(n,\delta)}(\theta,x)\\
&=h_{\delta+\varepsilon}-E_x^{\pi^{*\delta}}\left[h_\delta\exp\left(\theta\int_0^{T_\delta} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^{*\delta}(\xi_{t-}^{(n)},t))dt\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= h_\delta\Big[h_\varepsilon - E_x^{\pi^{*\delta}}\Big[\exp\Big(\theta\int_0^{T_\delta} e^{-\alpha t}c_n(\xi_t^{(n)}, \pi^{*\delta}(\xi_{t-}^{(n)}, t))dt\Big)\Big]\Big] \\
&\geq h_\delta\Big[h_\varepsilon - \exp\Big(\theta\int_0^{T_\delta} e^{-\alpha t} n dt\Big)\Big] \\
&= h_\delta\Big[h_\varepsilon - e^{n\theta\frac{(1-e^{-\alpha T_\delta})}{\alpha}}\Big] \geq 0.
\end{aligned}
$$

Case 3. $\theta \leq \delta$

$$\overline{\varphi}_\alpha^{(n,\delta+\varepsilon)}(\theta, x) - \overline{\varphi}_\alpha^{(n,\delta)}(\theta, x) = h_{\delta+\varepsilon} - h_\delta = h_\delta(h_\varepsilon - 1) = h_\delta(e^{\frac{n\varepsilon}{\alpha}} - 1) \geq 0.$$

Hence $\overline{\varphi}_\alpha^{(n,\delta)}(\theta, x)$ is increasing in δ for any $(\theta, x) \in [0,1]\times S$. Now from (2.4.12), we know that for each $x \in S$, $\overline{\varphi}_\alpha^{(n,\delta)}(\cdot, x)$ is Lipschitz continous in $\theta \in [0,1]$. Also, $\overline{\varphi}_\alpha^{(n,\delta)}(\theta, x)$ is increasing in δ for any $(\theta, x) \in [0,1]\times S$ and bounded above (since $\overline{\varphi}_\alpha^{(n,\delta)}(\theta, x) \leq e^{2n/\alpha},\ \ x \in S, \theta \in [\delta, 1]$), therefore there exists a function $\varphi_\alpha^{(n)}$ on $[0,1]\times S$ that is continuous with respect to $\theta \in [0,1]$, such that along a subsequence $\delta_m \to 0$, we have $\lim_{m\to\infty}\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta, x) = \varphi_\alpha^{(n)}(\theta, x)$ and for any fixed $x \in S$ this convergence is uniform in $\theta \in [0,1]$.

Let $\psi \in C_c^\infty(0,1)$, then we have

$$
\begin{aligned}
&-\int_0^1 \alpha\frac{d(\theta\psi)}{d\theta}(\theta)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta, x)d\theta = \int_0^1 \alpha\theta\frac{\partial\overline{\varphi}_\alpha^{(n,\delta_m)}}{\partial\theta}(\theta, x)\psi(\theta)d\theta \\
&= \int_0^1 \inf_{a\in A(x)}\Big[\theta c_n(x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta, x) + \int_S q^{(n)}(dy|x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta, y)\Big]\psi(\theta)d\theta \\
&- \int_0^{\delta_m} \inf_{a\in A(x)}\Big[\theta c_n(x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta, x) + \int_S q^{(n)}(dy|x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta, y)\Big]\psi(\theta)d\theta \\
&= \int_0^1 \inf_{a\in A(x)}\Big[\theta c_n(x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta, x) + \int_S q^{(n)}(dy|x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta, y)\Big]\psi(\theta)d\theta \\
&- \int_0^{\delta_m} \inf_{a\in A(x)}\Big[\theta c_n(x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta, x)\Big]\psi(\theta)d\theta. \qquad (2.4.13)
\end{aligned}
$$

Now take $\tau(x) := M_0 V(x)$ and define

$$Q^{(n)}(dy|x,a) := \delta_x(dy) + \frac{q^{(n)}(dy|x,a)}{\tau(x)}$$

for all $(x,a) \in K$ where $\delta_x(\cdot)$ is the Dirac measure concentrated at x. We see that under Assumption $(A1)$, $Q^{(n)}$ is a stochastic kernel on S given K. Then (2.4.13) can be written as

$$
\begin{aligned}
&-\int_0^1 \Big\{\frac{\alpha}{\tau(x)}\frac{d(\theta\psi)}{d\theta}\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta, x) - \overline{\varphi}_\alpha^{(n,\delta_m)}(\theta, x)\psi(\theta)\Big\}d\theta \\
&= \int_0^1 \inf_{a\in A(x)}\Big[\frac{\theta}{\tau(x)}c_n(x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta, x) + \int_S Q^{(n)}(dy|x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta, y)\Big]\psi(\theta)d\theta
\end{aligned}
$$

$$-\frac{1}{\tau(x)}\int_0^{\delta_m}\inf_{a\in A(x)}\left[\theta c_n(x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)\right]\psi(\theta)d\theta. \tag{2.4.14}$$

Now

$$\begin{aligned}
&\left|\inf_{a\in A(x)}\left[\frac{\theta}{\tau(x)}c_n(x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)+\int_S Q^{(n)}(dy|x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y)\right]\psi(\theta)\right| \\
&\le|\psi(\theta)|\sup_{a\in A(x)}\left[\frac{\theta}{\tau(x)}|c_n(x,a)||\overline{\varphi}_\alpha^{(n,\delta)}(\theta,x)|+\int_S Q^{(n)}(dy|x,a)|\overline{\varphi}_\alpha^{(n,\delta)}(\theta,y)|\right] \\
&\le\frac{\alpha^2}{\alpha^2-\rho_0\rho_1\theta}e^{\theta L_0/\alpha}\sup_{a\in A(x)}\left[\frac{\theta}{\tau(x)}nV^{\frac{\rho_1\theta}{\alpha}}(x)+\int_S Q^{(n)}(dy|x,a)V^{\frac{\rho_1\theta}{\alpha}}(y)\right]|\psi(\theta)| \\
&\le\frac{\alpha^2}{\alpha^2-\rho_0\rho_1\theta}e^{\theta L_0/\alpha}\sup_{a\in A(x)}\left[\frac{\theta}{\tau(x)}nV(x)+\int_S Q^{(n)}(dy|x,a)V(y)\right]|\psi(\theta)| \\
&\le\frac{\alpha^2}{\alpha^2-\rho_0\rho_1\theta}e^{\theta L_0/\alpha}\left[\frac{\theta}{\tau(x)}nV(x)+V(x)+\rho_0\frac{V(x)}{\tau(x)}\right]|\psi(\theta)| \\
&=\frac{\alpha^2}{\alpha^2-\rho_0\rho_1\theta}e^{\theta L_0/\alpha}\left[\frac{\theta}{\tau(x)}nV(x)+V(x)+\frac{\rho_0}{M_0}\right]|\psi(\theta)|. \qquad (2.4.15)
\end{aligned}$$

Under Assumption $(A3)$, there exists $a_m^*\in A(x)$ such that

$$\begin{aligned}
&\inf_{a\in A(x)}\left[\frac{\theta}{\tau(x)}c_n(x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)+\int_S Q^{(n)}(dy|x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y)\right] \\
&=\left[\frac{\theta}{\tau(x)}c_n(x,a_m^*)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)+\int_S Q^{(n)}(dy|x,a_m^*)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y)\right].
\end{aligned}$$

Since for each fixed $x\in S$, $A(x)$ is compact, there exists a subsequence of $\{m\}$, by abuse of notation, we denote the same sequence and $a^*\in A(x)$ such that $\lim_{m\to\infty}a_m^*=a^*$. Now, from (2.4.14), for any $a\in A(x)$, we have

$$\begin{aligned}
&-\int_0^1\left\{\frac{\alpha}{\tau(x)}\frac{d(\theta\psi)}{d\theta}\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)-\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)\psi(\theta)\right\}d\theta \\
&=\int_0^1\left[\frac{\theta}{\tau(x)}c_n(x,a_m^*)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)+\int_S Q^{(n)}(dy|x,a_m^*)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y)\right]\psi(\theta)d\theta \\
&\quad-\frac{1}{\tau(x)}\int_0^{\delta_m}\inf_{a\in A(x)}\left[\theta c_n(x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)\right]\psi(\theta)d\theta. \qquad (2.4.16)
\end{aligned}$$

So, by Lemma 8.3.7 in Hernandez-Lerma and Lassere (1999) [69] taking limit as $m\to\infty$ in (2.4.16), we get

$$\begin{aligned}
&-\int_0^1\left\{\frac{\alpha}{\tau(x)}\frac{d(\theta\psi)}{d\theta}(\theta)\varphi_\alpha^{(n)}(\theta,x)-\varphi_\alpha^{(n)}(\theta,x)\psi(\theta)\right\}d\theta \\
&\ge\int_0^1\left[\frac{\theta}{\tau(x)}c_n(x,a^*)\varphi_\alpha^{(n)}(\theta,x)+\int_S Q^{(n)}(dy|x,a^*)\varphi_\alpha^{(n)}(\theta,y)\right]\psi(\theta)d\theta.
\end{aligned}$$

Hence

$$-\int_0^1\left\{\frac{\alpha}{\tau(x)}\frac{d(\theta\psi)}{d\theta}(\theta)\varphi_\alpha^{(n)}(\theta,x)-\varphi_\alpha^{(n)}(\theta,x)\psi(\theta)\right\}d\theta$$

$$\geq \inf_{a\in A(x)} \int_0^1 \left[\frac{\theta}{\tau(x)} c_n(x,a)\varphi_\alpha^{(n)}(\theta,x) + \int_S Q^{(n)}(dy|x,a)\varphi_\alpha^{(n)}(\theta,y)\right]\psi(\theta)d\theta. \qquad (2.4.17)$$

But

$$\begin{aligned}
&-\int_0^1 \left\{\frac{\alpha}{\tau(x)}\frac{d(\theta\psi)}{d\theta}\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x) - \overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)\psi(\theta)\right\}d\theta \\
&\leq \int_0^1 \left[\frac{\theta}{\tau(x)} c_n(x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x) + \int_S Q^{(n)}(dy|x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y)\right]\psi(\theta)d\theta \\
&\quad - \frac{1}{\tau(x)}\int_0^{\delta_m} \inf_{a\in A(x)}\left[\theta c_n(x,a)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)\right]\psi(\theta)d\theta.
\end{aligned}$$

By analogous arguments, we get

$$\begin{aligned}
&-\int_0^1 \left\{\frac{\alpha}{\tau(x)}\frac{d(\theta\psi)}{d\theta}(\theta)\varphi_\alpha^{(n)}(\theta,x) - \varphi_\alpha^{(n)}(\theta,x)\psi(\theta)\right\}d\theta \\
&\leq \inf_{a\in A(x)} \int_0^1 \left[\frac{\theta}{\tau(x)} c_n(x,a)\varphi_\alpha^{(n)}(\theta,x) + \int_S Q^{(n)}(dy|x,a)\varphi_\alpha^{(n)}(\theta,y)\right]\psi(\theta)d\theta. \qquad (2.4.18)
\end{aligned}$$

From (2.4.17) and (2.4.18), we get

$$\begin{aligned}
&-\int_0^1 \left\{\frac{\alpha}{\tau(x)}\frac{d(\theta\psi)}{d\theta}(\theta)\varphi_\alpha^{(n)}(\theta,x) - \varphi_\alpha^{(n)}(\theta,x)\psi(\theta)\right\}d\theta \\
&= \inf_{a\in A(x)} \int_0^1 \left[\frac{\theta}{\tau(x)} c_n(x,a)\varphi_\alpha^{(n)}(\theta,x) + \int_S Q^{(n)}(dy|x,a)\varphi_\alpha^{(n)}(\theta,y)\right]\psi(\theta)d\theta. \qquad (2.4.19)
\end{aligned}$$

Thus we obtain

$$\begin{aligned}
&-\int_0^1 \alpha\frac{d(\theta\psi)}{d\theta}(\theta)\varphi_\alpha^{(n)}(\theta,x)d\theta \\
&= \inf_{a\in A(x)} \int_0^1 \left[\theta c_n(x,a)\varphi_\alpha^{(n)}(\theta,x) + \int_S q^{(n)}(dy|x,a)\varphi_\alpha^{(n)}(\theta,y)\right]\psi(\theta)d\theta.
\end{aligned}$$

Hence

$$\alpha\theta\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) = \inf_{a\in A(x)}\left[\theta c_n(x,a)\varphi_\alpha^{(n)}(\theta,x) + \int_S q^{(n)}(dy|x,a)\varphi_\alpha^{(n)}(\theta,y)\right] \text{ a.e. } \theta\in[0,1]$$

in the sense of distribution. Now for $\theta \in [\delta_m, 1]$, by using (2.4.4) and Proposition 2.2.3, we have

$$\begin{aligned}
\varphi_\alpha^{(n,\delta_m)}(\theta,x) &= \inf_{\pi\in\Pi_{Ad}} E_x^\pi\left[e^{n\delta_m/\alpha}\exp\left(\theta\int_0^{T_{\delta_m}(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right] \\
&\leq e^{n\delta_m/\alpha}\inf_{\pi\in\Pi_{Ad}} E_x^\pi\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right] \\
&\leq e^{n\delta_m/\alpha}\inf_{\pi\in\Pi_{Ad}} E_x^\pi\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c(\xi_t^{(n)},\pi(t))dt\right)\right]
\end{aligned}$$

$$\leq e^{n\delta_m/\alpha}\frac{\alpha^2 e^{\theta L_0/\alpha}}{\alpha^2-\rho_0\rho_1\theta}(V(x))^{\frac{\rho_1\theta}{\alpha}}.$$

Note that $\varphi_\alpha^{(n,\delta_m)} \to \varphi_\alpha^{(n)}$ as $m \to \infty$. Thus, letting $m \to \infty$ in the above equation, we obtain

$$1 \leq \varphi_\alpha^{(n)}(\theta,x) \leq \frac{\alpha^2 e^{\theta L_0/\alpha}}{\alpha^2-\rho_0\rho_1\theta}(V(x))^{\frac{\rho_1\theta}{\alpha}}. \tag{2.4.20}$$

By using (2.4.1), (2.4.2), (2.4.20), and the DE satisfied by $\varphi_\alpha^{(n)}$ (that is just proven), we see that $\varphi_\alpha^{(n)} \in L_V^\infty([0,1]\times S)$ and it is a solution of (2.4.11). Thus by closely mimicking the arguments as in Theorem 2.3.1, one can easily get the stochastic representation of the solution $\varphi_\alpha^{(n)}$, that is

$$\varphi_\alpha^{(n)}(\theta,x) = \inf_{\pi\in\Pi_{Ad}} E_x^\pi\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right]. \tag{2.4.21}$$

Step 2: In this step we prove Theorem 2.4.1, by passing to the limit as $n \to \infty$. Now we will prove that for each $x \in S$, $\{\varphi_\alpha^{(n)}\}_{n\geq 1}$ is equicontinuous on $[0,1]$. We consider the following expression, for any given $\pi \in \Pi_{Ad}$, $x \in S$, $\theta, \theta_0 \in [0,1]$:

$$\left|E_x^\pi\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right] - E_x^\pi\left[\exp\left(\theta_0\int_0^\infty e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right]\right|$$
$$\leq K_1,$$

where

$$\begin{aligned}
K_1 &= E_x^\pi\Bigg[\exp\left((\theta\wedge\theta_0)\int_0^\infty e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right) \\
&\quad\times\left(\exp\left(|\theta_0-\theta|\int_0^\infty e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)-1\right)\Bigg] \\
&\leq E_x^\pi\Bigg[\exp\left((\theta\wedge\theta_0)\int_0^\infty e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right) \\
&\quad\times\left(\exp\left(\int_0^\infty e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)-1\right)|\theta_0-\theta|\Bigg] \\
&\leq E_x^\pi\left[\exp\left(\int_0^\infty e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\times\left(\exp\left(\int_0^\infty e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)|\theta_0-\theta|\right)\right] \\
&= |\theta_0-\theta|\times E_x^\pi\left[\exp\left(2\int_0^\infty e^{-\alpha t}c_n(\xi_t^{(n)},\pi(t))dt\right)\right] \\
&\leq |\theta_0-\theta|\times\frac{\alpha e^{2L_0/\alpha}}{\alpha-\rho_2}\left(V^2(x)+\frac{b_0}{\rho_2}\right).
\end{aligned}$$

Here, the first inequality is according to $e^{bz}-1 \leq (e^b-1)z$ for all $z \in [0,1]$ and $b > 0$ and the last inequality follows from (2.3.7). Therefore, we have

$$|\varphi_\alpha^{(n)}(\theta_0,x)-\varphi_\alpha^{(n)}(\theta,x)| \leq \sup_{\pi\in\Pi_{Ad}}|\theta_0-\theta|\times\frac{\alpha e^{2L_0/\alpha}}{\alpha-\rho_2}\left(V^2(x)+\frac{b_0}{\rho_2}\right)$$

$$= |\theta_0 - \theta| \times \frac{\alpha e^{2L_0/\alpha}}{\alpha - \rho_2}\left(V^2(x) + \frac{b_0}{\rho_2}\right). \tag{2.4.22}$$

By measurable selection theorem, [17, Proposition 7.33], there exists a measurable function $f^{*n} : [0,1] \times S \to A$ such that

$$\begin{aligned} &\inf_{a \in A(x)}\left[\theta c_n(x,a)\varphi_\alpha(\theta,x) + \int_S q^{(n)}(dy|x,a)\varphi_\alpha(\theta,y)\right] \\ &= \left[\theta c_n(x, f^{*n}(\theta,x))\varphi_\alpha(\theta,x) + \int_S q^{(n)}(dy|x, f^{*n}(\theta,x))\varphi_\alpha(\theta,y)\right]. \end{aligned} \tag{2.4.23}$$

Let

$$\pi^{*n} : S \times \mathbb{R}_+ \to A$$

be defined by

$$\pi^{*n}(x,t) := f^{*n}(\theta e^{-\alpha t}, x).$$

Hence by equation (2.4.11), we have a.e. $\theta \in [0,1]$ and $\forall x \in S$, we have

$$\begin{cases} \alpha\theta\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) & = \left[\int_S q^{(n)}(dy|x, f^{*n}(\theta,x))\varphi_\alpha^{(n)}(\theta,y) + \theta c_n(x, f^{*n}(\theta,x))\varphi_\alpha^{(n)}(\theta,x)\right] \\ 1 \le \varphi_\alpha^{(n)}(\theta,x) & \le \frac{\alpha^2 e^{\theta L_0/\alpha}}{\alpha^2 - \rho_0\rho_1\theta}(V(x))^{\frac{\rho_1\theta}{\alpha}} \quad \forall\ (\theta,x) \in [0,1] \times S. \end{cases} \tag{2.4.24}$$

Since $c_n \ge 0$, by (2.4.21), we see $\varphi_\alpha^{(n)}(\theta,x)$ is increasing in θ. Also we know that $\varphi_\alpha^{(n)}(\theta,x)$ is differentiable a.e. with respect to $\theta \in [0,1]$. So

$$\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) \ge 0 \quad \text{for a.e. } \theta. \tag{2.4.25}$$

So, by (2.4.1), (2.4.2) and (2.4.24), for all $x \in S$ and for a.e. θ, we have

$$\begin{cases} -\alpha\theta\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) + \left[\int_S q^{(n-1)}(dy|x, f^{*n}(\theta,x))\varphi_\alpha^{(n)}(\theta,y) + \theta c_{n-1}(x, f^{*n}(\theta,x))\varphi_\alpha^{(n)}(\theta,x)\right] \le 0 \\ \qquad \text{if } x \in S_{n-1} \end{cases} \tag{2.4.26}$$

and

$$\begin{cases} -\alpha\theta\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) + \left[\int_S q^{(n-1)}(dy|x, f^{*n}(\theta,x))\varphi_\alpha^{(n)}(\theta,y) + \theta c_{n-1}(x, f^{*n}(\theta,x))\varphi_\alpha^{(n)}(\theta,x)\right] \\ = -\alpha\theta\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) \le 0 \\ \qquad \text{if } x \notin S_{n-1} \text{ (by (2.4.25)).} \end{cases} \tag{2.4.27}$$

So, by the Dynkin formula, we get

$$E_x^{\pi^{*n}}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c_{n-1}(\xi_t^{(n-1)}, \pi^{*n}(\xi_{t-}^{(n-1)}, t))dt\right)\right] \le \varphi_\alpha^{(n)}(\theta,x) \text{ for all } (\theta,x) \in [0,1] \times S. \tag{2.4.28}$$

Also using (2.4.11) and Dynkin formula (see (2.3.8) and (2.3.13)), we have

$$\varphi_\alpha^{(n-1)}(\theta,x) \leq E_x^{\pi^{*n}}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c_{n-1}(\xi_t^{(n-1)},\pi^{*n}(\xi_{t-}^{(n-1)},t))dt\right)\right]. \tag{2.4.29}$$

By (2.4.28) and (2.4.29), we have $\varphi_\alpha^{(n-1)}(\theta,x) \leq \varphi_\alpha^{(n)}(\theta,x)$.

Hence $\varphi_\alpha^{(n)}(\theta,x)$ is increasing in n for any $(\theta,x) \in [0,1]\times S$. Now from (2.4.22), we know that for each $x \in S$, $\varphi^{(n)}(\cdot,x)$ is Lipschitz continuous in $\theta \in [0,1]$. Also, $\varphi_\alpha^{(n)}(\theta,x)$ is increasing as $n \to \infty$ for any $(\theta,x) \in [0,1]\times S$ and bounded above (by (2.4.20)), therefore there exists a function φ_α on $[0,1]\times S$ that is continuous with respect to $\theta \in [0,1]$, such that along a subsequence $n_k \to \infty$, we have $\lim_{n_k\to\infty}\varphi_\alpha^{(n_k)}(\theta,x) = \varphi_\alpha(\theta,x)$ and this convergence is uniform in $\theta \in [0,1]$ for each fixed $x \in S$. Moreover, by (2.4.20), we have

$$1 \leq \varphi_\alpha(\theta,x) \leq \frac{\alpha^2 e^{\theta L_0/\alpha}}{\alpha^2-\rho_0\rho_1\theta}(V(x))^{\frac{\rho_1\theta}{\alpha}}. \tag{2.4.30}$$

As the proof of equation (2.4.11) in step 1 (starting from the first equality of (2.4.13)), we see that φ_α is a solution to the HJB equation (2.3.1). Also by (2.4.30), we can conclude that $\varphi_\alpha \in L_V^\infty([0,1]\times S)$. Finally, the uniqueness of $\varphi_\alpha(\theta,x)$ follows from the stochastic representation in Theorem 2.3.1. □

2.5 The existence of optimal control

In this section, we present the main result of this chapter. Here we show the existence of an optimal control.

Theorem 2.5.1. *Suppose that Assumptions* (A1)-(A3) *are satisfied. Then, the following assertions hold.*

1. *The HJB equation* (2.3.1) *has a unique solution* $\varphi_\alpha \in L_V^\infty([0,1]\times S)$ *and the solution admits the following representation*

$$\begin{aligned}1 \leq \varphi_\alpha(\theta,x) &= \inf_{\pi\in\Pi_{Ad}} E_x^\pi\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c(\xi_t,\pi(t))dt\right)\right]\\ &\leq \frac{\alpha^2 e^{\theta L_0/\alpha}}{\alpha^2-\rho_0\rho_1\theta}(V(x))^{\frac{\rho_1\theta}{\alpha}}.\end{aligned}$$

2. *There exists a measurable function* $f^* : [0,1]\times S \to A$ *such that*

$$\alpha\theta\frac{\partial\varphi_\alpha}{\partial\theta}(\theta,x) = \left[\int_S q(dy|x,f^*(\theta,x))\varphi_\alpha(\theta,y) + \theta c(x,f^*(\theta,x))\varphi_\alpha(\theta,x)\right]$$
$$a.e.\ \theta \in [0,1]. \tag{2.5.1}$$

3. *Furthermore an optimal Markov control for the cost criterion (2.2.2) exists and is given by*
$$\tilde{\pi}^*(x,t) := f^*(\theta e^{-\alpha t}, x),$$
where f^* satisfies (2.5.1).

Proof. Part (1) follows from Theorems 2.3.1 and 2.4.1.
To prove (2), by [69], we first observe that the function

$$\int_S q(dy|x,a)\varphi_\alpha(\theta,y) + \theta c(x,a)\varphi_\alpha(\theta,x)$$

is continuous in $a \in A(x)$ for each given $(\theta,x) \in [0,1] \times S$. Thus, by the measurable selection theorem [17, Proposition 7.33] there exists a measurable function f^* satisfying (2.5.1), and so (2) follows.
For part (3), take any f^* that satisfies (2.5.1). Then by Theorem 2.3.1, we have $\inf_{\pi \in \Pi_{Ad}} \tilde{J}_\alpha(\theta,x,\pi) = \tilde{J}_\alpha(\theta,x,\tilde{\pi}^*) = \varphi_\alpha(\theta,x)$, which together with (2.2.2), (2.2.3) and part (1), we have $\inf_{\pi \in \Pi_{Ad}} \mathscr{J}_\alpha(\theta,x,\pi) = \mathscr{J}_\alpha(\theta,x,\tilde{\pi}^*) = \dfrac{1}{\theta}\ln \tilde{J}_\alpha(\theta,x,\tilde{\pi}^*) = \dfrac{1}{\theta}\ln\varphi_\alpha(\theta,x)$. Hence $\tilde{\pi}^*$ is an optimal Markov control. ☐

Now we prove the converse of the Theorem 2.5.1.

Theorem 2.5.2. *Grant Assumptions (A1)-(A3). Suppose there exists an optimal Markov control for the cost criterion (2.2.2) and is given by*

$$\hat{\pi}^*(x,t) := \tilde{f}^*(\theta e^{-\alpha t}, x),$$

for some measurable function $\tilde{f}^*$. Then we prove that $\tilde{f}^*$ is a minimizing selector of (2.3.1).

Proof. Since $\hat{\pi}^*$ is optimal for the cost criterion (2.2.2), therefore we have

$$\inf_{\pi \in \Pi_{Ad}} \tilde{J}_\alpha(\theta,x,\pi) = \tilde{J}_\alpha(\theta,x,\hat{\pi}^*) = \tilde{J}^*_\alpha(\theta,x). \tag{2.5.2}$$

Now for $\tilde{f}^*$ by Theorem 2.4.1, there exists a unique solution $\psi_\alpha \in L^\infty_V([0,1]\times S)$ for the equation

$$\alpha\theta\frac{\partial\psi_\alpha}{\partial\theta}(\theta,x) = \left[\int_S q(dy|x,\tilde{f}^*(\theta,x))\psi_\alpha(\theta,y) + \theta c(x,\tilde{f}^*(\theta,x))\psi_\alpha(\theta,x)\right], \tag{2.5.3}$$

for each $x \in S$ and a.e. $\theta \in [0,1]$, satisfying $1 \le \psi_\alpha(\theta,x) \le \frac{\alpha^2 e^{\theta L_0/\alpha}}{\alpha^2-\rho_0\rho_1\theta}(V(x))^{\frac{\rho_1\theta}{\alpha}}$ for all $(\theta,x) \in [0,1]\times S$.

Now by Theorem 2.3.1, we know that

$$\begin{aligned} 1 \le \psi_\alpha(\theta,x) &= E_x^{\hat{\pi}^*}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c(\xi_t,\hat{\pi}^*(\xi_{t-},t))dt\right)\right] \\ &\le \frac{\alpha^2 e^{\theta L_0/\alpha}}{\alpha^2-\rho_0\rho_1\theta}(V(x))^{\frac{\rho_1\theta}{\alpha}}. \end{aligned} \tag{2.5.4}$$

From (2.5.2) and (2.5.4), we get

$$\psi_\alpha(\theta,x) = \inf_{\pi\in\Pi_{Ad}} \tilde{J}_\alpha(\theta,x,\pi) = \tilde{J}_\alpha(\theta,x,\hat{\pi}^*) = \tilde{J}^*_\alpha(\theta,x) \text{ for } (\theta,x)\in[0,1]\times S. \tag{2.5.5}$$

So, in view of Theorem 2.3.1, by equations (2.3.1), (2.5.3), and (2.5.5), we conclude that $\tilde{f}^*$ is a minimizing selector of (2.3.1). ☐

When the transition and cost rates are bounded, the existence of an optimal control is ensured by Theorem 2.5.1.

Corollary 2.5.1. *Grant Assumption* $(A3)$*((i)-(ii)). Also, assume that the transition and cost rates are bounded. Then, there exist a unique solution* φ_α *and an optimal control for the HJB equation (2.3.1).*

Proof. Suppose there exist constants L_1 and b_1, such that $\sup_{(x,a)\in K} q_x(a) \le L_1$ and $\sup_{(x,a)\in K} c(x,a) \le b_1$. First we take the Lyapunov function $V(x)\equiv P$, for all $x\in S$, $P\ge 1$, a constant. Now $\int_S V(y)q(dy|x,a) = \int_S V^2(y)q(dy|x,a) = 0$, for all $(x,a)\in K$. Now, take $\rho_0=\alpha$, $M_0=L_1$, any real number, $\rho_1\in(0,\alpha)$, and $L_0=b_1$. Then Assumption $(A1)$ is verified. Now for all $x\in S$, take any constants $\rho_2\in(0,\alpha)$ and $b_0\in(0,\infty)$. Then Assumption $(A2)$ holds. Also $\int_S V(y)q(dy|x,a)$ is continuous in $a\in A(x)$. So, Assumption $(A3)$ is also true. Then, by Theorem 2.5.1, we have a unique solution φ_α and an optimal control for the HJB equation (2.3.1). ☐

2.6 Application and example

In this section, we verify the above assumptions with one example, where the transition and cost rates are unbounded.

Example 2.6.1. ***The Gaussian Model:*** *Suppose a hunter is hunting outside his house for his manager. Suppose the house is at state 0. A positive state represents the distance from the house to the right, and a negative state represents the distance from the house to the left. Let* $S=\mathbb{R}$*. If the current position is* $x\in S$*, the hunter takes an action* $a\in A(x)$*, then after an exponentially distributed travel time with rate* $\lambda(x,a)>0$*, the hunter reaches the new position, and the travel distance follows the normal distribution* *with mean* x *and variance* σ*. (Or we can interpret* $\lambda(x,a)$ *as the total jump intensity that is an arbitrary measurable positive-valued function on* $S \times A$*, and the distribution of the state after a jump from* $x \in S$ *is normal with the variance* σ *and expectation* x*.) Also assume that the hunter receives a payoff* $c(x,a)$ *from his manager for each unit of time he spends there. Let us consider the model as* $A_2 := \{S, (A, A(x), x \in S), c(x,a), q(dy|x,a)\}$*, where* $S = (-\infty, \infty)$*. For each* $D \in \mathcal{B}(S)$*, the transition rate is*

$$q(D|x,a) = \lambda(x,a)\left[\int_{y \in D} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-x)^2}{2\sigma^2}} dy - \delta_x(D)\right], \; x \in S, a \in A(x), \sigma > 0. \quad (2.6.1)$$

To ensure the existence of an optimal Markov control for the model, we consider the following hypotheses.

(I) For each fixed $x \in S$*,* $\lambda(x,a)$ *is continuous in* $a \in A(x)$ *and there exists a positive constant* M_1 *such that* $0 < \sup_{a \in A(x)} \lambda(x,a) \le M_1(x^2+1)$ *and* $M_1 < \frac{\alpha}{6\sigma^2(\sigma^2+1)}$*.*

(II) For each $x \in S$*, the cost rate* $c(x,a)$ *is nonnegative and continuous in* $a \in A(x)$ *and there exists a constant* $0 < \rho_1 < \min\{\alpha, \frac{\alpha^2}{M_1\sigma^2}\}$ *such that*

$$\sup_{a \in A(x)} c(x,a) \le \rho_1 \log(1+x^2).$$

(III) For each fixed $x \in S$*,* $A(x)$ *is a compact subset of the Borel spaces* A*.*

Proposition 2.6.2. *Under conditions (I)-(III), the above controlled system satisfies the Assumptions* $(A1)$*-*$(A3)$*. Hence by Theorem 2.5.1, there exists an optimal Markov control for this model.*

Proof. We know $\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(y-x)^{2k+1}e^{-\frac{(y-x)^2}{2\sigma^2}}dy = 0$ and $\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(y-x)^{2k}e^{-\frac{(y-x)^2}{2\sigma^2}}dy = 1 \cdot 3 \cdots (2k-1)\sigma^{2k}$ for all $k = 0, 1 \cdots$.
We first verify Assumption $(A1)$. Let $V(x) = x^2+1$.

$$\begin{aligned}\int_S V(y)q(dy|x,a) &= \lambda(x,a)\left[\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(y^2+1)e^{-\frac{(y-x)^2}{2\sigma^2}}dy - (x^2+1)\right] \\ &= \lambda(x,a)\sigma^2 \\ &\le M_1\sigma^2\mathcal{V}(x). \end{aligned} \quad (2.6.2)$$

Let $\rho_0 = M_1\sigma^2$. Then $\int_S V(y)q(dy|x,a) \le \rho_0 V(x)$. Now

$$q^*(x) = \sup_{a \in A(x)} q_x(a) = \sup_{a \in A(x)} \lambda(x,a) \le M_1(x^2+1) = M_1 V(x) \; \forall \; x \in S.$$

Now by condition (II), we can write

$$\sup_{a \in A(x)} c(x,a) \le \rho_1 \log(1+x^2) + M_1.$$

Observe that by condition (II), $0 < \rho_1 < \min\{\alpha, \rho_0^{-1}\alpha^2\}$. Hence Assumption $(A1)$ is verified with $M_0 = L_0 = M_1$.
Next we verify Assumption $(A2)$.
For any $x \in S$, $a \in A(x)$,

$$\begin{aligned}
\int_S q(dy|x,a)V^2(y) &= \int_S q(dy|x,a)(1+y^2)^2 \\
&= \lambda(x,a)\left[\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(y^2+1)^2 e^{-\frac{(y-x)^2}{2\sigma^2}}dy - (x^2+1)^2\right] \\
&= \lambda(x,a)[1\cdot 3\sigma^4 + \sigma^2(2+6x^2)] \\
&\le M_1(x^2+1)[3\sigma^4 + \sigma^2(2+6x^2)] \\
&= M_1\sigma^2(x^2+1)(3\sigma^2+2+6x^2) \\
&\le 6M_1\sigma^2(x^2+1)(x^2+1)(\sigma^2+1) \\
&= 6M_1V^2(x)\sigma^2(\sigma^2+1) \\
&\le \rho_2 V^2(x) + 1
\end{aligned}$$

where $\rho_2 = 6M_1\sigma^2(\sigma^2+1)$, and $b_0 = 1$. Then by condition (I), we have $0 < \rho_2 < \alpha$. Hence, Assumption $(A2)$ is verified. Now by conditions (I) and (II) $c(x,a)$ is continuous in $a \in A(x)$. Observe that by condition (I) and (2.6.2), $\int_S V(y)q(dy|x,a)$ is continuous in $a \in A(x)$. Hence Assumption $(A3)$ is also verified. So, by Theorem 2.5.1, we see that there exists an optimal Markov control for this model. □

Remark 2.6.3. *As we mention in the introduction, there are many real-life applications, where the underlying system dynamic is modeled as a CTMDP, with a Borel state and action spaces as well as cost and transition rates are unbounded, see such a cash-flow problem in [65], [100, p. 112]. Also, there are lots of real-life examples like infrastructure surveillance models [100, p. 115-116], queueing model [100, p. 192], where we see that the state space is uncountable, can be formulated in our set-up.*

CHAPTER 3

Continuous-time Zero-Sum Games for Markov Decision Processes with Discounted Risk-Sensitive Cost Criterion

3.1 Introduction

In this chapter, we study two-person infinite-horizon discounted-cost risk-sensitive zero-sum stochastic games for continuous-time Markov decision processes (CTMDPs) with unbounded transition and reward rates on countable state space. Finite horizon risk-sensitive CTMDP is considered in [47], [53], [108], while for infinite-horizon risk-sensitive CTMDP see, [47], [49], [54], [81], [82], and the references therein. Infinite horizon risk-sensitive CTMDP for piecewise deterministic Markov decision processes has been studied in [54]. In [47], [82], [94], the transition and cost rates are assumed to be bounded. In [49], [62], the authors have studied infinite-horizon risk-sensitive CTMDP for unbounded transition and cost rates.

Discrete-time zero-sum risk-sensitive stochastic games on a countable state space have been considered by several authors, see [12], [41] and references therein. In [16], the authors studied zero-sum risk-sensitive stochastic games for discrete-time MDP with general state space. But the corresponding literature in the context of zero-sum risk-sensitive stochastic games for continuous-time MDP is rather limited. Some exceptions are [45], [109].

In this chapter, we have extended the risk-sensitive discounted control problem of article [62] to a zero-sum risk-sensitive discounted stochastic game problem for CTMDPs with countable state space. Here, the maximizer (player 1) tries to maximize his/her infinite-horizon risk-sensitive rewards over his strategies whereas the minimizer (player 2) tries to minimize the same over his/her strategies over their admissible strategies. This game model is primarily been formulated from the viewpoint of the minimizer player

(player 2) who is risk-averse. The maximizer is a virtual player (player 1) who is antagonistic to the minimizing player (player 2), see Chapter 1. As the system evolves, the rewards are accumulated and the performance of a strategy is measured by discounted risk-sensitive cost criterion, which in our present case is defined by (3.2.4), below. Such a model is relevant in worst-case scenarios, e.g., in financial applications when a risk-averse investor is trying to maximize his long-term portfolio gain against the market which, by default, is the minimizer in this case. The main objective of this work is to prove the existence of a value of the game and a saddle-point equilibrium and give a characterization of the optimal strategies in terms of the corresponding Hamilton-Jacobi-Isaacs (HJI) equation.

In [109], a zero-sum risk-sensitive stochastic game on the finite horizon has been studied for the Markov jump process with Borel state space, the costs are bounded and transition rates are unbounded. The corresponding game on the infinite-horizon has been studied in [45] for CTMDP with bounded transition and cost rates. This boundedness requirement, however, imposes some restrictions in applications, for instance in queueing control and population processes, where the transition and reward/cost rates are usually unbounded.

To the best of our knowledge, this chapter is the first work that deals with infinite-horizon zero-sum risk-sensitive stochastic games for continuous-time MDP with unbounded cost and transition rates and with admissible (feedback) policies. A natural technique to solve the zero-sum stochastic game problem under study is to characterize the value function as a solution to the HJI equation. Our results follow this approach. In Section 3.4, we provide sufficient conditions for the existence of a solution to the HJI equation in Theorem 3.4.1 and show that the solution of this HJI equation is in fact unique and coincides with the value function of the zero-sum game problem under consideration. Moreover, the existence of a saddle-point equilibrium in the class of Markov strategies is proven in Theorem 3.5.1.

The rest of this chapter is structured as follows. Section 3.2 deals with the description of the problem. In Section 3.3, we prove the stochastic representation of the solution of HJI equation (3.3.1). The existence of a unique solution to the HJI equation is proven in Section 3.4. In Section 3.5, we prove the value and saddle-point equilibrium in the class of Markov strategies for the discounted-cost risk-sensitive zero-sum game. In Section 3.6, we illustrate our theory and assumptions by an illustrative example. The content of this chapter is based on the published article [51].

3.2 The game model

In this section, we introduce the continuous-time two-person zero-sum stochastic game model which consists of the following elements

$$\{S, A, B, (A(i) \subset A, B(i) \subset B, i \in S), q(\cdot|i,a,b), c(i,a,b)\}, \tag{3.2.1}$$

- S, called the state space, is assumed to be the set of all nonnegative integers endowed with the discrete topology.

- A and B are the action sets for players 1 and 2, respectively. The action spaces A and B are assumed to be Borel spaces with the Borel σ-algebras $\mathcal{B}(A)$ and $\mathcal{B}(B)$, respectively.

- For each $i \in S$, $A(i) \in \mathcal{B}(A)$ and $B(i) \in \mathcal{B}(B)$ denote the sets of admissible actions for players 1 and 2 in state i, respectively. Let $K := \{(i,a,b)|i \in S, a \in A(i), b \in B(i)\}$, which is a Borel subset of $S \times A \times B$.

- Given any $(i,a,b) \in K$, the transition rate $q(j|i,a,b)$ is a signed kernel on S such that $q(j|i,a,b) \geq 0$ for all $j, i \in S$ with $j \neq i$. Moreover, we assume that $q(j|i,a,b)$ satisfies the following uniformly conservative and stable conditions: for any $i \in S$,

$$\begin{aligned} &\sum_{j\in S} q(j|i,a,b) = 0 \text{ uniformly in } (a,b) \in A(i) \times B(i) \quad \text{and} \\ &q^*(i) := \sup_{(a,b)\in A(i)\times B(i)} q_i(a,b) < \infty, \end{aligned} \tag{3.2.2}$$

  where $q_i(a,b) := -q(i|i,a,b) \geq 0$.

- Finally, the measurable function $c : K \to \mathbb{R}_+$ denotes the reward rate function for player 1 (or the cost rate function for player 2).

The game evolves as follows. The players observe continuously the current state of the system. When the system is in state $i \in S$ at time $t \geq 0$, the players independently choose actions $a_t \in A(i)$ and $b_t \in B(i)$ according to some strategies, respectively. As a consequence of this, the following happens:

- player 1 receives an immediate reward at rate $c(i, a_t, b_t)$ and player 2 incurs a cost at rate $c(i, a_t, b_t)$; and

- the system stays in state i for a random time, with rate of leaving i given by $q_i(a_t, b_t)$, and then jumps to a new state $j \neq i$ with the probability determined by $\dfrac{q(\cdot|i,a_t,b_t)}{q_i(a_t,b_t)}$ (see Proposition B.8 in [59, p. 205] for details).

When the state of the system transits to the new state j, the above procedure is repeated.

Thus, the goal of player 1 is to maximize his/her rewards, whereas that of player 2 is to minimize his/her costs with respect to some performance criterion $J_\alpha(\cdot,\cdot,\cdot,\cdot)$, which in our present case is defined by (3.2.4), below. Such a model is relevant in worst-case scenarios, e.g., in financial applications when a risk-averse investor is trying to maximize his long-term portfolio gain against the market which, by default, is the minimizer in this case.

To formalize what is described above, below we describe the construction of continuous-time Markov decision processes (CTMDPs) under possibly admissible (feedback) strategies. To construct the underlying CTMDPs (as in [63], [75], [99]) we define $S_\Delta := S \cup \{\Delta\}$ (with some $\Delta \notin S$), $\mathbb{R}_+ := (0,\infty)$. Now define the measurable space $(\Omega, \mathcal{B}(\Omega))$ with $\Omega := (S \times \mathbb{R}_+)^\infty \cup \{(i_0', \theta_1, i_1', \cdots, \theta_k, i_k', \infty, \Delta, \infty, \Delta, \cdots) | k \geq 0, i_l' \in S, \theta_l \in \mathbb{R}_+ \; \forall 0 \leq l \leq k\}$ and the Borel σ-algebra $\mathcal{B}(\Omega)$ of Ω. For each $\omega := (i_0', \theta_1, i_1', \cdots, \theta_k, i_k', \cdots) \in \Omega$, define $T_0(\omega) := 0$, $T_n(\omega) := T_{n-1}(\omega) + \theta_n$, $T_\infty(\omega) := \lim_{n\to\infty} T_n(\omega)$. Using $\{T_k\}$, we define the state process $\{\xi_t\}_{t\geq 0}$ as

$$\xi_t(\omega) := \sum_{k\geq 0} I_{\{T_k \leq t < T_{k+1}\}} i_k' + I_{\{t \geq T_\infty\}} \Delta, \text{ for } t \geq 0. \tag{3.2.3}$$

Here, I_E denotes the indicator function of a set E, and we use the convention that $0 + z =: z$ and $0z =: 0$ for all $z \in S_\Delta$. The process after T_∞ is regarded to be absorbed in the state Δ. Thus, let $q(\cdot|\Delta, a_\Delta, b_\Delta) :\equiv 0$, $A_\Delta := A \cup \{a_\Delta\}$, $B_\Delta := B \cup \{b_\Delta\}$, $A(\Delta) := \{a_\Delta\}$, $B(\Delta) := \{b_\Delta\}$, $c(\Delta, a, b) :\equiv 0$ for all $(a,b) \in A_\Delta \times B_\Delta$, where a_Δ, b_Δ are isolated points. Moreover, let $\mathscr{F}_t := \sigma(\{T_k \leq s, \xi_{T_k} \in S\} : 0 \leq s \leq t, k \geq 0)$ for all $t \geq 0$, $\mathscr{F}_{s-} =: \bigvee_{t<s} \mathscr{F}_t$, and $\mathscr{P} := \sigma(C \times \{0\}, D \times (s,\infty) : C \in \mathscr{F}_0, D \in \mathscr{F}_{s-})$ which denotes the σ-algebra of predictable sets on $\Omega \times [0,\infty)$ related to $\{\mathscr{F}_t\}_{t\geq 0}$.

To complete the specification of a risk-sensitive stochastic game problem, we need, of course, to introduce an optimality criterion. This requires defining the class of strategies as below.

Definition 3.2.1. *An admissible feedback strategy for player 1, denoted by* $\pi^1 = \{\pi^1(t)\}_{t\geq 0}$, *is a transition probability* $\pi^1(da|\omega,t)$ *from* $(\Omega \times [0,\infty), \mathscr{P})$ *onto* $(A_\Delta, \mathcal{B}(A_\Delta))$, *such that* $\pi^1(A(\xi_{t-}(\omega))|\omega,t) = 1$. *Using appropriate projections of the transition kernel* π^1, *an admissible feedback strategy for player 1, determines and is, in turn, determined by a sequence* $\{\pi_k^1, k \geq 0\}$ *of the stochastic kernel on* A *such that*

$$\begin{aligned}\pi^1(t)(\omega) &= \pi^1(da|\omega,t)\\ &= I_{\{t=0\}}(t)\pi_0^1(da|i_0',0) + \sum_{k\geq 0} I_{\{T_k < t \leq T_{k+1}\}} \pi_k^1(da|i_0', \theta_1, i_1', \ldots, \theta_k, i_k', t - T_k)\\ &\quad + I_{\{t \geq T_\infty\}} \delta_{a_\Delta}(da),\end{aligned}$$

where $\pi_0^1(da|i_0',0)$ *is a stochastic kernel on* A *given* S *such that* $\pi_0^1(A(i_0')|i_0',0) = 1$, $\pi_k^1(k \geq 1)$ *are stochastic kernels on* A *given* $(S \times (0,\infty))^{k+1}$ *such that* $\pi_k^1(A(i_k')|i_0',\theta_1,i_1',\cdots,\theta_k,i_k',t-T_k) = 1$, *and* $\delta_{a_\Delta}(da)$ *denotes the Dirac measure at the point* a_Δ.

For more details see [64, Definition 2.1, Remark 2.2], [99], [119]. The set of all admissible feedback strategies for player 1 is denoted by Π_{Ad}^1. A strategy $\pi^1 \in \Pi_{Ad}^1$ for player 1, is called a Markov if $\pi^1(t)(\omega) = \pi^1(\xi_{t-}(\omega),t)$ i.e., $\pi^1(da|\omega,t) = \pi^1(da|\xi_{t-}(\omega),t)$ for every $\omega \in \Omega$ and $t \geq 0$, where $\xi_{t-}(\omega) := \lim_{s\uparrow t}\xi_s(\omega)$. For notational simplicity, we would not write ω anywhere throughout the rest of this chapter. We denote by Π_M^1 the family of all Markov strategies for player 1. The sets of admissible feedback strategies Π_{Ad}^2 and Markov strategies Π_M^2 for player 2 are defined analogously.

Remark 3.2.2. *In the definition of strategies we do not include the entire history of the game, i.e., past and present states, past sojourn times, and past actions taken by the players. In our game model, each player's admissible strategies include only past and present states and past sojourn times. Hence such strategies are called feedback strategies. If players use general strategies (i.e., history-dependent non-anticipative strategies) there may not be a probability measure over plays; see Proposition 1 in [88]. See also [85]. Thus it is imperative for us to confine our attention to specific classes of strategies. In this chapter, we restrict our attention to feedback strategies only, i.e., at any point of time each player has access to past and present states and past sojourn times. Though the past and present states and past sojourn times implicitly contain the past actions of the players, explicit inclusions thereof in the strategies run into unassailable technical issues as explained clearly in [88]. The inclusion of history-dependent strategies is infeasible even for one-player games; see Proposition 1 in [88].*

For any compact metric space Y, let $\mathcal{P}(Y)$ denote the space of probability measures on Y with Prohorov topology. For each $i,j \in S$, $\mu \in \mathcal{P}(A(i))$ and $\nu \in \mathcal{P}(B(i))$, the associated transition and cost rates are defined, respectively, as follows:

$$q(j|i,\mu,\nu) := \int_{B(i)}\int_{A(i)} q(j|i,a,b)\mu(da)\nu(db),$$

$$c(i,\mu,\nu) := \int_{B(i)}\int_{A(i)} c(i,a,b)\mu(da)\nu(db).$$

(by abuse of notation we use the same notation q and c). Under Assumption $(A1)$, below, for any initial state $i \in S$ and any pair of strategies $(\pi^1,\pi^2) \in \Pi_{Ad}^1 \times \Pi_{Ad}^2$, Theorem 4.27 in [76] yields the existence of a unique probability measure denoted by $P_i^{\pi^1,\pi^2}$ on $(\Omega,\mathscr{F})$. Let $E_i^{\pi^1,\pi^2}$ be the expectation operator with respect to $P_i^{\pi^1,\pi^2}$.

Fix any discounted factor $\alpha > 0$. For any $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$ and $i \in S$, the risk-sensitive discounted criterion is defined as

$$J_\alpha(\theta, i, \pi^1, \pi^2) := \frac{1}{\theta} \log\left\{ E_i^{\pi^1,\pi^2}\left[\exp\left(\theta \int_0^\infty e^{-\alpha t} c(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right]\right\}, \tag{3.2.4}$$

provided that the integral is well defined, where ξ_t is the Markov process corresponding to $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$ and $\theta \in (0, 1]$ denotes a risk-sensitive parameter and the limiting case of $\theta \to 0$ is the risk-neutral case.

We also need the following concepts. The functions on S are defined as

$$L(i) := \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} J_\alpha(\theta, i, \pi^1, \pi^2) \quad \text{and} \quad U(i) := \inf_{\pi^2 \in \Pi^2_{Ad}} \sup_{\pi^1 \in \Pi^1_{Ad}} J_\alpha(\theta, i, \pi^1, \pi^2)$$

are called, respectively, the lower value and the upper value of the game. It is clear that $L(i) \leq U(i)$ for all $i \in S$.

Definition 3.2.3. *If $L(i) = U(i)$ for all $i \in S$, then the common function is called the value of the game and is denoted by $J^*_\alpha(\theta, i)$.*

Definition 3.2.4. *Suppose that the game has a value J^*_α. Then a strategy π^{*1} in Π^1_{Ad} is said to be optimal for player 1 if $\inf_{\pi^2 \in \Pi^2_{Ad}} J_\alpha(\theta, i, \pi^{*1}, \pi^2) = J^*_\alpha(\theta, i)$ for all $i \in S$. Similarly, $\pi^{*2} \in \Pi^2_{Ad}$ is optimal for player 2 if $\sup_{\pi^1 \in \Pi^1_{Ad}} J_\alpha(\theta, i, \pi^1, \pi^{*2}) = J^*_\alpha(\theta, i)$ for all $i \in S$.*
If $\pi^{*k} \in \Pi^k_{Ad}$ is optimal for player k (k=1,2), then (π^{*1}, π^{*2}) is called a pair of optimal strategies and also called a saddle-point equilibrium.

The objective of this chapter is to provide conditions for the existence of a pair of optimal strategies and introduce an HJI characterization of such pairs.

To avoid the explosion of the state process $\{\xi_t, t \geq 0\}$, we need the following Assumption imposed on the transition rates, which had been widely used in CTMDPs; see, for instance, [53], [62], [63], [64] and the references therein.

(*A*1) There exists a Lyapunov function $V : S \to [1, \infty)$ such that

(i) $\sum_{j \in S} V(j) q(j|i, a, b) \leq \rho_1 V(i)$ for all $(i, a, b) \in K$ with some constant $\rho_1 \geq 0$;

(ii) $\sup_{(a,b) \in A(i) \times B(i)} q_i(a, b) \leq L_1 V(i)$ for all $i \in S$ with some positive constant L_1.

Proposition 3.2.5. *Under Assumption (A1), for any pair of strategies $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$ and $i \in S$, the following assertions hold:*

(a) $P_i^{\pi^1,\pi^2}(T_\infty = \infty) = 1$ and $P_i^{\pi^1,\pi^2}(\xi_t \in S) = 1$ for all $t \geq 0$;

(b) $E_i^{\pi^1,\pi^2}[V(\xi_t)] \le e^{\rho_1 t}V(i)$ *for all* $t \ge 0$.

Proof. The proof follows as in ([63, Theorem 3.1]). □

Since logarithm is an increasing function, instead of studying $J_\alpha(\theta, i, \pi^1, \pi^2)$, we will consider $\hat{J}_\alpha(\theta, i, \pi^1, \pi^2)$ on $[0,1] \times S \times \Pi^1_{Ad} \times \Pi^2_{Ad}$ defined by

$$\hat{J}_\alpha(\theta, i, \pi^1, \pi^2) := E_i^{\pi^1,\pi^2}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right]. \tag{3.2.5}$$

Obviously, $\hat{J}_\alpha(\theta, i, \pi^1, \pi^2) \ge 1$ for $(\theta, i) \in (0,1] \times S$ and $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$, and we have (π^{*1}, π^{*2}) is optimal if and only if

$$\hat{J}_\alpha(\theta, i, \pi^{*1}, \pi^{*2}) = \sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}} \hat{J}_\alpha(\theta, i, \pi^1, \pi^2) = \inf_{\pi^2\in\Pi^2_{Ad}}\sup_{\pi^1\in\Pi^1_{Ad}} \hat{J}_\alpha(\theta, i, \pi^1, \pi^2) =: \hat{J}^*_\alpha(\theta, i) \forall i \in S.$$

To ensure the finiteness of $J_\alpha(\theta, i, \pi^1, \pi^2)$, we enforce a logarithmic growth condition for the cost function c as below.

(*A2*) There exist constants $L_2 \ge 0$ and $0 < \rho_2 < \min\{\alpha, \rho_1^{-1}\alpha^2\}$ such that

$$\sup_{(a,b)\in A(i)\times B(i)} c(i, a, b) \le \rho_2 \log V(i) + L_2 \quad \forall i \in S,$$

where V and ρ_1 are introduced in Assumption $(A1)$.

Proposition 3.2.6. *Suppose that Assumptions $(A1)$ and $(A2)$ are satisfied. Then, for any pair of strategies $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$ and $(\theta, i) \in [0,1] \times S$, we have*

$$\hat{J}_\alpha(\theta, i, \pi^1, \pi^2) \le \frac{\alpha^2}{\alpha^2 - \rho_1\rho_2\theta} e^{\theta L_2/\alpha} V^{\rho_2\theta/\alpha}(i) \le \frac{\alpha^2}{\alpha^2 - \rho_1\rho_2} e^{L_2/\alpha} V(i).$$

Moreover, we obtain

$$J^*_\alpha(\theta, i) \le \log\left(\frac{\alpha^2}{\alpha^2 - \rho_1\rho_2}\right) + \frac{L_2}{\alpha} + \frac{\rho_2}{\alpha}\log V(i) \quad \forall \theta \in (0,1], i \in S. \tag{3.2.6}$$

Proof. For the proof, see Proposition 2.2.3 (3), Chapter 2, p. 17-18. □

Let us introduce some frequently used notations. We define the spaces $A_{as}([0,1]\times S)$, $(B^\infty_W([0,1] \times S), \|\cdot\|^\infty_W)$, and $L^\infty_W([0,1] \times S)$, for any real-valued function $W \ge 1$ on S, as defined in Chapter 2, p. 19.

To study the admissible (feedback) strategy, we need an extension of Dynkin's formula. To that end, we imposed the following condition.

(*A*3) The function V^2 satisfies the Foster-Lyapunov condition with respect to constants $0 < \rho_3 < \alpha$ and $L_3 \geq 0$

$$\sum_{j \in S} q(j|i,a,b)V^2(j) \leq \rho_3 V^2(i) + L_3 \;\; \forall \; (i,a,b) \in K,$$

where V is introduced in Assumption $(A1)$.
To ensure the existence of pair of optimal strategies, in addition to Assumptions $(A1)$-$(A3)$, we also need the following continuity and compactness conditions.

(*A*4) The following conditions hold:

(i) for each $i \in S$, the sets $A(i)$ and $B(i)$ are compact;

(ii) for any fixed $i, j \in S$, $q(j|i,a,b)$ and $c(i,a,b)$ are continuous in $(a,b) \in A(i) \times B(i)$;

(iii) for any given $i \in S$, the convergence of $\sum_{j \in S} V(j)q(j|i,a,b)$ holds uniformly in $(a,b) \in A(i) \times B(i)$, where V is introduced in Assumption $(A1)$.

By closely mimicking the arguments in [56], one can easily get the following result, which will be used in subsequent sections; we omit the details.

Lemma 3.2.7. *Under Assumptions $(A1)$-$(A4)$, the functions*

$$c(i,\mu,\nu) \;\; \textit{and} \;\; \sum_{j \in S} q(j|i,\mu,\nu)u(\cdot,j)$$

are continuous at (μ,ν) on $\mathcal{P}(A(i)) \times \mathcal{P}(B(i))$ for each fixed $u \in L^{\infty}_{V}([0,1] \times S)$ and $i \in S$.

3.3 Stochastic representation of a solution to the HJI equation

In this section, we analyze the cost evolution criterion. We carry out our analysis via equation (3.2.5). For a.e. $\theta \in [0,1]$, we consider the following HJI equations:

$$\begin{cases} \alpha\theta \frac{d\varphi_\alpha}{d\theta}(\theta,i) & = \sup\limits_{\mu \in \mathcal{P}(A(i))} \inf\limits_{\nu \in \mathcal{P}(B(i))} \Big[\sum\limits_{j \in S} q(j|i,\mu,\nu)\varphi_\alpha(\theta,j) + \theta c(i,\mu,\nu)\varphi_\alpha(\theta,i) \Big], \\ & = \inf\limits_{\nu \in \mathcal{P}(B(i))} \sup\limits_{\mu \in \mathcal{P}(A(i))} \Big[\sum\limits_{j \in S} q(j|i,\mu,\nu)\varphi_\alpha(\theta,j) + \theta c(i,\mu,\nu)\varphi_\alpha(\theta,i) \Big], \\ & \quad i \in S, \;\; \text{a.e. } \theta \in [0,1], \\ & 1 \leq \varphi_\alpha(\theta,i) \leq \frac{\alpha^2}{\alpha^2 - \rho_1\rho_2\theta} e^{\theta L_2/\alpha} V^{\frac{\rho_2\theta}{\alpha}}(i) \;\; \text{for} \;\; (\theta,i) \in [0,1] \times S, \end{cases} \tag{3.3.1}$$

where the upper bound of $\varphi_\alpha(\theta,i)$ is inspired by Proposition 3.2.6. Note that $\varphi_\alpha(\theta,i)$ is differentiable almost everywhere with respect to the first variable, and second equality follows from Fan's minimax theorem, see [33, Theorem 3].

In section 3.4, we address the existence of a solution to the HJI equation. In the next theorem, we show that if the HJI equation has a solution then it is equal to the value function corresponding to (3.2.5).

Theorem 3.3.1. *Under Assumptions (A1)-(A4), suppose that the HJI equation (3.3.1) has a solution* $\varphi_\alpha \in L^\infty_V([0,1]\times S)$ *satisfying the bounds. Then, for all* $(\theta,i)\in[0,1]\times S$, *we have*

$$\begin{aligned}&\varphi_\alpha(\theta,i)\\ &= \sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^1,\pi^2}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c(\xi_t,\pi^1(t),\pi^2(t))dt\right)\right]\\ &= \inf_{\pi^2\in\Pi^2_{Ad}}\sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^2}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c(\xi_t,\pi^1(t),\pi^2(t))dt\right)\right]\end{aligned} \tag{3.3.2}$$

which means $\varphi_\alpha(\theta,i)=\hat{J}^*_\alpha(\theta,i)$ *for all* $(\theta,i)\in[0,1]\times S$.

Proof. Let

$$\bar{\mu}:(0,1)\times S\to\mathcal{P}(A) \text{ and } \bar{\nu}:(0,1)\times S\to\mathcal{P}(B),$$

be measurable functions such that

$$\begin{aligned}&\inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\left[\theta c(i,\mu,\nu)\varphi_\alpha(\theta,i)+\sum_{j\in S}q(j|i,\mu,\nu)\varphi_\alpha(\theta,j)\right]\\ &=\sup_{\mu\in\mathcal{P}(A(i))}\left[\theta c(i,\mu,\bar{\nu}(\theta,i))\varphi_\alpha(\theta,i)+\sum_{j\in S}q(j|i,\mu,\bar{\nu}(\theta,i))\varphi_\alpha(\theta,j)\right]\end{aligned} \tag{3.3.3}$$

and

$$\begin{aligned}&\sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[\theta c(i,\mu,\nu)\varphi_\alpha(\theta,i)+\sum_{j\in S}q(j|i,\mu,\nu)\varphi_\alpha(\theta,j)\right]\\ &=\inf_{\nu\in\mathcal{P}(B(i))}\left[\theta c(i,\bar{\mu}(\theta,i),\nu)\varphi_\alpha(\theta,i)+\sum_{j\in S}q(j|i,\bar{\mu}(\theta,i),\nu)\varphi_\alpha(\theta,j)\right].\end{aligned} \tag{3.3.4}$$

The existence of such measurable maps is ensured by the measurable selection theorem in [17, Proposition 7.33], or [89, Theorem 2.2]. Let

$$\pi^{*1}:\mathbb{R}_+\times S\to\mathcal{P}(A) \text{ and } \pi^{*2}:\mathbb{R}_+\times S\to\mathcal{P}(B)$$

be defined by

$$\pi^{*1}(i,t)=\bar{\mu}(\theta e^{-\alpha t},i) \text{ and } \pi^{*2}(i,t)=\bar{\nu}(\theta e^{-\alpha t},i).$$

Given any admissible feedback pair of strategies $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$ and $\theta \in [0,1]$, let $\{\xi_t, t \geq 0\}$ be the corresponding process, and define $\theta(t) := \theta e^{-\alpha t}$ and $h : [0,\infty) \times S \times \Omega \to [0,\infty)$ by

$$h(t,i,\omega) := \exp\left(\int_0^t \theta(s) c(\xi_s, \pi^1(s), \pi^2(s)) ds\right) \varphi_\alpha(\theta(t), i).$$

In order to use the extension of Dynkin's formula, (for details see [62, Theorem 3.3] or [53, Theorem 3.1]) to the function h, it suffices to verify that

$$E_i^{\pi^1,\pi^2}\left[\exp\left(\int_0^t 2e^{-\alpha s} c(\xi_s, \pi^1(s), \pi^2(s)) ds\right)\right] \leq \frac{\alpha e^{2L_2/\alpha}}{\alpha - \rho_3}\left(V^2(i) + \frac{L_3}{\rho_3}\right) \tag{3.3.5}$$

$\forall i \in S$ and $t \in (0,\infty)$. Let $\gamma(ds) := \alpha e^{-\alpha s} ds$. Then under Assumptions $(A1)$-$(A3)$, we have

$$\begin{aligned}
& E_i^{\pi^1,\pi^2}\left[\exp\left(\int_0^t 2e^{-\alpha s} c(\xi_s, \pi^1(s), \pi^2(s)) ds\right)\right] \\
& \leq E_i^{\pi^1,\pi^2}\left[\exp\left(\int_0^\infty \frac{2}{\alpha} c(\xi_s, \pi^1(s), \pi^2(s)) \gamma(ds)\right)\right] \\
& \leq E_i^{\pi^1,\pi^2}\left[\int_0^\infty \exp\left(\frac{2}{\alpha} c(\xi_s, \pi^1(s), \pi^2(s))\right) \gamma(ds)\right] \\
& \text{(by Jensen's inequality)} \\
& \leq E_i^{\pi^1,\pi^2}\left[\int_0^\infty \exp\left(\frac{2}{\alpha}(\rho_2 \log V(\xi_s) + L_2)\right) \gamma(ds)\right] \\
& \text{(by Assumption (A2))} \\
& = e^{2L_2/\alpha}\left[\int_0^\infty E_i^{\pi^1,\pi^2}\left(V(\xi_s)^{\frac{2\rho_2}{\alpha}}\right) \gamma(ds)\right] \\
& \leq \alpha e^{2L_2/\alpha}\left(V^2(i) + \frac{L_3}{\rho_3}\right)\left[\int_0^\infty e^{\rho_3 s - \alpha s} ds\right] \\
& = \frac{\alpha e^{2L_2/\alpha}}{\alpha - \rho_3}\left(V^2(i) + \frac{L_3}{\rho_3}\right),
\end{aligned}$$

which implies that (3.3.5) holds. Thus, using the extension of Dynkin's formula to the function h, we have

$$\begin{aligned}
& E_i^{\pi^1,\pi^2}[h(t, \xi_v, \omega)] - \varphi_\alpha(\theta, i) \\
& = E_i^{\pi^1,\pi^2}\bigg\{\int_0^t \exp\left(\int_0^s \theta(v) c(\xi_v, \pi^1(v), \pi^2(v)) dv\right) \\
& \quad \times \bigg[-\alpha\theta(s) \frac{d\varphi_\alpha}{d\theta}(\theta(s), \xi_s) + \sum_{j \in S} q(j|\xi_s, \pi^1(s), \pi^2(s)) \varphi_\alpha(\theta(s), j) \\
& \qquad + \theta(s) c(\xi_s, \pi^1(s), \pi^2(s)) \varphi_\alpha(\theta(s), \xi_s)\bigg] ds\bigg\}.
\end{aligned} \tag{3.3.6}$$

Now from (3.3.1), (3.3.4) and (3.3.6), we have

$$\varphi_\alpha(\theta,i) \le E_i^{\pi^{*1},\pi^2}\left[\exp\left(\int_0^t \theta(s)c(\xi_s,\pi^{*1}(\xi_{s-},s),\pi^2(s))ds\right)\varphi_\alpha(\theta(t),\xi_t)\right]. \tag{3.3.7}$$

Given any $p>1$, let $q>1$ such that $\frac{1}{p}+\frac{1}{q}=1$, by Hölder's inequality we have

$$\begin{aligned}
&\varphi_\alpha(\theta,i)\\
&\le E_i^{\pi^{*1},\pi^2}\left[\exp\left(\int_0^t \theta(s)c(\xi_s,\pi^{*1}(\xi_{s-},s),\pi^2(s))ds\right)\varphi_\alpha(\theta(t),\xi_t)\right]\\
&\le \left\{E_i^{\pi^{*1},\pi^2}\left[\exp\left(p\int_0^t \theta(s)c(\xi_s,\pi^{*1}(\xi_{s-},s),\pi^2(s))ds\right)\right]\right\}^{1/p}\times\left\{E_i^{\pi^{*1},\pi^2}[\varphi_\alpha^q(\theta(t),\xi_t)]\right\}^{1/q}\\
&=: I_1(p,t)\cdot I_2(q,t).
\end{aligned} \tag{3.3.8}$$

For $I_2(q,t):=\{E_i^{\pi^{*1},\pi^2}[\varphi_\alpha^q(\theta(t),\xi_t)]\}^{1/q}$, by the upper bound of φ_α in (3.3.1), we have

$$\varphi_\alpha(\theta(t),\xi_t)=\varphi_\alpha(\theta e^{-\alpha t},\xi_t)\le\frac{\alpha^2}{\alpha^2-\theta e^{-\alpha t}\rho_1\rho_2}\exp\left(\frac{\theta e^{-\alpha t}L_2}{\alpha}\right)V^{\frac{\theta e^{-\alpha t}\rho_2}{\alpha}}(\xi_t).$$

If $t>\alpha^{-1}\log(\theta q\rho_2/\alpha)$ then $\theta e^{-\alpha t}q\rho_2/\alpha<1$. Hence, by Jensen's inequality and Proposition 3.2.5(b), we obtain

$$\begin{aligned}
I_2(q,t)\le&\left\{E_i^{\pi^{*1},\pi^2}\left[\left(\frac{\alpha^2}{\alpha^2-\theta e^{-\alpha t}\rho_1\rho_2}\right)^q\exp\left(\frac{q\theta e^{-\alpha t}L_2}{\alpha}\right)V^{\frac{q\theta e^{-\alpha t}\rho_2}{\alpha}}(\xi_t)\right]\right\}^{1/q}\\
&\le\frac{\alpha^2}{\alpha^2-\theta e^{-\alpha t}\rho_1\rho_2}\exp\left(\frac{\theta e^{-\alpha t}L_2}{\alpha}\right)[E^{\pi^{*1},\pi^2}(V(x_t))]^{\frac{\theta e^{-\alpha t}\rho_2}{\alpha}}\\
&\le\frac{\alpha^2}{\alpha^2-\theta e^{-\alpha t}\rho_1\rho_2}\exp\left(\frac{\theta e^{-\alpha t}}{\alpha}(L_2+\rho_1\rho_2 t)\right)V(i)^{\frac{\theta e^{-\alpha t}\rho_2}{\alpha}}=:I_3(t).
\end{aligned} \tag{3.3.9}$$

By letting $t\to\infty$ we obtain

$$\begin{aligned}
&I_1(p,t)\to\left\{E_i^{\pi^{*1},\pi^2}\left[\exp\left(p\int_0^\infty \theta(s)c(\xi_s,\pi^{*1}(\xi_{s-},s),\pi^2(s))ds\right)\right]\right\}^{1/p}\\
&\text{and}\\
&I_3(t)\to 1.
\end{aligned} \tag{3.3.10}$$

Combining (3.3.8), (3.3.9) and (3.3.10), we obtain

$$\varphi_\alpha(\theta,i)\le\left\{E_i^{\pi^{*1},\pi^2}\left[\exp\left(p\theta\int_0^\infty e^{-\alpha t}c(\xi_t,\pi^{*1}(\xi_{t-},t),\pi^2(t))dt\right)\right]\right\}^{1/p},$$

for $p>1$. Then, passing to the limit as $p\downarrow 1$, we obtain

$$\varphi_\alpha(\theta,i)\le E_i^{\pi^{*1},\pi^2}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c(\xi_t,\pi^{*1}(\xi_{t-},t),\pi^2(t))dt\right)\right].$$

Since $\pi^2 \in \Pi^2_{Ad}$ is arbitrary strategy, we have

$$\begin{aligned}\varphi_\alpha(\theta, i) &\leq \inf_{\pi^2 \in \Pi^2_{Ad}} E_i^{\pi^{*1}, \pi^2}\left[\exp\left(\theta \int_0^\infty e^{-\alpha t} c(\xi_t, \pi^{*1}(\xi_{t-}, t), \pi^2(t)) dt\right)\right] \\ &\leq \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} E_i^{\pi^1, \pi^2}\left[\exp\left(\theta \int_0^\infty e^{-\alpha t} c(\xi_t, \pi^1(t), \pi^2(t)) dt\right)\right]. \qquad (3.3.11)\end{aligned}$$

Using analogous arguments we can show that

$$E_i^{\pi^1, \pi^{*2}}\left[\exp\left(\int_0^t \theta(s) c(\xi_s, \pi^1(s), \pi^{*2}(\xi_{s-}, s)) ds\right) \varphi_\alpha(\theta(t), \xi_t)\right] \leq \varphi_\alpha(\theta, i). \qquad (3.3.12)$$

Now, using the lower bound of φ_α in (3.3.1) and Fatou's lemma, we obtain

$$\begin{aligned}&\liminf_{t\to\infty} E_i^{\pi^1, \pi^{*2}}\left[\exp\left(\int_0^t \theta(s) c(\xi_s, \pi^1(s), \pi^{*2}(\xi_{s-}, s)) ds\right) \varphi_\alpha(\theta(t), \xi_t)\right] \\ &\geq \liminf_{t\to\infty} E_i^{\pi^1, \pi^{*2}}\left[\exp\left(\int_0^t \theta(s) c(\xi_s, \pi^1(s), \pi^{*2}(\xi_{s-}, s)) ds\right)\right] \\ &\geq E_i^{\pi^1, \pi^{*2}}\left[\liminf_{t\to\infty} \exp\left(\int_0^t \theta(s) c(\xi_s, \pi^1(s), \pi^{*2}(\xi_{s-}, s)) ds\right)\right] \\ &= \hat{J}_\alpha(\theta, i, \pi^1, \pi^{*2}). \qquad (3.3.13)\end{aligned}$$

From (3.3.12) and (3.3.13), we have

$$\hat{J}_\alpha(\theta, i, \pi^1, \pi^{*2}) \leq \varphi_\alpha(\theta, i).$$

Since $\pi^1 \in \Pi^1_{Ad}$ is arbitrary, we have $\sup_{\pi^1 \in \Pi^1_{Ad}} \hat{J}_\alpha(\theta, i, \pi^1, \pi^{*2}) \leq \varphi_\alpha(\theta, i)$. Thus

$$\inf_{\pi^2 \in \Pi^2_{Ad}} \sup_{\pi^1 \in \Pi^1_{Ad}} \hat{J}_\alpha(\theta, i, \pi^1, \pi^2) \leq \sup_{\pi^1 \in \Pi^1_{Ad}} \hat{J}_\alpha(\theta, i, \pi^1, \pi^{*2}) \leq \varphi_\alpha(\theta, i). \qquad (3.3.14)$$

From (3.3.11) and (3.3.14), we have (3.3.2).

□

3.4 The existence of solution to the HJI equation

In this Section, we prove that the equation (3.3.1) is the HJI equation for the α discounted cost given by (3.2.5) and the equation (3.3.1) has a solution in $L^\infty_V([0,1] \times S)$. We now proceed to make a rigorous analysis of the above. First, we truncate our transition and cost rates which plays a crucial role to derive the HJI equations and find the value of the game. Fix any k, $n \geq 1$, $0 < \delta < 1$, and let $S_k = \{0, 1, \cdots, k\}$ be the subset of S. For each $i, j \in S$ and $a \in A(i)$, $b \in B(i)$, define

$$q^{(k)}(j|i, a, b) := q(j|i, a, b) I_{S_k}(i), \quad \text{for } i \neq j, \quad q_i^{(k)}(a, b) := \sum_{j \neq i} q^{(k)}(j|i, a, b);$$

$$c_n(i, a, b) := c(i, a, b) \wedge n.$$

Lemma 3.4.1. *Suppose Assumptions (A1), (A2) and (A4) are satisfied. Then, there exists a unique function* $\varphi_\alpha^{(n,\delta,k)}$ *(depending on* n*,* δ*,* k*) in* $L_V^\infty([0,1]\times S)$ *for which the followings are true :*

1. $\varphi_\alpha^{(n,\delta,k)}$ *is a bounded solution to the following ordinary differential equations (ODEs) for all* $i \in S$ *and a.e.* $\theta \in (\delta, 1]$ *:*

$$\begin{cases} \alpha\theta \frac{d\varphi_\alpha^{(n,\delta,k)}}{d\theta}(\theta,i) & = \sup\limits_{\mu\in\mathcal{P}(A(i))}\inf\limits_{\nu\in\mathcal{P}(B(i))}\left[\theta c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta,k)}(\theta,i) + \sum\limits_{j\in S} q^{(k)}(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta,k)}(\theta,j)\right] \\ & = \inf\limits_{\nu\in\mathcal{P}(B(i))}\sup\limits_{\mu\in\mathcal{P}(A(i))}\left[\theta c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta,k)}(\theta,i) + \sum\limits_{j\in S} q^{(k)}(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta,k)}(\theta,j)\right] \\ \varphi_\alpha^{(n,\delta,k)}(\delta,i) & = e^{n\delta/\alpha}. \end{cases} \tag{3.4.1}$$

2. $\varphi_\alpha^{(n,\delta,k)}(\theta,i)$ *has a stochastic representation as follows: for each* $i \in S$ *and a.e.* $\theta \in (\delta, 1]$,

$$\begin{aligned} \varphi_\alpha^{(n,\delta,k)}(\theta,i) &= \sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^1,\pi^2}\left[e^{n\delta/\alpha}\exp\left(\theta\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(k)},\pi^1(t),\pi^2(t))dt\right)\right] \\ &= \inf_{\pi^2\in\Pi^2_{Ad}}\sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^2}\left[e^{n\delta/\alpha}\exp\left(\theta\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(k)},\pi^1(t),\pi^2(t))dt\right)\right], \end{aligned} \tag{3.4.2}$$

where $T_\delta(\theta) := \alpha^{-1}\log(\theta/\delta)$ *and* $\xi_t^{(k)}$ *is the process corresponding to the* $q^{(k)}(j|i,a,b)$*.*

Proof. (1) Due to the definition of $q_i^{(k)}(a,b)$, we have the Lyapunov function $V(\cdot)\equiv 1$ such that $\sum_{j\in S} q^{(k)}(j|i,a,b)V(j) \le \rho_1 V(i)$ for $i\in S$, and $\overline{q}^{(k)} := \sup_{(i,a,b)\in K} q_i^{(k)}(a,b) < \infty$. Define a nonlinear operator T on $B_1^\infty([0,1]\times S)$ as follows:

$$Tu(\theta,i) = e^{\delta n/\alpha} + \frac{1}{\alpha}\int_\delta^\theta \sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[\frac{1}{s}\sum_{j\in S} q^{(k)}(j|i,\mu,\nu)u(s,j) + c_n(i,\mu,\nu)u(s,i)\right]ds,$$

where $u \in B_1^\infty([0,1]\times S)$ and $(\theta,i)\in[\delta,1]\times S$. Then by Fan's minimax theorem, see [33, Theorem 3] we have

$$Tu(\theta,i) = e^{\delta n/\alpha} + \frac{1}{\alpha}\int_\delta^\theta \inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\left[\frac{1}{s}\sum_{j\in S} q^{(k)}(j|i,\mu,\nu)u(s,j) + c_n(i,\mu,\nu)u(s,i)\right]ds.$$

Then by analogous arguments as in [Lemma 2.4.1, Chapter 2, p. 24-25] and by Banach's fixed point theorem, there exists a unique bounded function $\varphi_\alpha^{(n,\delta,k)} \in B_1^\infty([0,1]\times S)$ (depending on (n,δ,k)) such that $T\varphi_\alpha^{(n,\delta,k)} = \varphi_\alpha^{(n,\delta,k)}$; that is,

$\varphi_\alpha^{(n,\delta,k)}(\theta,i)$

$$= e^{\delta n/\alpha} + \frac{1}{\alpha}\int_\delta^\theta \sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[\frac{1}{s}\sum_{j\in S} q^{(k)}(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta,k)}(s,j) + c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta,k)}(s,i)\right]ds$$
$$= e^{\delta n/\alpha} + \frac{1}{\alpha}\int_\delta^\theta \inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\left[\frac{1}{s}\sum_{j\in S} q^{(k)}(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta,k)}(s,j) + c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta,k)}(s,i)\right]ds.$$

Also note from the above equation that $\varphi_\alpha^{(n,\delta,k)}(\delta,i) = e^{\delta n/\alpha}$ and $\varphi_\alpha^{(n,\delta,k)} \in L_1^\infty([0,1]\times S)$.
(2) Let

$$\bar{\mu} : (0,1)\times S \to \mathcal{P}(A) \text{ and } \bar{\nu} : (0,1)\times S \to \mathcal{P}(B),$$

be measurable functions such that

$$\inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\left[\theta c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta,k)}(\theta,i) + \sum_{j\in S} q^{(k)}(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta,k)}(\theta,j)\right]$$
$$= \sup_{\mu\in\mathcal{P}(A(i))}\left[\theta c_n(i,\mu,\bar{\nu}(\theta,i))\varphi_\alpha^{(n,\delta,k)}(\theta,i) + \sum_{j\in S} q^{(k)}(j|i,\mu,\bar{\nu}(\theta,i))\varphi_\alpha^{(n,\delta,k)}(\theta,j)\right] \quad (3.4.3)$$

and

$$\sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[\theta c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta,k)}(\theta,i) + \sum_{j\in S} q^{(k)}(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta,k)}(\theta,j)\right]$$
$$= \inf_{\nu\in\mathcal{P}(B(i))}\left[\theta c_n(i,\bar{\mu}(\theta,i),\nu)\varphi_\alpha^{(n,\delta,k)}(\theta,i) + \sum_{j\in S} q^{(k)}(j|i,\bar{\mu}(\theta,i),\nu)\varphi_\alpha^{(n,\delta,k)}(\theta,j)\right]. \quad (3.4.4)$$

The existence of such measurable maps is ensured by the measurable selection theorem in [17, Proposition 7.33], or [89, Theorem 2.2]. Let

$$\pi^{*1} : \mathbb{R}_+ \times S \to \mathcal{P}(A) \text{ and } \pi^{*2} : \mathbb{R}_+ \times S \to \mathcal{P}(B)$$

be defined by

$$\pi^{*1}(i,t) = \bar{\mu}(\theta e^{-\alpha t}, i) \text{ and } \pi^{*2}(i,t) = \bar{\nu}(\theta e^{-\alpha t}, i).$$

Let $\theta(t) := \theta e^{-\alpha t}$ for $t \in [0,\infty)$. Since c_n and $\varphi_\alpha^{(n,\delta,k)}$ are bounded, by Dynkin's formula we get

$$E_i^{\pi^1,\pi^2}\left[\exp\left(\int_0^{T_\delta(\theta)} \theta(s)c_n(\xi_s^{(k)},\pi^1(s),\pi^2(s))ds\right)\varphi_\alpha^{(n,\delta,k)}\left(\theta(T_\delta),\xi_{T_\delta}^{(k)}\right)\right] - \varphi_\alpha^{(n,\delta,k)}(\theta,i)$$
$$= E_i^{\pi^1,\pi^2}\left\{\int_0^{T_\delta(\theta)}\left[-\alpha\theta(s)\frac{d\varphi_\alpha^{(n,\delta,k)}}{d\theta}(\theta(s),\xi_s^{(k)}) + \sum_{j\in S} q^{(k)}(j|\xi_s^{(k)},\pi^1(s),\pi^2(s))\varphi_\alpha^{(n,\delta,k)}(\theta(s),j)\right.\right.$$
$$\left.\left. + \theta(s)c_n(\xi_s^{(k)},\pi^1(s),\pi^2(s))\varphi_\alpha^{(n,\delta,k)}(\theta(s),\xi_s^{(k)})\right]\exp\left(\int_0^s \theta(v)c_n(\xi_v^{(k)},\pi^1(v),\pi^2(v))dv\right)ds\right\}. \quad (3.4.5)$$

By using the fact that $\varphi_\alpha^{(n,\delta,k)}$ is a solution to (3.4.1) and equation (3.4.4), we obtain

$$E_i^{\pi^{*1},\pi^2}\left[\exp\left(\int_0^{T_\delta(\theta)} \theta(s)c_n(\xi_s^{(k)},\pi^{*1}(\xi_{s-}^{(k)},s),\pi^2(s))ds\right)\varphi_\alpha^{(n,\delta,k)}\left(\theta(T_\delta),\xi_{T_\delta}^{(k)}\right)\right]$$

$$\geq \varphi_\alpha^{(n,\delta,k)}(\theta,i).$$

Since $\pi^2 \in \Pi^2_{Ad}$ is an arbitrary strategy for player 2 and $\varphi_\alpha^{(n,\delta,k)}(\theta(T_\delta(\theta)),\xi^{(k)}_{T_\delta}) = e^{n\delta/\alpha}$, we have

$$\begin{aligned}
&\varphi_\alpha^{(n,\delta,k)}(\theta,i)\\
&\leq \inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^{*1},\pi^2}\left[e^{n\delta/\alpha}\exp\left(\int_0^{T_\delta(\theta)} \theta(s)c_n(\xi^{(k)}_s,\pi^{*1}(\xi^{(k)}_s,s),\pi^2(s))ds\right)\right]\\
&\leq \sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^1,\pi^2}\left[e^{n\delta/\alpha}\exp\left(\int_0^{T_\delta(\theta)} \theta(s)c_n(\xi^{(k)}_s,\pi^1(s),\pi^2(s))ds\right)\right].
\end{aligned}\tag{3.4.6}$$

Using analogous arguments, we can show that

$$\begin{aligned}
&\varphi_\alpha^{(n,\delta,k)}(\theta,i)\\
&\geq \sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^{*2}}\left[e^{n\delta/\alpha}\exp\left(\int_0^{T_\delta(\theta)} \theta(s)c_n(\xi^{(k)}_s,\pi^1(s),\pi^{*2}(\xi^{(k)}_{s-},s))ds\right)\right].
\end{aligned}$$

Therefore

$$\begin{aligned}
&\varphi_\alpha^{(n,\delta,k)}(\theta,i)\\
&\geq \inf_{\pi^2\in\Pi^2_{Ad}}\sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^2}\left[e^{n\delta/\alpha}\exp\left(\int_0^{T_\delta(\theta)} \theta(s)c_n(\xi^{(k)}_s,\pi^1(s),\pi^2(s))ds\right)\right].
\end{aligned}\tag{3.4.7}$$

Therefore, from (3.4.6) and (3.4.7), we obtain (3.4.2). This completes the proof. □

Theorem 3.4.1. *Under Assumptions (A1)-(A4), the HJI equation (3.3.1) has a unique solution* $\varphi_\alpha \in L^\infty_V([0,1]\times S)$ *such that* $1 \leq \varphi_\alpha(\theta,i) \leq \frac{\alpha^2 e^{\theta L_2/\alpha}}{\alpha^2-\rho_1\rho_2\theta} V^{\frac{\rho_2\theta}{\alpha}}(i)$ *for all* $(\theta,i) \in [0,1]\times S$.

Proof. First note that, $\varphi_\alpha^{(n,\delta,k)}$ is the solution to the equation (3.4.1), which depends on three parameters n, δ, k. Hence, the proof of this theorem consists of three steps corresponding to these three parameters.

Step 1: In the first step, we construct a solution $\varphi_\alpha^{(n,\delta)}$ from $\varphi_\alpha^{(n,\delta,k)}$ by passing to the limit as $k\to\infty$, such that $\varphi_\alpha^{(n,\delta)}(\cdot,i)$ is absolutely continuous and satisfies the ODEs:

$$\begin{cases}
\alpha\theta\frac{d\varphi_\alpha^{(n,\delta)}}{d\theta}(\theta,i)\\
\quad = \sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[\sum_{j\in S} q(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(\theta,j) + \theta c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(\theta,i)\right]\\
\quad = \inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\left[\sum_{j\in S} q(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(\theta,j) + \theta c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(\theta,i)\right],\\
\quad i\in S,\ \text{a.e. } \theta\in[\delta,1],\\
\varphi_\alpha^{(n,\delta)}(\delta,i) = e^{n\delta/\alpha}\quad \forall i\in S.
\end{cases}\tag{3.4.8}$$

For proving (3.4.8), let $\varphi_\alpha^{(n,\delta,k)}$ be the solution to the equation (3.4.1), for given $k, n \geq 1, \delta > 0$. By Lemma 3.4.1, we have

$$\begin{aligned}
&\varphi_\alpha^{(n,\delta,k)}(\theta, i)\\
&= \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} E_i^{\pi^1,\pi^2}\left[e^{n\delta/\alpha} \exp\left(\theta \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t^{(k)}, \pi^1(t), \pi^2(t)) dt\right)\right]\\
&= \inf_{\pi^2 \in \Pi^2_{Ad}} \sup_{\pi^1 \in \Pi^1_{Ad}} E_i^{\pi^1,\pi^2}\left[e^{n\delta/\alpha} \exp\left(\theta \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t^{(k)}, \pi^1(t), \pi^2(t)) dt\right)\right]\\
&\leq e^{2n/\alpha}, \text{for all } i \in S, a.e.\ \theta \in [\delta, 1].
\end{aligned}$$

Fix $n \geq 1$ and $0 < \delta < 1$, and define

$$F_k(\theta, i) := \frac{1}{V(i)} \varphi_\alpha^{(n,\delta,k)}(\theta, i), \quad (\theta, i) \in [\delta, 1] \times S.$$

Next, we prove that $\{F_k, k \geq 1\}$ is equicontinuous on $[\delta, 1] \times S$. By (3.4.1), we have

$$\begin{aligned}
F_k(\theta, i) = &\frac{1}{V(i)}\Bigg[e^{n\delta/\alpha} + \frac{1}{\alpha}\int_\delta^\theta \sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \Bigg[\frac{1}{s}\sum_{j \in S} q^{(k)}(j|i, \mu, \nu)\varphi_\alpha^{(n,\delta,k)}(s, j)\\
&+ c_n(i, \mu, \nu)\varphi_\alpha^{(n,\delta,k)}(s, i)\Bigg] ds\Bigg]\\
= &\frac{1}{V(i)}\Bigg[e^{n\delta/\alpha} + \frac{1}{\alpha}\int_\delta^\theta \inf_{\nu \in \mathcal{P}(B(i))} \sup_{\mu \in \mathcal{P}(A(i))} \Bigg[\frac{1}{s}\sum_{j \in S} q^{(k)}(j|i, \mu, \nu)\varphi_\alpha^{(n,\delta,k)}(s, j)\\
&+ c_n(i, \mu, \nu)\varphi_\alpha^{(n,\delta,k)}(s, i)\Bigg] ds\Bigg]. \qquad (3.4.9)
\end{aligned}$$

For each $(s, i) \in [\delta, 1] \times S$, $(\mu, \nu) \in \mathcal{P}(A(i)) \times \mathcal{P}(B(i))$, and $k \geq 1$, let

$$h_k(s, i, \mu, \nu) := \left[\frac{1}{s}\sum_{j \in S} q^{(k)}(j|i, \mu, \nu)\varphi_\alpha^{(n,\delta,k)}(s, j) + c_n(i, \mu, \nu)\varphi_\alpha^{(n,\delta,k)}(s, i)\right],$$

$$H_k(s, i) := \frac{1}{V(i)} \sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} h_k(s, i, \mu, \nu).$$

By Assumption $(A1)$, we have

$$\begin{aligned}
|H_k(s, i)| &\leq \frac{1}{V(i)} \sup_{\mu \in \mathcal{P}(A(i))} \sup_{\nu \in \mathcal{P}(B(i))} \left[\frac{1}{s}\sum_{j \in S} |q^{(k)}(j|i, \mu, \nu)| \varphi_\alpha^{(n,\delta,k)}(s, j) + n\varphi_\alpha^{(n,\delta,k)}(s, i)\right]\\
&\leq \frac{1}{V(i)} \sup_{\mu \in \mathcal{P}(A(i))} \sup_{\nu \in \mathcal{P}(B(i))} \left[\frac{1}{s} 2q_i^{(k)}(\mu, \nu) + n\right] e^{2n/\alpha}\\
&\leq \left[\frac{2L_1}{s} + n\right] e^{2n/\alpha} \quad (\text{since } V(i) \geq 1).
\end{aligned}$$

Hence, for a given $\varepsilon_0 > 0$, take

$$\eta = \min\left\{\frac{\varepsilon_0\alpha}{2n}e^{-2n/\alpha}, \frac{\delta\varepsilon_0\alpha}{4L_1}e^{-2n/\alpha}, \frac{1}{2}\right\}$$

such that for all $(\theta, i), (\theta_0, i_0) \in [\delta, 1] \times S$ satisfying $|\theta - \theta_0| < \eta$ and $|i - i_0| < \eta$, we have

$$\begin{aligned}
|F_k(\theta, i) - F_k(\theta_0, i_0)| &= |F_k(\theta, i) - F_k(\theta_0, i)| \\
&= \left|\frac{1}{\alpha}\int_\delta^\theta H_k(s, i)ds - \frac{1}{\alpha}\int_\delta^{\theta_0} H_k(s, i)ds\right| \\
&= \frac{1}{\alpha}\left|\int_{\theta_0}^\theta H_k(s, i)ds\right| \\
&\le \frac{e^{2n/\alpha}}{\alpha}[2L_1|\log\theta - \log\theta_0| + n|\theta - \theta_0|] < \varepsilon_0 \ \text{ for all } \ k \ge 1,
\end{aligned}$$

which implies that $\{F_k, k \ge 1\}$ is equicontinuous on $[\delta, 1] \times S$. Hence, by the Ascoli-Arzelà theorem, there exist a continuous function F and a subsequence $\{F_{k_m}, m \ge 1\}$ such that

$$\lim_{m\to\infty} F_{k_m}(\theta, i) = F(\theta, i), \ |F(\theta, i)| \le e^{2n/\alpha} \text{ for all } (\theta, i) \in [\delta, 1] \times S.$$

Define $\varphi_\alpha^{(n,\delta)}(\theta, i) = F(\theta, i)V(i)$ for all $(\theta, i) \in [\delta, 1] \times S$. Then we have

$$\lim_{m\to\infty} \varphi_\alpha^{(n,\delta,k_m)}(\theta, i) = \lim_{m\to\infty} F_{k_m}(\theta, i)V(i) = \varphi_\alpha^{(n,\delta)}(\theta, i) \le e^{2n/\alpha} \ \forall \ (\theta, i) \in [\delta, 1] \times S.$$

Set

$$H(s, i) := \frac{1}{V(i)} \sup_{\mu\in\mathcal{P}(A(i))} \inf_{\nu\in\mathcal{P}(B(i))} \left[\frac{1}{s}\sum_{j\in S} q(j|i, \mu, \nu)\varphi_\alpha^{(n,\delta)}(s, j) + c_n(i, \mu, \nu)\varphi_\alpha^{(n,\delta)}(s, i)\right],$$
$$(s, i) \in [\delta, 1] \times S.$$

We can show that $\lim_{m\to\infty} H_{k_m}(s, i) = H(s, i)$. For any $i, j \in S$ and $(a, b) \in A(i) \times B(i)$, we have $\lim_{m\to\infty} q^{(k_m)}(j|i, a, b) = q(j|i, a, b)$; then take an upper bound function defined by

$$g_i(j) = \begin{cases} e^{2n/\alpha}V(i)^{-1}q(j|i, a, b) & \text{if } \ j \ne i, \\ e^{2n/\alpha}V(i)^{-1}q_i(a, b) & \text{if } \ j = i. \end{cases}$$

By Assumption $(A1)$, we have

$$|V(i)^{-1}q^{(k_m)}(j|i, a, b)\varphi_\alpha^{(n,\delta,k_m)}(s, j)| \le g_i(j) \ \text{ for all } s \in [\delta, 1], i, j \in S$$

and

$$\sum_{j\in S} g_i(j) = e^{2n/\alpha}V(i)^{-1}\left(\sum_{j\ne i} q(j|i, a, b) + q_i(a, b)\right) \le 2L_1e^{2n/\alpha} \ \text{ for all } i \in S.$$

Using dominated convergence theorem, we obtain

$$\lim_{m\to\infty}\sum_{j\in S}\frac{1}{V(i)}q^{(k_m)}(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta,k_m)}(s,j)=\frac{1}{V(i)}\sum_{j\in S}q(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,j). \tag{3.4.10}$$

Now by Assumption $(A4)$, there exist $\mu_{k_m},\mu^*\in\mathcal{P}(A(i))$ such that

$$\begin{aligned}\limsup_{m\to\infty}H_{k_m}(s,i)&=\frac{1}{V(i)}\limsup_{m\to\infty}\inf_{\nu\in\mathcal{P}(B(i))}h_{k_m}(s,i,\mu_{k_m},\nu)\\&\le\frac{1}{V(i)}\limsup_{m\to\infty}h_{k_m}(s,i,\mu_{k_m},\nu)\\&\le\frac{1}{V(i)}\left[\frac{1}{s}\sum_{j\in S}q(j|i,\mu^*,\nu)\varphi_\alpha^{(n,\delta)}(s,j)+c_n(i,\mu^*,\nu)\varphi_\alpha^{(n,\delta)}(s,i)\right].\end{aligned}$$

The last inequality follows from the extended Fatou's Lemma in [69, Lemma 8.3.7]. Since, $\nu\in\mathcal{P}(B(i))$ is arbitrary, we have

$$\limsup_{m\to\infty}H_{k_m}(s,i)\le\inf_{\nu\in\mathcal{P}(B(i))}\frac{1}{V(i)}\left[\frac{1}{s}\sum_{j\in S}q(j|i,\mu^*,\nu)\varphi_\alpha^{(n,\delta)}(s,j)+c_n(i,\mu^*,\nu)\varphi_\alpha^{(n,\delta)}(s,i)\right].$$

Thus,

$$\limsup_{m\to\infty}H_{k_m}(s,i)\le\sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\frac{1}{V(i)}\left[\frac{1}{s}\sum_{j\in S}q(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,j)+c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,i)\right]. \tag{3.4.11}$$

Using analogous arguments we can show that

$$\liminf_{m\to\infty}H_{k_m}(s,i)\ge\inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\frac{1}{V(i)}\left[\frac{1}{s}\sum_{j\in S}q(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,j)+c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,i)\right]. \tag{3.4.12}$$

Therefore from (3.4.11) and (3.4.12), we have $\lim_{m\to\infty}H_{k_m}(s,i)=H(s,i)$.

Note that $|H_k(s,i)|\le\left[\frac{2L_1}{s}+n\right]e^{2n/\alpha}$. By letting $k_m\to\infty$, in (3.4.9) and using dominated convergence theorem, we get

$$\begin{aligned}&\varphi_\alpha^{(n,\delta)}(\theta,i)\\&=e^{n\delta/\alpha}+\frac{1}{\alpha}\int_\delta^\theta\sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[\frac{1}{s}\sum_{j\in S}q(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,j)+c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,i)\right]ds\\&=e^{n\delta/\alpha}+\frac{1}{\alpha}\int_\delta^\theta\inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\left[\frac{1}{s}\sum_{j\in S}q(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,j)+c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,i)\right]ds\end{aligned} \tag{3.4.13}$$

and

$$\varphi_\alpha^{(n,\delta)}(\delta,i)=\lim_{k_m\to\infty}\varphi_\alpha^{(n,\delta,k_m)}(\delta,i)=e^{n\delta/\alpha}.$$

Moreover,

$$\begin{aligned}&\int_\delta^1 \sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[\left|\frac{1}{s}\sum_{j\in S} q(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,j)+c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,i)\right|\right]ds\\ &\le e^{2n/\alpha}[2L_1V(i)\log(\delta^{-1})+n(1-\delta)]<\infty,\end{aligned}$$

which implies that $\sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[s^{-1}\sum_{j\in S} q(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,j)+c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,i)\right]$ is Lebesgue integrable on $[\delta,1]$. By (3.4.13) and the fundamental theorem of Lebesgue integral calculus [9, Theorem 4.4.1], $\varphi_\alpha^{(n,\delta)}(\cdot,i)$ is absolutely continuous on $[\delta,1]$ and has a derivative $\frac{d\varphi_\alpha^{(n,\delta)}}{d\theta}(\cdot,i)$ a.e. satisfying the ODEs (3.4.8).

Step 2. In the second step, using the Ascoli-Arzelà theorem, we construct a solution $\varphi_\alpha^{(n)}(\cdot,i)$ from $\varphi_\alpha^{(n,\delta)}(\cdot,i)$ by passing to the limit as $\delta\to 0$, such that $\varphi_\alpha^{(n)}(\cdot,i)$ is absolutely continuous function and satisfies the flowing ODEs:

$$\begin{cases}\alpha\theta\frac{d\varphi_\alpha^{(n)}}{d\theta}(\theta,i) &= \sup\limits_{\mu\in\mathcal{P}(A(i))}\inf\limits_{\nu\in\mathcal{P}(B(i))}\left[\sum\limits_{j\in S} q(j|i,\mu,\nu)\varphi_\alpha^{(n)}(\theta,j)+\theta c_n(i,\mu,\nu)\varphi_\alpha^{(n)}(\theta,i)\right]\\ &= \inf\limits_{\nu\in\mathcal{P}(B(i))}\sup\limits_{\mu\in\mathcal{P}(A(i))}\left[\sum\limits_{j\in S} q(j|i,\mu,\nu)\varphi_\alpha^{(n)}(\theta,j)+\theta c_n(i,\mu,\nu)\varphi_\alpha^{(n)}(\theta,i)\right],\\ & \quad i\in S,\ \text{a.e. } \theta\in[0,1],\\ 1\le\varphi_\alpha^{(n)}(\theta,i) &\le \frac{\alpha^2 e^{\theta L_2/\alpha}}{\alpha^2-\rho_1\rho_2\theta}V^{\frac{\rho_2\theta}{\alpha}}(i)\quad \forall\ (\theta,i)\in[0,1]\times S.\end{cases} \tag{3.4.14}$$

Given $0<\delta<1$ and $1\le n<\infty$, as in the proof of Lemma 3.4.1, by (3.4.8) we can obtain the representation of $\varphi_\alpha^{(n,\delta)}$ as follows:

$$\begin{aligned}&\varphi_\alpha^{(n,\delta)}(\theta,i)\\ &=\inf_{\pi^2\in\Pi^2_{Ad}}\sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^2}\left[e^{n\delta/\alpha}\exp\left(\theta\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t,\pi^1(t),\pi^2(t))dt\right)\right]\\ &=\sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^1,\pi^2}\left[e^{n\delta/\alpha}\exp\left(\theta\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t,\pi^1(t),\pi^2(t))dt\right)\right]\\ &\le e^{2n/\alpha},\quad i\in S,\theta\in[\delta,1].\end{aligned} \tag{3.4.15}$$

The last inequality follows from the fact that $\sup_{(i,a,b)\in K} c_n(i,a,b)\le n$.

Next, we extend the domain of $\varphi_\alpha^{(n,\delta)}$ in (3.4.15) to $[0,1]\times S$ by

$$\overline{\varphi}_\alpha^{(n,\delta)}(\theta,i)=\begin{cases}\varphi_\alpha^{(n,\delta)}(\theta,i), & \delta\le\theta\le 1\ \forall i\in S\\ e^{n\delta/\alpha}, & 0\le\theta<\delta\ \forall i\in S,\end{cases}$$

and then show that $\{\overline{\varphi}^{(n,\delta)}, \delta \in [0,1]\}$ is equicontinuous on $[0,1] \times S$. We consider the following expression, for any given $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$, $i \in S$, $\theta, \theta_0 \in [\delta, 1]$:

$$\begin{aligned}
&\Bigg| E_i^{\pi^1,\pi^2}\left[e^{n\delta/\alpha} \exp\left(\theta \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right] \\
&\qquad - E_i^{\pi^1,\pi^2}\left[e^{n\delta/\alpha} \exp\left(\theta_0 \int_0^{T_\delta(\theta_0)} e^{-\alpha t} c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right]\Bigg| \\
&\qquad \leq K_1 + K_2,
\end{aligned}$$

where

$$\begin{aligned}
K_1 := &\Bigg| E_i^{\pi^1,\pi^2}\left[e^{n\delta/\alpha} \exp\left(\theta \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right] \\
&\qquad - E_i^{\pi^1,\pi^2}\left[e^{n\delta/\alpha} \exp\left(\theta_0 \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right]\Bigg|,
\end{aligned}$$

and

$$\begin{aligned}
K_2 := &\Bigg| E_i^{\pi^1,\pi^2}\left[e^{n\delta/\alpha} \exp\left(\theta_0 \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right] \\
&\qquad - E_i^{\pi^1,\pi^2}\left[e^{n\delta/\alpha} \exp\left(\theta_0 \int_0^{T_\delta(\theta_0)} e^{-\alpha t} c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right]\Bigg|.
\end{aligned}$$

Fix $n \geq 1$; we have

$$\int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t, \pi^1(t), \pi^2(t))dt \leq n \int_0^{T_\delta(\theta)} e^{-\alpha t} dt \leq \frac{n}{\alpha}$$

and

$$\begin{aligned}
\int_{T_\delta(\theta\wedge\theta_0)}^{T_\delta(\theta\vee\theta_0)} e^{-\alpha t} c_n(\xi_t, \pi^1(t), \pi^2(t))dt &\leq n \int_{T_\delta(\theta\wedge\theta_0)}^{T_\delta(\theta\vee\theta_0)} e^{-\alpha t} dt \\
&= \frac{n}{\alpha}[\exp(-\alpha T_\delta(\theta \wedge \theta_0)) - \exp(-\alpha T_\delta(\theta \vee \theta_0))] \leq \frac{\delta n |\theta_0 - \theta|}{\alpha \theta \theta_0},
\end{aligned}$$

where $c \wedge d := min\{c, d\}$ and $c \vee d := max\{c, d\}$. Hence, we obtain

$$\begin{aligned}
K_1 &= e^{n\delta/\alpha} E_i^{\pi^1,\pi^2}\Bigg[\exp\left((\theta \wedge \theta_0) \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right) \\
&\qquad \times \left(\exp\left(|\theta_0 - \theta| \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right) - 1\right)\Bigg] \\
&\leq e^{2n/\alpha} E_i^{\pi^1,\pi^2}\left[\exp\left(|\theta_0 - \theta| \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right) - 1\right] \\
&\leq e^{2n/\alpha}\left(\exp\left(\frac{n}{\alpha}|\theta_0 - \theta|\right) - 1\right) \\
&\leq e^{2n/\alpha}\left(e^{n/\alpha} - 1\right)|\theta_0 - \theta|.
\end{aligned}$$

Here, the last inequality follows from the fact that $e^{bx}-1 \leq (e^b-1)x$ for all $x \in [0,1]$ and $b>0$. Similarly for K_2 we have

$$\begin{aligned}
K_2 &= e^{n\delta/\alpha} E_i^{\pi^1,\pi^2}\Bigg[\exp\bigg(\theta_0 \int_0^{T_\delta(\theta\wedge\theta_0)} e^{-\alpha t} c_n(\xi_t, \pi^1(t), \pi^2(t))dt\bigg) \\
&\quad \times \bigg(\exp\bigg(\theta_0 \int_{T_\delta(\theta\wedge\theta_0)}^{T_\delta(\theta\vee\theta_0)} e^{-\alpha t} c_n(\xi_t, \pi^1(t), \pi^2(t))dt\bigg) - 1\bigg)\Bigg] \\
&\leq e^{2n/\alpha} E_i^{\pi^1,\pi^2}\bigg[\exp\bigg(\theta_0 \int_{T_\delta(\theta\wedge\theta_0)}^{T_\delta(\theta\vee\theta_0)} e^{-\alpha t} c_n(\xi_t, \pi^1(t), \pi^2(t))dt\bigg) - 1\bigg] \\
&\leq e^{2n/\alpha}\bigg(\exp\bigg(\frac{n\delta|\theta-\theta_0|}{\alpha\theta}\bigg) - 1\bigg) \\
&\leq e^{2n/\alpha}\bigg(e^{n/\alpha}-1\bigg)|\theta_0-\theta|.
\end{aligned}$$

Hence using the definition of $\overline{\varphi}_\alpha^{(n,\delta)}$, for all $(\theta,i) \in [0,1]\times S$, we have

$$|\overline{\varphi}_\alpha^{(n,\delta)}(\theta_0,i) - \overline{\varphi}_\alpha^{(n,\delta)}(\theta,i)| \leq 2e^{2n/\alpha}(e^{n/\alpha}-1)|\theta-\theta_0|, \tag{3.4.16}$$

which implies that $\{\overline{\varphi}_\alpha^{(n,\delta)}, \delta \in [0,1]\}$ is equicontinuous on $[0,1]\times S$. Hence, by Ascoli-Arzelà theorem, there exists a sequence $\delta_m \to 0$ (as $m\to\infty$) and a continuous function $\varphi_\alpha^{(n)}$ such that, for each $i \in S$, $\lim_{m\to\infty} \overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,i) = \varphi_\alpha^{(n)}(\theta,i)$ uniformly in $\theta\in[0,1]$ and $\varphi_\alpha^{(n)}(\cdot,i) \in C([0,1])$.

For any $m \geq 1$, define the following function:

$$G_m(\theta,i) := \begin{cases} \sup\limits_{\mu\in\mathcal{P}(A(i))} \inf\limits_{\nu\in\mathcal{P}(B(i))} \Big[\sum\limits_{j\in S} q(j|i,\mu,\nu)\varphi_\alpha^{(n,\delta_m)}(\theta,j) + \quad \theta c_n(i,\mu,\nu)\varphi_\alpha^{(n,\delta_m)}(\theta,i)\Big], \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \delta_m \leq \theta \leq 1 \quad \forall i \in S \\ 0, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad 0 \leq \theta < \delta_m \quad \forall i \in S. \end{cases}$$

Given any $i \in S$, by the definition of $\overline{\varphi}_\alpha^{(n,\delta_m)}$, it has a derivative $\frac{\overline{\varphi}_\alpha^{(n,\delta_m)}}{d\theta}(\cdot,i)$ a.e. Since (3.4.8), we have

$$\frac{d}{d\theta}\bigg(\alpha\theta\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,i)\bigg) = \alpha\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,i) + G_m(\theta,i), \quad \text{a.e. } \theta\in[0,1]. \tag{3.4.17}$$

Similarly, define

$$\begin{aligned} G(\theta,i) = \sup_{\mu\in\mathcal{P}(A(i))} \inf_{\nu\in\mathcal{P}(B(i))} &\Big[\sum_{j\in S} q(j|i,\mu,\nu)\varphi_\alpha^{(n)}(\theta,j) + \theta c_n(i,\mu,\nu)\varphi_\alpha^{(n)}(\theta,i)\Big] \\ &\forall(\theta,i)\in[0,1]\times S. \end{aligned}$$

Then, we have

$$\sup_{\theta\in[0,1]} |G(\theta,i) - G_m(\theta,i)| \leq \sup_{\theta\in[0,1]} \sup_{\mu\in\mathcal{P}(A(i))} \sup_{\nu\in\mathcal{P}(B(i))} \bigg|\sum_{j\in S} q(j|i,\mu,\nu)\bigg(\varphi_\alpha^{(n)}(\theta,j) - \overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,j)\bigg)\bigg|$$

$$+\sup_{\theta\in[0,1]}\sup_{\mu\in\mathcal{P}(A(i))}\sup_{\nu\in\mathcal{P}(B(i))}\left|\theta c_n(i,\mu,\nu)\left(\varphi_\alpha^{(n)}(\theta,i)-\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,i)\right)\right|$$
$$=: K_3(i)+K_4(i).$$

Note that $\sum_{j\in S}q(j|i,\mu,\nu)=0$ uniformly in $(\mu,\nu)\in\mathcal{P}(A(i))\times\mathcal{P}(B(i))$, therefore for any $\varepsilon>0$, there exists an integer $M_0:=M_0(\varepsilon,i)$ such that $\sup\limits_{(\mu,\nu)\in\mathcal{P}A(i))\times\mathcal{P}(B(i))}\sum\limits_{k\geq M_0}|q(k|i,\mu,\nu)|$ $<\varepsilon/(16e^{2n/\alpha})$. Since $\lim_{m\to\infty}\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,i)=\varphi_\alpha^{(n)}(\theta,i)$ uniformly in $\theta\in[0,1]$, for the ε given above, there exists an integer $M_1:=M_1(\varepsilon,i)$ such that for all $m>M_1$ and $0\leq j\leq M_0$, it holds that $\sup_{\theta\in[0,1]}|\varphi_\alpha^{(n)}(\theta,j)-\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,j)|<\varepsilon/(16q^*(i)M_0)$. Hence for all $m>M_0\vee M_1$, we have

$$\begin{aligned}K_3(i)&\leq\sup_{\theta\in[0,1]}\sup_{\mu\in\mathcal{P}(A(i))}\sup_{\nu\in\mathcal{P}(B(i))}\sum_{j\in S}|q(j|i,\mu,\nu)|\left|\varphi_\alpha^{(n)}(\theta,j)-\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,j)\right|\\&\leq 2e^{2n/\alpha}\sup_{\mu\in\mathcal{P}(A(i))}\sup_{\nu\in\mathcal{P}(B(i))}\sum_{j\geq M_0}|q(j|i,\mu,\nu)|\\&\qquad+\sup_{\theta\in[0,1]}\sup_{\mu\in\mathcal{P}(A(i))}\sup_{\nu\in\mathcal{P}(B(i))}\sum_{j=0}^{M_0-1}|q(j|i,\mu,\nu)|\left|\varphi_\alpha^{(n)}(\theta,j)-\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,j)\right|\\&<\frac{1}{4}\varepsilon.\end{aligned}$$

For the ε given above, there exists an integer $M_2:=M_2(\varepsilon,i)$ such that for all $m>M_2$,

$$\sup_{\theta\in[0,1]}\left|\varphi_\alpha^{(n)}(\theta,i)-\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,i)\right|<\frac{\varepsilon}{4n}\wedge\frac{\varepsilon}{2\alpha};$$

then we have

$$\begin{aligned}K_4(i)&\leq\sup_{\theta\in[0,1]}\sup_{\mu\in\mathcal{P}(A(i))}\sup_{\nu\in\mathcal{P}(B(i))}\theta|c_n(i,\mu,\nu)|\left|\varphi_\alpha^{(n)}(\theta,i)-\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,i)\right|\\&<\sup_{\theta\in[0,1]}n\left|\varphi_\alpha^{(n)}(\theta,i)-\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,i)\right|<\frac{1}{4}\varepsilon.\end{aligned}$$

Therefore, for $m>max\{M_0,M_1,M_2\}$, we have

$$\begin{aligned}&\sup_{\theta\in[0,1]}\left|\alpha\varphi_\alpha^{(n)}(\theta,i)+G(\theta,i)-\alpha\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,i)-G_m(\theta,i)\right|\\&\leq\alpha\sup_{\theta\in[0,1]}\left|\varphi_\alpha^{(n)}(\theta,i)-\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,i)\right|+\sup_{\theta\in[0,1]}|G(\theta,i)-G_m(\theta,i)|\\&<\frac{1}{2}\varepsilon+\frac{1}{4}\varepsilon+\frac{1}{4}\varepsilon=\varepsilon.\end{aligned}$$

By (3.4.17), we have $\frac{d}{d\theta}(\alpha\theta\overline{\varphi}_\alpha^{(n,\delta_m)}(\cdot,i))\to\alpha\varphi_\alpha^{(n)}(\cdot,i)+G(\cdot,i)$, a.e. $\theta\in[0,1]$ when $m\to\infty$. Therefore, for any $i\in S$ and $n\geq 1$, $\alpha\theta\varphi_\alpha^{(n)}(\cdot,i)$ is differentiable a.e., and

$$\alpha\varphi_\alpha^{(n)}(\theta,i)+\alpha\theta\frac{d\varphi_\alpha^{(n)}(\theta,i)}{d\theta}=\alpha\varphi_\alpha^{(n)}(\theta,i)+G(\theta,i),\ \text{a.e. }\theta\in[0,1].$$

Now for $\theta \in [\delta_m, 1]$, by using (3.4.15) and Proposition 3.2.6, we have

$$\begin{aligned}
&\varphi_\alpha^{(n,\delta_m)}(\theta, i)\\
&= \inf_{\pi^2\in\Pi^2_{Ad}} \sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^2}\left[e^{n\delta_m/\alpha}\exp\left(\theta\int_0^{T_{\delta_m}(\theta)} e^{-\alpha t}c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right]\\
&\le e^{n\delta_m/\alpha}\inf_{\pi^2\in\Pi^2_{Ad}} \sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^2}\left[\exp\left(\theta\int_0^{\infty} e^{-\alpha t}c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right]\\
&\le e^{n\delta_m/\alpha}\inf_{\pi^2\in\Pi^2_{Ad}} \sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^2}\left[\exp\left(\theta\int_0^{\infty} e^{-\alpha t}c(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right]\\
&\le e^{n\delta_m/\alpha}\frac{\alpha^2 e^{\theta L_2/\alpha}}{\alpha^2-\rho_1\rho_2\theta}V^{\frac{\rho_2\theta}{\alpha}}(i).
\end{aligned}$$

Note that $\varphi_\alpha^{(n,\delta_m)} \to \varphi_\alpha^{(n)}$ as $m \to \infty$. Thus, letting $m \to \infty$ in the above equation, we obtain

$$1 \le \varphi_\alpha^{(n)}(\theta, i) \le \frac{\alpha^2 e^{\theta L_2/\alpha}}{\alpha^2-\rho_1\rho_2\theta}V^{\frac{\rho_2\theta}{\alpha}}(i). \tag{3.4.18}$$

Hence, we obtain $\varphi_\alpha^{(n)} \in L_V^\infty([0,1]\times S)$, and (3.4.14) holds.
Thus by closely mimicking the arguments in Theorem 3.3.1, one can easily get the stochastic representation of the solution $\varphi_\alpha^{(n)}$, that is

$$\begin{aligned}
\varphi_\alpha^{(n)}(\theta, i) &= \inf_{\pi^2\in\Pi^2_{Ad}} \sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^2}\left[\exp\left(\theta\int_0^{\infty} e^{-\alpha t}c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right]\\
&= \sup_{\pi^1\in\Pi^1_{Ad}} \inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^1,\pi^2}\left[\exp\left(\theta\int_0^{\infty} e^{-\alpha t}c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right].
\end{aligned}\tag{3.4.19}$$

Step 3: In the final step we prove Theorem 3.4.1, by passing to the limit as $n \to \infty$. Now we will prove that $\{\varphi_\alpha^{(n)}\}_{n\ge1}$ is equicontinuous on $[0,1]\times S$. We consider the following expression, for any given $(\pi^1, \pi^2) \in \Pi^1_{Ad}\times\Pi^2_{Ad}$, $i \in S$, $\theta, \theta_0 \in [0,1]$:

$$\begin{aligned}
&\Bigg|E_i^{\pi^1,\pi^2}\left[\exp\left(\theta\int_0^{\infty} e^{-\alpha t}c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right]\\
&\qquad - E_i^{\pi^1,\pi^2}\left[\exp\left(\theta_0\int_0^{\infty} e^{-\alpha t}c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right)\right]\Bigg|\\
&\le K_1,
\end{aligned}$$

where

$$\begin{aligned}
K_1 = E_i^{\pi^1,\pi^2}\Bigg[&\exp\left((\theta\wedge\theta_0)\int_0^{\infty} e^{-\alpha t}c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right)\\
&\times\left(\exp\left(|\theta_0-\theta|\int_0^{\infty} e^{-\alpha t}c_n(\xi_t, \pi^1(t), \pi^2(t))dt\right)-1\right)\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&\le E_i^{\pi^1,\pi^2}\Bigg[\exp\bigg((\theta\wedge\theta_0)\int_0^\infty e^{-\alpha t}c_n(\xi_t,\pi^1(t),\pi^2(t))dt\bigg)\\
&\qquad\times\bigg(\exp\bigg(\int_0^\infty e^{-\alpha t}c_n(\xi_t(\omega),\pi^1(t),\pi^2(t))dt\bigg)-1\bigg)|\theta_0-\theta|\Bigg]\\
&\le E_i^{\pi^1,\pi^2}\Bigg[\exp\bigg(\int_0^\infty e^{-\alpha t}c_n(\xi_t,\pi^1(t),\pi^2(t))dt\bigg)\\
&\qquad\times\bigg(\exp\bigg(\int_0^\infty e^{-\alpha t}c_n(\xi_t,\pi^1(t),\pi^2(t))dt\bigg)|\theta_0-\theta|\bigg)\Bigg]\\
&=|\theta_0-\theta|\times E_i^{\pi^1,\pi^2}\Bigg[\exp\bigg(2\int_0^\infty e^{-\alpha t}c_n(\xi_t,\pi^1(t),\pi^2(t))dt\bigg)\Bigg]\\
&\le|\theta_0-\theta|\times\frac{\alpha e^{2L_2/\alpha}}{\alpha-\rho_3}\bigg(V^2(i)+\frac{L_3}{\rho_3}\bigg).
\end{aligned}
$$

Here, the first inequality is according to $e^{bx}-1\le(e^b-1)x$ for all $x\in[0,1]$ and $b>0$ and the last inequality follows from (3.3.5). Therefore, we have

$$
\begin{aligned}
|\varphi_\alpha^{(n)}(\theta_0,i)-\varphi_\alpha^{(n)}(\theta,i)|&\le\sup_{\pi^1\in\Pi^1_{Ad}}\sup_{\pi^2\in\Pi^2_{Ad}}|\theta_0-\theta|\times\frac{\alpha e^{2L_2/\alpha}}{\alpha-\rho_3}\bigg(V^2(i)+\frac{L_3}{\rho_3}\bigg)\\
&=|\theta_0-\theta|\times\frac{\alpha e^{2L_2/\alpha}}{\alpha-\rho_3}\bigg(V^2(i)+\frac{L_3}{\rho_3}\bigg) \qquad (3.4.20)
\end{aligned}
$$

which means that $\{\varphi_\alpha^{(n)},n\in\mathbb{N}\}$ is equicontinuous on $[0,1]\times S$. According to the Ascoli-Arzelà theorem, there exists a sequence $n_k\to\infty$ and a continuous function φ_α such that, for each $i\in S$,

$$
\lim_{k\to\infty}\varphi_\alpha^{(n_k)}(\theta,i)=\varphi_\alpha(\theta,i)\ \text{ uniformly in }\theta\in[0,1]\text{ and }\varphi_\alpha(\cdot,i)\in C([0,1]).
$$

Moreover, by taking $n_k\to\infty$ in (3.4.18), we have

$$
1\le\varphi_\alpha(\theta,i)\le\frac{\alpha^2e^{\theta L_2/\alpha}}{\alpha^2-\rho_1\rho_2\theta}V^{\frac{\rho_2\theta}{\alpha}}(i).
$$

Define

$$
L_{n_k}(\theta,i)=\sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\bigg[\sum_{j\in S}q(j|i,\mu,\nu)\varphi_\alpha^{(n_k)}(\theta,j)+\theta c_{n_k}(i,\mu,\nu)\varphi_\alpha^{(n_k)}(\theta,i)\bigg],
$$

$$
L(\theta,i)=\sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\bigg[\sum_{j\in S}q(j|i,\mu,\nu)\varphi_\alpha(\theta,j)+\theta c(i,\mu,\nu)\varphi_\alpha(\theta,i)\bigg].
$$

Then by Fan's Minimax Theorem [33, Theorem 3], we have

$$
L_{n_k}(\theta,i)=\inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\bigg[\sum_{j\in S}q(j|i,\mu,\nu)\varphi_\alpha^{(n_k)}(\theta,j)+\theta c_{n_k}(i,\mu,\nu)\varphi_\alpha^{(n_k)}(\theta,i)\bigg],
$$

and

$$L(\theta,i)=\inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\left[\sum_{j\in S}q(j|i,\mu,\nu)\varphi_\alpha(\theta,j)+\theta c(i,\mu,\nu)\varphi_\alpha(\theta,i)\right].$$

Since $\varphi_\alpha^{(n_k)}$ is the solution of (3.4.14), then $\varphi_\alpha^{(n_k)}$ has a derivative $\frac{d\varphi_\alpha^{(n_k)}}{d\theta}$ a.e. and satisfies

$$\frac{d}{d\theta}\left[\alpha\theta\varphi_\alpha^{(n_k)}(\theta,i)\right]=\alpha\varphi_\alpha^{(n_k)}(\theta,i)+L_{n_k}(\theta,i),\quad \text{a.e. } \theta\in[0,1]. \tag{3.4.21}$$

Fix any $i\in S$; $\{\alpha\varphi_\alpha^{(n_k)}(\cdot,i)+L_{n_k}(\cdot,i)\}_{n_k\geq 1}$ is a sequence of functions on $[0,1]$. Then, we should prove that $\{\alpha\varphi_\alpha^{(n_k)}(\cdot,i)+L_{n_k}(\cdot,i)\}_{n_k\geq 1}$ uniformly converges to $\alpha\varphi_\alpha(\cdot,i)+L(\cdot,i)$ on $[0,1]$. Note that $\{\varphi_\alpha^{(n_k)}(\cdot,i)\}_{n_k\geq 1}$ converges to $\varphi_\alpha(\cdot,i)$ uniformly in $\theta\in[0,1]$. Thus by closely mimicking the arguments in step 2, one can prove that for any $\varepsilon>0$, there exists a N_0 such that for $n_k>N_0$, we have

$$\begin{aligned}&\sup_{\theta\in[0,1]}\left|\alpha\varphi_\alpha(\theta,i)+L(\theta,i)-\alpha\varphi_\alpha^{(n_k)}(\theta,i)-L_{n_k}(\theta,i)\right|\\&\leq\alpha\sup_{\theta\in[0,1]}\left|\varphi_\alpha(\theta,i)-\varphi_\alpha^{(n_k)}(\theta,i)\right|+\sup_{\theta\in[0,1]}|L(\theta,i)-L_{n_k}(\theta,i)|<\varepsilon.\end{aligned}$$

That implies that $\{\alpha\varphi_\alpha^{(n_k)}(\cdot,i)+L_{n_k}(\cdot,i)\}_{n_k\geq 1}$ uniformly converges to $\alpha\varphi_\alpha(\cdot,i)+L(\cdot,i)$. Therefore, for any $i\in S$, taking limit $n_k\to\infty$ in (3.4.21), we have

$$\alpha\varphi_\alpha(\theta,i)+\alpha\theta\frac{d\varphi_\alpha(\theta,i)}{d\theta}=\alpha\varphi_\alpha(\theta,i)+L(\theta,i),\quad\text{a.e. }\theta\in[0,1],$$

which implies that

$$\begin{aligned}\alpha\theta\frac{d\varphi_\alpha}{d\theta}(\theta,i)&=\sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[\sum_{j\in S}q(j|i,\mu,\nu)\varphi_\alpha(\theta,j)+\theta c(i,\mu,\nu)\varphi_\alpha(\theta,i)\right]\\&=\inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\left[\sum_{j\in S}q(j|i,\mu,\nu)\varphi_\alpha(\theta,j)+\theta c(i,\mu,\nu)\varphi_\alpha(\theta,i)\right],\\&\quad i\in S,\ \text{a.e. }\theta\in[0,1].\end{aligned}\tag{3.4.22}$$

By (3.4.22), Assumptions ($A1$) and ($A2$), and the upper bound of φ_α given above, we say $\varphi_\alpha\in L_V^\infty([0,1]\times S)$ and is a solution to the HJI equation (3.3.1) satisfying the bounds. Finally, the uniqueness follows from the probability expression in Theorem 3.3.1. □

3.5 The existence of saddle-point equilibrium

In this section, we present the main result of this chapter, namely, that the α discounted risk-sensitive zero-sum game problem admits a value and shows the existence of a saddle-point equilibrium.

Theorem 3.5.1. *Suppose that Assumptions (A1)-(A4) are satisfied. Then, the following assertions hold.*

1. *The HJI equation (3.3.1) has a unique solution* $\varphi_\alpha \in L^\infty_V([0,1]\times S)$ *and the solution admits the following representation*

$$\begin{aligned} 1 \le \varphi_\alpha(\theta,i) &= \sup_{\pi^1\in\Pi^1_{Ad}} \inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^1,\pi^2}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c(\xi_t,\pi^1(t),\pi^2(t))dt\right)\right] \\ &= \inf_{\pi^2\in\Pi^2_{Ad}} \sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^2}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c(\xi_t,\pi^1(t),\pi^2(t))dt\right)\right] \\ &\le \frac{\alpha^2 e^{\theta L_2/\alpha}}{\alpha^2-\rho_1\rho_2\theta} V^{\frac{\rho_2\theta}{\alpha}}(i). \end{aligned}$$

2. *There exists a pair of measurable functions* (μ^*,ν^*) *such that* μ^* *is the outer maximizing selector in (3.3.1), and* ν^* *is the outer minimizing selector in (3.3.1). The optimal equation is*

$$\begin{aligned} \alpha\theta\frac{d\varphi_\alpha}{d\theta}(\theta,i) &= \sup_{\mu\in\mathcal{P}(A(i))}\left[\sum_{j\in S} q(j|i,\mu,\nu^*)\varphi_\alpha(\theta,j) + \theta c(i,\mu,\nu^*)\varphi_\alpha(\theta,i)\right] \\ &= \inf_{\nu\in\mathcal{P}(B(i))}\left[\sum_{j\in S} q(j|i,\mu^*,\nu)\varphi_\alpha(\theta,j) + \theta c(i,\mu^*,\nu)\varphi_\alpha(\theta,i)\right] \end{aligned}$$

a.e. $\theta\in[0,1]$. (3.5.1)

3. *For any pair of measurable functions* (μ^*,ν^*) *as in part (2), the corresponding pair of Markov strategies* $(\hat{\pi}^{*1},\hat{\pi}^{*2})$ *where* $\hat{\pi}^{*1}(i,t) := \mu^*(\theta e^{-\alpha t},i), \hat{\pi}^{*2}(i,t) := \nu^*(\theta e^{-\alpha t},i), (t\ge 0, i\in S)$, *is optimal. So,* $(\hat{\pi}^{*1},\hat{\pi}^{*2}) \in \Pi^1_M\times\Pi^2_M$ *is a saddle-point equilibrium point.*

Proof. Part (1) follows from Theorems 3.3.1 and 3.4.1. To prove (2), for each given $(\theta,i)\in[0,1]\times S$, by Lemma 3.2.7 we have the continuity of the function

$$F(i,\theta,\mu,\nu) := \sum_{j\in S} q(j|i,\mu,\nu)\varphi_\alpha(\theta,j) + \theta c(i,\mu,\nu)\varphi_\alpha(\theta,i)$$

in $(\mu,\nu)\in\mathcal{P}(A(i))\times\mathcal{P}(B(i))$. Thus, the measurable selection theorem [17, Proposition 7.33], or [89, Theorem 2.2] ensured the existence of a pair of measurable functions (μ^*,ν^*) satisfying (3.5.1), and so (2) follows. Moreover for any (μ^*,ν^*), satisfying (3.5.1), if we define $\hat{\pi}^{*1}(i,t) := \mu^*(\theta e^{-\alpha t},i), \hat{\pi}^{*2}(i,t) := \nu^*(\theta e^{-\alpha t},i), (t\ge 0, i\in S)$, from the proof of Theorem 3.3.1, we have
$\hat{J}_\alpha(i,\theta,\hat{\pi}^{*1},\hat{\pi}^{*2}) = \sup_{\pi^1\in\Pi^1_{Ad}} \hat{J}_\alpha(i,\theta,\pi^1,\hat{\pi}^{*2}) = \inf_{\pi^2\in\Pi^2_{Ad}} \hat{J}_\alpha(i,\theta,\hat{\pi}^{*1},\pi^2) = \varphi_\alpha(\theta,i)$, which together with (3.2.4), (3.2.5) and part (1), implies that $(\hat{\pi}^{*1},\hat{\pi}^{*2})$ is a saddle-point equilibrium. ☐

Remark 3.5.1. *Note that* $(\hat{\pi}^{*1}, \hat{\pi}^{*2}) \in \Pi^1_M \times \Pi^2_M$ *as defined in theorem above is a saddle-point equilibrium for the cost criterion (3.2.4) and we have* $U(i) = L(i) = J^*_\alpha(\theta, i) = \frac{1}{\theta}\log\varphi_\alpha(\theta, i)$. *Thus the value of the game exists.*

3.6 Example

In this section, we verify the above assumptions with an example, in which the transition rate is unbounded and cost rate is nonnegative and unbounded.

Example 3.6.1. *Consider a controlled birth-death system in which the state variable denotes the total population at each time* $t \geq 0$. *There are 'natural' birth and death rates, say* λ *and* $\tilde{\mu}$, *respectively, and death parameters* h_1 *controlled by player 1 and birth parameters* h_2 *controlled by player 2. When the state of the system is* $i \in S := \{0, 1, \cdots\}$, *player 1 takes an action* a *from a given set* $A(i)$, *which may increase* $(h_1(i,a) \geq 0)$ *or decrease* $(h_1(i,a) \leq 0)$ *the death rate. These actions produce a payoff denoted by* $c_1(i,a)$ *per unit time. Similarly, if the state is* $i \in S$, *player 2 takes an action* b *from a set* $B(i)$ *to decrease* $(h_2(i,b) \leq 0)$ *or to increase* $(h_2(i,b) \geq 0)$ *birth rate, and these actions result in a payoff denoted by* $c_2(i,b)$ *per unit time. In addition, assuming that player 1 'owns' the system, he/she gets a reward* $r(i) := p \cdot i$ *for each unit of time during which the system remains in the state* $i \in S$, *where* $p > 0$ *is a fixed reward fee per customer. We next formulate this model as a continuous-time Markov game. The corresponding transition rate* $q(j|i,a,b)$ *and reward rate* $c(i,a,b)$ *for player 1 are given as follows: for* $(0,a,b) \in K$ *(*K *as in the game model (3.2.1)).*

$$q(1|0,a,b) = -q(0|0,a,b) := h_2(0,b)$$

and, for $(i,a,b) \in K$ *with* $i \geq 1$,

$$q(j|i,a,b) = \begin{cases} \tilde{\mu}i + h_1(i,a), & \text{if } j = i-1 \\ -\lambda i - \tilde{\mu}i - h_1(i,a) - h_2(i,b), & \text{if } j = i \\ \lambda i + h_2(i,b), & \text{if } j = i+1 \\ 0, & \text{otherwise.} \end{cases}$$

$$c(i,a,b) := p \cdot i - c_1(i,a) + c_2(i,b) \quad \text{for} \quad (i,a,b) \in K. \tag{3.6.1}$$

We assume that player 1 is maximizing player and player 2 is minimizing player. Player 1 gets an immediately reward rate $c(i,a,b)$ *and he/she tries to maximize his/her total payoff through the infinite horizon continuous-time discounted cost criterion and at the same time, player 2 incurs an immediately cost rate* $c(i,a,b)$ *and he/she tries to minimize his/her total payoff through the same cost criterion.*

We want to find conditions under which there exists a pair of optimal strategies. To do so, we use the following assumptions.

(I) Let $\tilde{\mu} > \lambda$, $\tilde{\mu}i + h_1(i,a) > 0$, *and* $\lambda i + h_2(i,b) > 0$ *for all* $(i,a,b) \in K$ *with* $i \geq 1$*; and assume that* $h_1(0,a) = 0$ *and* $h_2(0,b) > 0$ *for all* $(a,b) \in A(i) \times B(i)$.

(II) For each $i \in S$, $A(i)$ *and* $B(i)$ *are compact subsets of a Polish space.*

(III) The functions $h_1(i,a)$, $h_2(i,b)$, $c_1(i,a)$, *and* $c_2(i,b)$ *are bounded in the supremum norm and continuous with their respective variables for each fixed* $i \in S$. *Hence there exist nonnegative constants* L_2, B_1, *and* B_2 *such that*

$$\sup_{(i,a,b)\in K} (|c_1(i,a)| + |c_2(i,b)|) \leq L_2,$$

$$\sup_{(i,a)\in S\times A} |h_1(i,a)| \leq B_1, \textit{ and } \sup_{(i,b)\in S\times B} |h_1(i,b)| \leq B_2.$$

(IV) Set $C = B_1 + B_2$ *and assume that* $C < \frac{\alpha\lambda}{\tilde{\mu}-\lambda}$, *and* $c_2(i,b) - c_1(i,a) \geq 0$ *for all* $i \in S$, $a \in A(i)$, *and* $b \in B(i)$. *Moreover, there exists a constant* ρ_2 *such that* $0 < \rho_2 < \min\left\{\alpha, \frac{\alpha^2\sqrt{\lambda}}{(\sqrt{\tilde{\mu}}-\sqrt{\lambda})C}\right\}$ *and* $p \leq \frac{\rho_2}{2}\log(\tilde{\mu}/\lambda)$.

Proposition 3.6.2. *Under conditions (I)-(IV), the above controlled birth-death system satisfies the Assumptions (A1)-(A4). Hence by Theorem 3.5.1, there exists a pair of optimal strategies.*

Proof. Take a Lyapunov function as

$$V(i) := \left(\frac{\tilde{\mu}}{\lambda}\right)^{\frac{i}{2}} \quad \text{for} \quad i \in S. \tag{3.6.2}$$

By condition (I), we have $V(i) \geq 1$ for all $i \in S$. Define $\rho_1 = (\frac{\sqrt{\tilde{\mu}}}{\sqrt{\lambda}} - 1)C$. So, by above conditions, we have $\rho_1 > 0$. Now for each $i \in S$, and $(a,b) \in A(i) \times B(i)$, we have

$$\begin{aligned}
\sum_{j\in S} q(j|i,a,b)V(j) &= q(i-1|i,a,b)V(i-1) + V(i)q(i|i,a,b) + V(i+1)q(i+1|i,a,b) \\
&= (\lambda i + h_2(i,b))(V(i+1) - V(i)) + (\tilde{\mu}i + h_1(i,a))(V(i-1) - V(i)) \\
&= (\frac{\sqrt{\tilde{\mu}}}{\sqrt{\lambda}} - 1)\big((\lambda - \sqrt{\lambda\tilde{\mu}})i + h_2(i,b) - \frac{\sqrt{\lambda}}{\sqrt{\tilde{\mu}}}h_1(i,a)\big)V(i) \\
&\leq (\frac{\sqrt{\tilde{\mu}}}{\sqrt{\lambda}} - 1)\Big(|h_1(i,a)| + |h_2(i,b)|\Big)V(i) \\
&\leq (\frac{\sqrt{\tilde{\mu}}}{\sqrt{\lambda}} - 1)(B_1 + B_2)V(i) \\
&= C(\frac{\sqrt{\tilde{\mu}}}{\sqrt{\lambda}} - 1)V(i)
\end{aligned}$$

$$= \rho_1 V(i). \tag{3.6.3}$$

Now let $L_1 := \max\left\{2(B_1+B_2), 4\frac{(\lambda+\tilde{\mu})}{\log(\frac{\tilde{\mu}}{\lambda})}\right\} > 0$. So, we have

$$\begin{aligned}
\sup_{(a,b)\in A(i)\times B(i)} q_i(a,b) &\leq \sup_{(a,b)\in A(i)\times B(i)} \left\{\lambda i + \tilde{\mu} i + |h_1(i,a)| + |h_2(i,b)|\right\}\\
&\leq i(\lambda+\tilde{\mu}) + (B_1+B_2)\\
&\leq L_1 V(i) \ \forall\ i \in S.
\end{aligned}$$

Hence Assumption $(A1)$ is verified. Now we verify Assumption $(A2)$

$$\begin{aligned}
\sup_{(a,b)\in A(i)\times B(i)} c(i,a,b) &\leq p\cdot i + \sup_{a\in A(i)} |c_1(i,a)| + \sup_{b\in B(i)} |c_2(i,b)|\\
&\leq i\frac{\rho_2}{2}\log(\frac{\tilde{\mu}}{\lambda}) + L_2\\
&= \rho_2 \log V(i) + L_2 \ \forall\ i \in S.
\end{aligned}$$

By condition (IV), it is obvious that $0 < \rho_2 < \min\{\alpha, \rho_1^{-1}\alpha^2\}$. So, Assumption $(A2)$ holds.

Now we want to verify Assumption $(A3)$.

$$\begin{aligned}
\sum_{j\in S} q(j|i,a,b)V^2(j) &= V^2(i-1)q(i-1|i,a,b) + V^2(i)q(i|i,a,b) + V^2(i+1)q(i+1|i,a,b)\\
&= (\lambda i + h_2(i,b))(V^2(i+1) - V^2(i)) + (\tilde{\mu} i + h_1(i,a))(V^2(i-1) - V^2(i))\\
&= (\frac{\tilde{\mu}}{\lambda} - 1)\big(h_2(i,b) - \frac{\lambda}{\tilde{\mu}}h_1(i,a)\big)V^2(i)\\
&\leq (\frac{\tilde{\mu}}{\lambda} - 1)\Big(|h_1(i,a)| + |h_2(i,b)|\Big)V^2(i)\\
&\leq (\frac{\tilde{\mu}}{\lambda} - 1)(B_1+B_2)V^2(i)\\
&= (\frac{\tilde{\mu}}{\lambda} - 1)CV^2(i).
\end{aligned}$$

Now take $\rho_3 = (\frac{\tilde{\mu}}{\lambda} - 1)C$ and $L_3 = 0$. By condition (IV), we have $0 < \rho_3 < \alpha$. So, Assumption $(A3)$ is verified. By condition (III), we see that $q(j|i,a,b)$ and $c(i,a,b)$ are continuous in $(a,b) \in A(i)\times B(i)$. Now since $h_1(i,a)$ and $h_2(i,b)$ are continuous with their respective variables for each fixed $i \in S$, by (3.6.3), we can confirm that the convergence of $\sum_{j\in S} q(j|i,a,b)V(j)$ is uniform in $(a,b) \in A(i)\times B(i)$. Hence Assumption $(A4)$ is also verified. Hence by Theorem 3.5.1, we say that there exists an optimal pair of Markov strategies for this controlled Birth-Death process. □

CHAPTER 4

Continuous-time zero-sum games for Markov decision processes with discounted risk-sensitive cost criterion on a general state space

4.1 Introduction

In this chapter, we study two-person infinite-horizon discounted-cost risk-sensitive zero-sum stochastic games for CTMDPs with unbounded transition and reward rates. The main objective of this work is (1) to prove the existence of the value of the game and saddle-point equilibrium (2) to give a characterization of the optimal strategies in terms of the corresponding Hamilton-Jacobi-Isaacs (HJI) equation (3) to provide an important example.

Risk-sensitive cost criterion plays an important role in many applications including mathematical finance see, Whittle [114] and references therein. The analysis of risk-sensitive control is technically more involved because of the exponential nature of the cost. Though this criterion has been studied extensively in the literature for MDPs see [21], [30], [32], [47], [53], [62], [81], [82], [94], [108], the corresponding game version has a few literatures. In a risk-sensitive zero-sum game, the goal of player 1 is to maximize his/her rewards, whereas that of player 2 is to minimize his/her costs with respect to some performance criterion $J_\alpha(\cdot,\cdot,\cdot,\cdot)$, which is defined by (4.2.4), below. Such a model is relevant in worst-case scenarios, e.g., in financial applications when a risk-averse investor is trying to maximize his long-term portfolio gain against the market which, by default, is the minimizer in this case. Infinite horizon zero-sum risk-sensitive stochastic games on a countable state space have been studied in [12]. The paper [16] extended the results of [12] for general state space. Both in the papers [12] and [16], the authors considered the discrete-time case and bounded cost function. Let us also mention the recent work of [95], which studies the infinite-horizon zero-sum risk-sensitive game for CTMDPs with general state space and bounded cost and transition rates. But this boundedness requirement restricts our domain of application since in many real-life situations, we see that the reward/cost and transition rates are unbounded for example in queueing control and population processes. In [48] and [117], the authors studied finite horizon zero-sum risk-sensitive continuous-time stochastic games with unbounded cost and transition rates. In the paper, [51], the authors studied infinite-horizon continuous-time risk-sensitive zero-sum games for MDPs on countable state space with unbounded cost and transition rates. In this chapter, we have generalized the results of Chapter 3 (that is a zero-sum stochastic game problem on countable space) to a zero-sum stochastic game problem on general state space. In other words, here we extend the work [49] from a single control problem to a game problem.

To the best of our knowledge, this chapter is the first work that deals with infinite-horizon continuous-time zero-sum risk-sensitive stochastic games with unbounded cost and transition rates with feedback strategy strategies and the state space is general Borel space. We know that there are lots of literature on countable state space in the game model. But a few for general state space. There are lots of real-life situations for example when we work in chemical reactions, the Gaussian model,s etc., where the state spaces are uncountable. Hence, our model can play a significant role in those situations. Here, we solve the zero-sum stochastic game problem by characterizing the value function as a solution to the HJI equation. In particular, we prove that the solution of the HJI equation is in fact unique and coincides with the value function of the zero-sum game problem under consideration. Also, we prove the existence of a saddle-point equilibrium in the class of Markov strategies via HJI equation.

The rest of this chapter is organized as follows. Section 4.2 deals with the description of the problem. In Section 4.3, we prove the stochastic representation of the solution of HJI equation (4.3.1). The existence of the unique solution to the HJI equation is proven in Section 4.4. In Section 4.5, we prove the value and saddle-point equilibrium in the class of Markov strategies. In Section 4.6, we illustrate our theory and assumptions with an example. The content of this chapter is based on the published article [50].

4.2 The game model

In this section, we first introduce the continuous-time zero-sum stochastic game model

$$\hat{\mathbb{M}} := \{S, A, B, (A(x) \subset A, B(x) \subset B, x \in S), q(\cdot|x,a,b), c(x,a,b)\}, \tag{4.2.1}$$

which consists of the following elements:

- S is the state space. We assume that S is a Borel space with the Borel σ-algebra $\mathcal{B}(S)$.

- The action spaces for players 1 and 2, are respectively, denoted by A and B. Here the spaces A and B are assumed to be Borel spaces with the Borel σ-algebras $\mathcal{B}(A)$ and $\mathcal{B}(B)$, respectively.

- For each $x \in S$, $A(x) \in \mathcal{B}(A)$ and $B(x) \in \mathcal{B}(B)$ denote the sets of admissible actions for players 1 and 2 in state x, respectively. We assume the following.
  (A0)(a) The admissible action spaces $A(x)$ and $B(x)$ are compact for each x.

- The set of all admissible state action pairs is denoted by $K := \{(x, a, b)|x \in S, a \in A(x), b \in B(x)\}$, is a Borel subset of $S \times A \times B$. Given any $(x, a, b) \in K$, the transition rate $q(\cdot|x, a, b)$ is a signed kernel on S such that $q(D|x, a, b) \geq 0$ for all $x \notin D \in \mathcal{B}(S)$. Moreover, we assume that $q(\cdot|x, a, b)$ satisfies the following conservative and stable conditions:
  (A0)(b) for any $x \in S$,

$$\begin{aligned} & q(S|x, a, b) = 0 \quad \text{and} \\ & q^*(x) := \sup_{(a,b)\in A(x)\times B(x)} q_x(a, b) < \infty, \end{aligned} \tag{4.2.2}$$

  where $q_x(a, b) := -q(\{x\}|x, a, b) \geq 0$.

- At last, the nonnegative measurable function $c : K \to \mathbb{R}_+ \cup \{0\}$ denotes the reward rate function for player 1 (or the cost rate function for player 2).

The game evolves as follows. The players observe continuously the current state of the system. When at time $t \geq 0$, the system is in state $x \in S$, the players independently choose actions $a_t \in A(x)$ and $b_t \in B(x)$, respectively, according to some strategies. As a consequence of this, the following happens:

- player 1 receives an immediate reward at rate $c(x, a_t, b_t)$ and player 2 incurs a cost at rate $c(x, a_t, b_t)$; and

- the system stays in state x for a random time, with rate of leaving x given by $q_x(a_t, b_t)$, and then jumps to a set D, $x \notin D$ with the probability determined by $\dfrac{q(D|x, a_t, b_t)}{q_x(a_t, b_t)}$ (see Proposition B.8 in [59, p. 205] for details).

When the state of the system transits to the new state $y \neq x$, the above procedure is repeated.

Now, we describe the construction of CTMDPs under possible feedback strategy strategies. To construct the underlying CTMDPs (as in [63], [75], [99]) we introduce some notations: let $S_\Delta := S \cup \{\Delta\}$ (with some isolated point $\Delta \notin S$), $\Omega_0 := (S \times (0, \infty))^\infty$, $\Omega := \Omega_0 \cup \{(x'_0, \eta_1, x'_1, \cdots, \eta_k, x'_k, \infty, \Delta, \infty, \Delta, \cdots)|$

$x_0^{'} \in S, x_l^{'} \in S, \eta_l \in (0,\infty)$, for each $1 \le l \le k, k \ge 1\}$ (endowed with the product topology), and let $\mathscr{F}$ be the Borel σ-algebra on Ω. Then we obtain the measurable space $(\Omega, \mathscr{F})$. Now, for each $\omega := (x_0^{'}, \eta_1, x_1^{'}, \cdots, \eta_k, x_k^{'}, \cdots) \in \Omega$, define $T_0(\omega) := 0$, $T_n(\omega) := T_{n-1}(\omega) + \eta_n$, $T_\infty(\omega) := \lim_{n\to\infty} T_n(\omega)$. Using the definition of $\{T_k\}$, we define the state process $\{\xi_t\}_{t\ge 0}$ as

$$\xi_t(\omega) := \sum_{k\ge 0} I_{\{T_k \le t < T_{k+1}\}} x_k^{'} + I_{\{t \ge T_\infty\}}\Delta, \text{ for } t \ge 0, \tag{4.2.3}$$

where, I_E denotes the indicator function of a set E, and for all $z \in S_\Delta$, we use the convention, $0 + z =: z$ and $0z =: 0$. It is obvious to note that $\xi_t(\omega)$ is right-continuous on $[0,\infty)$. From (4.2.3), we conclude that $T_k(\omega)$ $(k \ge 1)$ denotes the k-th jump instance of $\{\xi_t, t \ge 0\}$, $\xi_{k-1}(\omega) = x_{k-1}^{'}$ is the state on $[T_{k-1}(\omega), T_k(\omega))$, $\eta_k = T_k(\omega) - T_{k-1}(\omega)$ is the sojourn time at state $x_{k-1}^{'}$, and the sample path $\{\xi_t(\omega), t \ge 0\}$ has at most denumerable states $x_k^{'} (k = 0, 1, \cdots)$. The process after T_∞ is assumed to be absorbed in the state Δ. Let $A(\Delta) := \{a_\Delta\}$, $B(\Delta) := \{b_\Delta\}$, where a_Δ, b_Δ are isolated points. Then, let $q(\cdot|\Delta, a_\Delta, b_\Delta) :\equiv 0$, $c(\Delta, a, b) :\equiv 0$, for all $(a, b) \in A_\Delta \times B_\Delta$, where $A_\Delta := A \cup A(\Delta)$, $B_\Delta := B \cup B(\Delta)$. Also, suppose $\mathscr{F}_t := \sigma(\{T_k \le s, \xi_{T_k} \in S\} : 0 \le s \le t, k \ge 0)$ for all $t \ge 0$, and $\mathscr{P} := \sigma(\{C \times \{0\}, D \in \mathscr{F}_0\} \cup \{C \times (s,\infty), D \in \mathscr{F}_{s-}\})$ which denotes the σ-algebra of predictable sets on $\Omega \times [0,\infty)$ related to $\{\mathscr{F}_t\}_{t\ge0}$, where $\mathscr{F}_{s-} =: \bigvee_{0\le t<s} \mathscr{F}_t$. Now to define the risk-sensitive cost criterion, we introduce below the definition of strategy.

Definition 4.2.1. *An admissible feedback strategy for player 1, denoted by* $\pi^1 = \{\pi^1(t)\}_{t\ge0}$, *is a transition probability* $\pi^1(da|\omega, t)$ *from* $(\Omega \times [0,\infty), \mathscr{P})$ *onto* $(A_\Delta, \mathcal{B}(A_\Delta))$, *such that* $\pi^1(A(\xi_{t-}(\omega))|\omega, t) = 1$. *Using appropriate projections of the transition kernel* π^1, *an admissible feedback strategy for player 1, determines and is, in turn, determined by a sequence* $\{\pi_k^1, k \ge 0\}$ *of the stochastic kernel on* A *such that*

$$\begin{aligned}\pi^1(t)(\omega) &= \pi^1(da|\omega, t)\\ &= I_{\{t=0\}}(t)\pi_0^1(da|x_0^{'}, 0) + \sum_{k\ge0} I_{\{T_k < t \le T_{k+1}\}} \pi_k^1(da|x_0^{'}, \theta_1, x_1^{'}, \ldots, \theta_k, x_k^{'}, t - T_k)\\ &\quad + I_{\{t \ge T_\infty\}} \delta_{a_\Delta}(da),\end{aligned}$$

where $\pi_0^1(da|x_0^{'}, 0)$ *is a stochastic kernel on* A *given* S *such that* $\pi_0^1(A(x_0^{'})|x_0^{'}, 0) = 1$, $\pi_k^1 (k \ge 1)$ *are stochastic kernels on* A *given* $(S \times (0,\infty))^{k+1}$ *such that* $\pi_k^1(A(x_k^{'})|x_0^{'}, \theta_1, x_1^{'}, \cdots, \theta_k, x_k^{'}, t - T_k) = 1$, *and* $\delta_{a_\Delta}(da)$ *denotes the Dirac measure at the point* a_Δ.

For more details see [64, Definition 2.1, Remark 2.2], [99], [119]. The set of all admissible feedback strategies for player 1 is denoted by Π_{Ad}^1. A strategy $\pi^1 \in \Pi_{Ad}^1$ for player 1, is called a Markov if $\pi^1(t)(\omega) = \pi^1(\xi_{t-}(\omega), t)$ i.e., $\pi^1(da|\omega, t) = \pi^1(da|\xi_{t-}(\omega), t)$ for every $\omega \in \Omega$ and $t \geq 0$, where $\xi_{t-}(\omega) := \lim_{s \uparrow t} \xi_s(\omega)$. For notational simplicity, we would not write ω anywhere throughout the rest of this chapter. We denote by Π^1_M the family of all Markov strategies for player 1. The sets of admissible feedback strategies Π^2_{Ad} and Markov strategies Π^2_M for player 2 are defined analogously.

Remark 4.2.2. *In the definition of strategies we do not include the entire history of the game, i.e., past and present states, past sojourn times, and past actions taken by the players. Instead, we are dealing with feedback strategies that include only past and present states and past sojourn times for each player's admissible strategies, for details, see [Chapter 3, Remark 3.2.2].*

For any compact metric space Y, let $\mathcal{P}(Y)$ denote the space of probability measures on Y with Prohorov topology. Then under Assumption $(A0)(a)$, Proposition 7.22 in Bertsekas and Shreve [17, p. 130] implies that $\mathcal{P}(Y)$ is compact and metrizable. Hence, for each $x \in S$, $\mathcal{P}(A(x))$ and $\mathcal{P}(B(x))$ are compact and metrizable. For any pair of strategies $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$ and any initial state $x \in S$, by Theorem 4.27 in [76], there exists a unique probability measure denoted by P^{π^1,π^2}_x on $(\Omega, \mathscr{F})$. Let E^{π^1,π^2}_x be the expectation operator with respect to P^{π^1,π^2}_x. For each $x \in S$, $\mu \in \mathcal{P}(A(x))$ and $\nu \in \mathcal{P}(B(x))$, the associated transition and cost rates are defined, respectively, as follows:

$$q(dy|x, \mu, \nu) := \int_{B(x)} \int_{A(x)} q(dy|x, a, b)\mu(da)\nu(db),$$
$$c(x, \mu, \nu) := \int_{B(x)} \int_{A(x)} c(x, a, b)\mu(da)\nu(db).$$

(by abuse of notation we use the same notation q and c). We first fix any discounted factor $\alpha > 0$ and risk-sensitive parameter $\theta \in (0, 1]$. For any $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$ and $x \in S$, the risk-sensitive discounted criterion is defined as

$$J_\alpha(\theta, x, \pi^1, \pi^2) := \frac{1}{\theta} \log\left\{ E^{\pi^1,\pi^2}_x \left[\exp\left(\theta \int_0^\infty e^{-\alpha t} c(\xi_t, \pi^1(t), \pi^2(t)) dt \right) \right] \right\}, \tag{4.2.4}$$

provided that the integral is well defined, where $\{\xi_t\}_{t \geq 0}$ is the MDP corresponding to $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$.

Now we define the functions on S as

$$U(x) := \inf_{\pi^2 \in \Pi^2_{Ad}} \sup_{\pi^1 \in \Pi^1_{Ad}} J_\alpha(\theta, x, \pi^1, \pi^2) \quad \text{and} \quad L(x) := \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} J_\alpha(\theta, x, \pi^1, \pi^2).$$

The functions $U(x)$ and $L(x)$ are called the upper value and lower value, respectively, of the game. It is clear that $L(x) \leq U(x)$ for all $x \in S$.

Definition 4.2.3. *If the two functions $U(x)$ and $L(x)$ are equal, i.e., $U(x) = L(x)$ for all $x \in S$, then the common function is called the value of the game and denoted by $J^*_\alpha(\theta, x)$.*

Definition 4.2.4. *Consider that the game has a value J^*_α. Then a strategy π^{*1} in Π^1_{Ad} is said to be optimal for player 1 if*

$$\inf_{\pi^2 \in \Pi^2_{Ad}} J_\alpha(\theta, x, \pi^{*1}, \pi^2) = J^*_\alpha(\theta, x) \ \forall x \in S.$$

Similarly, $\pi^{*2} \in \Pi^2_{Ad}$ is optimal for player 2 if

$$\sup_{\pi^1 \in \Pi^1_{Ad}} J_\alpha(\theta, x, \pi^1, \pi^{*2}) = J^*_\alpha(\theta, x) \ \forall x \in S.$$

If $\pi^{*k} \in \Pi^k_{Ad}$ is optimal for player k (k=1,2), then (π^{*1}, π^{*2}) is called a pair of optimal strategies and also called a saddle-point equilibrium.

Now, since logarithm is an increasing function, we study $\hat{J}_\alpha(\theta, x, \pi^1, \pi^2)$ instead of studying $J_\alpha(\theta, x, \pi^1, \pi^2)$ on $[0,1] \times S \times \Pi^1_{Ad} \times \Pi^2_{Ad}$. We define $\hat{J}_\alpha(\theta, x, \pi^1, \pi^2)$ as

$$\hat{J}_\alpha(\theta, x, \pi^1, \pi^2) := E^{\pi^1,\pi^2}_x \left[\exp\left(\theta \int_0^\infty e^{-\alpha t} c(\xi_t, \pi^1(t), \pi^2(t)) dt \right) \right]. \tag{4.2.5}$$

It is easy to observe that (π^{*1}, π^{*2}) is optimal if and only if for all $x \in S$,

$$\inf_{\pi^2 \in \Pi^2_{Ad}} \hat{J}_\alpha(\theta, x, \pi^{*1}, \pi^2) = \sup_{\pi^1 \in \Pi^1_{Ad}} \hat{J}_\alpha(\theta, x, \pi^1, \pi^{*2}) =: \hat{J}^*_\alpha(\theta, x).$$

We now introduce some frequently used notations.

Notations: Let us introduce some frequently used notations. We define the spaces $C^\infty_c(a, b)$, $A_{as}([0,1] \times S)$, $(B^\infty_W([0,1] \times S), \| \cdot \|^\infty_W)$, and $L^\infty_W([0,1] \times S)$, for any real-valued function $W \geq 1$ on S, as defined in Chapter 2, p. 19.

We need the following conditions for non-explosion of the state process $\{\xi_t, t \geq 0\}$, and finiteness of $J_\alpha(\theta, x, \pi^1, \pi^2)$, which are widely used in CTMDPs; see, for instance, [53], [62], [63], [64].

(*A*1) There exists a real-valued Borel measurable function $V \geq 1$ on S such that

(i) for some constant $\hat{\rho}_0 > 0$, the Lyapunov stability condition, $\int_S V(y) q(dy|x, a, b) \leq \hat{\rho}_0 V(x)$ holds $\forall (x, a, b) \in K$;

(ii) for some constant $M_0 > 0$, the non-explosion condition, $\sup_{(a,b) \in A(x) \times B(x)} q_x(a, b) \leq M_0 V(x)$ holds $\forall x \in S$;

(iii) for some constants $\hat{L}_0 \geq 0$ and $\hat{\rho}_1$, the logarithm growth condition, $\sup_{(a,b)\in A(x)\times B(x)} c(x,a,b) \leq \hat{\rho}_1 \log V(x) + \hat{L}_0$ holds $\forall x \in S$, where $0 < \hat{\rho}_1 < \min\{\alpha, \hat{\rho}_0^{-1}\alpha^2\}$.

Proposition 4.2.5. *Grant Assumptions (A0)(a), (A0)(b) and (A1). Then for any* $(\theta, x) \in [0,1] \times S$ *and pair of strategies* $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$, *the following results are true:*

(a) $P^{\pi^1,\pi^2}_x(\xi_t \in S) = 1$, $P^{\pi^1,\pi^2}_x(\xi_0 = x) = 1$, *and* $P^{\pi^1,\pi^2}_x(T_\infty = \infty) = 1$, *for all* $t \geq 0$*;*

(b) $E^{\pi^1,\pi^2}_x[V(\xi_t)] \leq e^{\hat{\rho}_0 t} V(x)$ *for all* $t \geq 0$*;*

(c)

$$J^*_\alpha(\theta, x) \leq \log\left(\frac{\alpha^2}{\alpha^2 - \hat{\rho}_0\hat{\rho}_1}\right) + \frac{\hat{L}_0}{\alpha} + \frac{\hat{\rho}_1}{\alpha}\log V(x) \quad \forall \theta \in (0,1], x \in S. \tag{4.2.6}$$

Proof. For parts (a) and (b), see, [61] and ([63, Theorem 3.1]).
Proof of part (c): See Proposition 2.2.3 (3), Chapter 2, p. 17-18.

□

(A2) The Borel measurable function $V^2 \geq 1$ on S, satisfies the following condition $\int_S q(dy|x,a,b)V^2(y) \leq \hat{\rho}_2 V^2(x) + b_1 \ \ \forall\ (x,a,b) \in K$, for some constants $0 < \hat{\rho}_2 < \alpha$, and $b_1 \geq 0$, where V is as in Assumption $(A1)$.

An extension of Itô-Dynkin's formula has been developed in [53] under the above assumption, which can be used to study the admissible feedback strategy case. Also, we need the following continuity conditions to ensure the existence of pair of optimal strategies.

(A3) For V as introduced in Assumption $(A1)$, the followings hold:

(i) the functions $q(\cdot|x,a,b)$ and $c(x,a,b)$ are continuous in $(a,b) \in A(x) \times B(x)$, for any fixed $x \in S$;

(ii) the function $\displaystyle\int_S V(y)q(dy|x,a,b)$ is continuous in $(a,b) \in A(x) \times B(x)$, for any fixed $x \in S$.

4.3 Probabilistic representation of a solution to the HJI equation

We continue our analysis through the criterion (4.2.5). Using dynamic programming heuristics, we consider the following HJI equations satisfying the bounds:

$$\begin{cases} \alpha\theta\frac{\partial\varphi_\alpha}{\partial\theta}(\theta,x) & = \sup\limits_{\mu\in\mathcal{P}(A(x))}\inf\limits_{\nu\in\mathcal{P}(B(x))}\left[\int_S q(dy|x,\mu,\nu)\varphi_\alpha(\theta,y)+\theta c(x,\mu,\nu)\varphi_\alpha(\theta,x)\right] \\ & = \inf\limits_{\nu\in\mathcal{P}(B(x))}\sup\limits_{\mu\in\mathcal{P}(A(x))}\left[\int_S q(dy|x,\mu,\nu)\varphi_\alpha(\theta,y)+\theta c(x,\mu,\nu)\varphi_\alpha(\theta,x)\right], \\ & \quad x\in S,\ \text{a.e. } \theta\in[0,1], \\ & 1\le\varphi_\alpha(\theta,x)\le\frac{\alpha^2}{\alpha^2-\hat\rho_0\hat\rho_1\theta}e^{\theta\hat L_0/\alpha}V^{\frac{\hat\rho_1\theta}{\alpha}}(x)\ \text{ for }\ (\theta,x)\in[0,1]\times S, \end{cases} \tag{4.3.1}$$

where the upper bound of $\varphi_\alpha(\theta,x)$ is inspired by Proposition 4.2.5.

The next theorem is our verification theorem, here we show that if the HJI equation has a solution then it is equal to the value function corresponding to the criterion (4.2.5).

Theorem 4.3.1. *Under Assumptions (A0)(a), (A0)(b), (A1)-(A3), suppose that the HJI equation (4.3.1) has a solution* $\varphi_\alpha\in L^\infty_V([0,1]\times S)$ *satisfying the bounds. Then, for all* $(\theta,x)\in[0,1]\times S$, *we have* $\varphi_\alpha(\theta,x)=\hat J^*_\alpha(\theta,x)$ *that is,*

$$\begin{aligned} \varphi_\alpha(\theta,x) &= \sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}} E^{\pi^1,\pi^2}_x\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c(\xi_t,\pi^1(t),\pi^2(t))dt\right)\right] \\ &= \inf_{\pi^2\in\Pi^2_{Ad}}\sup_{\pi^1\in\Pi^1_{Ad}} E^{\pi^1,\pi^2}_x\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c(\xi_t,\pi^1(t),\pi^2(t))dt\right)\right]. \end{aligned} \tag{4.3.2}$$

Proof. For the proof, see Theorem 3.3.1, Chapter 3, p. 48-51. □

4.4 The existence of solution to the HJI equation

In this Section, we prove that the equation (4.3.1) is the HJI equation for the α discounted cost (4.2.5) and the equation (4.3.1) has a solution in $L^\infty_V([0,1]\times S)$. We give detail proof of the above. To do so, we first truncate our transition and cost rates which plays a crucial role to find the solution of the HJI equations.

Theorem 4.4.1. *Grant Assumptions (A0)(a), (A0)(b), (A1)-(A3). Then the HJI equation (4.3.1) has a unique solution* $\varphi_\alpha\in L^\infty_V([0,1]\times S)$ *satisfying the bounds* $1\le\varphi_\alpha(\theta,x)\le\frac{\alpha^2e^{\theta\hat L_0/\alpha}}{\alpha^2-\hat\rho_0\hat\rho_1\theta}V^{\frac{\hat\rho_1\theta}{\alpha}}(x)$ *for all* $(\theta,x)\in[0,1]\times S$.

Proof. Let $0<\delta<1$ and $n\ge1$ be fixed constants. For each $n\ge1$, $x\in S$, $a\in A(x)$, $b\in B(x)$, let $S_n:=\{x\in S|V_0(x)\le n\}$, $A_n(x):=A(x)$, $B_n(x):=B(x)$, $K_n:=$

$\{(x,a,b)|x\in S_n, a\in A_n(x), b\in B_n(x)\}$. Moreover for each $x\in S$, $a\in A_n(x)$, $b\in B_n(x)$, define truncated cost and transition rates are

$$c_n(x,a,b) := \begin{cases} c(x,a,b)\wedge \ \min\{n, \hat{\rho}_1 \ln V(x)+\hat{L}_0\} \ \text{ if } x\in S_n, \\ 0 \ \text{ if } \ x\notin S_n \end{cases} \tag{4.4.1}$$

and

$$q^{(n)}(dy|x,a,b) := \begin{cases} q(dy|x,a,b) \ \text{ if } x\in S_n, \\ 0 \ \text{ if } \ x\notin S_n. \end{cases} \tag{4.4.2}$$

By the definition of S_n and Assumption (*A*1)(ii), we say that $q_x^{(n)}(a,b) := \int_{S/\{x\}} q^{(n)}(dy|x,a,b)$ is bounded and so we can choose a suitable Lyapunov function V to satisfy $\int_S q^{(n)}(dy|x,a,b)V(y)\le \hat{\rho}_0 V(x)$, and $\tilde{q}^{(n)} := \sup_{(x,a,b)\in K} q_x^{(n)}(a,b) < \infty$. Now, we prove our theorem in three steps. In the first step, we show that there exists a unique function $\varphi_\alpha^{(n,\delta)}$ (depending on n, δ) in $L_V^\infty([0,1]\times S)$ that satisfies a truncated HJI equation (4.4.3), given below. In the second step we construct a solution $\varphi_\alpha^{(n)}(\cdot,x)$ by passing the limit as $\delta\to 0$. In the last step, we take limit $n\to\infty$ and show that the HJI equation (4.3.1) has a solution.

Step 1: In this first step, we show the following parts. Let $\xi_t^{(n)}$ be the process corresponding to the $q^{(n)}(\cdot|x,a,b)$ and $T_\delta(\theta) := \alpha^{-1}\log(\theta/\delta)$.

1. $\varphi_\alpha^{(n,\delta)} \in B_1([0,1]\times S)$ is a bounded solution to the following first order differential equations (DEs) satisfying the boundary condition as:

$$\begin{cases} \alpha\theta\frac{\partial\varphi_\alpha^{(n,\delta)}}{\partial\theta}(\theta,x) & = \sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\left[\int_S q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n,\delta)}(\theta,y)+\theta c_n(x,\mu,\nu)\varphi_\alpha^{(n,\delta)}(\theta,x)\right] \\ & = \inf_{\nu\in\mathcal{P}(B(x))}\sup_{\mu\in\mathcal{P}(A(x))}\left[\int_S q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n,\delta)}(\theta,y)+\theta c_n(x,\mu,\nu)\varphi_\alpha^{(n,\delta)}(\theta,x)\right] \\ \varphi_\alpha^{(n,\delta)}(\delta,x) & = e^{n\delta/\alpha}, \end{cases} \tag{4.4.3}$$

   where $\theta\in(\delta,1]$ a.e., and $x\in S$.

2. $\varphi_\alpha^{(n,\delta)}(\theta,x)$ has the following probabilistic representation:

$$\begin{aligned} \varphi_\alpha^{(n,\delta)}(\theta,x) &= \sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}} E_x^{\pi^1,\pi^2}\left[e^{n\delta/\alpha}\exp\left(\theta\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\right)\right] \\ &= \inf_{\pi^2\in\Pi^2_{Ad}}\sup_{\pi^1\in\Pi^1_{Ad}} E_x^{\pi^1,\pi^2}\left[e^{n\delta/\alpha}\exp\left(\theta\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\right)\right] \\ &= \sup_{\pi^1\in\Pi^1_{Ad}} E_x^{\pi^1,\pi^{*2}_\delta}\left[e^{n\delta/\alpha}\exp\left(\theta\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^{*2}_\delta(\xi_{t-},t))dt\right)\right] \\ &= \inf_{\pi^2\in\Pi^2_{Ad}} E_x^{\pi^{*1}_\delta,\pi^2}\left[e^{n\delta/\alpha}\exp\left(\theta\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^{*1}_\delta(\xi_{t-},t),\pi^2(t))dt\right)\right], \end{aligned} \tag{4.4.4}$$

where $(\pi_\delta^{*1}, \pi_\delta^{*2}) \in \Pi_M^1 \times \Pi_M^2$, some pair of strategies and $x \in S$, $\theta \in (\delta, 1]$ a.e.

Proof of part 1: Let us fix a function $u \in B_1([0,1]\times S)$. Then since c_n is bounded, we have

$$\begin{aligned}
&\sup_{\theta\in[\delta,1]}\sup_{x\in S}\left|\int_\delta^\theta \sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\left\{\frac{1}{s}\sup_{x\in S}\left[\int_S q^{(n)}(dy|x,\mu,\nu)u(s,y)\right]+c_n(x,\mu,\nu)\sup_{x\in S}u(s,x)\right\}ds\right| \\
&\le \|u\|_1^\infty\left\{\int_\delta^1 \sup_{\mu\in\mathcal{P}(A(x))}\sup_{\nu\in\mathcal{P}(B(x))}\frac{1}{s}\sup_{x\in S}\Big(2q_x^{(n)}(\mu,\nu)\Big)ds+n(1-\delta)\right\} \\
&\le \|u\|_1^\infty\left[(-2)\tilde{q}^{(n)}\log\delta+n(1-\delta)\right].
\end{aligned}$$

So, we can define a nonlinear operator $\hat{T}: B_1([0,1]\times S)\to B_1([0,1]\times S)$ as follows:

$$\hat{T}u(\theta,x) = e^{\delta n/\alpha}+\frac{1}{\alpha}\int_\delta^\theta \sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\left[\frac{1}{s}\int_S q^{(n)}(dy|x,\mu,\nu)u(s,y)+c_n(x,\mu,\nu)u(s,x)\right]ds.$$

It is easy to see that the map $\hat{T}$ is well-defined.

Then by analogous arguments as in [Lemma 2.4.1, Chapter 2, p. 24-25] and by Banach fixed point theorem, there exists a unique bounded function $\varphi_\alpha^{(n,\delta)} \in B_1([0,1]\times S)$ (depending on (n,δ)) such that $\hat{T}\varphi_\alpha^{(n,\delta)} = \varphi_\alpha^{(n,\delta)}$; that is,

$$\begin{aligned}
&\varphi_\alpha^{(n,\delta)}(\theta,x) \\
&= e^{\delta n/\alpha}+\frac{1}{\alpha}\int_\delta^\theta \sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\left[\frac{1}{s}\int_S q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,y)+c_n(x,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,x)\right]ds \\
&= e^{\delta n/\alpha}+\frac{1}{\alpha}\int_\delta^\theta \inf_{\nu\in\mathcal{P}(B(x))}\sup_{\mu\in\mathcal{P}(A(x))}\left[\frac{1}{s}\int_S q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,y)+c_n(x,\mu,\nu)\varphi_\alpha^{(n,\delta)}(s,x)\right]ds,
\end{aligned}$$

where the second equality holds by Fan's minimax theorem, see [33, Theorem 3]. Hence by using the above equation, (4.4.1), and (4.4.2) we say $\varphi_\alpha^{(n,\delta)} \in L_V^\infty([0,1]\times S)$ and it satisfies equation (4.4.3).

Proof of part 2: First notice that $\varphi_\alpha^{(n,\delta)}(\theta(T_\delta(\theta)),\xi_{T_\delta}^{(n)}) = e^{n\delta/\alpha}$. Now let us consider two measurable functions

$$\hat{\mu}:(0,1)\times S\to\mathcal{P}(A) \text{ and } \hat{\nu}:(0,1)\times S\to\mathcal{P}(B),$$

such that

$$\begin{aligned}
&\sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\left[\int_S q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n,\delta)}(\theta,y)+\theta c_n(x,\mu,\nu)\varphi_\alpha^{(n,\delta)}(\theta,x)\right] \\
&= \inf_{\nu\in\mathcal{P}(B(x))}\left[\int_S q^{(n)}(dy|x,\hat{\mu}(\theta,x),\nu)\varphi_\alpha^{(n,\delta)}(\theta,y)+\theta c_n(x,\hat{\mu}(\theta,x),\nu)\varphi_\alpha^{(n,\delta)}(\theta,x)\right] \qquad (4.4.5)
\end{aligned}$$

and

$$\begin{aligned}
&\inf_{\nu\in\mathcal{P}(B(x))}\sup_{\mu\in\mathcal{P}(A(x))}\left[\int_S q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n,\delta)}(\theta,y)+\theta c_n(x,\mu,\nu)\varphi_\alpha^{(n,\delta)}(\theta,x)\right]\\
&=\sup_{\mu\in\mathcal{P}(A(x))}\left[\int_S q^{(n)}(dy|x,\mu,\hat{\nu}(\theta,x))\varphi_\alpha^{(n,\delta)}(\theta,y)+\theta c_n(x,\mu,\hat{\nu}(\theta,x))\varphi_\alpha^{(n,\delta)}(\theta,x)\right].
\end{aligned}\tag{4.4.6}$$

The measurable selection theorem in [89, Theorem 2.2] ensures the existence of such measurable maps. Let

$$\pi_\delta^{*1}:\mathbb{R}_+\times S\to\mathcal{P}(A)\ \text{and}\ \ \pi_\delta^{*2}:\mathbb{R}_+\times S\to\mathcal{P}(B)\tag{4.4.7}$$

be defined by $\pi_\delta^{*1}(x,t)=\hat{\mu}(\theta e^{-\alpha t},x)$ and $\pi_\delta^{*2}(x,t)=\hat{\nu}(\theta e^{-\alpha t},x)$. Here $(\pi_\delta^{*1},\pi_\delta^{*2})\in\Pi_M^1\times\Pi_M^2$. Let $\theta(t):=\theta e^{-\alpha t}$ for $t\in[0,\infty)$. Since c_n and $\varphi_\alpha^{(n,\delta)}$ are bounded, by using Dynkin's formula from [53, Theorem 3.1] we obtain

$$\begin{aligned}
&E_x^{\pi^1,\pi^2}\Bigg[\int_0^{T_\delta(\theta)}\Bigg[-\alpha\theta(s)\frac{\partial\varphi_\alpha^{(n,\delta)}}{\partial\theta}(\theta(s),\xi_s^{(n)})+\int_S q^{(n)}(dy|\xi_s^{(n)},\pi^1(s),\pi^2(s))\varphi_\alpha^{(n,\delta)}(\theta(s),y)\\
&+\theta(s)c_n(\xi_s^{(n)}(\omega),\pi^1(s),\pi^2(s))\varphi_\alpha^{(n,\delta)}(\theta(s),\xi_s^{(n)})\Bigg]\exp\left(\int_0^s\theta(v)c_n(\xi_v^{(n)},\pi^1(v),\pi^2(v))dv\right)ds\Bigg]\\
&=E_x^{\pi^1,\pi^2}\left[\exp\left(\int_0^{T_\delta(\theta)}\theta(s)c_n(\xi_s^{(n)},\pi^1(s),\pi^2(s))ds\right)\varphi_\alpha^{(n,\delta)}\left(\theta(T_\delta),\xi_{T_\delta}^{(n)}\right)\right]-\varphi_\alpha^{(n,\delta)}(\theta,x).
\end{aligned}\tag{4.4.8}$$

From (4.4.3), (4.4.6), and (4.4.8), we have

$$E_x^{\pi^1,\pi_\delta^{*2}}\left[e^{n\delta/\alpha}\exp\left(\int_0^{T_\delta(\theta)}\theta(s)c_n(\xi_s^{(n)},\pi^1(s),\pi_\delta^{*2}(\xi_{s-},s))ds\right)\right]\le\varphi_\alpha^{(n,\delta)}(\theta,x).$$

Therefore, we obtain

$$\begin{aligned}
\varphi_\alpha^{(n,\delta)}(\theta,x)&\ge\sup_{\pi^1\in\Pi_{Ad}^1}E_x^{\pi^1,\pi_\delta^{*2}}\left[e^{n\delta/\alpha}\exp\left(\int_0^{T_\delta(\theta)}\theta(s)c_n(\xi_s^{(n)},\pi^1(s),\pi_\delta^{*2}(\xi_{s-},s))ds\right)\right]\\
&\ge\inf_{\pi^2\in\Pi_{Ad}^2}\sup_{\pi^1\in\Pi_{Ad}^1}E_x^{\pi^1,\pi^2}\left[e^{n\delta/\alpha}\exp\left(\int_0^{T_\delta(\theta)}\theta(s)c_n(\xi_s^{(n)},\pi^1(s),\pi^2(s))ds\right)\right].
\end{aligned}\tag{4.4.9}$$

Using (4.4.3), (4.4.5), and (4.4.8), by analogous arguments, one can prove that

$$\begin{aligned}
\varphi_\alpha^{(n,\delta)}(\theta,x)&\le\inf_{\pi^2\in\Pi_{Ad}^2}E_x^{\pi_\delta^{*1},\pi^2}\left[e^{n\delta/\alpha}\exp\left(\int_0^{T_\delta(\theta)}\theta(s)c_n(\xi_s^{(n)},\pi_\delta^{*1}(\xi_{s-},s),\pi^2(s))ds\right)\right]\\
&\le\sup_{\pi^1\in\Pi_{Ad}^1}\inf_{\pi^2\in\Pi_{Ad}^2}E_x^{\pi^1,\pi^2}\left[e^{n\delta/\alpha}\exp\left(\int_0^{T_\delta(\theta)}\theta(s)c_n(\xi_s^{(n)},\pi^1(s),\pi^2(s))ds\right)\right].
\end{aligned}\tag{4.4.10}$$

Therefore, from (4.4.9) and (4.4.10), we obtain (4.4.4).

Step 2: In this step, we prove the existence of an absolutely continuous function $\varphi_\alpha^{(n)}(\cdot,x)$ from $\varphi_\alpha^{(n,\delta)}(\cdot,x)$ by passing to the limit as $\delta \to 0$, which satisfies the following first-order DEs:

$$\begin{cases} \alpha\theta\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) & = \sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\left[\int_S q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,y)+\theta c_n(x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,x)\right] \\ & = \inf_{\nu\in\mathcal{P}(B(x))}\sup_{\mu\in\mathcal{P}(A(x))}\left[\int_S q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,y)+\theta c_n(x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,x)\right], \\ & \quad x\in S,\ \text{a.e. } \theta\in[0,1], \\ 1\le\varphi_\alpha^{(n)}(\theta,x) & \le \frac{\alpha^2 e^{\theta\hat{L}_0/\alpha}}{\alpha^2-\hat{\rho}_0\hat{\rho}_1\theta}V^{\frac{\hat{\rho}_1\theta}{\alpha}}(x) \quad \forall\,(\theta,x)\in[0,1]\times S. \end{cases} \tag{4.4.11}$$

Since $\sup_{(x,a,b)\in K} c_n(x,a,b) \le n$, for given n and $0<\delta<1$, by (4.4.4) we have $\varphi_\alpha^{(n,\delta)}(\theta,x) \le e^{2n/\alpha}$, $x\in S, \theta\in[\delta,1]$. Next, we extend the domain of $\varphi_\alpha^{(n,\delta)}$ to $[0,1]\times S$ by

$$\overline{\varphi}_\alpha^{(n,\delta)}(\theta,x)=\begin{cases} e^{n\delta/\alpha}, & 0\le\theta<\delta\ \forall x\in S \\ \varphi_\alpha^{(n,\delta)}(\theta,x), & \delta\le\theta\le1\ \forall x\in S. \end{cases}$$

Note that $e^{bz}-1\le(e^b-1)z$ holds for $z\in[0,1]$ and $b>0$. Thus, for any fix $n\ge1$; we have

$$\begin{aligned} \int_{T_\delta(\theta\wedge\theta_0)}^{T_\delta(\theta\vee\theta_0)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt &\le n\int_{T_\delta(\theta\wedge\theta_0)}^{T_\delta(\theta\vee\theta_0)} e^{-\alpha t}dt \\ &= \frac{n}{\alpha}[\exp(-\alpha T_\delta(\theta\wedge\theta_0))-\exp(-\alpha T_\delta(\theta\vee\theta_0))] \\ &\le \frac{\delta n|\theta_0-\theta|}{\alpha\theta\theta_0}, \end{aligned} \tag{4.4.12}$$

where $a\wedge b:=\min\{a,b\}$ and $a\vee b:=\max\{a,b\}$ and

$$\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt \le n\int_0^{T_\delta(\theta)} e^{-\alpha t}dt \le \frac{n}{\alpha}. \tag{4.4.13}$$

Now by (4.4.12) and (4.4.13), we have

$$\begin{aligned} \hat{P}_1 :=& e^{n\delta/\alpha}E_x^{\pi^1,\pi^2}\left[\exp\left(\theta_0\int_0^{T_\delta(\theta\wedge\theta_0)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\right)\right. \\ &\left.\times\left(\exp\left(\theta_0\int_{T_\delta(\theta\wedge\theta_0)}^{T_\delta(\theta\vee\theta_0)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\right)-1\right)\right] \\ \le& e^{2n/\alpha}E_x^{\pi^1,\pi^2}\left[\exp\left(\theta_0\int_{T_\delta(\theta\wedge\theta_0)}^{T_\delta(\theta\vee\theta_0)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\right)-1\right] \\ \le& e^{2n/\alpha}\left(\exp\left(\frac{n\delta|\theta-\theta_0|}{\alpha\theta}\right)-1\right) \end{aligned}$$

$$\leq e^{2n/\alpha}\Big(e^{n/\alpha}-1\Big)|\theta_0-\theta|,$$

and

$$\begin{aligned}\hat{P}_2 :=& e^{n\delta/\alpha}E_x^{\pi^1,\pi^2}\Bigg[\exp\Big((\theta\wedge\theta_0)\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\Big)\\ &\times\Big(\exp\Big(|\theta_0-\theta|\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\Big)-1\Big)\Bigg]\\ \leq& e^{2n/\alpha}E_x^{\pi^1,\pi^2}\Bigg[\exp\Big(|\theta_0-\theta|\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\Big)-1\Bigg]\\ \leq& e^{2n/\alpha}\Big(\exp\Big(\frac{n}{\alpha}|\theta_0-\theta|\Big)-1\Big)\\ \leq& e^{2n/\alpha}\Big(e^{n/\alpha}-1\Big)|\theta_0-\theta|.\end{aligned}$$

Then for any given $\theta,\theta_0\in[\delta,1]$, $(\pi^1,\pi^2)\in\Pi^1_{Ad}\times\Pi^2_{Ad}$, $x\in S$, we have

$$\begin{aligned}&\Bigg|E_x^{\pi^1,\pi^2}\Bigg[e^{n\delta/\alpha}\exp\Big(\theta_0\int_0^{T_\delta(\theta_0)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\Big)\Bigg]\\ &\quad - E_x^{\pi^1,\pi^2}\Bigg[e^{n\delta/\alpha}\exp\Big(\theta\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\Big)\Bigg]\Bigg|\\ \leq& \Bigg|E_x^{\pi^1,\pi^2}\Bigg[e^{n\delta/\alpha}\exp\Big(\theta_0\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\Big)\Bigg]\\ &\quad - E_x^{\pi^1,\pi^2}\Bigg[e^{n\delta/\alpha}\exp\Big(\theta_0\int_0^{T_\delta(\theta_0)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\Big)\Bigg]\Bigg|\\ &\quad + \Bigg|E_x^{\pi^1,\pi^2}\Bigg[e^{n\delta/\alpha}\exp\Big(\theta\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\Big)\Bigg]\\ &\quad - E_x^{\pi^1,\pi^2}\Bigg[e^{n\delta/\alpha}\exp\Big(\theta_0\int_0^{T_\delta(\theta)} e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\Big)\Bigg]\Bigg|\\ =& \hat{P}_1+\hat{P}_2.\end{aligned}$$

Hence for all $\theta,\theta_0\in[0,1]$, $x\in S$, we have

$$|\overline{\varphi}_\alpha^{(n,\delta)}(\theta_0,x)-\overline{\varphi}_\alpha^{(n,\delta)}(\theta,x)|\leq 2e^{2n/\alpha}(e^{n/\alpha}-1)|\theta-\theta_0|. \tag{4.4.14}$$

Now we want to show that $\overline{\varphi}_\alpha^{(n,\delta)}$ is increasing in δ for any (θ,x). By measurable selection theorem we get $\pi_\delta^{*1}\in\Pi^1_M$ as like in equation (4.4.7), corresponding to $\overline{\varphi}_\alpha^{(n,\delta)}$. Consider a fixed constant $\varepsilon>0$ to be small enough. For any $\pi^2\in\Pi^2_{Ad}$, the followings cases hold.

Case 1. $\theta\leq\delta$

$$\overline{\varphi}_\alpha^{(n,\delta+\varepsilon)}(\theta,x)-\overline{\varphi}_\alpha^{(n,\delta)}(\theta,x)=e^{\frac{n(\delta+\varepsilon)}{\alpha}}-e^{\frac{n\delta}{\alpha}}=e^{\frac{n\delta}{\alpha}}(e^{\frac{n\varepsilon}{\alpha}}-1)\geq 0.$$

Case 2. $\delta < \theta \leq \delta + \varepsilon$

$$\begin{aligned}
&\overline{\varphi}_{\alpha}^{(n,\delta+\varepsilon)}(\theta, x) - \overline{\varphi}_{\alpha}^{(n,\delta)}(\theta, x)\\
&= e^{\frac{n(\delta+\varepsilon)}{\alpha}} - \inf_{\pi^2\in\Pi^2_{Ad}} E_x^{\pi_\delta^{*1},\pi^2}\left[e^{\frac{n\delta}{\alpha}} \exp\left(\theta \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi_\delta^{*1}(\xi_{t-}, t), \pi^2(t))dt\right)\right]\\
&\geq e^{\frac{n(\delta+\varepsilon)}{\alpha}} - E_x^{\pi_\delta^{*1},\pi^2}\left[e^{\frac{n\delta}{\alpha}} \exp\left(\theta \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi_\delta^{*1}(\xi_{t-}, t), \pi^2(t))dt\right)\right]\\
&= e^{\frac{n\delta}{\alpha}}\left[e^{\frac{n\varepsilon}{\alpha}} - E_x^{\pi_\delta^{*1},\pi^2}\left[\exp\left(\theta \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi_\delta^{*1}(\xi_{t-}, t), \pi^2(t))dt\right)\right]\right]\\
&\geq e^{\frac{n\delta}{\alpha}}\left[e^{\frac{n\varepsilon}{\alpha}} - \exp\left(\theta \int_0^{T_\delta(\theta)} e^{-\alpha t} n dt\right)\right]\\
&= e^{\frac{n\delta}{\alpha}}\left[e^{\frac{n\varepsilon}{\alpha}} - e^{\theta n \frac{(1-e^{-\alpha T_\delta(\theta)})}{\alpha}}\right]\\
&\geq 0.
\end{aligned}$$

Case 3. If $\theta > \delta + \varepsilon$ then

$$\begin{aligned}
&\overline{\varphi}_{\alpha}^{(n,\delta+\varepsilon)}(\theta, x) - \overline{\varphi}_{\alpha}^{(n,\delta)}(\theta, x)\\
&= \sup_{\pi^1\in\Pi^1_{Ad}} \inf_{\pi^2\in\Pi^2_{Ad}} E_x^{\pi^1,\pi^2}\left[e^{\frac{n(\delta+\varepsilon)}{\alpha}} \exp\left(\theta \int_0^{T_{\delta+\varepsilon}(\theta)} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi^1(t), \pi^2(t))dt\right)\right]\\
&\quad - \inf_{\pi^2\in\Pi^2_{Ad}} E_x^{\pi_\delta^{*1},\pi^2}\left[e^{\frac{n\delta}{\alpha}} \exp\left(\theta \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi_\delta^{*1}(\xi_{t-}, t), \pi^2(t))dt\right)\right]\\
&\geq \inf_{\pi^2\in\Pi^2_{Ad}} E_x^{\pi_\delta^{*1},\pi^2}\left[e^{\frac{n(\delta+\varepsilon)}{\alpha}} \exp\left(\theta \int_0^{T_{\delta+\varepsilon}(\theta)} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi_\delta^{*1}(\xi_{t-}, t), \pi^2(t))dt\right)\right]\\
&\quad - \inf_{\pi^2\in\Pi^2_{Ad}} E_x^{\pi_\delta^{*1},\pi^2}\left[e^{\frac{n\delta}{\alpha}} \exp\left(\theta \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi_\delta^{*1}(\xi_{t-}, t), \pi^2(t))dt\right)\right]\\
&\geq \inf_{\pi^2\in\Pi^2_{Ad}} \left[E_x^{\pi_\delta^{*1},\pi^2}\left[e^{\frac{n(\delta+\varepsilon)}{\alpha}} \exp\left(\theta \int_0^{T_{\delta+\varepsilon}(\theta)} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi_\delta^{*1}(\xi_{t-}, t), \pi^2(t))dt\right)\right]\right.\\
&\quad \left. - E_x^{\pi_\delta^{*1},\pi^2}\left[e^{\frac{n\delta}{\alpha}} \exp\left(\theta \int_0^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi_\delta^{*1}(\xi_{t-}, t), \pi^2(t))dt\right)\right]\right]\\
&= \inf_{\pi^2\in\Pi^2_{Ad}} e^{\frac{n\delta}{\alpha}} E_x^{\pi_\delta^{*1},\pi^2}\left[\exp\left(\theta \int_0^{T_{\delta+\varepsilon}(\theta)} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi_\delta^{*1}(\xi_{t-}, t), \pi^2(t))dt\right)\right.\\
&\quad \left. \times \left\{e^{\frac{n\varepsilon}{\alpha}} - \exp\left(\theta \int_{T_{\delta+\varepsilon}(\theta)}^{T_\delta(\theta)} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi_\delta^{*1}(\xi_{t-}, t), \pi^2(t))dt\right)\right\}\right]\\
&\geq \inf_{\pi^2\in\Pi^2_{Ad}} e^{\frac{n\delta}{\alpha}} E_x^{\pi_\delta^{*1},\pi^2}\left[\exp\left(\theta \int_0^{T_{\delta+\varepsilon}(\theta)} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi_\delta^{*1}(\xi_{t-}, t), \pi^2(t))dt\right)\right.\\
&\quad \left. \times \left\{e^{\frac{n\varepsilon}{\alpha}} - \exp\left(\theta \int_{T_{\delta+\varepsilon}(\theta)}^{T_\delta(\theta)} e^{-\alpha t} n dt\right)\right\}\right]\\
&= \inf_{\pi^2\in\Pi^2_{Ad}} e^{\frac{n\delta}{\alpha}} E_x^{\pi_\delta^{*1},\pi^2}\left[\exp\left(\theta \int_0^{T_{\delta+\varepsilon}(\theta)} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi_\delta^{*1}(\xi_{t-}, t), \pi^2(t))dt\right)\right.
\end{aligned}$$

$$\times \left\{ e^{\frac{n\varepsilon}{\alpha}} - \exp\left(\frac{n\theta(e^{-\alpha T_{\delta+\varepsilon}(\theta)} - e^{-\alpha T_{\delta}(\theta)})}{\alpha} \right) \right\} \Bigg]$$
$$= 0.$$

Hence $\varphi_\alpha^{(n,\delta)}(\theta,x)$ is increasing in δ for any $(\theta,x) \in [0,1] \times S$. So, by this fact and in view of (4.4.14), there exists a function $\varphi_\alpha^{(n)}$ on $[0,1]\times S$ that is continuous with respect to $\theta \in [0,1]$, such that along a subsequence $\delta_m \to 0$, we have $\lim_{m\to\infty} \varphi_\alpha^{(n,\delta_m)}(\theta,x) = \varphi_\alpha^{(n)}(\theta,x)$. Also for any fixed $x \in S$ this convergence is uniform in $\theta \in [0,1]$. Now for $\theta \in [\delta_m, 1]$, by using (4.4.4) and Proposition 4.2.5, we have

$$\begin{aligned}
&\varphi_\alpha^{(n,\delta_m)}(\theta,x)\\
&= \inf_{\pi^2\in\Pi^2_{Ad}} \sup_{\pi^1\in\Pi^1_{Ad}} E_x^{\pi^1,\pi^2}\left[e^{n\delta_m/\alpha} \exp\left(\theta \int_0^{T_{\delta_m}(\theta)} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi^1(t), \pi^2(t))dt\right)\right]\\
&\le e^{n\delta_m/\alpha} \inf_{\pi^2\in\Pi^2_{Ad}} \sup_{\pi^1\in\Pi^1_{Ad}} E_x^{\pi^1,\pi^2}\left[\exp\left(\theta \int_0^{\infty} e^{-\alpha t} c_n(\xi_t^{(n)}, \pi^1(t), \pi^2(t))dt\right)\right]\\
&\le e^{n\delta_m/\alpha} \inf_{\pi^2\in\Pi^2_{Ad}} \sup_{\pi^1\in\Pi^1_{Ad}} E_x^{\pi^1,\pi^2}\left[\exp\left(\theta \int_0^{\infty} e^{-\alpha t} c(\xi_t^{(n)}, \pi^1(t), \pi^2(t))dt\right)\right]\\
&\le e^{n\delta_m/\alpha} \frac{\alpha^2 e^{\theta \hat{L}_0/\alpha}}{\alpha^2 - \hat{\rho}_0\hat{\rho}_1\theta} V^{\frac{\hat{\rho}_1\theta}{\alpha}}(x).
\end{aligned}$$

Thus, letting $m \to \infty$ in the above equation, we obtain

$$1 \le \varphi_\alpha^{(n)}(\theta,x) \le \frac{\alpha^2 e^{\theta\hat{L}_0/\alpha}}{\alpha^2 - \hat{\rho}_0\hat{\rho}_1\theta} V^{\frac{\hat{\rho}_1\theta}{\alpha}}(x). \tag{4.4.15}$$

Now by measurable selection theorem in [89, Theorem 2.2], there exists a pair of strategies $(\mu_m^*, \nu_m^*) \in \mathcal{P}(A(x)) \times \mathcal{P}(B(x))$ such that (4.4.3) can be written as

$$\begin{cases}
\alpha\theta \frac{\partial \overline{\varphi}_\alpha^{(n,\delta_m)}}{\partial\theta}(\theta,x) &= \sup\limits_{\mu\in\mathcal{P}(A(x))} \left[\int_S q^{(n)}(dy|x,\mu,\nu_m^*)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y) + \theta c_n(x,\mu,\nu_m^*)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)\right]\\
&= \inf\limits_{\nu\in\mathcal{P}(B(x))} \left[\int_S q^{(n)}(dy|x,\mu_m^*,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y) + \theta c_n(x,\mu_m^*,\nu)\varphi_\alpha^{(n,\delta_m)}(\theta,x)\right]\\
\overline{\varphi}_\alpha^{(n,\delta_m)}(\delta_m,x) &= e^{n\delta_m/\alpha}.
\end{cases} \tag{4.4.16}$$

Since for each fixed $x \in S$, $\mathcal{P}(A(x))$ and $\mathcal{P}(B(x))$ are compact, there exists a subsequence (by abuse of notation, we denoted by the same sequence) and a pair of strategies $(\mu^*, \nu^*) \in \mathcal{P}(A(x)) \times \mathcal{P}(B(x))$ such that $\lim_{m\to\infty} \mu_m^* = \mu^*$ and $\lim_{m\to\infty} \nu_m^* = \nu^*$. Define

$$Q^{(n)}(dy|x,a,b) := \delta_x(dy) + \frac{q^{(n)}(dy|x,a,b)}{M_0 V(x)} \tag{4.4.17}$$

for all $(x,a,b) \in K$ where $\delta_x(\cdot)$ is the Dirac measure concentrated at x. We see that under Assumption (A1), $Q^{(n)}$ is a stochastic kernel on S given K. Now let $\hat{\psi} \in C_c^\infty(0,1)$.

Hence $\hat{\psi}(1) = 0$. First, see

$$\begin{aligned}
&\left| \sup_{\mu\in\mathcal{P}(A(x))} \inf_{\nu\in\mathcal{P}(B(x))} \left[\int_S Q^{(n)}(dy|x,\mu,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y) + \frac{\theta}{M_0V(x)} c_n(x,\mu,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x) \right] \hat{\psi}(\theta) \right| \\
&\leq \sup_{\mu\in\mathcal{P}(A(x))} \sup_{\nu\in\mathcal{P}(B(x))} \left[\int_S Q^{(n)}(dy|x,\mu,\nu)|\overline{\varphi}_\alpha^{(n,\delta)}(\theta,y)| + \frac{\theta}{M_0V(x)} |c_n(x,\mu,\nu)||\overline{\varphi}_\alpha^{(n,\delta)}(\theta,x)| \right] |\hat{\psi}(\theta)| \\
&\leq \|\overline{\varphi}_\alpha^{(n,\delta_m)}\|_V \sup_{\mu\in\mathcal{P}(A(x))} \sup_{\nu\in\mathcal{P}(B(x))} \left[\int_S Q(dy|x,\mu,\nu)V(y) + \frac{\theta}{M_0V(x)} nV(x) \right] |\hat{\psi}(\theta)| \\
&\leq \|\overline{\varphi}_\alpha^{(n,\delta_m)}\|_V \left[\hat{\rho}_0 \frac{V(x)}{M_0V(x)} + V(x) + \frac{\theta}{M_0V(x)} nV(x) \right] |\hat{\psi}(\theta)| \\
&= \|\overline{\varphi}_\alpha^{(n,\delta_m)}\|_V \left[\frac{\theta}{M_0} n + V(x) + \frac{\hat{\rho}_0}{M_0} \right] |\hat{\psi}(\theta)|. \qquad (4.4.18)
\end{aligned}$$

Now,

$$\begin{aligned}
&-\int_0^1 \alpha \frac{d(\theta\hat{\psi})}{d\theta}(\theta)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)d\theta = \int_0^1 \alpha\theta \frac{\partial\overline{\varphi}_\alpha^{(n,\delta_m)}}{\partial\theta}(\theta,x)\hat{\psi}(\theta)d\theta \\
&= \int_0^1 \sup_{\mu\in\mathcal{P}(A(x))} \left[\int_S q^{(n)}(dy|x,\mu,\nu_m^*)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y) + \theta c_n(x,\mu,\nu_m^*)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x) \right] \hat{\psi}(\theta)d\theta \\
&- \int_0^{\delta_m} \sup_{\mu\in\mathcal{P}(A(x))} \inf_{\nu\in\mathcal{P}(B(x))} \left[\int_S q^{(n)}(dy|x,\mu,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y) + \theta c_n(x,\mu,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x) \right] \hat{\psi}(\theta)d\theta \\
&= \int_0^1 \inf_{\nu\in\mathcal{P}(B(x))} \left[\int_S q^{(n)}(dy|x,\mu_m^*,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y) + \theta c_n(x,\mu_m^*,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x) \right] \hat{\psi}(\theta)d\theta \\
&- \int_0^{\delta_m} \sup_{\mu\in\mathcal{P}(A(x))} \inf_{\nu\in\mathcal{P}(B(x))} \left[\int_S q^{(n)}(dy|x,\mu,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y) + \theta c_n(x,\mu,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x) \right] \hat{\psi}(\theta)d\theta \\
&= \int_0^1 \sup_{\mu\in\mathcal{P}(A(x))} \left[\int_S q^{(n)}(dy|x,\mu,\nu_m^*)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y) + \theta c_n(x,\mu,\nu_m^*)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x) \right] \hat{\psi}(\theta)d\theta \\
&- \int_0^{\delta_m} \sup_{\mu\in\mathcal{P}(A(x))} \inf_{\nu\in\mathcal{P}(B(x))} \left[\theta c_n(x,\mu,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x) \right] \hat{\psi}(\theta)d\theta \\
&= \int_0^1 \inf_{\nu\in\mathcal{P}(B(x))} \left[\int_S q^{(n)}(dy|x,\mu_m^*,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y) + \theta c_n(x,\mu_m^*,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x) \right] \hat{\psi}(\theta)d\theta \\
&- \int_0^{\delta_m} \sup_{\mu\in\mathcal{P}(A(x))} \inf_{\nu\in\mathcal{P}(B(x))} \left[\theta c_n(x,\mu,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x) \right] \hat{\psi}(\theta)d\theta. \qquad (4.4.19)
\end{aligned}$$

Then by (4.4.17), eq. (4.4.19) can be written as

$$\begin{aligned}
&-\int_0^1 \left\{ \frac{\alpha}{M_0V(x)} \frac{d(\theta\psi)}{d\theta} \overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x) - \overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)\hat{\psi}(\theta) \right\} d\theta \\
&= \int_0^1 \sup_{\mu\in\mathcal{P}(A(x))} \left[\int_S Q^{(n)}(dy|x,\mu,\nu_m^*)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y) + \frac{\theta}{M_0V(x)} c_n(x,\mu,\nu_m^*)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x) \right] \hat{\psi}(\theta)d\theta \\
&\quad - \frac{1}{M_0V(x)} \int_0^{\delta_m} \sup_{\mu\in\mathcal{P}(A(x))} \inf_{\nu\in\mathcal{P}(B(x))} \left[\theta c_n(x,\mu,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x) \right] \hat{\psi}(\theta)d\theta
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^1 \inf_{\nu\in\mathcal{P}(B(x))}\Big[\int_S Q^{(n)}(dy|x,\mu_m^*,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y)+\frac{\theta}{M_0V(x)}c_n(x,\mu_m^*,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)\Big]\hat{\psi}(\theta)d\theta\\
&\quad -\frac{1}{M_0V(x)}\int_0^{\delta_m}\sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\Big[\theta c_n(x,\mu,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)\Big]\hat{\psi}(\theta)d\theta. \qquad (4.4.20)
\end{aligned}
$$

Now, from (4.4.20), for any $\nu\in\mathcal{P}(B(x))$, we have

$$
\begin{aligned}
&-\int_0^1\Big\{\frac{\alpha}{M_0V(x)}\frac{d(\theta\psi)}{d\theta}\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)-\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)\hat{\psi}(\theta)\Big\}d\theta\\
&\le\int_0^1\Big[\int_S Q^{(n)}(dy|x,\mu_m^*,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,y)+\frac{\theta}{M_0V(x)}c_n(x,\mu_m^*,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)\Big]\hat{\psi}(\theta)d\theta\\
&\quad-\frac{1}{M_0V(x)}\int_0^{\delta_m}\sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\Big[\theta c_n(x,\mu,\nu)\overline{\varphi}_\alpha^{(n,\delta_m)}(\theta,x)\Big]\hat{\psi}(\theta)d\theta. \qquad (4.4.21)
\end{aligned}
$$

Under our given Assumptions from [56, Lemma 7.2], we know that the functions $c(x,\mu,\nu)$, and $\int_S q(dy|x,\mu,\nu)u(\cdot,y)$ are continuous at (μ,ν) on $\mathcal{P}(A(x))\times\mathcal{P}(B(x))$ for each fixed $x\in S$ and $u\in L_V^\infty([0,1]\times S)$. So, by using dominated convergent theorem and taking the limit as $m\to\infty$ in (4.4.21), we obtain

$$
\begin{aligned}
&-\int_0^1\Big\{\frac{\alpha}{M_0V(x)}\frac{d(\theta\hat{\psi})}{d\theta}(\theta)\varphi_\alpha^{(n)}(\theta,x)-\varphi_\alpha^{(n)}(\theta,x)\hat{\psi}(\theta)\Big\}d\theta\\
&\le\int_0^1\Big[\int_S Q^{(n)}(dy|x,\mu^*,\nu)\varphi_\alpha^{(n)}(\theta,y)+\frac{\theta}{M_0V(x)}c_n(x,\mu^*,\nu)\varphi_\alpha^{(n)}(\theta,x)\Big]\hat{\psi}(\theta)d\theta.
\end{aligned}
$$

Since $\nu\in\mathcal{P}(B(x))$ is arbitrary, therefore

$$
\begin{aligned}
&-\int_0^1\Big\{\frac{\alpha}{M_0V(x)}\frac{d(\theta\psi)}{d\theta}(\theta)\varphi_\alpha^{(n)}(\theta,x)-\varphi_\alpha^{(n)}(\theta,x)\hat{\psi}(\theta)\Big\}d\theta\\
&\le\inf_{\nu\in\mathcal{P}(B(x))}\int_0^1\Big[\int_S Q^{(n)}(dy|x,\mu^*,\nu)\varphi_\alpha^{(n)}(\theta,y)+\frac{\theta}{M_0V(x)}c_n(x,\mu^*,\nu)\varphi_\alpha^{(n)}(\theta,x)\Big]\hat{\psi}(\theta)d\theta.
\end{aligned}
$$

Hence

$$
\begin{aligned}
&-\int_0^1\Big\{\frac{\alpha}{M_0V(x)}\frac{d(\theta\psi)}{d\theta}(\theta)\varphi_\alpha^{(n)}(\theta,x)-\varphi_\alpha^{(n)}(\theta,x)\hat{\psi}(\theta)\Big\}d\theta\\
&\le\sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\int_0^1\Big[\int_S Q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,y)+\frac{\theta}{M_0V(x)}c_n(x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,x)\Big]\hat{\psi}(\theta)d\theta. \qquad (4.4.22)
\end{aligned}
$$

Similarly, by using Fatou's Lemma from [69, Lemma 8.3.7] and taking $m\to\infty$ in (4.4.20), we have

$$
\begin{aligned}
&-\int_0^1\Big\{\frac{\alpha}{M_0V(x)}\frac{d(\theta\hat{\psi})}{d\theta}(\theta)\varphi_\alpha^{(n)}(\theta,x)-\varphi_\alpha^{(n)}(\theta,x)\hat{\psi}(\theta)\Big\}d\theta\\
&\ge\sup_{\mu\in\mathcal{P}(A(x))}\int_0^1\Big[\int_S Q^{(n)}(dy|x,\mu,\nu^*)\varphi_\alpha^{(n)}(\theta,y)+\frac{\theta}{M_0V(x)}c_n(x,\mu,\nu^*)\varphi_\alpha^{(n)}(\theta,x)\Big]\hat{\psi}(\theta)d\theta.
\end{aligned}
$$

Therefore we have

$$\begin{aligned}
&-\int_0^1 \left\{ \frac{\alpha}{M_0 V(x)} \frac{d(\theta\hat{\psi})}{d\theta}(\theta)\varphi_\alpha^{(n)}(\theta,x) - \varphi_\alpha^{(n)}(\theta,x)\hat{\psi}(\theta)\right\} d\theta \\
&\geq \inf_{\nu\in\mathcal{P}(B(x))}\sup_{\mu\in\mathcal{P}(A(x))}\int_0^1\left[\int_S Q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,y) + \frac{\theta}{M_0V(x)}c_n(x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,x)\right]\hat{\psi}(\theta)d\theta.
\end{aligned} \tag{4.4.23}$$

Thus by (4.4.17), (4.4.22), and (4.4.23), we obtain

$$\begin{aligned}
&-\int_0^1 \alpha\frac{d(\theta\hat{\psi})}{d\theta}(\theta)\varphi_\alpha^{(n)}(\theta,x)d\theta \\
&= \sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\int_0^1\left[\int_S q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,y) + \theta c_n(x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,x)\right]\hat{\psi}(\theta)d\theta \\
&= \inf_{\nu\in\mathcal{P}(B(x))}\sup_{\mu\in\mathcal{P}(A(x))}\int_0^1\left[\int_S q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,y) + \theta c_n(x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,x)\right]\hat{\psi}(\theta)d\theta.
\end{aligned}$$

Hence for a.e. $\theta\in[0,1]$, in the sense of distribution, we have

$$\begin{aligned}
\alpha\theta\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) &= \sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\left[\int_S q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,y) + \theta c_n(x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,x)\right] \\
&= \inf_{\nu\in\mathcal{P}(B(x))}\sup_{\mu\in\mathcal{P}(A(x))}\left[\int_S q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,y) + \theta c_n(x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,x)\right].
\end{aligned}$$

When $\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}$ does not exist for some (θ,x), we define

$$\alpha\theta\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) = \sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\left[\int_S q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,y) + \theta c_n(x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,x)\right].$$

Hence $\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\cdot,x)$ exists almost everywhere. By using the first order DEs satisfied by $\varphi_\alpha^{(n)}$ (that is just proven), (4.4.1), (4.4.2), and (4.4.15), we say that $\varphi_\alpha^{(n)}\in L_V^\infty([0,1]\times S)$ and it is a solution of (4.4.11). Thus by closely mimicking the arguments in Theorem 4.3.1, one can easily get the stochastic representation of the solution $\varphi_\alpha^{(n)}$, that is

$$\begin{aligned}
\varphi_\alpha^{(n)}(\theta,x) &= \inf_{\pi^2\in\Pi_{Ad}^2}\sup_{\pi^1\in\Pi_{Ad}^1}E_x^{\pi^1,\pi^2}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\right)\right] \\
&= \sup_{\pi^1\in\Pi_{Ad}^1}\inf_{\pi^2\in\Pi_{Ad}^2}E_x^{\pi^1,\pi^2}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\right)\right].
\end{aligned} \tag{4.4.24}$$

Step 3: In this step, we complete the proof by passing to the limit as $n\to\infty$. First we prove that for each $x\in S$, $\{\varphi_\alpha^{(n)}\}_{n\geq1}$ is equicontinuous on $[0,1]$. Note that $e^{bz}-1\leq(e^b-1)z$ for all $z\in[0,1]$ and $b>0$. Now, for any given $\theta,\theta_0\in[\delta,1]$, $(\pi^1,\pi^2)\in\Pi_{Ad}^1\times\Pi_{Ad}^2$, $x\in S$, by [eq. (2.3.7), Chapter 2], we get

$$|\varphi_\alpha^{(n)}(\theta_0,x)-\varphi_\alpha^{(n)}(\theta,x)|$$

$$
\begin{aligned}
&\le \sup_{\pi^1\in\Pi^1_{Ad}} \sup_{\pi^2\in\Pi^2_{Ad}} \Bigg| E_x^{\pi^1,\pi^2}\bigg[\exp\bigg(\theta\int_0^\infty e^{-\alpha t} c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\bigg)\bigg] \\
&\qquad - E_x^{\pi^1,\pi^2}\bigg[\exp\bigg(\theta_0\int_0^\infty e^{-\alpha t} c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\bigg)\bigg]\Bigg| \\
&\le \sup_{\pi^1\in\Pi^1_{Ad}} \sup_{\pi^2\in\Pi^2_{Ad}} E_x^{\pi^1,\pi^2}\bigg[\exp\bigg((\theta\wedge\theta_0)\int_0^\infty e^{-\alpha t} c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\bigg) \\
&\qquad \times\bigg(\exp\bigg(|\theta_0-\theta|\int_0^\infty e^{-\alpha t} c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\bigg)-1\bigg)\bigg] \\
&\le \sup_{\pi^1\in\Pi^1_{Ad}} \sup_{\pi^2\in\Pi^2_{Ad}} E_x^{\pi^1,\pi^2}\bigg[\exp\bigg((\theta\wedge\theta_0)\int_0^\infty e^{-\alpha t} c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\bigg) \\
&\qquad \times\bigg(\exp\bigg(\int_0^\infty e^{-\alpha t} c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\bigg)-1\bigg)|\theta_0-\theta|\bigg] \\
&\le \sup_{\pi^1\in\Pi^1_{Ad}} \sup_{\pi^2\in\Pi^2_{Ad}} E_x^{\pi^1,\pi^2}\bigg[\exp\bigg(\int_0^\infty e^{-\alpha t} c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\bigg) \\
&\qquad \times\bigg(\exp\bigg(\int_0^\infty e^{-\alpha t} c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\bigg)|\theta_0-\theta|\bigg)\bigg] \\
&= |\theta_0-\theta|\times E_x^{\pi^1,\pi^2}\bigg[\exp\bigg(2\int_0^\infty e^{-\alpha t} c_n(\xi_t^{(n)},\pi^1(t),\pi^2(t))dt\bigg)\bigg] \\
&\le |\theta_0-\theta|\times\frac{\alpha e^{2\hat{L}_0/\alpha}}{\alpha-\hat\rho_2}\bigg(V^2(x)+\frac{b_1}{\hat\rho_2}\bigg). \qquad (4.4.25)
\end{aligned}
$$

So, for each $x\in S$, $\varphi^{(n)}(\cdot,x)$ is Lipschitz continuous in $\theta\in[0,1]$. Now, by measurable selection theorem in [89, Theorem 2.2], there exists a measurable function $\nu_n^*:[0,1]\times S\to\mathcal{P}(B)$ such that

$$
\begin{aligned}
&\inf_{\nu\in\mathcal{P}(B(x))}\sup_{\mu\in\mathcal{P}(A(x))}\bigg[\int_S q^{(n)}(dy|x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,y)+\theta c_n(x,\mu,\nu)\varphi_\alpha^{(n)}(\theta,x)\bigg] \\
&=\sup_{\mu\in\mathcal{P}(A(x))}\bigg[\int_S q^{(n)}(dy|x,\mu,\nu_n^*(\theta,x))\varphi_\alpha^{(n)}(\theta,y)+\theta c_n(x,\mu,\nu_n^*(\theta,x))\varphi_\alpha^{(n)}(\theta,x)\bigg]. \qquad (4.4.26)
\end{aligned}
$$

Hence by equation (4.4.11), we have a.e. $\theta\in[0,1]$ and $\forall x\in S$, we have

$$
\begin{cases}
\alpha\theta\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) &= \sup_{\mu\in\mathcal{P}(A(x))}\Big[\int_S q^{(n)}(dy|x,\mu,\nu_n^*(\theta,x))\varphi_\alpha^{(n)}(\theta,y)+\theta c_n(x,\mu,\nu_n^*(\theta,x))\varphi_\alpha^{(n)}(\theta,x)\Big] \\
1\le\varphi_\alpha^{(n)}(\theta,x) &\le \frac{\alpha^2 e^{\theta\hat{L}_0/\alpha}}{\alpha^2-\hat\rho_0\hat\rho_1\theta}(V(x))^{\frac{\hat\rho_1\theta}{\alpha}} \quad \forall\ (\theta,x)\in[0,1]\times S.
\end{cases}
\qquad (4.4.27)
$$

Since $c_n\ge 0$, by (4.4.24), we say $\varphi_\alpha^{(n)}(\theta,x)$ is increasing in θ. Also we know that $\varphi_\alpha^{(n)}(\theta,x)$ is differentiable a.e. with respect to $\theta\in[0,1]$. So

$$
\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x)\ge 0 \ \text{ for a.e. } \theta. \qquad (4.4.28)
$$

So, for $x \notin S_{n-1}$ and a.e. $\theta \in [0,1]$,

$$\begin{cases} & -\alpha\theta\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) + \left[\int_S q^{(n-1)}(dy|x,\mu,\nu_n^*(\theta,x))\varphi_\alpha^{(n)}(\theta,y) + \theta c_{n-1}(x,\mu,\nu_n^*(\theta,x))\varphi_\alpha^{(n)}(\theta,x)\right] \\ & = -\alpha\theta\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) \le 0. \end{cases} \tag{4.4.29}$$

Now, by (4.4.1), (4.4.2), (4.4.26), and (4.4.27), for all $x \in S_{n-1}$ and a.e. $\theta \in [0,1]$, we have

$$-\alpha\theta\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) + \left[\int_S q^{(n-1)}(dy|x,\mu,\nu_n^*(\theta,x))\varphi_\alpha^{(n)}(\theta,y) + \theta c_{n-1}(x,\mu,\nu_n^*(\theta,x))\varphi_\alpha^{(n)}(\theta,x)\right] \le 0. \tag{4.4.30}$$

So, for all $x \in S$, by (4.4.29) and (4.4.30), we have

$$-\alpha\theta\frac{\partial\varphi_\alpha^{(n)}}{\partial\theta}(\theta,x) + \left[\int_S q^{(n-1)}(dy|x,\mu,\nu_n^*(\theta,x))\varphi_\alpha^{(n)}(\theta,y) + \theta c_{n-1}(x,\mu,\nu_n^*(\theta,x))\varphi_\alpha^{(n)}(\theta,x)\right] \le 0. \tag{4.4.31}$$

Let

$$\pi_n^{*2} : \mathbb{R}_+ \times S \to \mathcal{P}(B)$$

be defined by $\pi_n^{*2}(x,t) = \nu_n^*(\theta e^{-\alpha t}, x)$. Here $\pi_n^{*2} \in \Pi_M^2$. Now, for any $\pi^1 \in \Pi_{Ad}^1$, by Feynman-Kac formula, we get

$$E_x^{\pi^1,\pi_n^{*2}}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c_{n-1}(\xi_t^{(n-1)},\pi^1(t),\pi_n^{*2}(\xi_{t-},t))dt\right)\right] \le \varphi_\alpha^{(n)}(\theta,x).$$

Since $\pi^1 \in \Pi_{Ad}^1$ is arbitrary

$$\sup_{\pi^1\in\Pi_{Ad}^1} E_x^{\pi^1,\pi_n^{*2}}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c_{n-1}(\xi_t^{(n-1)},\pi^1(t),\pi_n^{*2}(\xi_{t-},t))dt\right)\right] \le \varphi_\alpha^{(n)}(\theta,x). \tag{4.4.32}$$

Also using (4.4.11) and Feynman-Kac formula (see, Chapter 2 or [53, Theorem 3.1]), we have

$$\begin{aligned} \varphi_\alpha^{(n-1)}(\theta,x) &= \inf_{\pi^2\in\Pi_{Ad}^2}\sup_{\pi^1\in\Pi_{Ad}^1} E_x^{\pi^1,\pi^2}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c_{n-1}(\xi_t^{(n-1)},\pi^1(t),\pi^2(t))dt\right)\right] \\ &\le \sup_{\pi^1\in\Pi_{Ad}^1} E_x^{\pi^1,\pi_n^{*2}}\left[\exp\left(\theta\int_0^\infty e^{-\alpha t}c_{n-1}(\xi_t^{(n-1)},\pi^1(t),\pi_n^{*2}(\xi_{t-},t))dt\right)\right]. \end{aligned} \tag{4.4.33}$$

By (4.4.32) and (4.4.33), we have $\varphi_\alpha^{(n-1)}(\theta,x) \le \varphi_\alpha^{(n)}(\theta,x)$.

Hence $\varphi_\alpha^{(n)}(\theta,x)$ is increasing in n for any $(\theta,x) \in [0,1]\times S$. So, there exists a function φ_α on $[0,1]\times S$ that is continuous with respect to $\theta \in [0,1]$, such that along a subsequence $n_k \to \infty$, we have $\lim_{n_k \to \infty} \varphi_\alpha^{(n_k)}(\theta, x) = \varphi_\alpha(\theta, x)$. Also, for each fixed $x \in S$, this convergence is uniform in $\theta \in [0,1]$. Now, by (4.4.15), we have

$$1 \leq \varphi_\alpha(\theta, x) \leq \frac{\alpha^2 e^{\theta \hat{L}_0/\alpha}}{\alpha^2 - \hat{\rho}_0 \hat{\rho}_1 \theta} V^{\frac{\hat{\rho}_1 \theta}{\alpha}}(x). \tag{4.4.34}$$

As in the proof of equation (4.4.11) in step 2 (starting from the first equality of (4.4.19)), we say that φ_α is a solution to the HJI equation (4.3.1). Also by (4.4.34), we conclude that $\varphi_\alpha \in L_V^\infty([0,1] \times S)$. Finally, the uniqueness of $\varphi_\alpha(\theta, x)$ follows from the stochastic representation in Theorem 4.3.1. □

4.5 The existence of saddle-point equilibrium

In this section, we prove our main result of this chapter, namely, the existence of a saddle-point equilibrium for the α discounted risk-sensitive zero-sum game problem.

Theorem 4.5.1. *Grant Assumptions (A0)(a), (A0)(b), (A1)-(A3). Then, the following assertions hold.*

1. *There exist a solution $\varphi_\alpha \in L_V^\infty([0,1] \times S)$ and a pair of measurable mappings (f^{*1}, f^{*2}) satisfying*

$$\begin{aligned}
\alpha\theta \frac{\partial \varphi_\alpha}{\partial \theta}(\theta, x) &= \sup_{\mu \in \mathcal{P}(A(x))} \inf_{\nu \in \mathcal{P}(B(x))} \left[\int_S q(dy|x, \mu, \nu) \varphi_\alpha(\theta, y) + \theta c(x, \mu, \nu) \varphi_\alpha(\theta, x) \right] \\
&= \inf_{\nu \in \mathcal{P}(B(x))} \sup_{\mu \in \mathcal{P}(A(x))} \left[\int_S q(dy|x, \mu, \nu) \varphi_\alpha(\theta, y) + \theta c(x, \mu, \nu) \varphi_\alpha(\theta, x) \right] \\
&= \sup_{\mu \in \mathcal{P}(A(x))} \left[\int_S q(dy|x, \mu, f^{*2}) \varphi_\alpha(\theta, y) + \theta c(x, \mu, f^{*2}) \varphi_\alpha(\theta, x) \right] \\
&= \inf_{\nu \in \mathcal{P}(B(x))} \left[\int_S q(dy|x, f^{*1}, \nu) \varphi_\alpha(\theta, y) + \theta c(x, f^{*1}, \nu) \varphi_\alpha(\theta, x) \right]
\end{aligned}$$

$$a.e.\ \theta \in [0,1]. \tag{4.5.1}$$

2. *The solution of (4.5.1) is unique and has the following probabilistic representation*

$$\begin{aligned}
1 \leq \varphi_\alpha(\theta, x) &= \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} E_x^{\pi^1, \pi^2} \left[\exp\left(\theta \int_0^\infty e^{-\alpha t} c(\xi_t, \pi^1(t), \pi^2(t)) dt \right) \right] \\
&= \inf_{\pi^2 \in \Pi^2_{Ad}} \sup_{\pi^1 \in \Pi^1_{Ad}} E_x^{\pi^1, \pi^2} \left[\exp\left(\theta \int_0^\infty e^{-\alpha t} c(\xi_t, \pi^1(t), \pi^2(t)) dt \right) \right] \\
&\leq \frac{\alpha^2 e^{\theta \hat{L}_0/\alpha}}{\alpha^2 - \hat{\rho}_0 \hat{\rho}_1 \theta} V^{\frac{\hat{\rho}_1 \theta}{\alpha}}(x).
\end{aligned} \tag{4.5.2}$$

3. *There exists an optimal Markov saddle-point equilibrium point $(\hat{\pi}^{*1}, \hat{\pi}^{*2}) \in \Pi^1_M \times \Pi^1_M$ for the cost criterion (4.2.4).*

Proof. (1) By Theorem 4.4.1, there exists a solution $\varphi_\alpha \in L^\infty_V([0,1] \times S)$ of the HJI equation (4.5.1). Now by measurable selection theorem in [89, Theorem 2.2], there exists a pair of measurable mappings (f^{*1}, f^{*2}) satisfying (4.5.1).

(2) By the verification theorem in Theorem 4.3.1, we conclude that the solution of (4.5.1) is unique and has the stochastic representation given by (4.5.2).

(3) Let the pair of strategies (f^{*1}, f^{*2}) satisfies (4.5.1). Now let us define

$$\hat{\pi}^{*1} : \mathbb{R}_+ \times S \to \mathcal{P}(A)$$

and

$$\hat{\pi}^{*2} : \mathbb{R}_+ \times S \to \mathcal{P}(B)$$

where

$$\hat{\pi}^{*k}(x,t) := f^{*k}(\theta e^{-\alpha t}, x),$$

for $k = 1, 2$. Then from the proof of Theorem 4.3.1, we have

$$\sup_{\pi^1 \in \Pi^1_{Ad}} \hat{J}_\alpha(x, \theta, \pi^1, \hat{\pi}^{*2}) = \inf_{\pi^2 \in \Pi^2_{Ad}} \hat{J}_\alpha(x, \theta, \hat{\pi}^{*1}, \pi^2) = \varphi_\alpha(\theta, x).$$

Using this, (4.2.4), (4.2.5) and part (2), we have

$$\sup_{\pi^1 \in \Pi^1_{Ad}} J_\alpha(x, \theta, \pi^1, \hat{\pi}^{*2}) = \inf_{\pi^2 \in \Pi^2_{Ad}} J_\alpha(x, \theta, \hat{\pi}^{*1}, \pi^2) = J^*_\alpha(\theta, x) = \frac{1}{\theta} \log \varphi_\alpha(\theta, x).$$

Hence $(\hat{\pi}^{*1}, \hat{\pi}^{*2})$ is a saddle-point equilibrium for the cost criterion (4.2.4). □

Remark 4.5.1. *It is obvious to note that for the pair of strategies* $(\hat{\pi}^{*1}, \hat{\pi}^{*2}) \in \Pi^1_M \times \Pi^2_M$ *as obtained in Theorem 4.5.1,* $U(x) = L(x) = J^*_\alpha(\theta, x) = \frac{1}{\theta} \log \varphi_\alpha(\theta, x)$. *Hence we conclude that the value of the game exists.*

When the transition and cost rates are bounded, the existence of a saddle-point equilibrium is ensured by Theorem 4.5.1.

Corollary 4.5.2. *Grant Assumption* $(A3)$*((i)-(ii)). Also, assume that the transition and cost rates are bounded. Then, the HJI equation (4.3.1) has a unique solution* φ_α *satisfying (4.5.1) and there exists a saddle-point equilibrium.*

Proof. Suppose there exist constant P_1 and P_2, such that $\sup_{(x,a,b) \in K} q_x(a,b) \leq P_1$ and $\sup_{(x,a,b) \in K} c(x,a,b) \leq P_2$. First we take the Lyapunov function $V(x) \equiv P_3$, for all $x \in S$, $P_3 \geq 1$, a constant. Now $\int_S V(y) q(dy|x,a,b) = \int_S V^2(y) q(dy|x,a,b) = 0$, for all $(x,a,b) \in K$. Now, take $\hat{\rho}_0 = \alpha$, $M_0 = P_1$, any real number, $\hat{\rho}_1 \in (0, \alpha)$, and $\hat{L}_0 = P_2$. Then Assumption $(A1)$ is verified. Now, for any constants $\hat{\rho}_2 \in (0, \alpha)$, $b_1 \in (0, \infty)$, Assumption $(A2)$ holds. Also $\int_S V(y) q(dy|x,a,b)$ is continuous in $(a,b) \in A(x) \times B(x)$. So, Assumption $(A3)$ is also true. Then, by Theorem 4.5.1, we have a unique solution φ_α satisfying (4.5.1), and there exists a saddle-point equilibrium. □

4.6 Example

In this section, we verify our assumptions with an example, in which the transition rate is unbounded and the cost rate is nonnegative and unbounded.

Example 4.6.1. *Let us consider a zero-sum game model as*

$$\hat{\mathbb{M}} := \{S, (A, A(x), x \in S), (B, B(x), x \in S), c(x, a, b), q(dy|x, a, b)\}.$$

Suppose our state space is $S = (-\infty, \infty)$ *and transition rate is*

$$q(\hat{D}|x, a, b) = \hat{\lambda}(x, a, b)\left[\int_{y \in \hat{D}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-x)^2}{2\sigma^2}} dy - \delta_x(\hat{D})\right], \tag{4.6.1}$$

where $x \in S, \hat{D} \in \mathscr{B}(S), (a, b) \in A(x) \times B(x)$*. For the existence of a saddle-point equilibrium for our model, we consider the following conditions.*

(I) $A(x)$ *and* $B(x)$ *are compact subsets of the Borel spaces* A *and* B*, respectively, for each fixed* $x \in S$*.*

(II) Suppose there exists a positive constant M_0 *such that* $M_0 < \frac{\alpha}{6\sigma^2(\sigma^2+1)}$*. Also, assume that for each fixed* $x \in S$*,* $\hat{\lambda}(x, a, b)$ *is continuous in* $(a, b) \in A(x) \times B(x)$*, and satisfies* $0 < \sup_{(a,b) \in A(x) \times B(x)} \hat{\lambda}(x, a, b) \leq M_0(x^2 + 1)$*.*

(III) Consider that for each $x \in S$*, the reward rate function* $c(x, a, b)$ *for player 1 (or cost rate for player 2) is nonnegative and continuous in* $(a, b) \in A(x) \times B(x)$*. Also, assume that there exist constants* $\hat{L}_0 \geq 0$ *and* $0 < \hat{\rho}_1 < \min\{\alpha, \frac{\alpha^2}{M_0\sigma^2}\}$ *such that*

$$\sup_{(a,b) \in A(x) \times B(x)} c(x, a, b) \leq \hat{\rho}_1 \log(1 + x^2) + \hat{L}_0.$$

Proposition 4.6.2. *Under conditions (I)-(III), the above controlled system satisfies the Assumptions* (A0)(a)*,* (A0)(b)*,* (A1)*-*(A3)*. Hence by Theorem 4.5.1, there exists a saddle-point equilibrium for our model.*

Proof. We first recall our known results: for all $k = 0, 1 \cdots .,$ we have $(1) : \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (y - x)^{2k+1} e^{-\frac{(y-x)^2}{2\sigma^2}} dy = 0$; $(2) : \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (y - x)^{2k} e^{-\frac{(y-x)^2}{2\sigma^2}} dy = 1 \cdot 3 \cdots (2k - 1)\sigma^{2k}$.
Let us consider a Lyapunov function $V(x) = x^2 + 1$. Then for all $x \in S$, $V(x) \geq 1$. Now, we see for any $x \in S$, $(a, b) \in A(x) \times B(x)$,

(i)

$$\int_S V(y) q(dy|x, a, b) = \hat{\lambda}(x, a, b)\left[\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (y^2 + 1) e^{-\frac{(y-x)^2}{2\sigma^2}} dy - (x^2 + 1)\right]$$

$$
\begin{aligned}
&= \hat{\lambda}(x,a,b)\sigma^2 \\
&\leq M_0\sigma^2 V(x) = \hat{\rho}_0 V(x),
\end{aligned} \tag{4.6.2}
$$

where $\hat{\rho}_0 = M_0\sigma^2$.

(ii)

$$
\begin{aligned}
q^*(x) &= \sup_{(a,b)\in A(x)\times B(x)} q_x(a,b) = \sup_{(a,b)\in A(x)\times B(x)} -q(\{x\}|x,a,b) \\
&= \sup_{(a,b)\in A(x)\times B(x)} \hat{\lambda}(x,a,b) \leq M_0(x^2+1) = M_0V(x).
\end{aligned}
$$

(iii) By condition (III), we have

$$
\sup_{(a,b)\in A(x)\times B(x)} c(x,a,b) \leq \hat{\rho}_1 \log(1+x^2) + \hat{L}_0,
$$

where, $0 < \hat{\rho}_1 < \min\{\alpha, \hat{\rho}_0^{-1}\alpha^2\}$.

Hence Assumption $(A1)$ is verified.

For all $x \in S$, $V^2(x) \geq 1$. Now for any $x \in S$, $(a,b) \in A(x) \times B(x)$, we have

(i)

$$
\begin{aligned}
\int_S V^2(y)q(dy|x,a,b) &= \hat{\lambda}(x,a,b)\left[\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(y^2+1)^2 e^{-\frac{(y-x)^2}{2\sigma^2}}dy - (x^2+1)^2\right] \\
&= \lambda(x,a,b)[1\cdot 3\sigma^4 + \sigma^2(2+6x^2)] \\
&\leq M_0(x^2+1)[3\sigma^4 + \sigma^2(2+6x^2)] \\
&= M_0\sigma^2(x^2+1)(3\sigma^2+2+6x^2) \\
&\leq 6M_0\sigma^2(x^2+1)(x^2+1)(\sigma^2+1) \\
&= 6M_0V^2(x)\sigma^2(\sigma^2+1) \\
&\leq \rho_2 V^2(x) + 1
\end{aligned}
$$

where $\hat{\rho}_2 = 6M_0\sigma^2(\sigma^2+1)$, and $b_0 \in [1,\infty)$. Then by condition (II), we have $0 < \hat{\rho}_2 < \alpha$. Hence, Assumption $(A2)$ is verified.

(i) By eq. (4.6.1) and conditions (II) and (III), $q(\cdot|x,a,b)$ and $c(x,a,b)$ are continuous in $(a,b) \in A(x) \times B(x)$.

(ii) By (4.6.2), $\int_S V(y)q(dy|x,a,b)$ is continuous in $(a,b) \in A(x) \times B(x)$.

Hence Assumption $(A3)$ is also verified. So, by Theorem 4.5.1, we say that there exists a saddle-point equilibrium for our model. □

CHAPTER 5

Discrete-time zero-sum games for Markov chains with risk-sensitive average cost criterion

5.1 Introduction

We address a risk-sensitive discrete-time zero-sum game with a long-run (or ergodic) cost criterion where the underlying state dynamics are given by a controlled Markov process determined by a prescribed transition kernel. The state space is a denumerable/compact set, actions spaces are Borel spaces and the cost function is possibly unbounded for the countable state space case and for the compact state space case, it is a real-valued and bounded function.

The analysis of stochastic systems with the risk-sensitive ergodic criterion can be traced back to the seminal papers by Jacobson in [73] and Howard and Matheson in [71]. The literature on risk-sensitive MDP under ergodic cost criterion is quite extensive, e.g., [2], [15], [21], [24], [25], [26], [30], [31], [47], [52], [67], [71], [81], [82], [94], [111], [114]. In [25], the authors studied risk-sensitive ergodic cost criterion for discrete-time Markov decision processes (DTMDPs) with bounded cost using a simultaneous Doeblin condition on a countable state space. Also, see [2], [24], and the references therein for multiplicative ergodic theory. These papers used the eigenvalue approach to study risk-sensitive ergodic control problems. The authors in [47], [94] used the results of [78], [79] to study their risk-sensitive ergodic control problems. Also, in the context of controlled diffusions, the eigenvalue approach is used in [5], [6], [7], [19] to study the risk-sensitive ergodic control problems. The articles [12], [16] address zero-sum risk-sensitive stochastic games for discrete-time Markov decision processes with discounted as well as ergodic cost criteria. The analysis of the ergodic cost criterion in [12] is carried out using vanishing discount asymptotics. The results of the article [12] are extended to the general state space case in [16]. In [16], the ergodic cost criterion is studied under a local minorization property and a Lyapunov condition. The analogous results in continuous-time setup are carried out in [45]. The corresponding nonzero-sum risk-sensitive ergodic stochastic games for discrete-time Markov decision processes are studied in [13], [113].

In this chapter, we study the stochastic game problems for ergodic cost criterion by analyzing the principal eigenpair of the associated Shapley equation. The analysis of our ergodic game problems is inspired by the work of [2] and [21]. In [21], the authors studied risk-sensitive discrete/continuous-time ergodic control problems for controlled Markov processes with countable state space . They established the existence of a principal eigenpair of the associated ergodic HJB equation. For this, they first studied the corresponding Dirichlet eigenvalue problems on finite sets and then pass to the limit by increasing the finite sets to countable state space. In [2], authors used a novel technique to provide a variational formula for infinite-horizon risk-sensitive reward on a compact state and action spaces. They build a nonlinear version of Kreĭn-Rutman theorem to study the corresponding ergodic HJB equation which leads to the existence of optimal ergodic control. We have extended the results of [21] to a discrete-time zero-sum game in this chapter. Also, in this chapter, we have analyzed the same problem (zero-sum game problem) for bounded cost on compact state space by getting inspired by the work of [2]. Moreover, this chapter can be seen as an extension of the results of [12] to the case with unbounded cost. This is carried out under a certain Lyapunov type stability condition. Also, we have extended the results of [12] to a compact state space case.

Under a certain condition, in this chapter, using a nonlinear version of the Kreĭn-Rutman theorem, we establish the existence of a principal eigenpair to the associated Shapley equations for both countable/compact state space cases and show that the principal eigenvalues are the values of the corresponding games. Also, we establish the existence of a saddle-point equilibrium via the outer maximizing/minimizing selectors of the associated Shapley equations. Additionally, we give a complete characterization of all possible saddle-point strategies in the space of stationary Markov strategies.

The rest of this chapter is arranged as follows. Section 5.2 deals with problem descriptions and preliminaries. In Section 5.3, we study Dirichlet eigenvalue problems. In Section 5.4, we show that the risk-sensitive optimality equation (i.e., Shapley equation) has a solution, obtain the value of the game and saddle-point equilibrium in the class of stationary Markov strategies. We also completely characterize all possible saddle-point strategies in the class of stationary strategies in this section. In Section 5.5, we present an illustrative example. In the next section, we study the same problem on compact state space. Section 5.7 concludes the chapter with some concluding remarks. The content of this chapter is based on the published article [41].

5.2 The game model

In this section, we introduce a discrete-time zero-sum stochastic game model which consists of the following elements

$$\{S, A, B, (A(i) \subset A, B(i) \subset B, i \in S), P(\cdot|i, a, b), c(i, a, b)\}. \tag{5.2.1}$$

- $S := \{0, 1, \cdots\}$, called the state space, endowed with the discrete topology of our controlled Markov processes $\boldsymbol{\xi} := \{\xi_0, \xi_1, \cdots\}$.
- A and B are action spaces for players 1 and 2, respectively. The action spaces A and B are assumed to be Borel spaces with the Borel σ-algebras $\mathcal{B}(A)$ and $\mathcal{B}(B)$, respectively.
- For each $i \in S$, $A(i) \in \mathcal{B}(A)$ and $B(i) \in \mathcal{B}(B)$ denote the sets of admissible actions for players 1 and 2, respectively when the system is at state i. For any metric space Y, let $\mathcal{P}(Y)$ denote the space of probability measures on $\mathcal{B}(Y)$ with Prohorov topology.
- Next $P : K \to \mathcal{P}(S)$ is a transition (stochastic) kernel, where $K := \{(i, a, b)|i \in S, a \in A(i), b \in B(i)\}$, a Borel subset of $S \times A \times B$. We assume that the function $P(j|i, a, b)$ is continuous in $(a, b) \in A(i) \times B(i)$ for any fixed $i, j \in S$.
- Finally, the function $c : K \to \mathbb{R}_+$ denotes the cost function which is assumed to be continuous in $(a, b) \in A(i) \times B(i)$ for any fixed $i \in S$.

The game evolves as follows. When the state $i \in S$ at time $t \in \mathbb{N}_0 := \{0, 1, \cdots\}$, players independently choose actions $a_t \in A(i)$ and $b_t \in B(i)$ according to some strategies, respectively. As a consequence of this, the following happens:

- player 1 incurs an immediate cost $c(i, a_t, b_t)$ and player 2 receives a reward $c(i, a_t, b_t)$;
- the system moves to a new state $j \neq i$ with the probability determined by $P(j|i, a_t, b_t)$.

When the state of the system transits to a new state j, the above procedure repeats. Both the players have full information of past and present states and past actions of both players. The goal of player 1 is to maximize his/her accumulated costs, whereas that of player 2 is to minimize the same with respect to some performance criterion $\mathscr{J}^{\cdot,\cdot}(\cdot, \cdot)$, which in our present case is defined by (5.2.3), below. At each stage, the players choose their actions on the basis of accumulated information. The available information for decision making at time $t \in \mathbb{N}_0$, i.e., the history of the process up to time t is given by

$$h_t := (i_0', (a_0, b_0), i_1', (a_1, b_1), \cdots, i_{t-1}', (a_{t-1}, b_{t-1}), i_t'),$$

where $H_0 = S$, $H_t = H_{t-1} \times (A \times B \times S), \cdots, H_\infty = (A \times B \times S)^\infty$ are the history spaces. An admissible strategy for player 1 is a sequence $\pi^1 := \{\pi^1(t) : H_t \to \mathcal{P}(A)\}_{t \in \mathbb{N}_0}$ of stochastic kernels satisfying $\pi^1(A(\xi_t)|h_t, t) = 1$, for all $h_t \in H_t$; $t \geq 0$, where $\{\xi_t\}$ is the state process. The set of all such strategies for player 1 is denoted by Π_{Ad}^1. A strategy for player 1 is called a Markov strategy if

$$\pi^1(t)(h_{t-1}, a, b, i) = \pi^1(t)(h_{t-1}', a', b', i)$$

i.e., $\pi^1(t)(\cdot|h_{t-1}, a, b, i) = \pi^1(t)(\cdot|h_{t-1}', a', b', i)$ for all $h_{t-1}, h_{t-1}' \in H_{t-1}, a, a' \in A, b, b' \in B, i \in S, t \in \mathbb{N}_0$. Thus a Markov strategy for player 1 can be identified with a sequence of maps, denoted by $\pi^1 \equiv \{\pi^1(t) : S \to \mathcal{P}(A)\}_{t \in \mathbb{N}_0}$. A Markov strategy $\{\pi^1(t)\}$ is called stationary Markov for player 1, if it does not have any explicit time dependence, i.e., $\pi^1(t)(h_t) = \tilde{\phi}(i_t')$ i.e., $\pi^1(\cdot|h_t, t) = \tilde{\phi}(\cdot|i_t')$ for all $h_t \in H_t$ for some mapping $\tilde{\phi}$ satisfying $\tilde{\phi}(A(i)|i) = 1$ for all $i \in S$. The sets of all Markov strategies and all stationary Markov strategies for player 1, are denoted by Π_M^1 and Π_{SM}^1, respectively. Similarly, the set of all admissible strategies, Markov strategies, and stationary Markov strategies for player 2 are defined similarly and denoted by Π_{Ad}^2, Π_M^2, and Π_{SM}^2, respectively. For each $i, j \in S$, $\mu \in \mathcal{P}(A(i))$ and $\nu \in \mathcal{P}(B(i))$, the cost function c and the transition kernel P are extended as follows:

$$\begin{aligned} c(i, \mu, \nu) &:= \int_{B(i)} \int_{A(i)} c(i, a, b) \mu(da) \nu(db), \\ P(j|i, \mu, \nu) &:= \int_{B(i)} \int_{A(i)} P(j|i, a, b) \mu(da) \nu(db), \end{aligned}$$

(by abuse of notation we use the same notation c and P). For a given initial distribution $\tilde{\pi}_0 \in \mathcal{P}(S)$ and a pair of strategies $(\pi^1, \pi^2) \in \Pi_{Ad}^1 \times \Pi_{Ad}^2$, by Tulcea's Theorem (see Proposition 7.28 of [17]), there exists unique probability measure $P_{\tilde{\pi}_0}^{\pi^1, \pi^2}$ on $(\Omega, \mathcal{B}(\Omega))$, where $\Omega = (S \times A \times B)^\infty$. When $\tilde{\pi}_0 = \delta_i$, $i \in S$ this probability measure is simply written by $P_i^{\pi^1, \pi^2}$ satisfying

$$P_i^{\pi^1, \pi^2}(\xi_0 = i) = 1 \text{ and } P_i^{\pi^1, \pi^2}(\xi_{t+1} \in D | H_t, \pi^1(t), \pi^2(t)) = P(D|\xi_t, \pi^1(t), \pi^2(t)) \; \forall \; D \in \mathcal{B}(S). \tag{5.2.2}$$

Let $E_i^{\pi^1, \pi^2}$ denote the expectation with respect to the probability measure $P_i^{\pi^1, \pi^2}$. Now from [68, p. 6], we know that under any $(\pi^1, \pi^2) \in \Pi_M^1 \times \Pi_M^2$, the corresponding stochastic process $\{\xi_t\}$ is strong Markov.

We now introduce some useful notations.

Notations:
For any finite set $\mathscr{D} \subset S$, we define $\mathcal{B}_{\mathscr{D}} = \{f : S \to \mathbb{R} \mid f$ is Borel measurable and $f(i) = 0 \ \forall\, i \in \mathscr{D}^c\}$, $\mathcal{B}^+_{\mathscr{D}} \subset \mathcal{B}_{\mathscr{D}}$ denotes the cone of all nonnegative functions vanishing outside $\mathscr{D}$. Given any real-valued function $\mathcal{V} \geq 1$ on S, we define a Banach space $(L^\infty_{\mathcal{V}}(S), \|\cdot\|^\infty_{\mathcal{V}})$ of $\mathcal{V}$-weighted functions by

$$L^\infty_{\mathcal{V}}(S) = \left\{ f : S \to \mathbb{R} \mid \|f\|^\infty_{\mathcal{V}} := \sup_{i\in S} \frac{|f(i)|}{\mathcal{V}(i)} < \infty \right\}.$$

For any ordered Banach space $\tilde{\mathcal{X}}$, a subset $\tilde{\mathcal{C}} \subset \tilde{\mathcal{X}}$ and $x, y \in \tilde{\mathcal{X}}$, we define $\succeq$ as $x \succeq y \Leftrightarrow x - y \in \tilde{\mathcal{C}}$, i.e., the partial ordering in $\tilde{\mathcal{X}}$ with respect to the cone $\tilde{\mathcal{C}}$. For any subset $\mathscr{B} \subset S$, $\tau(\mathscr{B}) = \inf\{t : \xi_t \in \mathscr{B}\}$, i.e., the first entry time of ξ_t to $\mathscr{B}$. Also, for any subset $\mathscr{D} \subset S$, $\tau(\mathscr{D}) := \inf\{t > 0 : \xi_t \notin \mathscr{D}\}$ denotes the first exit time from $\mathscr{D}$.

We now introduce the cost evaluation criterion.

Ergodic cost criterion: Now we define the risk-sensitive ergodic cost criterion for zero-sum discrete-time games. Let $\theta > 0$ be the risk-sensitive parameter. For each $i \in S$ and any $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$, the risk-sensitive ergodic cost criterion is given by

$$\mathscr{J}(i, c, \pi^1, \pi^2) := \limsup_{T\to\infty} \frac{1}{T} \log E^{\pi^1,\pi^2}_i \left[e^{\theta \sum_{t=0}^{T-1} c(\xi_t, \pi^1(t), \pi^2(t))} \right]. \tag{5.2.3}$$

Since the risk-sensitive parameter remains the same throughout, we assume without loss of generality that $\theta = 1$. The lower value and upper value of the game, are functions on S, defined as
$L(i) := \sup_{\pi^1\in\Pi^1_{Ad}} \inf_{\pi^2\in\Pi^2_{Ad}} \mathscr{J}(i, c, \pi^1, \pi^2)$ and $U(i) := \inf_{\pi^2\in\Pi^2_{Ad}} \sup_{\pi^1\in\Pi^1_{Ad}} \mathscr{J}(i, c, \pi^1, \pi^2)$ respectively. It is easy to see that

$$L(i) \leq U(i) \text{ for all } i \in S.$$

If $L(i) = U(i)$ for all $i \in S$, then the common function is called the value of the game and is denoted by $\mathscr{J}^*(i)$. A strategy π^{*1} in Π^1_{Ad} is said to be optimal for player 1 if

$$\mathscr{J}(i, c, \pi^{*1}, \pi^2) \geq \inf_{\pi^2\in\Pi^2_{Ad}} \sup_{\pi^1\in\Pi^1_{Ad}} \mathscr{J}(i, c, \pi^1, \pi^2) = U(i) \ \forall\, i \in S, \ \forall \pi^2 \in \Pi^2_{Ad}.$$

Similarly, $\pi^{*2} \in \Pi^2_{Ad}$ is optimal for player 2 if

$$\mathscr{J}(i, c, \pi^1, \pi^{*2}) \leq \sup_{\pi^1\in\Pi^1_{Ad}} \inf_{\pi^2\in\Pi^2_{Ad}} \mathscr{J}(i, c, \pi^1, \pi^2) = L(i) \ \forall\, i \in S, \ \forall \pi^1 \in \Pi^1_{Ad}.$$

If $\pi^{*k} \in \Pi^k_{Ad}$ is optimal for player k (k=1,2), then (π^{*1}, π^{*2}) is called a pair of optimal strategies. The pair of strategies (π^{*1}, π^{*2}) at which this value is attained i.e., if

$$\mathscr{J}(i, c, \pi^1, \pi^{*2}) \leq \mathscr{J}(i, c, \pi^{*1}, \pi^{*2}) \leq \mathscr{J}(i, c, \pi^{*1}, \pi^2), \ \ \forall \pi^1 \in \Pi^1_{Ad}, \ \forall \pi^2 \in \Pi^2_{Ad},$$

then the pair (π^{*1}, π^{*2}) is called a saddle-point equilibrium, and then π^{*1} and π^{*2} are optimal for player 1 and player 2, respectively.

Following [12], the Shapley equation for the above problem is given by

$$\begin{aligned} e^{\rho}\psi(i) &= \sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\Big[e^{c(i,\mu,\nu)}\sum_{j\in S}\psi(j)P(j|i,\mu,\nu)\Big] \\ &= \inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\Big[e^{c(i,\mu,\nu)}\sum_{j\in S}\psi(j)P(j|i,\mu,\nu)\Big],\ i\in S. \end{aligned}$$

In the above equation, ρ is a scalar and ψ is an appropriate function.

Our goal is to establish the existence of a saddle-point equilibrium among the class of admissible history-dependent strategies and provide its complete characterization. We now describe briefly our technique for establishing the existence of a saddle-point equilibrium. We first construct an increasing sequence of bounded subsets of the state space S. Then we apply Kreĭn-Rutman theorem [4] on each bounded subset to obtain a bounded solution of the corresponding Dirichlet eigenvalue problem, i.e., a solution to the above equation on each finite subset with the condition that the solution is zero in the complement of that subset. Using a suitable Lyapunov stability condition (to be stated shortly), we pass to the limit and show that the risk-sensitive zero-sum ergodic optimality equation admits a principal eigenpair. Subsequently, we establish a stochastic representation of the principal eigenfunction. This enables us to characterize all possible saddle-point equilibria in the space of stationary Markov strategies. To this end, we make certain assumptions. First, we define a norm-like function which is used in our assumptions.

Definition 5.2.1. *A function $f : S \to \mathbb{R}$ is said to be norm-like if for every $k \in \mathbb{R}$, the set $\{i \in S : f(i) \leq k\}$ is either empty or finite.*

Since the cost function (i.e., $c(i, a, b)$) may be unbounded, to guarantee the finiteness of $\mathscr{J}(i, c, \pi^1, \pi^2)$, we use the following assumption.

(*A*1) We assume that the Markov chain $\{\xi_t\}_{t\geq 0}$ is irreducible under every pair of stationary Markov strategies $(\pi^1, \pi^2) \in \Pi^1_{SM} \times \Pi^2_{SM}$. Also, assume that there exists a constant $\tilde{C} > 0$, a real-valued function $V \geq 1$ on S and, a finite set $\mathscr{K}$ such that one of the following hold.

(a) **If the running cost is bounded:** For some positive constant $\gamma > \|c\|_\infty$, we have the following blanket stability condition

$$\sup_{(a,b)\in A(i)\times B(i)}\sum_{j\in S}V(j)P(j|i,a,b) \leq \tilde{C}I_{\mathscr{K}}(i) + e^{-\gamma}V(i)\ \forall i \in S, \tag{5.2.4}$$

where $\|c\|_\infty := \sup_{(i,a,b)\in K} c(i,a,b)$.

(b) **If the running cost is unbounded:** For some real-valued nonnegative norm-like function ℓ on S it holds that

$$\sup_{(a,b)\in A(i)\times B(i)} \sum_{j\in S} V(j)P(j|i,a,b) \leq \tilde{C} I_{\mathscr{K}}(i) + e^{-\ell(i)} V(i) \ \forall i \in S, \tag{5.2.5}$$

where the function $\ell(\cdot) - \max_{(a,b)\in A(\cdot)\times B(\cdot)} c(\cdot,a,b)$ is norm-like.

Assumption $(A1)$ and its variants are key conditions of the standard ergodicity hypothesis, see [21], [59], [87]. In this context, [25] used the Doeblin condition, a stronger assumption than a variant of Assumption $(A1)$(a) to study ergodic control problems. The condition (5.2.5) plays important role in studing the ergodic optimal control problems with unbounded running cost. We show that, (5.2.5) implies (5.2.3) is finite. A similar condition is also used in [10, Theorem 1.2], [24, Theorem 2.2] in the study of multiplicative ergodicity. Also, we refer [27], [28], [29], [113] to see the importance of Lyapunov stability assumption in studying stochastic control problem .

Let $i_0 \in S$ be a fixed state, we call it the reference state. Consider an increasing sequence of finite subsets $\mathscr{D}_n \subset S$ such that $\cup_{n=1}^\infty \mathscr{D}_n = S$ and $i_0 \in \mathscr{D}_n$ for all $n \in \mathbb{N}$. Recall that $\tau(\mathscr{D}_n) := \inf\{t > 0 : \xi_t \notin \mathscr{D}_n\}$, is the first exit time from $\mathscr{D}_n$. For our game problem, we wish to establish the existence of a saddle-point equilibrium in the space of stationary Markov strategies. To ensure the existence of saddle-point equilibrium, we make the following assumptions.

(*A*2)

(i) The admissible action spaces $A(i)(\subset A)$ and $B(i)(\subset B)$ are compact for each $i \in S$.

(ii) We assume that for any n and any pair $i, j \in \mathscr{D}_n$, the probability of hitting j from i before exiting $\mathscr{D}_n$ is bounded from below by some $\delta_{ij,n} > 0$ under all stationary Markov strategies i.e.,

$$\inf_{(\pi^1,\pi^2)\in \Pi^1_{SM}\times\Pi^2_{SM}} P_i^{\pi^1,\pi^2}(\hat{\tau}_j < \tau(\mathscr{D}_n)) \geq \delta_{ij,n}, \tag{5.2.6}$$

where $\hat{\tau}_j$ denotes the hitting time to j i.e., for any pair $i, j \in \mathscr{D}_n$, under any pair of strategies $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$, there exists $i_1, i_2, \cdots, i_m \in \mathscr{D}_n$ satisfying

$$P(j|i_m, \pi^{*1}(i_m), \pi^{*2}(i_m))P(i_m|i_{m-1}, \pi^{*1}(i_{m-1}), \pi^{*2}(i_{m-1})) \cdots P(i_1|i, \pi^{*1}(i), \pi^{*2}(i)) > 0. \tag{5.2.7}$$

(iii) $(i,a,b) \to \sum_{j\in S} V(j)P(j|i,a,b)$ is continuous in $(a,b) \in A(i) \times B(i)$, where V is the Lyapunov function defined in Assumption $(A1)$.

Remark 5.2.2. *(1) Assumption* $(A2)$*(i) and* $(A2)$*(iii) are standard continuity-compactness assumption.*

(2) Under Assumption $(A2)$*(i), for each* $i \in S$*, by in [17, Proposition 7.22, p. 130], we know that* $\mathcal{P}(A(i))$ *and* $\mathcal{P}(B(i))$ *are compact and metrizable. Note that* $\pi^1 \in \Pi^1_{SM}$ *can be identified with a map* $\pi^1 : S \to \mathcal{P}(A)$ *such that* $\pi^1(\cdot|i) \in \mathcal{P}(A(i))$ *for each* $i \in S$*. Thus, we have* $\Pi^1_{SM} = \Pi_{i\in S}\mathcal{P}(A(i))$*. Similarly,* $\Pi^2_{SM} = \Pi_{i\in S}\mathcal{P}(B(i))$*. Therefore by Tychonoff theorem, the sets* Π^1_{SM} *and* Π^2_{SM} *are compact metric spaces endowed with the product topology. Also, it is clear that these sets are convex.*

(3) Instead of using (5.2.6), we can assume $\inf_{(\pi^1,\pi^2)\in\Pi^1_{SM}\times\Pi^2_{SM}} P_i^{\pi^1,\pi^2}(\hat{\tau}_j < \tau(\mathscr{D}_n)) > 0$*. Then this weaker condition also implies that* $\psi_n > 0$*, (see Lemma 5.3.4).*

Using generalized Fatou's lemma as in [35], [69, Lemma 8.3.7], from Assumption $(A2)$ one can easily get the following result, which will be used in subsequent sections; we omit the details.

Lemma 5.2.3. *Under Assumptions* $(A1)$ *and* $(A2)$*, the functions* $\sum_{j\in S} P(j|i,\mu,\nu)f(j)$ *and* $c(i,\mu,\nu)$ *are continuous at* (μ,ν) *on* $\mathcal{P}(A(i)) \times \mathcal{P}(B(i))$ *for each fixed* $f \in L^\infty_V(S)$ *and* $i \in S$*.*

5.3 Dirichlet eigenvalue problems

We begin this section by stating a version of the nonlinear Kreĭn-Rutman theorem from [4, Section 3.1], (cf. [80]) which plays a crucial role in our analysis of the Dirichlet eigenvalue problems.

Theorem 5.3.1. *Let* $\tilde{\mathcal{X}}$ *be an ordered Banach space and* $\tilde{\mathcal{C}}$ *a nonempty closed (cone) subset of* $\tilde{\mathcal{X}}$ *satisfying* $\tilde{\mathcal{X}} = \tilde{\mathcal{C}} - \tilde{\mathcal{C}}$*. Let* $\tilde{T} : \tilde{\mathcal{X}} \to \tilde{\mathcal{X}}$ *be a 1-homogeneous, order-preserving, continuous, and compact map satisfying the property that for some nonzero* $\zeta \in \tilde{\mathcal{C}}$ *and* $\hat{N} > 0$*, we have* $\hat{N}\tilde{T}(\zeta) \succeq \zeta$*. Then there exists a nontrivial* $\hat{f} \in \tilde{\mathcal{C}}$ *and a scalar* $\tilde{\lambda} > 0$*, such that* $\tilde{T}\hat{f} = \tilde{\lambda}\hat{f}$*.*

In the following lemma, we establish a few important estimates which will play a crucial role in our analysis.

Lemma 5.3.1. *Suppose that Assumption* $(A1)$ *holds. Let* $\mathscr{B} \supset \mathscr{K}$ *be a finite subset of* S *and let* $\hat{\tau}(\mathscr{B}) = \inf\{t : \xi_t \in \mathscr{B}\}$*, be the first entry time of* ξ_t *to* $\mathscr{B}$*. Then for any pair of strategies* $(\pi^1,\pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$ *we have the following:*

(i) If Assumption (A1)(a) holds: Then

$$E_i^{\pi^1,\pi^2}\Big[e^{\gamma\hat{\tau}(\mathscr{B})}V(\xi_{\hat{\tau}(\mathscr{B})})\Big] \le V(i)\ \forall\ i\in\mathscr{B}^c. \tag{5.3.1}$$

(ii) If Assumption (A1)(b) holds:

$$E_i^{\pi^1,\pi^2}\Big[e^{\sum_{s=0}^{\hat{\tau}(\mathscr{B})-1}\ell(\xi_s)}V(\xi_{\hat{\tau}(\mathscr{B})})\Big] \le V(i)\ \forall\ i\in\mathscr{B}^c. \tag{5.3.2}$$

Proof. This result is proved in [21, Lemma 2.3] for one controller case. The proof for the two controller cases is analogous. □

Now we prove the following existence result which is useful in establishing the existence of a Dirichlet eigenpair.

Proposition 5.3.2. *Suppose Assumption (A2) holds. Take any function* $\bar{c}: K\to\mathbb{R}$ *which is continuous in* $(a,b)\in A(i)\times A(i)$ *for each fixed* $i\in S$*, satisfying the relation* $\bar{c}<-\delta$ *in* $\mathscr{D}_n$*, where* $\delta>0$ *is a constant and* $\mathscr{D}_n$ *is a finite set as described previously. Then for any* $g\in\mathcal{B}_{\mathscr{D}_n}$*, there exists a unique solution* $\varphi\in\mathcal{B}_{\mathscr{D}_n}$ *to the following nonlinear equation*

$$\begin{aligned}\varphi(i) &= \inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\Big[e^{\bar{c}(i,\mu,\nu)}\sum_{j\in S}\varphi(j)P(j|i,\mu,\nu)+g(i)\Big]\\ &= \sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\Big[e^{\bar{c}(i,\mu,\nu)}\sum_{j\in S}\varphi(j)P(j|i,\mu,\nu)+g(i)\Big]\ \forall i\in\mathscr{D}_n.\end{aligned} \tag{5.3.3}$$

Moreover, we have

$$\begin{aligned}\varphi(i) &= \inf_{\pi^2\in\Pi^2_{Ad}}\sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^2}\Big[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1}\bar{c}(\xi_s,\pi^1(s),\pi^2(s))}g(\xi_t)\Big]\\ &= \sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^1,\pi^2}\Big[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1}\bar{c}(\xi_s,\pi^1(s),\pi^2(s))}g(\xi_t)\Big]\ \forall i\in S,\end{aligned} \tag{5.3.4}$$

where $\tau(\mathscr{D}_n):=\inf\{t>0:\xi_t\notin\mathscr{D}_n\}$*, first exit time from* $\mathscr{D}_n$*.*

Proof. Let $g\in\mathcal{B}_{\mathscr{D}_n}$. Define a map $\hat{T}:\mathcal{B}_{\mathscr{D}_n}\to\mathcal{B}_{\mathscr{D}_n}$ by

$$\begin{aligned}&\sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\Big[e^{\bar{c}(i,\mu,\nu)}\sum_{j\in S}\tilde{\phi}(j)P(j|i,\mu,\nu)+g(i)\Big] = \hat{T}\tilde{\phi}(i),\ i\in\mathscr{D}_n, \tilde{\phi}\in\mathcal{B}_{\mathscr{D}_n}\\ &\quad\text{and } \hat{T}\tilde{\phi}(i)=0 \text{ for } i\in\mathscr{D}_n^c.\end{aligned} \tag{5.3.5}$$

Now, let $\tilde{\phi}_1, \tilde{\phi}_2 \in \mathcal{B}_{\mathscr{D}_n}$. Then

$$(\hat{T}\tilde{\phi}_2(i) - \hat{T}\tilde{\phi}_1(i)) \le \max_{i\in\mathscr{D}_n} \sup_{\mu\in\mathcal{P}(A(i))} \sup_{\nu\in\mathcal{P}(B(i))} e^{\bar{c}(i,\mu,\nu)} \|\tilde{\phi}_2 - \tilde{\phi}_1\|_{\mathscr{D}_n}.$$

Similarly, we have

$$(\hat{T}\tilde{\phi}_1(i) - \hat{T}\tilde{\phi}_2(i)) \le \max_{i\in\mathscr{D}_n} \sup_{\mu\in\mathcal{P}(A(i))} \sup_{\nu\in\mathcal{P}(B(i))} e^{\bar{c}(i,\mu,\nu)} \|\tilde{\phi}_2 - \tilde{\phi}_1\|_{\mathscr{D}_n}.$$

Hence

$$\|\hat{T}\tilde{\phi}_1(i) - \hat{T}\tilde{\phi}_2(i)\|_{\mathscr{D}_n} \le \max_{i\in\mathscr{D}_n} \sup_{\mu\in\mathcal{P}(A(i))} \sup_{\nu\in\mathcal{P}(B(i))} e^{\bar{c}(i,\mu,\nu)} \|\tilde{\phi}_2 - \tilde{\phi}_1\|_{\mathscr{D}_n},$$

where for any function $f \in \mathcal{B}_{\mathscr{D}_n}$, $\|f\|_{\mathscr{D}_n} = \max\{|f(i)| : i \in \mathscr{D}_n\}$. Since $\bar{c} < 0$, it is easy to see that $\max_{i\in\mathscr{D}_n} \sup_{\mu\in\mathcal{P}(A(i))} \sup_{\nu\in\mathcal{P}(B(i))} e^{\bar{c}(i,\mu,\nu)} < 1$. Hence $\hat{T}$ is a contraction map. Thus by Banach's fixed point theorem, there exists a unique $\varphi \in \mathcal{B}_{\mathscr{D}_n}$ such that $\hat{T}(\varphi) = \varphi$. Now by applying Fan's minimax theorem in [33, Theorem 3], we get

$$\sup_{\mu\in\mathcal{P}(A(i))} \inf_{\nu\in\mathcal{P}(B(i))} \Big[e^{\bar{c}(i,\mu,\nu)} \sum_{j\in S} \varphi(j) P(j|i,\mu,\nu)\Big] = \inf_{\nu\in\mathcal{P}(B(i))} \sup_{\mu\in\mathcal{P}(A(i))} \Big[e^{\bar{c}(i,\mu,\nu)} \sum_{j\in S} \varphi(j) P(j|i,\mu,\nu)\Big].$$

Hence we conclude that (5.3.3) has a unique solution. Now let $(\pi_n^{*1}, \pi_n^{*2}) \in \Pi_{SM}^1 \times \Pi_{SM}^2$ be a mini-max selector of (5.3.3), i.e.,

$$\begin{aligned} \varphi(i) &= \inf_{\nu\in\mathcal{P}(B(i))} \Big[e^{\bar{c}(i,\pi_n^{*1}(i),\nu)} \sum_{j\in S} \varphi(j) P(j|i,\pi_n^{*1}(i),\nu) + g(i)\Big] \\ &= \sup_{\mu\in\mathcal{P}(A(i))} \Big[e^{\bar{c}(i,\mu,\pi_n^{*2}(i))} \sum_{j\in S} \varphi(j) P(j|i,\mu,\pi_n^{*2}(i)) + g(i)\Big]. \end{aligned} \tag{5.3.6}$$

Now by Dynkin's formula [113, Lemma 3.1], for any $(\pi^1, \pi^2) \in \Pi_{Ad}^1 \times \Pi_{Ad}^2$ and $N \in \mathbb{N}$, we have

$$\begin{aligned} &E_i^{\pi^1,\pi^2}\Big[e^{\sum_{t=0}^{N\wedge\tau(\mathscr{D}_n)-1} \bar{c}(\xi_t,\pi^1(t),\pi^2(t))} \varphi(\xi_{N\wedge\tau(\mathscr{D}_n)})\Big] - \varphi(i) \\ &= E_i^{\pi^1,\pi^2}\Big[\sum_{t=1}^{N\wedge\tau(\mathscr{D}_n)} e^{\sum_{r=0}^{t-1} \bar{c}(\xi_r,\pi^1(r),\pi^2(r))} \Big(\sum_{j\in S} \varphi(j) P(j|\xi_{t-1},\pi^1(t-1),\pi^2(t-1)) \\ &\quad - e^{-\bar{c}(\xi_{t-1},\pi^1(t-1),\pi^2(t-1))} \varphi(\xi_{t-1})\Big)\Big]. \end{aligned} \tag{5.3.7}$$

Then, using (5.3.6) and (5.3.7), we obtain

$$E_i^{\pi^1,\pi_n^{*2}}\Big[\sum_{t=0}^{N\wedge\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1} \bar{c}(\xi_s,\pi^1(s),\pi_n^{*2}(\xi_s))} g(\xi_t)\Big]$$

$$\leq -E_i^{\pi^1,\pi_n^{*2}}\left[e^{\sum_{s=0}^{N\wedge\tau(\mathscr{D}_n)-1}\bar{c}(\xi_s,\pi^1(s),\pi_n^{*2}(\xi_s))}\varphi(\xi_{N\wedge\tau(\mathscr{D}_n)})\right]+\varphi(i).$$

Since $\bar{c}<0$ and $\varphi\in\mathcal{B}_{\mathscr{D}_n}$, taking $N\to\infty$ in the above equation and using the dominated convergence theorem, we deduce that

$$\begin{aligned}
&E_i^{\pi^1,\pi_n^{*2}}\left[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1}\bar{c}(\xi_s,\pi^1(s),\pi_n^{*2}(\xi_s))}g(\xi_t)\right]\\
&\leq -E_i^{\pi^1,\pi_n^{*2}}\left[e^{\sum_{s=0}^{\tau(\mathscr{D}_n)-1}\bar{c}(\xi_s,\pi^1(s),\pi_n^{*2}(\xi_s))}\varphi(\xi_{\tau(\mathscr{D}_n)})\right]+\varphi(i).
\end{aligned}$$

Hence

$$\varphi(i)\geq E_i^{\pi^1,\pi_n^{*2}}\left[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1}\bar{c}(\xi_s,\pi^1(s),\pi_n^{*2}(\xi_s))}g(\xi_t)\right].$$

Since $\pi^1\in\Pi_{Ad}^1$ is arbitrary,

$$\begin{aligned}
\varphi(i)&\geq \sup_{\pi^1\in\Pi_{Ad}^1} E_i^{\pi^1,\pi_n^{*2}}\left[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1}\bar{c}(\xi_s,\pi^1(s),\pi_n^{*2}(\xi_s))}g(\xi_t)\right]\\
&\geq \inf_{\pi^2\in\Pi_{Ad}^2}\sup_{\pi^1\in\Pi_{Ad}^1} E_i^{\pi^1,\pi^2}\left[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1}\bar{c}(\xi_s,\pi^1(s),\pi^2(s))}g(\xi_t)\right]. \qquad (5.3.8)
\end{aligned}$$

By similar arguments, using (5.3.6), (5.3.7) and the dominated convergence theorem, we obtain

$$\begin{aligned}
\varphi(i)&\leq \inf_{\pi^2\in\Pi_{Ad}^2} E_i^{\pi_n^{*1},\pi^2}\left[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1}\bar{c}(\xi_s,\pi_n^{*1}(\xi_s),\pi^2(s))}g(\xi_t)\right]\\
&\leq \sup_{\pi^1\in\Pi_{Ad}^1}\inf_{\pi^2\in\Pi_{Ad}^2} E_i^{\pi^1,\pi^2}\left[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1}\bar{c}(\xi_s,\pi^1(s),\pi^2(s))}g(\xi_t)\right]. \qquad (5.3.9)
\end{aligned}$$

Now combining (5.3.8) and (5.3.9), we obtain (5.3.4). ☐

Next using Theorem 5.3.1, we show that for each $n\in\mathbb{N}$, Dirichlet eigenpair exists in $\mathscr{D}_n$. That is we establish the following result.

Lemma 5.3.3. *Suppose Assumptions (A1) and (A2) hold. Then there exists an eigenpair* $(\rho_n,\psi_n)\in\mathbb{R}\times\mathcal{B}_{\mathscr{D}_n}^{+}$, $\psi_n\gneq 0$ *on* $\mathscr{D}_n$, *for the following Dirichlet nonlinear eigenequation*

$$\begin{aligned}
e^{\rho_n}\psi_n(i)&=\inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\left[e^{c(i,\mu,\nu)}\sum_{j\in S}\psi_n(j)P(j|i,\mu,\nu)\right]\\
&=\sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[e^{c(i,\mu,\nu)}\sum_{j\in S}\psi_n(j)P(j|i,\mu,\nu)\right]. \qquad (5.3.10)
\end{aligned}$$

The eigenvalue of the above equation satisfies

$$\rho_n \leq \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} \mathscr{J}(i, c, \pi^1, \pi^2), \tag{5.3.11}$$

for all $i \in S$ such that $\psi_n(i) > 0$.

Proof. For some constant $\delta > 0$, let us define $c^{'}(i,\mu,\nu) = c(i,\mu,\nu) - k_n - \delta$ in $\mathscr{D}_n$, where $k_n = \sup_{(i,\mu,\nu) \in \mathscr{D}_n \times \mathcal{P}(A(i)) \times \mathcal{P}(B(i))} |c(i,\mu,\nu)|$. Then it is easy to see that $c^{'}(i,\mu,\nu) < -\delta$, $\forall (i,\mu,\nu) \in \mathscr{D}_n, \times \mathcal{P}(A(i)) \times \mathcal{P}(B(i))$. Now consider a mapping $\bar{T}_n : \mathcal{B}_{\mathscr{D}_n} \to \mathcal{B}_{\mathscr{D}_n}$ defined by

$$\bar{T}_n(g)(i) := \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} E_i^{\pi^1,\pi^2} \left[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1} c^{'}(\xi_s, \pi^1(s), \pi^2(s))} g(\xi_t) \right], \; i \in \mathscr{D}_n\,, \tag{5.3.12}$$

with $\bar{T}_n(g)(i) = 0$ for $i \in \mathscr{D}_n^c$, where $g \in \mathcal{B}_{\mathscr{D}_n}$.
From Proposition 5.3.2 it is clear that $\bar{T}_n$ is well defined. Since $c^{'} < -\delta$, for $g_1, g_2 \in \mathcal{B}_{\mathscr{D}_n}$, it follows that

$$\|\bar{T}_n(g_1) - \bar{T}_n(g_2)\|_{\mathscr{D}_n} \leq \alpha_1 \|g_1 - g_2\|_{\mathscr{D}_n},$$

for some constant $\alpha_1 > 0$. Hence the map $\bar{T}_n$ is continuous.
Let $g_1, g_2 \in \mathcal{B}_{\mathscr{D}_n}$ with $g_1 \succeq g_2$. Also, let $\bar{T}_n(g_k) = \varphi_k$, $k = 1, 2$. Thus φ_2 is a solution of

$$\begin{aligned}
\varphi_2(i) &= \sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \left[e^{c^{'}(i,\mu,\nu)} \sum_{j \in \mathscr{D}_n} \varphi_2(j) P(j|i,\mu,\nu) + g_2(i) \right] \\
&= \inf_{\nu \in \mathcal{P}(B(i))} \left[e^{c^{'}(i,\pi_n^{*1}(i),\nu)} \sum_{j \in \mathscr{D}_n} \varphi_2(j) P(j|i,\pi_n^{*1}(i),\nu) + g_2(i) \right] \; \forall i \in \mathscr{D}_n,
\end{aligned}$$

where $\pi_n^{*1} \in \Pi^1_{SM}$ is an outer maximizing selector. Therefore

$$\begin{aligned}
&\bar{T}_n(g_1)(i) - \bar{T}_n(g_2)(i) \\
&\quad = \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} E_i^{\pi^1,\pi^2} \left[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1} c^{'}(\xi_s, \pi^1(s), \pi^2(s))} g_1(\xi_t) \right] \\
&\qquad - \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} E_i^{\pi^1,\pi^2} \left[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1} c^{'}(\xi_s, \pi^1(s), \pi^2(s))} g_2(\xi_t) \right] \\
&\quad = \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} E_i^{\pi^1,\pi^2} \left[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1} c^{'}(\xi_s, \pi^1(s), \pi^2(s))} g_1(\xi_t) \right] \\
&\qquad - \inf_{\pi^2 \in \Pi^2_{Ad}} E_i^{\pi_n^{*1},\pi^2} \left[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1} c^{'}(\xi_s, \pi_n^{*1}(\xi_s), \pi^2(s))} g_2(\xi_t) \right] \\
&\quad \geq \inf_{\pi^2 \in \Pi^2_{Ad}} E_i^{\pi_n^{*1},\pi^2} \left[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1} c^{'}(\xi_s, \pi_n^{*1}(\xi_s), \pi^2(s))ds} g_1(\xi_t) \right]
\end{aligned}$$

$$
\begin{aligned}
&- \inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi_n^{*1},\pi^2}\left[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1} c'(\xi_s,\pi_n^{*1}(\xi_s),\pi^2(s))} g_2(\xi_t)\right]\\
\geq &\inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi_n^{*1},\pi^2}\left[\sum_{t=0}^{\tau(\mathscr{D}_n)-1} e^{\sum_{s=0}^{t-1} c'(\xi_s,\pi_n^{*1}(\xi_s),\pi^2(s))} (g_1(\xi_t)-g_2(\xi_t))\right].
\end{aligned}
$$

Hence $\bar{T}_n(g_1)(i) - \bar{T}_n(g_2)(i) \geq 0$ for all $i \in S$. This implies that $\bar{T}_n(g_1) \succeq \bar{T}_n(g_2)$. Choose a function $g \in \mathcal{B}_{\mathscr{D}_n}$ such that $g(i_0) = 1$ and $g(j) = 0$ for all $j \neq i_0$, where i_0 is a fixed state (see p. 106). Thus by (5.3.12), we have

$$\bar{T}_n(g)(i_0) \geq g(i_0) > 0.$$

Thus we have $\bar{T}_n(g) \succeq g$. Let $\{g_m\} \subset \mathcal{B}_{\mathscr{D}_n}$ be a bounded sequence. Then since $c' < 0$, from (5.3.12), we get $\|\bar{T}_n g_m\|_\infty \leq \alpha_2$, for some constant $\alpha_2 > 0$. So, by a diagonalization argument, there exists a subsequence m_k of m and a function $\phi \in \mathcal{B}_{\mathscr{D}_n}$ such that $\|\bar{T}_n g_{m_k} - \phi\|_{\mathscr{D}_n} \to 0$ as $k \to \infty$. Thus the map $\bar{T}_n$ is completely continuous. By the definition of the map $\bar{T}_n$, it is easy to see that $\bar{T}_n(\lambda g) = \lambda \bar{T}_n(g)$ for all $\lambda \geq 0$. Hence by Theorem 5.3.1, there exists a nontrivial $\psi_n \in \mathcal{B}^+_{\mathscr{D}_n}$ and a constant $\lambda'_{\mathscr{D}_n} > 0$ such that
$\bar{T}_n(\psi_n) = \lambda'_{\mathscr{D}_n}\psi_n$ i.e.,

$$\lambda'_{\mathscr{D}_n}\psi_n(i) = \sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[e^{c'(i,\mu,\nu)}\sum_{j\in\mathscr{D}_n}\lambda'_{\mathscr{D}_n}\psi_n(j)P(j|i,\mu,\nu) + \psi_n(i)\right] \ \forall i \in \mathscr{D}_n. \tag{5.3.13}$$

Since $\psi_n \geq 0$ and $\psi_n(i) > 0$, for some $i \in \mathscr{D}_n$, it follows from (5.3.13) that $\left[\frac{\lambda'_{\mathscr{D}_n}-1}{\lambda'_{\mathscr{D}_n}}\right] \geq 0$.

Next we prove (5.3.11). Now if $\left[\frac{\lambda'_{\mathscr{D}_n}-1}{\lambda'_{\mathscr{D}_n}}\right] = 0$, it is easy to show that (5.3.11) holds. Assume that $\left[\frac{\lambda'_{\mathscr{D}_n}-1}{\lambda'_{\mathscr{D}_n}}\right] > 0$. Let $\rho'_n = \log\left[\frac{\lambda'_{\mathscr{D}_n}-1}{\lambda'_{\mathscr{D}_n}}\right]$. Then from, (5.3.13), we get

$$e^{\rho'_n}\psi_n(i) = \sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[e^{c'(i,\mu,\nu)}\sum_{j\in\mathscr{D}_n}\psi_n(j)P(j|i,\mu,\nu)\right] \ \forall i \in \mathscr{D}_n. \tag{5.3.14}$$

Now multiplying both sides of (5.3.14) by $e^{k_n+\delta}$ and applying Fan's minimax theorem, (see [33, Theorem 3]), we obtain

$$
\begin{aligned}
e^{\rho_n}\psi_n(i) &= \sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[e^{c(i,\mu,\nu)}\sum_{j\in S}\psi_n(j)P(j|i,\mu,\nu)\right]\\
&= \inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\left[e^{c(i,\mu,\nu)}\sum_{j\in S}\psi_n(j)P(j|i,\mu,\nu)\right] \ \forall i \in \mathscr{D}_n\,,
\end{aligned} \tag{5.3.15}
$$

where $\rho_n = \rho'_n + k_n + \delta$, (where k_n is defined on p. 111).

Let $\pi^{*1}_n \in \Pi^1_{SM}$ be an outer maximizing selector of (5.3.10). Then we have

$$e^{\rho_n}\psi_n(i) = \inf_{\nu\in\mathcal{P}(B(i))}\Big[e^{c(i,\pi^{*1}_n(i),\nu)}\sum_{j\in S}\psi_n(j)P(j|i,\pi^{*1}_n(i),\nu)\Big] \ \forall i \in \mathscr{D}_n. \tag{5.3.16}$$

Therefore by using Dynkin's formula and (5.3.16), we obtain

$$\begin{aligned}\psi_n(i) &\leq E^{\pi^{*1}_n,\pi^2}_i\Big[e^{\sum_{s=0}^{T-1}(c(\xi_s,\pi^{*1}_n(\xi_s),\pi^2(s))-\rho_n)}\psi_n(\xi_T)I_{\{T<\tau(\mathscr{D}_n)\}}\Big]\\ &\leq (\sup_{\mathscr{D}_n}\psi_n)E^{\pi^{*1}_n,\pi^2}_i\Big[e^{\sum_{s=0}^{T-1}(c(\xi_s,\pi^{*1}_n(\xi_s),\pi^2(s))-\rho_n)}\Big].\end{aligned} \tag{5.3.17}$$

Now, taking logarithm on the both sides of (5.3.17), dividing by T and letting $T \to \infty$, for each $i \in S$ for which $\psi_n > 0$, we deduce that

$$\rho_n \leq \mathscr{J}(i,c,\pi^{*1}_n,\pi^2).$$

Since $\pi^2 \in \Pi^2_{Ad}$ is arbitrary, we get

$$\rho_n \leq \inf_{\pi^2\in\Pi^2_{Ad}}\mathscr{J}(i,c,\pi^{*1}_n,\pi^2) \leq \sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}}\mathscr{J}(i,c,\pi^1,\pi^2).$$

□

Now, we show that the sequence $\{\rho\}_n$ is bounded, and for each n, $\psi_n > 0$ on $\mathscr{D}_n$ and $\liminf_{n\to\infty}\rho_n \geq 0$.

Lemma 5.3.4. *Suppose Assumptions (A1) and (A2) hold. Then for each n, $\psi_n > 0$ on $\mathscr{D}_n$ and the sequence of eigenvalues $\{\rho_n\}_n$ of the above eq. (5.3.10) is bounded. Moreover, we have*

$$\liminf_{n\to\infty}\rho_n \geq 0. \tag{5.3.18}$$

Proof. We first prove that $\{\rho_n\}_n$ is bounded. Under Assumption (A1)(a) since $\|c\|_\infty < \gamma$, it is easy to see that $\mathscr{J}(i,c,\pi^1,\pi^2) \leq \gamma$. Under Assumption (A1)(b) since $\mathscr{K}$ is finite, there exists a constant k_1 such that (5.2.5) can be written as

$$\sup_{(a,b)\in A(i)\times B(i)}\sum_{j\in S}V(j)P(j|i,u,v) \leq e^{(k_1-\ell(i))}V(i) \ \forall i \in S. \tag{5.3.19}$$

Then by using (5.2.2) and successive conditioning, we get

$$E^{\pi^1,\pi^2}_i\Big[e^{\sum_{t=0}^{T-1}(\ell(\xi_t)-k_1)}V(\xi_T)\Big] \leq V(i) \ \forall i \in S. \tag{5.3.20}$$

Since, $V \geq 1$, from (5.3.20), we get

$$\mathscr{J}(i,\ell,\pi^1,\pi^2) \leq k_1 \quad \text{for all } i \in S.$$

Now since $\ell - \sup_{(a,b)\in A(i)\times B(i)} c(\cdot,a,b)$ is norm-like, there exists a constant k_2 such that for all $i \in S$, we have $\sup_{(a,b)\in A(i)\times B(i)} c(i,a,b) \leq \ell(i) + k_2$. Hence we get

$$\mathscr{J}(i,c,\pi^1,\pi^2) \leq k_1 + k_2 \quad \forall(\pi^1,\pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}, \forall i \in S. \tag{5.3.21}$$

Therefore using (5.3.11), it is clear that ρ_n has an upper bound.

Next we want to show that ρ_n is bounded below. To this end, first we claim that $\psi_n > 0$ on $\mathscr{D}_n$ for each n. Let $n \in \mathbb{N}$ be fixed. Suppose that the claim is not true, then there exists $\tilde{i} \in \mathscr{D}_n$ such that $\psi_n(\tilde{i}) = 0$. Also, since $\psi_n \gneq 0$ on $\mathscr{D}_n$, there exists $\hat{i} \in \mathscr{D}_n$ such that $\psi_n(\hat{i}) > 0$. Now, for any outer minimizing selector $\pi^{*2}_n \in \Pi^2_{SM}$ of (5.3.15), the eq. (5.3.16) can be rewritten as

$$0 = e^{\rho_n}\psi_n(\tilde{i}) = \left[e^{c(\tilde{i},\pi^{*1}_n(\tilde{i}),\pi^{*2}_n(\tilde{i}))}\sum_{j\in S}\psi_n(j)P(j|\tilde{i},\pi^{*1}_n(\tilde{i}),\pi^{*2}_n(\tilde{i}))\right]. \tag{5.3.22}$$

Again, in view of Assumption $(A2)$(ii), under any pair of strategies $(\pi^{*1}_n,\pi^{*2}_n) \in \Pi^1_{SM} \times \Pi^2_{SM}$, there exists $i_1, i_2, \cdots, i_m \in \mathscr{D}_n$ satisfying

$$P(\hat{i}|i_m,\pi^{*1}_n(i_m),\pi^{*2}_n(i_m))P(i_m|i_{m-1},\pi^{*1}_n(i_{m-1}),\pi^{*2}_n(i_{m-1}))\cdots P(i_1|\tilde{i},\pi^{*1}_n(\tilde{i}),\pi^{*2}_n(\tilde{i})) > 0. \tag{5.3.23}$$

Thus, from (5.3.22) and (5.3.23), we deduce that $\psi_n(\hat{i}) = \psi_n(i_1) = \cdots = \psi_n(i_m) = \psi_n(\tilde{i}) = 0$. But this contradicts the fact that ψ_n is nontrivial. Since, n is arbitrary, this establishes our claim. So, for all n we can pin ψ_n such that $\psi_n(i_0) = 1$, where i_0 is a reference state (defined as in p. 106).

Now, suppose that the sequence $\{\rho_n\}_n$ is not bounded below. Hence, along a subsequence $\rho_n \to -\infty$ as $n \to \infty$. So, $\rho_n < 0$ for all large enough n. Let $(\pi^{*1}_n,\pi^{*2}_n) \in \Pi^1_{SM} \times \Pi^2_{SM}$ be a mini-max selector of (5.3.10), thus we have

$$\begin{aligned}1 = \psi_n(i_0) &= e^{-\rho_n}\sup_{\mu\in\mathcal{P}(A(i_0))}\left[e^{c(i_0,\mu,\pi^{*2}_n(i_0))}\sum_{j\in S}\psi_n(j)P(j|i_0,\mu,\pi^{*2}_n(i_0))\right]\\ &= e^{-\rho_n}\left[e^{c(i_0,\pi^{*1}_n(i_0),\pi^{*2}_n(i_0))}\sum_{j\in S}\psi_n(j)P(j|i_0,\pi^{*1}_n(i_0),\pi^{*2}_n(i_0))\right].\end{aligned} \tag{5.3.24}$$

Since $\rho_n < 0$ for all large enough n, and our cost function c is nonnegative, it is easy to see that $c(i_0,\pi^{*1}_n(i_0),\pi^{*2}_n(i_0)) - \rho_n > 0$, for all large enough n. Assumption $(A2)$(ii), implies that for any $j \in \mathscr{D}_n$, under any pair of strategies $(\pi^1,\pi^2) \in \Pi^1_{SM} \times \Pi^2_{SM}$, there exists $i_1, i_2, \cdots, i_m \in \mathscr{D}_n$ satisfying

$$P(j|i_m,\pi^1(i_m),\pi^2(i_m))P(i_m|i_{m-1},\pi^1(i_{m-1}),\pi^2(i_{m-1}))\cdots P(i_1|i_0,\pi^1(i_0),\pi^2(i_0)) > 0. \tag{5.3.25}$$

We claim that if $j \in \mathscr{D}_n$, then

$$\inf_{(\pi^1,\pi^2)\in\Pi^1_{SM}\times\Pi^2_{SM}} P^{\pi^1,\pi^2}_{i_0}(\hat{\tau}_j \le n \wedge \tau(\mathscr{D}_n)) \ge k(j,n), \text{ for some constant } k(j,n) > 0. \tag{5.3.26}$$

If not, suppose there exists a pair $(\tilde{\pi}^1_k, \tilde{\pi}^2_k) \in \Pi^1_{SM} \times \Pi^2_{SM}$ such that $P^{\tilde{\pi}^1_k,\tilde{\pi}^2_k}_{i_0}(\check{\tau}_j \le n \wedge \tau(\mathscr{D}_n)) \to 0$ as $k \to \infty$. Now, since Π^1_{SM} and Π^2_{SM} are compact, there exist a further subsequence and $\tilde{\pi}^1 \in \Pi^1_{SM}$ and $\tilde{\pi}^2 \in \Pi^2_{SM}$, such that $\tilde{\pi}^1_k \to \tilde{\pi}^1$ and $\tilde{\pi}^2_k \to \tilde{\pi}^2$ as $k \to \infty$. By Assumption $(A2)$, we know that the law of $\boldsymbol{\xi}_k$ converges to $\boldsymbol{\xi}$, where $\boldsymbol{\xi}_k$ $(\boldsymbol{\xi})$ is the DTMDP governed by $(\tilde{\pi}^1_k, \tilde{\pi}^2_k)$ $((\tilde{\pi}^1, \tilde{\pi}^2)$ respectively). So, for every $p \le n$,

$$\begin{aligned}
&P^{\tilde{\pi}^1,\tilde{\pi}^2}_{i_0}(\xi_i \in \mathscr{D}_n\backslash\{i_0, j\}, \xi_p = j \text{ for all } i \le p-1) \\
&\quad= \lim_{k\to\infty} P^{\tilde{\pi}^1_k,\tilde{\pi}^2_k}_{i_0}(\xi_{k,i} \in \mathscr{D}_n\backslash\{i_0, j\}, \xi_{k,p} = j \text{ for all } i \le p-1) \\
&\quad\le \lim_{k\to\infty} P^{\tilde{\pi}^1_k,\tilde{\pi}^2_k}_{i_0}(\check{\tau}_j \le n \wedge \tau(\mathscr{D}_n)) = 0.
\end{aligned}$$

So, this contradicts (5.3.25). Hence, we must have (5.3.26).

From the monotonicity of $\tau(\mathscr{D}_n)$, it then follows that for $\mathscr{D}_n \supset \mathscr{D}_m \ni j$, we have

$$\inf_{(\pi^1,\pi^2)\in\Pi^1_{SM}\times\Pi^2_{SM}} P^{\pi^1,\pi^2}_{i_0}(\hat{\tau}_j \le m \wedge \tau(\mathscr{D}_n)) \ge \inf_{(\pi^1,\pi^2)\in\Pi^1_{SM}\times\Pi^2_{SM}} P^{\pi^1,\pi^2}_{i_0}(\hat{\tau}_j \le m \wedge \tau(\mathscr{D}_m)) \ge k(j,m). \tag{5.3.27}$$

Since for large enough n, $c(i_0, \pi^{*1}_n(i_0), \pi^{*2}_n(i_0)) - \rho_n > 0$, from (5.3.24), we have

$$\begin{aligned}
1 = \psi_n(i_0) &= E^{\pi^{*1}_n,\pi^{*2}_n}_{i_0}\left[e^{\sum_{t=0}^{m\wedge\tau(\mathscr{D}_n)\wedge\hat{\tau}_j-1}(c(\xi_t,\pi^{*1}_n(\xi_t),\pi^{*2}_n(\xi_t))-\rho_n)}\psi_n(\xi_{m\wedge\tau(\mathscr{D}_n)\wedge\hat{\tau}_j})\right] \\
&\ge E^{\pi^{*1}_n,\pi^{*2}_n}_{i_0}\left[\psi_n(\xi_{m\wedge\tau(\mathscr{D}_n)\wedge\hat{\tau}_j})\right] \\
&\ge \psi_n(j) \inf_{(\pi^1,\pi^2)\in\Pi^1_{SM}\times\Pi^2_{SM}} P^{\pi^1,\pi^2}_{i_0}(\hat{\tau}_j \le m \wedge \tau(\mathscr{D}_n)) \\
&\ge k(j,m)\psi_n(j) \text{ (using (5.3.27))}.
\end{aligned}$$

Choose $m = j+1$. Then for all $n > j$, we have $\psi_n(j) \le \frac{1}{k(j,j+1)}$, $\forall j \in S$. This implies that, $\{\psi_n\}$ has an upper bound. Thus by a standard diagonalization argument, there exists a subsequence (by abuse of notation denoting by the same sequence) and a bounded function $\psi \ge 0$ with $\psi(i_0) = 1$ such that $\psi_n(i) \to \psi(i)$, as $n \to \infty$ for all $i \in S$. Now, since Π^1_{SM} and Π^2_{SM} are compact, there exist a further subsequence and $\pi^{*1} \in \Pi^1_{SM}$ and $\pi^{*2} \in \Pi^2_{SM}$, such that $\pi^{*1}_n \to \pi^{*1}$ and $\pi^{*2}_n \to \pi^{*2}$ as $n \to \infty$. Since $c \ge 0$, (5.3.10) gives us

$$e^{\rho_n}\psi_n(i) \ge \left[\sum_{j\in S} \psi_n(j) P(j|i, \pi^{*1}_n(i), \pi^{*2}_n(i))\right]. \tag{5.3.28}$$

Hence, by taking $n \to \infty$, it follows that

$$\sum_{j\in S} \psi(j)P(j|i,\pi^{*1}(i),\pi^{*2}(i)) \leq 0,\ \ i \in S. \tag{5.3.29}$$

In view of (5.3.29), we claim that $\psi \equiv 0$. If not then there exists $\hat{i} \in S$ such that $\psi^*(\hat{i}) > 0$. Also, since $\psi \geq 0$ from (5.3.29), it is easy to see that there exists a point $\tilde{i} \in S$ for which $\psi(\tilde{i}) = 0$. Now, since $\{\xi_t\}$ is irreducible under any pair of strategies $(\pi^{*1},\pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$, there exists $i_1, i_2, \cdots, i_m \in S$ satisfying

$$P(\hat{i}|i_m,\pi^{*1}(i_m),\pi^{*2}(i_m))P(i_m|i_{m-1},\pi^{*1}(i_{m-1}),\pi^{*2}(i_{m-1}))\cdots P(i_1|\tilde{i},\pi^{*1}(\tilde{i}),\pi^{*2}(\tilde{i})) > 0.$$

Thus, from (5.3.29) we deduce that $\psi(\hat{i}) = \psi(i_1) = \cdots = \psi(i_m) = \psi(\tilde{i}) = 0$. But this contradicts to the fact that $\psi(\hat{i}) > 0$. This proves the claim . But since $\psi(i_0) = 1$, this is a contradiction. Therefore, we obtain that, $\{\rho_n\}$ is bounded below.

Now we show that $\rho^* = \liminf_{n\to\infty} \rho_n \geq 0$. If not, then on contrary, $\rho^* < 0$. So, for large enough n, $\rho_n < 0$. Since, our cost function c is nonnegative, for large enough n, $c(i,\mu,\nu) - \rho_n > 0$ for all $(\mu,\nu) \in \mathcal{P}(A(i)) \times \mathcal{P}(B(i))$. So, by repeating the above arguments, there exists a subsequence (by abuse of notation denoting by the same sequence) and a bounded function $\phi \geq 0$ with $\phi(i_0) = 1$ such that $\psi_n(i) \to \phi(i)$, as $n \to \infty$ for all $i \in S$. From (5.3.10), we have

$$\psi_n(i) \geq \left[\sum_{j\in S} \psi_n(j)P(j|i,\pi^{*1}_n(i),\pi^{*2}_n(i))\right], \tag{5.3.30}$$

where (π^{*1}_n,π^{*2}_n) is a mini-max selector of (5.3.10) . By Fatou's lemma, taking $n \to \infty$, we deduce that

$$\phi(i) \geq E^{\pi^{*1},\pi^{*2}}_i[\phi(\xi_1)]\ \forall i \in S,$$

for some pair of stationary strategies $(\pi^{*1},\pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$. Hence, $\{\phi(\xi_m),\mathscr{F}_m\}$ is supermartingale where $\{\xi_t\}$ is the Markov process under the pair of stationary strategies $(\pi^{*1},\pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$. So, by Doob's martingale convergence theorem $\phi(\xi_m) \to \hat{Y}$ almost surely, as $m \to \infty$. On the other hand by Assumption $(A1)$, we have $\{\xi_t\}$ is recurrent. Hence $\{\xi_t\}$ visits every state (in particular i_0) of S infinitely often. Since, $\phi(i_0) = 1$, $\{\phi(\xi_m)\}$ converges only if $\phi \equiv 1$. Now, taking limit $n \to \infty$ in (5.3.10), we obtain

$$1 = \phi(i) \geq e^{c(i,\pi^{*1}(i),\pi^{*2}(i))-\rho^*} > 1.$$

But this is a contradiction. Thus, $\liminf_{n\to\infty} \rho_n \geq 0$. □

5.4 Existence of risk-sensitive average optimal strategies

In this section, we prove the existence of a risk-sensitive average optimal stationary strategy using the Shapley equation. Now we state and prove our main result of this section.

Theorem 5.4.1. *Suppose Assumptions (A1) and (A2) hold. Then there exists a unique (upto a scalar multiplication) eigenpair* $(\rho^*, \psi^*) \in \mathbb{R}_+ \times L^\infty_V(S)$ *with* $\psi^* > 0$, *such that*

$$\begin{aligned} e^{\rho^*}\psi^*(i) &= \sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \left[e^{c(i,\mu,\nu)} \sum_{j \in S} \psi^*(j) P(j|i,\mu,\nu) \right] \\ &= \inf_{\nu \in \mathcal{P}(V(i))} \sup_{\mu \in \mathcal{P}(A(i))} \left[e^{c(i,\mu,\nu)} \sum_{j \in S} \psi^*(j) P(j|i,\mu,\nu) \right], \; i \in S. \end{aligned} \tag{5.4.1}$$

Moreover, we have the following

(i)

$$\rho^* = \inf_{i \in S} \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} \mathscr{J}(i, c, \pi^1, \pi^2) = \inf_{i \in S} \inf_{\pi^2 \in \Pi^2_{Ad}} \sup_{\pi^1 \in \Pi^1_{Ad}} \mathscr{J}(i, c, \pi^1, \pi^2). \tag{5.4.2}$$

(ii) If $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ *be a mini-max selector of (5.4.1), then* $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ *is a saddle-point equilibrium, i.e.,*

$$\mathscr{J}(i, c, \pi^1, \pi^{*2}) \le \mathscr{J}(i, c, \pi^{*1}, \pi^{*2}) = \rho^* \le \mathscr{J}(i, c, \pi^{*1}, \pi^2), \;\; \forall \pi^1 \in \Pi^1_{Ad}, \; \forall \pi^2 \in \Pi^2_{Ad}. \tag{5.4.3}$$

Thus the value of the game is independent of the initial state.

(iii) Let $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ *is a saddle-point equilibrium, then this pair is a mini-max selector of (5.4.1).*

The rest of this section is dedicated to the proof of Theorem 5.4.1.

Since $c \ge 0$, using Assumption $(A1)$, there exists a finite set $\mathscr{B}$ containing $\mathscr{K}$ such that we have the following:

- Under Assumption $(A1)$(a): since $\gamma > \|c\|_\infty$, from (5.3.11) we have $\rho_n \le \gamma$. Thus, for all large enough n it holds that

$$\left(\sup_{(a,b) \in A(i) \times B(i)} c(i,a,b) - \rho_n \right) < \gamma \;\forall i \in \mathscr{B}^c. \tag{5.4.4}$$

- Under Assumption $(A1)$(b): since the function $\ell(\cdot) - \max_{(a,b)\in A(\cdot)\times B(\cdot)} c(\cdot, a, b)$ is norm-like, for all large enough n it holds that

$$\left(\sup_{(a,b)\in A(i)\times B(i)} c(i,a,b) - \rho_n\right) < \ell(i) \ \forall i \in \mathscr{B}^c. \tag{5.4.5}$$

Now letting $n \to \infty$ from (5.3.10) we show that the limiting equation admits a positive eigenpair.

Lemma 5.4.1. *Suppose Assumptions $(A1)$ and $(A2)$ hold. Then there exists an eigenpair $(\rho^*, \psi^*) \in \mathbb{R}_+ \times L^\infty_V(S)$ with $\psi^* > 0$, such that*

$$\begin{aligned} e^{\rho^*}\psi^*(i) &= \sup_{\mu\in\mathcal{P}(A(i))} \inf_{\nu\in\mathcal{P}(B(i))} \left[e^{c(i,\mu,\nu)} \sum_{j\in S} \psi^*(j) P(j|i,\mu,\nu)\right] \\ &= \inf_{\nu\in\mathcal{P}(B(i))} \sup_{\mu\in\mathcal{P}(A(i))} \left[e^{c(i,\mu,\nu)} \sum_{j\in S} \psi^*(j) P(j|i,\mu,\nu)\right], \ i \in S. \end{aligned} \tag{5.4.6}$$

Furthermore, for any mini-max selector $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ of (5.4.6) we have the following:

(i)

$$\rho^* \le \inf_{i\in S} \sup_{\pi^1\in\Pi^1_{Ad}} \inf_{\pi^2\in\Pi^2_{Ad}} \mathscr{J}(i, c, \pi^1, \pi^2). \tag{5.4.7}$$

(ii) For any finite set $\mathscr{B}_1 \supset \mathscr{B}$, we have the following stochastic representation of the eigenfunction

$$\begin{aligned} \psi^*(i) &= \inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^{*1},\pi^2}\left[e^{\sum_{t=0}^{\hat\tau(\mathscr{B}_1)-1}(c(\xi_t,\pi^{*1}(\xi_t),\pi^2(t))-\rho^*)}\psi^*(\xi_{\hat\tau(\mathscr{B}_1)})\right] \\ &= \sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^{*2}}\left[e^{\sum_{t=0}^{\hat\tau(\mathscr{B}_1)-1}(c(\xi_t,\pi^1(t),\pi^{*2}(\xi_t))-\rho^*)}\psi^*(\xi_{\hat\tau(\mathscr{B}_1)})\right] \ \forall i \in \mathscr{B}_1^c. \end{aligned} \tag{5.4.8}$$

Proof. First we scale ψ_n in such a way that we obtain $\psi_n(i) \le V(i)$ for all $i \in S$. Set

$$\tilde\theta_n = \sup\{\alpha > 0 : (V - \alpha\psi_n) > 0 \text{ in } S\}.$$

Since ψ_n vanishes in $\mathscr{D}_n^c$ and $\psi_n > 0$ on $\mathscr{D}_n$, it follows that $\tilde\theta_n$ is finite. We claim that if we replace ψ_n by $\tilde\theta_n\psi_n$, then ψ_n touches V inside $\mathscr{B}$. If this is not true, then on the contrary, we assume that for some state $\hat i \in \mathscr{B}^c \cap \mathscr{D}_n$, $(V - \psi_n)(\hat i) = 0$ and $V - \psi_n > 0$ in $\mathscr{B} \cup \mathscr{D}_n^c$. Let π_n^{*1} be an outer maximizing selector of (5.3.10). Then under Assumption $(A1)$(b), applying Dynkin's formula (as in [113, Lemma 3.1]), we obtain

$$\psi_n(\hat i) \le E_{\hat i}^{\pi_n^{*1},\pi^2}\left[e^{\sum_{s=0}^{N\wedge\hat\tau(\mathscr{B})-1}(c(\xi_s,\pi_n^{*1}(\xi_s),\pi^2(s))-\rho_n)}\psi_n(\xi_{N\wedge\hat\tau(\mathscr{B})})I_{\{N\wedge\hat\tau(\mathscr{B})<\tau(\mathscr{D}_n)\}}\right]$$

$$\leq E_{\hat{i}}^{\pi_n^{*1},\pi^2}\left[e^{\sum_{s=0}^{N\wedge\hat{\tau}(\mathscr{B})-1}\ell(\xi_s)}\psi_n(\xi_{N\wedge\hat{\tau}(\mathscr{B})})I_{\{N\wedge\hat{\tau}(\mathscr{B})<\tau(\mathscr{D}_n)\}}\right].$$

Since $\psi_n \leq V$ (by our scaling), in view of Lemma 5.3.1, by the dominated convergence theorem taking $N \to \infty$, we get

$$\psi_n(\hat{i}) \leq E_{\hat{i}}^{\pi_n^{*1},\pi^2}\left[e^{\sum_{s=0}^{\hat{\tau}(\mathscr{B})-1}\ell(\xi_s)ds}\psi_n(\xi_{\hat{\tau}(\mathscr{B})})\right].$$

Combining this and (5.3.2), we get

$$0 = (V-\psi_n)(\hat{i}) \geq E_{\hat{i}}^{\pi_n^{*1},\pi^2}\left[e^{\sum_{s=0}^{\hat{\tau}(\mathscr{B})-1}\ell(\xi_s)ds}(V-\psi_n)(\xi_{\hat{\tau}(\mathscr{B})})\right] > 0.$$

But this is a contradiction. Thus ψ_n touches V inside $\mathscr{B}$. Using estimate as in (5.3.1), one can show that similar conclusion holds under Assumption $(A1)$(a) .

So, there exists a point $i^* \in \mathscr{B}$ such that $(V-\psi_n)(i^*) = 0$, for all large n. Since $\psi_n \leq V$, by diagonalization arguments, there exists a subsequence (here we use the same sequence by abuse of notation), and a function $\psi^* \leq V$ such that $\psi_n \to \psi^*$ as $n \to \infty$. Again, from Lemma 5.3.4, we know that the sequence $\{\rho_n\}$ is bounded and $\liminf_{n\to\infty} \rho_n \geq 0$, thus along a further subsequence we have $\rho_n \to \rho^*$ as $n \to \infty$ for some $\rho^* \geq 0$.

Also, we have $(V-\psi^*)(\hat{i}^*) = 0$ for some $\hat{i}^* \in \mathscr{B}$. By the continuity-compactness assumptions, for any mini-max selector $(\pi_n^{*1},\pi_n^{*2}) \in \Pi_{SM}^1 \times \Pi_{SM}^2$ of (5.3.10), we get

$$\begin{aligned} e^{\rho_n}\psi_n(i) &= \sup_{\mu\in\mathcal{P}(A(i))}\left[e^{c(i,\mu,\pi_n^{*2}(i))}\sum_{j\in S}\psi_n(j)P(j|i,\mu,\pi_n^{*2}(i))\right] \\ &= \inf_{\nu\in\mathcal{P}(B(i))}\left[e^{c(i,\pi_n^{*1}(i),\nu,)}\sum_{j\in S}\psi_n(j)P(j|i,\pi_n^{*1}(i),\nu)\right]. \end{aligned} \tag{5.4.9}$$

Note that since $\psi_n \in L_V^\infty(S)$, we have

$$\sum_{j\in S}\psi_n(j)P(j|i,a,b) \leq \sum_{j\in S}V(j)P(j|i,a,b) \ \forall (i,a,b) \in K. \tag{5.4.10}$$

Since Π_{SM}^1 and Π_{SM}^2 are compact there exists $(\pi^{*1},\pi^{*2}) \in \Pi_{SM}^1 \times \Pi_{SM}^2$ such that $\pi_n^{*1} \to \pi^{*1}$ and $\pi_n^{*2} \to \pi^{*2}$ as $n \to \infty$. Now, from (5.4.9) we obtain,

$$e^{\rho_n}\psi_n(i) \geq \left[e^{c(i,\mu,\pi_n^{*2}(i))}\sum_{j\in S}\psi_n(j)P(j|i,\mu,\pi_n^{*2}(i))\right]. \tag{5.4.11}$$

Then, using Lemma 5.2.3, taking $n \to \infty$ from (5.4.11), by the extended Fatou's lemma [35], [69, Lemma 8.3.7], we obtain

$$e^{\rho^*}\psi^*(i) \geq e^{c(i,\mu,\pi^{*2}(i))}\sum_{j\in S}\psi^*(j)P(j|i,\mu,\pi^{*2}(i)).$$

Thus

$$\begin{aligned} e^{\rho^*}\psi^*(i) &\geq \sup_{\mu\in\mathcal{P}(A(i))}\left[e^{c(i,\mu,\pi^{*2}(i))}\sum_{j\in S}\psi^*(j)P(j|i,\mu,\pi^{*2}(i))\right] \\ &\geq \inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\left[e^{c(i,\mu,\nu)}\sum_{j\in S}\psi^*(j)P(j|i,\mu,\nu)\right]. \end{aligned} \tag{5.4.12}$$

Also, from (5.4.9), we get

$$e^{\rho_n}\psi_n(i) \leq \left[e^{c(i,\pi_n^{*1}(i),\nu)}\sum_{j\in S}\psi_n(j)P(j|i,\pi_n^{*1}(i),\nu)\right].$$

Using (5.4.10), by the dominated convergence theorem, taking limit $n\to\infty$ in above equation, we deduce that

$$\begin{aligned} e^{\rho^*}\psi^*(i) &\leq \inf_{\nu\in\mathcal{P}(B(i))}\left[e^{c(i,\pi^{*1}(i),\nu)}\sum_{j\in S}\psi^*(j)P(j|i,\pi^{*1}(i),\nu)\right] \\ &\leq \sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[e^{c(i,\mu,\nu)}\sum_{j\in S}\psi^*(j)P(j|i,\mu,\nu)\right]. \end{aligned} \tag{5.4.13}$$

Hence by (5.4.12) and (5.4.13), we get (5.4.6). Since we have $(V-\psi^*)(\hat{i}^*)=0$ and $V\geq 1$, it follows that ψ^* is nontrivial.

Now, we claim that $\psi^*>0$. If not, then on contrary there exists a point $\tilde{i}\in S$ for which $\psi^*(\tilde{i})=0$. Again by continuity-compactness assumptions, there exists a mini-max selector (π^{*1},π^{*2}) such that (5.4.6) can be rewritten as

$$e^{\rho^*}\psi^*(i)=\left[e^{c(i,\pi^{*1}(i),\pi^{*2}(i))}\sum_{j\in S}\psi^*(j)P(j|i,\pi^{*1}(i),\pi^{*2}(i))\right] \quad \forall i\in S.$$

So, we get

$$0=e^{\rho^*}\psi^*(\tilde{i})=\left[e^{c(\tilde{i},\pi^{*1}(\tilde{i}),\pi^{*2}(\tilde{i}))}\sum_{j\in S}\psi^*(j)P(j|\tilde{i},\pi^{*1}(\tilde{i}),\pi^{*2}(\tilde{i}))\right]. \tag{5.4.14}$$

Since ψ^* is nontrivial, there exists $\hat{i}\in S$ such that $\psi^*(\hat{i})>0$. Again, since ξ is irreducible under any pair of strategies $(\pi^{*1},\pi^{*2})\in\Pi^1_{SM}\times\Pi^2_{SM}$, there exists $i_1,i_2,\cdots,i_n\in S$ satisfying

$$P(\hat{i}|i_n,\pi^{*1}(i_n),\pi^{*2}(i_n))P(i_n|i_{n-1},\pi^{*1}(i_{n-1}),\pi^{*2}(i_{n-1}))\cdots P(i_1|\tilde{i},\pi^{*1}(\tilde{i}),\pi^{*2}(\tilde{i}))>0.$$

Thus, from (5.4.14) we deduce that $\psi^*(\hat{i})=\psi^*(i_1)=\cdots=\psi^*(i_n)=\psi^*(\tilde{i})=0$. But this contradicts the fact that ψ^* is nontrivial. This establishes our claim.

Next we prove (5.4.7). Since $\psi_n>0$ on $\mathscr{D}_n$ for all n, using (5.3.11), we have $\rho^*=\lim_{n\to\infty}\rho_n\leq\sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}}\mathscr{J}(i,c,\pi^1,\pi^2)$ for all $i\in S$.

Finally, we prove the stochastic representation (5.4.8) of ψ^*. As before there exists a pair of strategies $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ satisfying

$$\begin{aligned} e^{\rho^*}\psi^*(i) &= \sup_{\mu\in\mathcal{P}(A(i))}\left[e^{c(i,\mu,\pi^{*2}(i))}\sum_{j\in S}\psi^*(j)P(j|i,\mu,\pi^{*2}(i))\right] \\ &= \inf_{\nu\in\mathcal{P}(B(i))}\left[e^{c(i,\pi^{*1}(i),\nu)}\sum_{j\in S}\psi^*(j)P(j|i,\pi^{*1}(i),\nu)\right]. \end{aligned} \tag{5.4.15}$$

Now for any finite set $\mathscr{B}_1 \supset \mathscr{B}$, applying Dynkin's formula (as in [113, Lemma 3.1]) from (5.4.15), we get

$$\psi^*(i) \le E_i^{\pi^{*1},\pi^2}\left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_1)\wedge N-1}(c(\xi_t,\pi^{*1}(\xi_t),\pi^2(t))-\rho^*)}\psi^*(\xi_{\hat{\tau}(\mathscr{B}_1)\wedge N})\right] \ \forall i \in \mathscr{B}_1^c.$$

Since $\psi^* \le V$, using estimates of Lemma 5.3.1, by the dominated convergence theorem taking $N \to \infty$, it follows that

$$\psi^*(i) \le E_i^{\pi^{*1},\pi^2}\left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_1)-1}(c(\xi_t,\pi^{*1}(\xi_t),\pi^2(t))-\rho^*)}\psi^*(\xi_{\hat{\tau}(\mathscr{B}_1)})\right] \ \forall i \in \mathscr{B}_1^c. \tag{5.4.16}$$

Hence

$$\begin{aligned} \psi^*(i) &\le \inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^{*1},\pi^2}\left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_1)-1}(c(\xi_t,\pi^{*1}(\xi_t),\pi^2(t))-\rho^*)}\psi^*(\xi_{\hat{\tau}(\mathscr{B}_1)})\right] \\ &\le \sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^1,\pi^2}\left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_1)-1}(c(\xi_t,\pi^1(t),\pi^2(t))-\rho^*)}\psi^*(\xi_{\hat{\tau}(\mathscr{B}_1)})\right], \ \forall i \in \mathscr{B}_1^c. \end{aligned} \tag{5.4.17}$$

Now using (5.4.15) and Dynkin's formula

$$\psi^*(i) \ge E_i^{\pi^1,\pi^{*2}}\left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_1)\wedge N-1}(c(\xi_t,\pi^1(t),\pi^{*2}(\xi_t))-\rho^*)}\psi^*(\xi_{\hat{\tau}(\mathscr{B}_1)\wedge N})\right] \ \forall i \in \mathscr{B}_1^c.$$

In view of Lemma 5.3.1 by Fatou's lemma taking $N \to \infty$, we get

$$\psi^*(i) \ge E_i^{\pi^1,\pi^{*2}}\left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_1)-1}(c(\xi_t,\pi^1(t),\pi^{*2}(\xi_t))-\rho^*)}\psi^*(\xi_{\hat{\tau}(\mathscr{B}_1)})\right], \ \forall i \in \mathscr{B}_1^c. \tag{5.4.18}$$

Hence,

$$\begin{aligned} \psi^*(i) &\ge \sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^{*2}}\left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_1)-1}(c(\xi_t,\pi^1(t),\pi^{*2}(\xi_t))-\rho^*)}\psi^*(\xi_{\hat{\tau}(\mathscr{B}_1)})\right] \\ &\ge \inf_{\pi^2\in\Pi^2_{Ad}}\sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^2}\left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_1)-1}(c(\xi_t,\pi^1(t),\pi^2(t))-\rho^*)}\psi^*(\xi_{\hat{\tau}(\mathscr{B}_1)})\right], \ \forall i \in \mathscr{B}_1^c. \end{aligned} \tag{5.4.19}$$

From (5.4.17) and (5.4.19), we get eq. (5.4.8). □

Next, we prove the existence of the value of the game. To this end we first perturb the cost function as follows:

- When Assumption $(A1)$(a) holds: Let $\alpha_3 > 0$, be a small number satisfying $\|c\|_\infty + \alpha_3 < \gamma$. Now we define $\tilde{c}_n(i,a,b) = c(i,a,b)I_{\mathscr{D}_n}(i) + (\|c\|_\infty + \alpha_3)I_{\mathscr{D}_n^c}\ \forall (a,b) \in A(i) \times B(i)$, $i \in S$. Note $\|\tilde{c}_n\|_\infty < \gamma$, where $\|\tilde{c}_n\|_\infty = \sup_{(i,a,b)\in K} \tilde{c}_n(i,a,b)$.
- When Assumption $(A1)$(b) holds: Define

  $$\tilde{c}_n(i,a,b) = c(i,a,b) + \frac{1}{n}[\ell(i) - \sup_{(a,b)\in A(i)\times B(i)} c(i,a,b)]_+ \ \forall (a,b) \in A(i) \times B(i), i \in S.$$

  Since the function $[\ell(\cdot) - \sup_{(a,b)\in A(\cdot)\times B(\cdot)} c(\cdot,a,b)]_+$ is norm-like function, we have $\ell - \sup_{(a,b)\in A(\cdot)\times B(\cdot)} \tilde{c}_n(\cdot,a,b)$ is norm-like for large enough n.

Theorem 5.4.2. *Suppose that Assumptions $(A1)$ and $(A2)$ hold. Let $(\pi^{*1},\pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ be any mini-max selector of (5.4.6), i.e. $(\pi^{*1},\pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ satisfies*

$$\begin{aligned} e^{\rho^*}\psi^*(i) &= \sup_{\mu\in\mathcal{P}(A(i))} \inf_{\nu\in\mathcal{P}(B(i))} \left[e^{c(i,\mu,\nu)} \sum_{j\in S} \psi^*(j)P(j|i,\mu,\nu)\right] \\ &= \inf_{\nu\in\mathcal{P}(B(i))} \sup_{\mu\in\mathcal{P}(A(i))} \left[e^{c(i,\mu,\nu)} \sum_{j\in S} \psi^*(j)P(j|i,\mu,\nu)\right] \\ &= \inf_{\nu\in\mathcal{P}(B(i))} \left[e^{c(i,\pi^{*1}(i),\nu,)} \sum_{j\in S} \psi^*(j)P(j|i,\pi^{*1}(i),\nu)\right] \\ &= \sup_{\mu\in\mathcal{P}(A(i))} \left[e^{c(i,\mu,\pi^{*2}(i))} \sum_{j\in S} \psi^*(j)P(j|i,\mu,\pi^{*2}(i))\right], \quad i \in S. \end{aligned} \tag{5.4.20}$$

Then we have

$$\begin{aligned} \rho^* &= \inf_{i\in S} \sup_{\pi^1\in\Pi^1_{Ad}} \inf_{\pi^2\in\Pi^2_{Ad}} \mathscr{J}(i,c,\pi^1,\pi^2) = \inf_{i\in S} \inf_{\pi^2\in\Pi^2_{Ad}} \sup_{\pi^1\in\Pi^1_{Ad}} \mathscr{J}(i,c,\pi^1,\pi^2) \\ &= \inf_{i\in S} \inf_{\pi^2\in\Pi^2_{Ad}} \mathscr{J}(i,c,\pi^{*1},\pi^2) = \inf_{i\in S} \sup_{\pi^1\in\Pi^1_{Ad}} \mathscr{J}^{\pi^1,\pi^{*2}}(i,c) = \mathscr{J}(i,c,\pi^{*1},\pi^{*2}). \end{aligned} \tag{5.4.21}$$

Proof. Arguing as Lemma 5.4.1, for the stationary strategy $\pi^{*2} \in \Pi^2_{SM}$, there exists an eigenpair $(\hat{\rho}_n, \hat{\psi}_n) \in \mathbb{R}_+ \times L^\infty_V(S)$ with $\hat{\psi}_n > 0$ satisfying

$$e^{\hat{\rho}_n}\hat{\psi}_n(i) = \sup_{\mu\in\mathcal{P}(A(i))} \left[e^{\tilde{c}_n(i,\mu,\pi^{*2}(i))} \sum_{j\in S} \hat{\psi}_n(j)P(j|i,\mu,\pi^{*2}(i))\right], \quad i \in S \tag{5.4.22}$$

such that

$$0 \le \hat{\rho}_n \le \sup_{\pi^1\in\Pi^1_{Ad}} \mathscr{J}(i,\tilde{c}_n,\pi^1,\pi^{*2}). \tag{5.4.23}$$

Also,

$$\hat{\psi}_n(i) = \sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^{*2}} \left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_1)-1}(\tilde{c}_n(\xi_t,\pi^1(t),\pi^{*2}(\xi_t))-\hat{\rho}_n)} \hat{\psi}_n(\xi_{\hat{\tau}(\mathscr{B}_1)})\right], \quad i \in \mathscr{B}_1^c, \tag{5.4.24}$$

for some finite set $\mathscr{B}_1$ containing $\mathscr{B}$.

Now as in Lemma 5.4.1, we have a finite set $\tilde{\mathscr{B}}_1$, depending on n, containing $\mathscr{K}$ such that:

- Under Assumption $(A1)$(a): From (5.4.23), we have $\hat{\rho}_n \leq \|\tilde{c}_n\|_\infty$. So, from the above definition of $\tilde{c}_n$, for $i \in \mathscr{D}_n^c$, we have $\tilde{c}_n(i,a,b) - \hat{\rho}_n \geq 0$ for all $(a,b) \in A(i) \times B(i)$. Consequently, we may take $\tilde{\mathscr{B}}_1 = \mathscr{D}_n$ such that $\tilde{c}_n(i,a,b) - \hat{\rho}_n \geq 0$ in $\tilde{\mathscr{B}}_1^c$ for all $(a,b) \in A(i) \times B(i)$.

- Under Assumption $(A1)$(b): since $\tilde{c}_n$ is norm-like function, we can choose suitable finite set $\tilde{\mathscr{B}}_1$ such that $(\tilde{c}_n(i,a,b) - \hat{\rho}_n) \geq 0$ in $\tilde{\mathscr{B}}_1^c$ for all $(a,b) \in A(i) \times B(i)$.

From (5.4.22), we obtain

$$\hat{\psi}_n(i) \geq \left[e^{(\tilde{c}_n(i,\mu,\pi^{*2}(i)) - \hat{\rho}_n)} \sum_{j \in S} \hat{\psi}_n(j) P(j|i,\mu,\pi^{*2}(i)) \right]. \tag{5.4.25}$$

By Dynkin's formula from (5.4.25), we deduce that

$$\hat{\psi}_n(i) \geq E_i^{\pi^1,\pi^{*2}} \left[e^{\sum_{t=0}^{\hat{\tau}(\tilde{\mathscr{B}}_1)\wedge N-1} (\tilde{c}_n(\xi_t,\pi^1(t),\pi^{*2}(\xi_{t-})) - \hat{\rho}_n)} \hat{\psi}_n(\xi_{\check{\tau}(\tilde{\mathscr{B}}_1)\wedge N}) \right].$$

Since $\tilde{c}_n(i,a,b) - \hat{\rho}_n \geq 0$, in $\tilde{\mathscr{B}}_1^c$, for all $(a,b) \in A(i) \times B(i)$, by Fatou lemma taking $N \to \infty$, we obtain

$$\hat{\psi}_n(i) \geq E_i^{\pi^1,\pi^{*2}} \left[e^{\sum_{t=0}^{\hat{\tau}(\tilde{\mathscr{B}}_1)-1} (\tilde{c}_n(\xi_t,\pi^1(t),\pi^{*2}(\xi_t)) - \hat{\rho}_n)} \hat{\psi}_n(\xi_{\hat{\tau}(\tilde{\mathscr{B}}_1)}) \right] \geq \min_{\tilde{\mathscr{B}}_1} \hat{\psi}_n \ \forall\ i \in \tilde{\mathscr{B}}_1^c.$$

So, $\hat{\psi}_n$ has a lower bound. Again by Dynkin's formula from (5.4.22), we get

$$\hat{\psi}_n(i) \geq E_i^{\pi^1,\pi^{*2}} \left[e^{\sum_{t=0}^{T\wedge\tau(\mathscr{D}_n)-1} (\tilde{c}_n(\xi_t,\pi^1(t),\pi^{*2}(\xi_t)) - \hat{\rho}_n)} \hat{\psi}_n(\xi_{T\wedge\tau(\mathscr{D}_n)}) \right].$$

By Fatou's lemma, taking $m \to \infty$, we obtain

$$\begin{aligned} \hat{\psi}_n(i) &\geq E_i^{\pi^1,\pi^{*2}} \left[e^{\sum_{t=0}^{T-1} (\tilde{c}_n(\xi_t,\pi^1(t),\pi^{*2}(\xi_t)) - \hat{\rho}_n)} \hat{\psi}_n(\xi_T) \right] \\ &\geq (\min_{\tilde{\mathscr{B}}_1} \hat{\psi}_n) E_i^{\pi^1,\pi^{*2}} \left[e^{\sum_{t=0}^{T-1} (\tilde{c}_n(\xi_t,\pi^1(t),\pi^{*2}(\xi_t)) - \hat{\rho}_n)} \right]. \end{aligned}$$

So, taking logarithm both sides, dividing by T and letting $T \to \infty$, we deduce that

$$\hat{\rho}_n \geq \mathscr{J}(i, \tilde{c}_n, \pi^1, \pi^{*2}).$$

Since $\pi^1 \in \Pi_{Ad}^1$ is arbitrary,

$$\hat{\rho}_n \geq \sup_{\pi^1 \in \Pi_{Ad}^1} \mathscr{J}(i, \tilde{c}_n, \pi^1, \pi^{*2}) \geq \sup_{\pi^1 \in \Pi_{Ad}^1} \mathscr{J}(i, c, \pi^1, \pi^{*2}).$$

Using this and (5.4.23), we get $\sup_{\pi^1\in\Pi^1_{Ad}} \mathscr{J}(i,c,\pi^1,\pi^{*2}) \le \sup_{\pi^1\in\Pi^1_{Ad}} \mathscr{J}(i,\tilde{c}_n,\pi^1,\pi^{*2}) = \hat{\rho}_n$ for all n. Now, by suitable scaling as in the proof of Lemma 5.4.1, it is easy to see that $\hat{\psi}_n \le V$ and it touches V. Also, we note from the definition of $\tilde{c}_n$ that $\hat{\rho}_n$ is a monotone decreasing sequence bounded below. Thus, using diagonalization arguments, there exists a subsequence (denoting the same sequence) and a pair $(\hat{\rho},\hat{\psi})$, $\hat{\psi} > 0$ such that $\hat{\rho}_n \to \hat{\rho}$ and $\hat{\psi}_n \to \hat{\psi}$ as $n\to\infty$. Now arguing as in the proof of Lemma 5.4.1, taking $n\to\infty$ in (5.4.22), we get

$$e^{\hat{\rho}}\hat{\psi}(i) = \sup_{\mu\in\mathcal{P}(A(i))}\left[e^{c(i,\mu,\pi^{*2}(i))}\sum_{j\in S}\hat{\psi}(j)P(j|i,\mu,\pi^{*2}(i))\right]. \tag{5.4.26}$$

Also, we have $\lim_{n\to\infty}\hat{\rho}_n = \hat{\rho} \ge \sup_{\pi^1\in\Pi^1_{Ad}}\mathscr{J}(i,c,\pi^1,\pi^{*2}) \ge \inf_{\pi^2\in\Pi^2_{Ad}}\sup_{\pi^1\in\Pi^1_{Ad}}\mathscr{J}(i,c,\pi^1,\pi^2) \ge \rho^*$.

We want to show that $\hat{\rho} = \rho^*$. By continuity-compactness assumptions, there exists $\hat{\pi}^{*1}$ such that (5.4.26) can be rewritten as

$$e^{\hat{\rho}}\hat{\psi}(i) = \left[e^{c(i,\hat{\pi}^{*1}(i),\pi^{*2}(i))}\sum_{j\in S}\hat{\psi}(j)P(j|i,\hat{\pi}^{*1}(i),\pi^{*2}(i))\right]. \tag{5.4.27}$$

By Dynkin's formula, for some $\mathscr{B}_2$ containing $\mathscr{B}$, we have

$$\hat{\psi}(i) = E_i^{\hat{\pi}^{*1},\pi^{*2}}\left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_2)\wedge N-1}(c(\xi_t,\hat{\pi}^{*1}(\xi_t),\pi^{*2}(\xi_t))-\hat{\rho})}\hat{\psi}(\xi_{\hat{\tau}(\mathscr{B}_2)\wedge N})\right],\ \forall i\in\mathscr{B}_2^c. \tag{5.4.28}$$

Using the estimates of Lemma 5.3.1 and the dominated convergence theorem, taking $N\to\infty$ in (5.4.28), we obtain

$$\hat{\psi}(i) = E_i^{\hat{\pi}^{*1},\pi^{*2}}\left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_2)-1}(c(\xi_t,\hat{\pi}^{*1}(\xi_t),\pi^{*2}(\xi_t))-\hat{\rho})}\hat{\psi}(\xi_{\hat{\tau}(\mathscr{B}_2)})\right],\ \forall i\in\mathscr{B}_2^c. \tag{5.4.29}$$

Since $\hat{\rho}\ge\rho^*$, from (5.4.8) we have

$$\psi^*(i) \ge E_i^{\hat{\pi}^{*1},\pi^{*2}}\left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_2)-1}(c(\xi_t,\hat{\pi}^{*1}(\xi_t),\pi^{*2}(\xi_t))-\hat{\rho})}\psi^*(\xi_{\hat{\tau}(\mathscr{B}_2)})\right]\ \forall i\in\mathscr{B}_2^c. \tag{5.4.30}$$

Hence, from (5.4.29) and (5.4.30), it follows that

$$\psi^*(i)-\hat{k}_1\hat{\psi}(i) \ge E_i^{\hat{\pi}^{*1},\pi^{*2}}\left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_2)-1}(c(\xi_t,\hat{\pi}^{*1}(\xi_t),\pi^{*2}(\xi_t))-\hat{\rho})}(\psi^*-\hat{k}_1\hat{\psi})(\xi_{\hat{\tau}(\mathscr{B}_2)})\right]\ \forall i\in\mathscr{B}_2^c. \tag{5.4.31}$$

Let $\hat{k}_1 = \min_{\mathscr{B}_2}\frac{\psi^*}{\hat{\psi}}$, thus we have $(\psi^*-\hat{k}_1\hat{\psi})\ge 0$ in $\mathscr{B}_2$ and for some $\hat{i}_0\in\mathscr{B}_2$, $(\psi^*-\hat{k}_1\hat{\psi})(\hat{i}_0)=0$. Therefore, from (5.4.31), we obtain that $(\psi^*-\hat{k}_1\hat{\psi})\ge 0$ in S. Now since $\hat{\rho}\ge\rho^*$, from (5.4.20) and (5.4.27), we deduce that

$$e^{\hat{\rho}}(\psi^*-\hat{k}_1\hat{\psi})(\hat{i}_0) \ge \left[e^{c(\hat{i}_0,\hat{\pi}^{*1}(i_0),\pi^{*2}(i_0))}\sum_{j\in S}(\psi^*-\hat{k}_1\hat{\psi})(j)P(j|\hat{i}_0,\hat{\pi}^{*1}(\hat{i}_0),\pi^{*2}(\hat{i}_0))\right].$$

This gives us

$$0 = \sum_{j \in S} (\psi^* - \hat{k}_1 \hat{\psi})(j) P(j|\hat{i}_0, \hat{\pi}^{*1}(\hat{i}_0), \pi^{*2}(\hat{i}_0)). \tag{5.4.32}$$

Thus, in view of the irreducibility property of the Markov chain under stationary Markov strategies, it follows that $\psi^* = \hat{k}_1 \hat{\psi}$ in S. Hence from (5.4.20) and (5.4.26), it is easy to see that $\hat{\rho} = \rho^*$ for all $i \in S$. Therefore, we obtain

$$\begin{aligned} \rho^* &= \inf_{i \in S} \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} \mathscr{J}(i, c, \pi^1, \pi^2) = \inf_{i \in S} \inf_{\pi^2 \in \Pi^2_{Ad}} \sup_{\pi^1 \in \Pi^1_{Ad}} \mathscr{J}(i, c, \pi^1, \pi^2) \\ &= \inf_{i \in S} \sup_{\pi^1 \in \Pi^1_{Ad}} \mathscr{J}(i, c, \pi^1, \pi^{*2}) \geq \inf_{i \in S} \mathscr{J}(i, c, \pi^{*1}, \pi^{*2}). \end{aligned} \tag{5.4.33}$$

Now arguing as in [21, Lemma 2.6], it follows that for $\pi^{*1} \in \Pi^1_{SM}$, there exists $(\psi', \rho') \in L^\infty_V(S) \times \mathbb{R}_+$, $\psi' > 0$ satisfying

$$e^{\rho'} \psi'(i) = \inf_{\nu \in \mathcal{P}(B(i))} \left[e^{c(i, \pi^{*1}(i), \nu)} \sum_{j \in S} \psi'(j) P(j|i, \pi^{*1}(i), \nu) \right], \tag{5.4.34}$$

with

$$\rho' = \inf_{i \in S} \inf_{\pi^2 \in \Pi^2_{Ad}} \mathscr{J}(i, c, \pi^{*1}, \pi^2). \tag{5.4.35}$$

Thus, we have

$$\rho' = \inf_{i \in S} \inf_{\pi^2 \in \Pi^2_{Ad}} \mathscr{J}(i, c, \pi^{*1}, \pi^2) \leq \inf_{i \in S} \mathscr{J}(i, c, \pi^{*1}, \pi^{*2}) \leq \rho^*. \tag{5.4.36}$$

For any minimizing selector $\tilde{\pi}^{*2}$ of (5.4.34), we obtain

$$e^{\rho'} \psi'(i) = \left[e^{c(i, \pi^{*1}(i), \tilde{\pi}^{*2}(i))} \sum_{j \in S} \psi'(j) P(j|i, \pi^{*1}(i), \tilde{\pi}^{*2}(i)) \right]. \tag{5.4.37}$$

Also, arguing as in Lemma 5.4.1, for some finite set $\mathscr{B}_3 \supset \mathscr{B}$, we deduce that

$$\psi'(i) = E_i^{\pi^{*1}, \tilde{\pi}^{*2}} \left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_3)-1} (c(\xi_t, \pi^{*1}(\xi_t), \tilde{\pi}^{*2}(\xi_t)) - \rho') dt} \psi'(\xi_{\hat{\tau}(\mathscr{B}_3)}) \right], \quad i \in \mathscr{B}_3^c. \tag{5.4.38}$$

From (5.4.20), we have

$$e^{\rho^*} \psi^*(i) \leq \left[e^{c(i, \pi^{*1}(i), \tilde{\pi}^{*2}(i))} \sum_{j \in S} \psi^*(j) P(j|i, \pi^{*1}(i), \tilde{\pi}^{*2}(i)) \right], \quad i \in S\,. \tag{5.4.39}$$

Also, from (5.4.8), it follows that

$$\psi^*(i) \leq E_i^{\pi^{*1}, \tilde{\pi}^{*2}} \left[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_3)-1} (c(\xi_t, \pi^{*1}(\xi_t), \tilde{\pi}^{*2}(\xi_t)) - \rho^*) dt} \psi^*(\xi_{\hat{\tau}(\mathscr{B}_3)}) \right] \; \forall i \in \mathscr{B}_3^c\,. \tag{5.4.40}$$

Therefore, by analogous arguments as above, using the irreducibility property of the Markov chain, we get $\psi^{'} = \hat{k}_2\psi^*$, for some positive constant $\hat{k}_2$. Thus, from (5.4.20) and (5.4.34), it follows that

$$\rho^* = \rho^{'}. \tag{5.4.41}$$

Hence, by (5.4.33), (5.4.36) and (5.4.41), we obtain (5.4.21). This completes the proof of the theorem. ☐

Now we prove the uniqueness of the eigenpair of the optimality equation (5.4.6) in the space $\mathbb{R}_+ \times L^\infty_V(S)$.

Lemma 5.4.2. *Suppose that Assumptions (A1) and (A2) hold. Then the eigenpair* $(\rho^*, \psi^*) \in \mathbb{R}_+ \times L^\infty_V(S)$ *is a unique solution of (5.4.6) (upto a scalar multiplication).*

Proof. Let $(\tilde{\rho}, \tilde{\psi}) \in \mathbb{R}_+ \times L^\infty_V(S)$, $\tilde{\psi} > 0$ be another solution of (5.4.6). Then using Fan's minimax theorem and (5.4.6), we get

$$\begin{aligned} e^{\tilde{\rho}}\tilde{\psi}(i) &= \sup_{\mu\in\mathcal{P}(A(i))} \inf_{\nu\in\mathcal{P}(B(i))} \Big[e^{c(i,\mu,\nu)} \sum_{j\in S} \tilde{\psi}(j) P(j|i,\mu,\nu) \Big] \\ &= \inf_{\nu\in\mathcal{P}(B(i))} \sup_{\mu\in\mathcal{P}(A(i))} \Big[e^{c(i,\mu,\nu)} \sum_{j\in S} \tilde{\psi}(j) P(j|i,\mu,\nu) \Big], \quad i \in S. \end{aligned} \tag{5.4.42}$$

There exists an outer minimizing selector $\tilde{\pi}^{*2} \in \Pi^2_{SM}$ such that (5.4.42) can be written

$$e^{\tilde{\rho}}\tilde{\psi}(i) = \sup_{\mu\in\mathcal{P}(A(i))} \Big[e^{c(i,\mu,,\tilde{\pi}^{*2}(i))} \sum_{j\in S} \tilde{\psi}(j) P(j|i,\mu,\tilde{\pi}^{*2}(i)) \Big], \quad i \in S. \tag{5.4.43}$$

We claim that $\rho^* = \tilde{\rho}$. If possible let us assume that $\tilde{\rho} < \rho^*$. Now from (5.4.8), for some finite set $\mathscr{B}_4 \supset \mathscr{B}$, we get

$$\psi(i) \leq E_i^{\pi^{*1},\tilde{\pi}^{*2}} \Big[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_4)-1}(c(\xi_t,\pi^{*1}(\xi_t),\tilde{\pi}^{*2}(\xi_t))-\rho^*)} \psi(\xi_{\hat{\tau}(\mathscr{B}_4)}) \Big] \ \forall i \in \mathscr{B}_4^c, \tag{5.4.44}$$

where π^{*1} is an outer maximizing selector of (5.4.6) as in (5.4.20). Since $\rho^* > \tilde{\rho}$, in view of (5.4.18), by Dynkin's formula and Fatou's lemma from (5.4.43), we deduce that

$$\tilde{\psi}(i) \geq E_i^{\pi^{*1},\tilde{\pi}^{*2}} \Big[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_4)-1}(c(\xi_t,\pi^{*1}(\xi_t),\tilde{\pi}^{*2}(\xi_t))-\rho^*)} \tilde{\psi}(\xi_{\hat{\tau}(\mathscr{B}_4)}) \Big] \ \forall i \in \mathscr{B}_4^c. \tag{5.4.45}$$

Therefore, by analogous arguments as above (see (5.4.29)-(5.4.32) or (5.4.37)-(5.4.41)), using irreducibility property of the Markov chain and (5.4.6), (5.4.43), (5.4.44) and (5.4.45) and by taking $\hat{k}_3 = \min_{\mathscr{B}_4} \frac{\tilde{\psi}}{\psi^*}$, we get $\tilde{\psi} = \hat{k}_3\psi^*$. Thus, from (5.4.6) and (5.4.42), it follows that $\rho^* = \tilde{\rho}$. Hence we arrive at a contradiction and it contradicts the fact

that $\tilde{\rho} < \rho^*$. Thus, we obtain $\rho^* \leq \tilde{\rho}$.
Next, if possible let $\rho^* < \tilde{\rho}$. Then by the analogous arguments, (by taking the outer maximizer of (5.4.42) and outer minimizer of (5.4.6)), we will arrive at a contradiction to the fact that $\rho^* < \tilde{\rho}$. Hence, we deduce that

$$\rho^* = \tilde{\rho} \text{ and } \psi^* = \hat{k}_4 \tilde{\psi}, \tag{5.4.46}$$

for some positive constant $\hat{k}_4$.
In particular, this implies that the eigenpair of (5.4.6) is unique upto a scalar multiplication. ☐

Remark 5.4.3. *In deriving (5.4.34), we used [21, Lemma 2.6]. It should be noted that the results of the paper [21] can be derived by using the assumptions of this chapter (see [21, Remark 2.3]).*

Next, we prove the converse of the above theorem. That is, any saddle-point equilibrium of our game problem will be a mini-max selector of the associated optimality equation.

Theorem 5.4.3. *Suppose Assumptions (A1) and (A2) hold. Suppose there exists a saddle-point equilibrium* $(\hat{\pi}^{*1}, \hat{\pi}^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$, *i.e., for all* $i \in S$,

$$\begin{aligned} &\mathscr{J}(i,c,\hat{\pi}^{*1},\hat{\pi}^{*2}) \leq \mathscr{J}(i,c,\hat{\pi}^{*1},\pi^2), \text{ for all } \pi^2 \in \Pi^2_{Ad}, \\ &\mathscr{J}(i,c,\hat{\pi}^{*1},\hat{\pi}^{*2}) \geq \mathscr{J}(i,c,\pi^1,\hat{\pi}^{*2}), \text{ for all } \pi^1 \in \Pi^1_{Ad}. \end{aligned} \tag{5.4.47}$$

Then $(\hat{\pi}^{*1}, \hat{\pi}^{*2})$ *is a mini-max selector of (5.4.6).*

Proof. By Theorem 5.4.2 and (5.4.47), we have

$$\begin{aligned} \rho^* = \inf_{\pi^2\in\Pi^2_{Ad}} \sup_{\pi^1\in\Pi^1_{Ad}} \mathscr{J}(i,c,\pi^1,\pi^2) &\leq \sup_{\pi^1\in\Pi^1_{Ad}} \mathscr{J}(i,c,\pi^1,\hat{\pi}^{*2}) \leq \mathscr{J}(i,c,\hat{\pi}^{*1},\hat{\pi}^{*2}) \\ &\leq \inf_{\pi^2\in\Pi^2_{Ad}} \mathscr{J}(i,c,\hat{\pi}^{*1},\pi^2) \leq \sup_{\pi^1\in\Pi^1_{Ad}} \inf_{\pi^2\in\Pi^2_{Ad}} \mathscr{J}(i,c,\pi^1,\pi^2) \\ &= \rho^*. \end{aligned}$$

This implies that $\rho^* = \mathscr{J}(i,c,\hat{\pi}^{*1},\hat{\pi}^{*2}) = \sup_{\pi^1\in\Pi^1_{Ad}} \mathscr{J}(i,c,\pi^1,\hat{\pi}^{*2}) = \inf_{\pi^2\in\Pi^2_{Ad}} \mathscr{J}(i,c,\hat{\pi}^{*1},\pi^2)$.
Now arguing as in Lemma 5.4.1 and Theorem 5.4.2, it follows that for $\hat{\pi}^{*1} \in \Pi^1_{SM}$ there exists $(\rho^{\hat{\pi}^{*1}}, \psi^*_{\hat{\pi}^{*1}}) \in \mathbb{R}_+ \times L^\infty_V$ with $\psi^*_{\hat{\pi}^{*1}} > 0$ such that

$$e^{\rho^{\hat{\pi}^{*1}}} \psi^*_{\hat{\pi}^{*1}}(i) = \inf_{\nu\in\mathcal{P}(B(i))} \Big[e^{c(i,\hat{\pi}^{*1}(i),\nu)} \sum_{j\in S} \psi^*_{\hat{\pi}^{*1}}(j) P(j|i,\hat{\pi}^{*1}(i),\nu) \Big], \tag{5.4.48}$$

and $\rho^{\hat{\pi}^{*1}} = \inf_{\pi^2 \in \Pi^2_{Ad}} \mathscr{J}(i, c, \hat{\pi}^{*1}, \pi^2) = \rho^*$. Thus for π^{*2} as in (5.4.20), we have

$$e^{\rho^{\hat{\pi}^{*1}}} \psi^*_{\hat{\pi}^{*1}}(i) \leq \Big[e^{c(i,\hat{\pi}^{*1}(i),\pi^{*2}(i))} \sum_{j \in S} \psi^*_{\hat{\pi}^{*1}}(j) P(j|i, \hat{\pi}^{*1}(i), \pi^{*2}(i)) \Big]. \tag{5.4.49}$$

Arguing as in Lemma 5.4.1, for some finite set $\mathscr{B}_5 \supset \mathscr{B}$ it follows that

$$\psi^*_{\hat{\pi}^{*1}}(i) \leq E_i^{\hat{\pi}^{*1},\pi^{*2}} \Big[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_5)-1} (c(\xi_t, \hat{\pi}^{*1}(\xi_t), \pi^{*2}(\xi_t)) - \rho^*)} \psi^*_{\hat{\pi}^{*1}}(\xi_{\hat{\tau}(\mathscr{B}_5)}) \Big] \ \forall i \in \mathscr{B}_5^c. \tag{5.4.50}$$

Also, from (5.4.20), we deduce that

$$e^{\rho^*} \psi^*(i) \geq \Big[e^{c(i,\hat{\pi}^{*1}(i),\pi^{*2}(i))} \sum_{j \in S} \psi^*(j) P(j|i, \hat{\pi}^{*1}(i), \pi^{*2}(i)) \Big]. \tag{5.4.51}$$

By Dynkin's formula and Fatou's lemma (as in Lemma 5.4.1), we obtain

$$\psi^*(i) \geq E_i^{\hat{\pi}^{*1},\pi^{*2}} \Big[e^{\sum_{t=0}^{\hat{\tau}(\mathscr{B}_5)-1} (c(\xi_t, \hat{\pi}^{*1}(\xi_t), \pi^{*2}(\xi_t)) - \rho^*)} \psi^*(\xi_{\hat{\tau}(\mathscr{B}_5)}) \Big] \ \forall i \in \mathscr{B}_5^c. \tag{5.4.52}$$

Now, in view of (5.4.50) and (5.4.52) and applying the same technique as before (as in the proof of Theorem 5.4.2), it follows that $\psi^* = \hat{k}_5 \psi^*_{\hat{\pi}^{*1}}$, for some constant $\hat{k}_5 > 0$. Hence from (5.4.20) and (5.4.48), it is easy to see that $\hat{\pi}^{*1}$ is an outer maximizing selector of (5.4.6). Similarly, one can show that $\hat{\pi}^{*2}$ is an outer minimizing selector of (5.4.6). This completes the proof. □

Now we are ready to prove Theorem 5.4.1.

Proof of Theorem 5.4.1:

Proof. Existence of an eigenpair (ρ^*, ψ^*) of eq. (5.4.1) follows from Lemma 5.4.1. Uniqueness of the eigenpair of equation (5.4.1) is proved in Lemma 5.4.2. Also, from Theorem 5.4.2, we have Theorem 5.4.1 (i) and Theorem 5.4.1 (ii). Theorem 5.4.1 (iii) follows from Theorem 5.4.3. This completes the proof. □

5.5 Example

We present here an illustrative example in which all our assumptions hold, and the cost function is nonnegative and unbounded.

Example 5.5.1. *Consider a controlled birth-and-death system in which the state variable stands for the total population size at time* $t \geq 0$. *Thus, the state space can be represented by* $S := \{0, 1, 2, \cdots\}$. *Suppose that there are two players, player 1 and player 2, and they can control death and birth, respectively. Depending on the number of populations* *in the system, player 1 can modify the number of deaths by choosing some action* a*, from the set* $A(i) = [\delta, L_1]$*. But this action results in a cost given by* $\tilde{c}_1(i,a) \geq 0$ *(or a reward* $\tilde{c}_1(i,a) \leq 0$*), if* i *is the state of the system. On the other hand, player 2 can modify the number of births by choosing some action* b *from the set* $B(i) = [\delta, L_1]$*. The action of player 2 incurs a cost given by* $\tilde{c}_2(i,b) \geq 0$ *(or a reward* $\tilde{c}_2(i,b) \leq 0$*). Also, in addition, assume that player 1 'owns' the system and he/she gets a reward* $r(i) := \hat{p} \cdot i$ *for each unit of time during which the system remains in the state* $i \in S$*, where* $\hat{p} > 0$ *is a fixed cost per population. We assume that player 1 is maximizing player and player 2 is minimizing player. Player 1 gets an immediately reward rate* $c(i,a,b)$ *and he/she tries to maximize his/her total payoff through the infinite horizon discrete-time ergodic cost criterion and at the same time, player 2 incurs an immediately cost rate* $c(i,a,b)$ *and he/she tries to minimize his/her total payoff through the same cost criterion.*

We next formulate this model as a discrete-time Markov game. The corresponding transition stochastic kernel $P(j|i,a,b)$ *and reward* $c(i,a,b)$ *for player 1 are given as follows: for* $(0,a,b) \in K$ *(*K *as in the game model (5.2.1)).*

$$\sum_{j \in S} P(j|0,a,b) = 1, \ \textit{and } P(j|0,a,b) = e^{-\frac{j^2}{3}-3} \ \forall \ j \geq 1. \tag{5.5.1}$$

Similarly, for $(1,a,b) \in K$*,*

$$P(j|1,a,b) = \begin{cases} 1 - \frac{3e^{-2}b}{2(L_1+L_2)}, & \textit{if } j = 0 \\ \frac{e^{-2}b}{2(L_1+L_2)}, & \textit{if } j = 1 \\ \frac{e^{-2}b}{2(L_1+L_2)} & \textit{if } j = 2 \\ \frac{e^{-2}b}{2(L_1+L_2)}, & \textit{if } j = 3 \\ 0, & \textit{otherwise.} \end{cases}$$

Also, for $(i,a,b) \in K$ *with* $i \geq 2$*,*

$$P(j|i,a,b) = \begin{cases} \frac{ae^{-i}}{2(L_1+L_2)}, & \textit{if } j = i-1 \\ \frac{ae^{-i}+be^{-2i}}{2(L_1+L_2)} & \textit{if } j = i \\ \frac{be^{-2i}}{2(L_1+L_2)}, & \textit{if } j = i+1 \\ 1 - \frac{2(ae^{-i}+be^{-2i})}{2(L_1+L_2)}, & \textit{if } j = 0 \\ 0, & \textit{otherwise.} \end{cases}$$

$$c(i,a,b) := \hat{p} \cdot i - \tilde{c}_1(i,a) + \tilde{c}_2(i,b) \quad \textit{for} \quad (i,a,b) \in K. \tag{5.5.2}$$

We make the following assumptions to ensure the existence of a pair of optimal strategies.

(I) *The functions* $\tilde{c}_1(i,a)$*, and* $\tilde{c}_2(i,b)$ *are continuous with their respective variables for each fixed* $i \in S$*.*

(II) *Suppose that* $\hat{p} \cdot i - \tilde{c}_1(i,a) + \tilde{c}_2(i,b) \geq 0$ *for* $(i,a,b) \in K$ *and* $\hat{p} < \frac{1}{6}$. *Also, assume that* f*, is a norm-like function, where* $f(i) := \min_{(a,b)\in A(i)\times B(i)} [\tilde{c}_1(i,a) - \tilde{c}_2(i,b)]$ *for all* $i \in S$.

(III) *We also consider an increasing sequence of finite subsets* $\mathscr{D}_n \subset S$ *such that* $\cup_{n=1}^{\infty} \mathscr{D}_n = S$ *and* $0 \in \mathscr{D}_n$ *for all* $n \in \mathbb{N}$. *We assume that for any pair* $i, j \in \mathscr{D}_n$*, the probability of hitting* j *from* i *before exiting* $\mathscr{D}_n$ *is bounded from below by some constant* $\delta_{ij,n} > 0$ *for any stationary Markov strategy.*

Proposition 5.5.2. *Under conditions (I)-(III), the above controlled system satisfies the Assumptions* (A1) *and* (A2). *Hence by Theorem 5.4.1, there exists a saddle-point equilibrium for this controlled model.*

Proof. Consider the Lyapunov function $V(i) := e^{\frac{i^2}{6}+1}$ for $i \in S$. Then $V(i) \geq 1$ for all $i \in S$. Now for each $i \geq 2$, and $(a,b) \in A(i) \times B(i)$, we have

$$
\begin{aligned}
&\sum_{j\in S} P(j|i,a,b)V(j) \\
&= P(i-1|i,a,b)V(i-1) + P(i|i,a,b)V(i) + P(i+1|i,a,b)V(i+1) + P(0|i,a,b)V(0) \\
&= \frac{1}{2(L_1+L_2)}\left[ae^{-i}e^{\frac{(i-1)^2}{6}+1} + e^{\frac{i^2}{6}+1}\left(ae^{-i}+be^{-2i}\right) + be^{-2i}e^{\frac{(i+1)^2}{6}+1}\right] + e\left(1 - \frac{2(ae^{-i}+be^{-2i})}{2(L_1+L_2)}\right) \\
&= e^{\frac{i^2}{6}+1}\left[\frac{ae^{-i}}{2(L_1+L_2)}e^{-\frac{i}{3}+\frac{1}{6}} + \left(\frac{ae^{-i}+be^{-2i}}{2(L_1+L_2)}\right) + \frac{be^{-2i}}{2(L_1+L_2)}e^{\frac{i}{3}+\frac{1}{6}} + e^{-\frac{i^2}{6}}\left(1 - \frac{2(ae^{-i}+be^{-2i})}{2(L_1+L_2)}\right)\right] \\
&\leq e^{\frac{i^2}{6}+1}e^{-\frac{i}{3}+\frac{1}{6}}\left[\frac{ae^{-i}}{2(L_1+L_2)} + \frac{a+b}{2(L_1+L_2)} + \frac{b}{2(L_1+L_2)} + \left(1 - \frac{2(ae^{-i}+be^{-2i})}{2(L_1+L_2)}\right)\right] \\
&\leq 4e^{\frac{i^2}{6}+1}e^{-\frac{i}{3}+\frac{1}{6}} \\
&\leq e^{(\frac{i^2}{6}+1)-\frac{1}{3}(i+3)+4} \\
&= V(i)e^{-\frac{1}{6}(i+3)-\frac{1}{6}(i+3)+4} \\
&\leq e^{-\frac{1}{6}(i+3)+4I_{\mathscr{M}}(i)}V(i) \leq e^{-\frac{1}{6}(i+3)}V(i) + \max_{j\in\mathscr{M}} V(j)e^4 I_{\mathscr{M}}(i) \leq e^{-\ell(i)}V(i) + \tilde{C}I_{\mathscr{M}}(i),
\end{aligned}
\tag{5.5.3}
$$

where $\ell(i) = \frac{1}{6}(i+3)$, $\mathscr{M} := \{i : 4 - \frac{1}{6}(i+3) > 0\}$, and $\tilde{C} = \max\{\max_{j\in\mathscr{M}} V(j)e^4, e^{-2}\sum_{i\geq 1}(e^{-\frac{i^2}{6}} - e^{-\frac{i^2}{3}}) + e\}$. It is clear that $0, 1 \in \mathscr{M}$. Also, we have

$$
\sum_{j\in S} P(j|0,a,b)V(j) = eP(0|0,a,b) + \sum_{j\geq 1} e^{-2}e^{-\frac{j^2}{6}} \leq \tilde{C}I_{\mathscr{M}}(i). \tag{5.5.4}
$$

By similar arguments as in (5.5.3), we have

$$
\sum_{j\in S} P(j|1,a,b)V(j)
$$

$$
\begin{aligned}
&= P(0|1,a,b)V(0) + P(1|1,a,b)V(1) + P(2|1,a,b)V(2) + P(3|1,a,b)V(3)\\
&= e\left(1-\frac{3e^{-2}b}{2(L_1+L_2)}\right) + e^{\frac{1}{6}+1}\left(\frac{e^{-2}b}{2(L_1+L_2)}\right) + e^{\frac{4}{6}+1}\left(\frac{e^{-2}b}{2(L_1+L_2)}\right) + e^{\frac{9}{6}+1}\left(\frac{e^{-2}b}{2(L_1+L_2)}\right)\\
&\le e^{-\frac{1}{6}(1+3)+4I_{\mathscr{M}}(1)}V(1) \le e^{-\frac{2}{3}}V(1) + \max_{j\in\mathscr{M}} V(j)e^4 I_{\mathscr{M}}(1) \le e^{-\ell(1)}V(1) + \tilde{C} I_{\mathscr{M}}(1).
\end{aligned} \tag{5.5.5}
$$

Now

$$
\ell(i) - \max_{(a,b)\in A(i)\times B(i)} c(i,a,b) = \frac{1}{2} + (\frac{1}{6} - \hat{p})i + \min_{(a,b)\in A(\cdot)\times B(\cdot)} [\tilde{c}_1(i,a) - \tilde{c}_2(i,b)]. \tag{5.5.6}
$$

We see from condition (II) and (5.5.6) that $\ell(i) - \sup_{(a,b)\in A(i)\times B(i)} c(i,a,b)$ is norm-like function. So, by condition (II), equations (5.5.3), (5.5.4), (5.5.5), and (5.5.6), Assumption $(A1)$ is satisfied. Now, by the transition probability defined above and condition (III), Assumption $(A2)$(ii) is verified. Next, Assumption $(A2)$(iii) is verified by (5.5.3), (5.5.4) and (5.5.5). Also, by the above construction of probability kernel, (5.5.2), and condition (I), $P(\cdot|i,a,b)$ and $c(i,a,b)$ are continuous in $(a,b) \in A(i) \times B(i)$ for all $i,j \in S$. Hence by Theorem 5.4.1, it follows that there exists a saddle-point equilibrium for this controlled model. □

5.6 Eigenvalue problem for compact state space case:

In this section, we extend our results to compact state space S without assuming any Lyapunov type stability assumptions since compact state space eliminates the need for any stability consideration. To this end, let us first introduce some notations. Let $\mathscr{C}^+(S) := \{f \in \mathscr{C}(S) : f(x) \ge 0 \ \forall x \in S\}$ denotes the closed cone of $\mathscr{C}(S)$, where $\mathscr{C}(S)$ denotes the Banach space of continuous maps $f : \mathscr{C}(S) \to \mathbb{R}$ with the supremum norm, denoted by $\|\cdot\|$. Thus $\mathscr{C}^+(S)$ defines a partial order on $\mathscr{C}(S)$, denoted $\succeq$, given by this: for any $f, g \in \mathscr{C}(S)$, we define $f \succeq g$ if $f - g \in \mathscr{C}^+(S)$, i.e., the partial ordering in $\mathscr{C}(S)$ with respect to the cone $\mathscr{C}^+(S)$. We write $f \succ g$ (equivalently, $g \prec f$) if $f \succeq g$, $f \neq g$, and we write $f \gg g$ (equivalently, $g \ll f$) if $f - g$ is a strictly positive function in $\mathscr{C}(S)$ i.e., if $f - g \in \text{interior}(\mathscr{C}^+(S))$.

For our analysis, we need to impose the following set of assumptions on the system.

$(A1')$

(i) The admissible action spaces $A(x)(\subset A)$ and $B(x)(\subset B)$ are compact for each $x \in S$.

(ii) The functions $P(D|x,a,b)$ and $c : K \to \mathbb{R}$ are continuous in $(x,a,b) \in S \times A(x) \times B(x)$, $D \subseteq S$.

(iii) The maps $(x,a,b) \to \int_S f(y)P(dy|x,a,b)$, $f \in \mathscr{C}(S)$ with $\|f\| \leq 1$ are equicontinuous.

(iv) We assume that the transition kernel $P(dy|x,a,b)$ of the Markov chain $\{\xi_t\}_{t\geq 0}$ has the full support for all $(x,a,b) \in K$ i.e., $\text{support}(P(dy|x,a,b)) = S \; \forall (x,a,b) \in K$.

Now, using the nonlinear version of the Kreĭn-Rutman theorem, see [2, Theorem 2.2], [4], [92], [93], we establish the existence of an eigenpair to the associated Shapley equation.

Theorem 5.6.1. *Suppose Assumption* $(A1^{'})$ *holds. Then there exists a unique eigenpair* $(\rho^*, \psi^*) \in \mathbb{R}_+ \times \mathscr{C}^+(S)$, $\psi^* \in$ *interior*$(\mathscr{C}^+(S))$ *(unique up to a multiplicative constant) for the following nonlinear eigenequation*

$$\begin{aligned} e^{\rho^*}\psi^*(x) &= \inf_{\nu\in\mathcal{P}(B(x))}\sup_{\mu\in\mathcal{P}(A(x))}\left[e^{c(x,\mu,\nu)}\int_S \psi^*(y)P(dy|x,\mu,\nu)\right] \\ &= \sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\left[e^{c(x,\mu,\nu)}\int_S \psi^*(y)P(dy|x,\mu,\nu)\right]. \end{aligned} \tag{5.6.1}$$

Furthermore, for any mini-max selector $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ *of (5.6.1) we have the following:*

$$\begin{aligned} \rho^* &= \inf_{x\in S}\sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}}\mathscr{J}(x,c,\pi^1,\pi^2) = \inf_{x\in S}\inf_{\pi^2\in\Pi^2_{Ad}}\sup_{\pi^1\in\Pi^1_{Ad}}\mathscr{J}(x,c,\pi^1,\pi^2) \\ &= \inf_{x\in S}\inf_{\pi^2\in\Pi^2_{Ad}}\mathscr{J}(x,c,\pi^{*1},\pi^2) = \inf_{x\in S}\sup_{\pi^1\in\Pi^1_{Ad}}\mathscr{J}(x,c,\pi^1,\pi^{*2}) = \mathscr{J}(x,c,\pi^{*1},\pi^{*2}), \end{aligned} \tag{5.6.2}$$

and consequently $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ *is a saddle-point equilibrium.*

Proof. Let us consider a mapping $\hat{T} : \mathscr{C}(S) \to \mathscr{C}(S)$ defined by

$$\hat{T}g(x) = \sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\left[e^{c(x,\mu,\nu)}\int_S g(y)P(dy|x,\mu,\nu)\right], \tag{5.6.3}$$

where $g \in \mathscr{C}(S)$ and $x \in S$.

Note that in view of Remark 5.2.2, the sets $\mathcal{P}(A(x))$ and $\mathcal{P}(B(x))$ are compact as well as convex. Also, the extreme points of $\mathcal{P}(A(x))$ and $\mathcal{P}(B(x))$ corresponds to the Dirac measures at points in $A(x)$ and $B(x)$ respectively. Hence,

$$\begin{aligned} \hat{T}g(x) &= \sup_{\mu\in\mathcal{P}(A(x))}\inf_{\nu\in\mathcal{P}(B(x))}\left[e^{c(x,\mu,\nu)}\int_S g(y)P(dy|x,\mu,\nu)\right] \\ &= \sup_{a\in A(x)}\inf_{b\in B(x)}\left[e^{c(x,a,b)}\int_S g(y)P(dy|x,a,b)\right], \text{ for details, see [2, p. 965].} \end{aligned} \tag{5.6.4}$$

Next, we show that the map $\hat{T}$ is well defined on $\mathscr{C}(S)$. So, it is sufficient to take the family $\{g \in \mathscr{C}(S) : \|g\| \leq R\}$, for some $R > 0$. Let $x, z \in S$ be arbitrary but fixed points. Then in view of (5.6.4), we have

$$|\hat{T}g(x) - \hat{T}g(z)|$$
$$\leq e^{\|c\|} \sup_{b\in B} \sup_{a\in A} \sup_{g:\|g\|\leq R} \left| \int_S g(y)P(dy|x,a,b) - \int_S g(y)P(dy|z,a,b) \right| + R \sup_{b\in B} \sup_{a\in A} \left| e^{c(x,a,b)} - e^{c(z,a,b)} \right|. \tag{5.6.5}$$

In view of Assumption $(A1^{'})$, it follows that the right hand side tends to zero when $x \to z$. Next, from the definition of the map $\hat{T}$, for any $g_1, g_2 \in \mathscr{C}(S)$, it follows that $\|\hat{T}(g_1) - \hat{T}(g_2)\| \leq e^{\|c\|}\|g_1 - g_2\|$. Hence the map $\hat{T}$ is Lipschitz continuous map from $\mathscr{C}(S) \to \mathscr{C}(S)$.

Now, using Assumptions $(A1^{'})$, we prove the following properties of $\hat{T}$.

Let $g_1 \succ g_2$ i.e., $g_1 \succeq g_2$, $g_1 \neq g_2$. Let $\pi^{*2} \in \Pi^2_{SM}$ be such that

$$\sup_{\mu\in\mathcal{P}(A(x))} \inf_{\nu\in\mathcal{P}(B(x))} \left[e^{c(x,\mu,\nu)} \int_S g_1(y)P(dy|x,\mu,\nu) \right]$$
$$= \sup_{\mu\in\mathcal{P}(A(x))} \left[e^{c(x,\mu,\pi^{*2}(x))} \int_S g_1(y)P(dy|x,\mu,\pi^{*2}(x)) \right] \; \forall x \in S. \tag{5.6.6}$$

Also, let $\pi^{*1} \in \Pi^1_{SM}$ be such that

$$\sup_{\mu\in\mathcal{P}(A(x))} \left[e^{c(x,\mu,\pi^{*2}(x))} \int_S g_2(y)P(dy|x,\mu,\pi^{*2}(x)) \right]$$
$$= \left[e^{c(x,\pi^{*1}(x),\pi^{*2}(x))} \int_S g_2(y)P(dy|x,\pi^{*1}(x),\pi^{*2}(x)) \right] \; \forall x \in S. \tag{5.6.7}$$

Then

$$\hat{T}(g_1)(x) - \hat{T}(g_2)(x)$$
$$\geq \left[e^{c(x,\pi^{*1}(x),\pi^{*2}(x))} \int_S (g_1(y) - g_2(y))P(dy|x,\pi^{*1}(x),\pi^{*2}(x)) \right]$$
$$\geq e^{\alpha_1} \int_S (g_1(y) - g_2(y))P(dy|x,\pi^{*1}(x),\pi^{*2}(x)) > 0,$$

since $g_1 \succ g_2$ and support$(P(dy|x,\pi^{*1}(x),\pi^{*2}(x))) = S$, $\forall x \in S$, where $\alpha_1 > 0$ is the greatest lower bound of c on S. So, $\hat{T}$ is strictly increasing.

By the definition of the map $\hat{T}$, it is easy to see that $\hat{T}(\lambda g) = \lambda\hat{T}(g)$ for all $\lambda > 0$. For $M > e^{-\alpha_1}$ and $g \in \mathscr{C}^+(S)$ defined by $g(\cdot) \equiv 1$, $M\hat{T}g > g$. Using Assumption $(A1^{'})$, in view of (5.6.5), by analogous arguments as in [2, p. 967-968], it is easy to say that the map $\hat{T} : \mathscr{C}(S) \to \mathscr{C}(S)$ is a compact operator. Hence by Theorem 5.3.1, there exists a nontrival $\psi^* \in \mathscr{C}^+(S)$ and a constant $e^{\rho^*} > 0$ such that $\hat{T}\psi^* = e^{\rho^*}\psi^*$ i.e,

$$e^{\rho^*}\psi^*(x) = \sup_{\mu\in\mathcal{P}(A(x))} \inf_{\nu\in\mathcal{P}(B(x))} \left[e^{c(x,\mu,\nu)} \int_S \psi^*(y)P(dy|x,\mu,\nu) \right] \; \forall x \in S.$$

Thus, by Fan's minimax theorem [33], we have

$$e^{\rho^*}\psi^*(x) = \inf_{\nu\in\mathcal{P}(B(x))}\sup_{\mu\in\mathcal{P}(A(x))}\left[e^{c(x,\mu,\nu)}\int_S \psi^*(y)P(dy|x,\mu,\nu)\right] \; \forall x \in S\,.$$

This implies that the pair (ρ^*, ψ^*) satisfies (5.6.1).
Now, we claim that $\psi^* > 0$. If not, then on contrary there exists a point $\tilde{x} \in S$ for which $\psi^*(\tilde{x}) = 0$. Again by continuity-compactness assumptions, there exists a mini-max selector (π^{*1}, π^{*2}) such that (5.6.1) can be rewritten as

$$e^{\rho^*}\psi^*(x) = \left[e^{c(x,\pi^{*1}(x),\pi^{*2}(x))}\int_S \psi^*(y)P(dy|x,\pi^{*1}(x),\pi^{*2}(x))\right] \quad \forall x \in S.$$

So, we get

$$0 = e^{\rho^*}\psi^*(\tilde{x}) = \left[e^{c(\tilde{x},\pi^{*1}(\tilde{x}),\pi^{*2}(\tilde{x}))}\int_S \psi^*(y)P(dy|\tilde{x},\pi^{*1}(\tilde{x}),\pi^{*2}(\tilde{x}))\right]. \tag{5.6.8}$$

Since ψ^* is nontrivial, there exists $\hat{x} \in S$ such that $\psi^*(\hat{x}) > 0$.
Now since support$(P(dy|x,a,b)) = S \; \forall (x,a,b) \in K$, it follows that

$$\left[\int_S \psi^*(y)P(dy|\tilde{x},\pi^{*1}(\tilde{x}),\pi^{*2}(\tilde{x}))\right] > 0.$$

Hence it contradicts (5.6.8). This establishes our claim.

In view of Assumption $(A1^{'})$(iv), the uniqueness of the eigenpair of equation (5.6.1) can be proved easily by the analogous arguments as in [Lemma 5.4.2, eqs. (5.4.42)-(5.4.46)]. To see this, suppose that $(\tilde{\rho}, \tilde{\psi}) \in \mathbb{R}_+ \times \mathscr{C}^+(S)$, $\tilde{\psi} > 0$ is an another solution of (5.6.1) and if possible let $\tilde{\rho} < \rho^*$. Let $\hat{k}_6 = \min_S \frac{\tilde{\psi}}{\psi^*}$, thus we have $(\tilde{\psi} - \hat{k}_6\psi^*) \geq 0$ in S and for some $\hat{x}_0 \in S$, $\tilde{\psi}(\hat{x}_0) - \hat{k}_6\psi^*(\hat{x}_0) = 0$. Let π^{*1} and $\tilde{\pi}^{*2}$ are outer maximizing and minimizing selectors of (5.6.1), corresponding to the eigenpair $(\tilde{\rho}, \tilde{\psi})$ and (ρ, ψ), respectively. Thus we obtain

$$e^{\rho^*}(\tilde{\psi} - \hat{k}_6\psi^*)(x) \geq \left[e^{c(x,\pi^{*1}(x),\tilde{\pi}^{*2}(x))}\int_S (\tilde{\psi} - \hat{k}_6\psi^*)(y)P(dy|x,\pi^{*1}(x),\tilde{\pi}^{*2}(x))\right]. \tag{5.6.9}$$

This implies that

$$0 = \int_S (\tilde{\psi} - \hat{k}_6\psi^*)(y)P(dy|\hat{x}_0,\pi^{*1}(\hat{x}_0),\tilde{\pi}^{*2}(\hat{x}_0)).$$

Then we claim that $\tilde{\psi} \equiv \hat{k}_6\psi^*$. If not, then since support$(P(dy|x,a,b)) = S \; \forall (x,a,b) \in K$, it follows that

$$\int_S (\tilde{\psi} - \hat{k}_6\psi^*)(y)P(dy|\hat{x}_0,\pi^{*1}(\hat{x}_0),,\tilde{\pi}^{*2}(\hat{x}_0)) > 0. \tag{5.6.10}$$

So, we arrive at a contradiction and thus, we have $\tilde{\psi} \equiv \hat{k}_6\psi^*$. Hence we obtain $\rho^* = \tilde{\rho}$, which is a contradiction to the fact that $\tilde{\rho} < \rho^*$. This implies that $\rho^* \leq \tilde{\rho}$. Again, by similar arguments, one can derive the analogous contradiction to the fact that $\rho^* < \tilde{\rho}$. Therefore, we deduce that $\rho^* = \tilde{\rho}$ and $\tilde{\psi} \equiv \hat{k}_7\psi^*$, for some $\hat{k}_7 > 0$. So, the eigenpair of equation (5.6.1) is unique (up to a scalar multiplication).

Let $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ be a pair of outer mini-max selector of (5.6.1), satisfying

$$\begin{aligned} e^{\rho^*}\psi^*(x) &= \sup_{\mu \in \mathcal{P}(B(x))}\left[e^{c(x,\mu,\pi^{*2}(x))}\int_S \psi^*(y)P(dy|x,\mu,\pi^{*2}(x))\right] \\ &= \inf_{\nu \in \mathcal{P}(B(x))}\left[e^{c(x,\pi^{*1}(x),\nu)}\int_S \psi^*(y)P(dy|x,\pi^{*1}(x),\nu)\right]. \end{aligned} \tag{5.6.11}$$

Therefore, by Dynkin's formula and (5.6.11) (as in [113, Lemma 3.1]), we obtain

$$\begin{aligned} \psi^*(x) &\leq E_x^{\pi^{*1},\pi^2}\left[e^{\sum_{s=0}^{T-1}(c(\xi_s,\pi^{*1}(\xi_s),\pi^1(s))-\rho^*)}\psi^*(\xi_T)\right] \\ &\leq (\sup_S \psi^*)E_x^{\pi^{*1},\pi^2}\left[e^{\sum_{s=0}^{T-1}(c(\xi_s,\pi^{*1}(\xi_s),\pi^2(s))-\rho^*)}\right]. \end{aligned} \tag{5.6.12}$$

Taking logarithm on both sides, dividing by T and letting $T \to \infty$, we deduce that

$$\rho^* \leq \inf_{\pi^2 \in \Pi^2_{Ad}} \mathscr{J}(x,c,\pi^{*1},\pi^2) \leq \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} \mathscr{J}(x,c,\pi^1,\pi^2). \tag{5.6.13}$$

Similarly, by Dynkin's formula and using (5.6.11), we get

$$\rho^* \geq \sup_{\pi^1 \in \Pi^1_{Ad}} \mathscr{J}(x,c,\pi^1,\pi^{*2}) \geq \inf_{\pi^2 \in \Pi^2_{Ad}} \sup_{\pi^1 \in \Pi^1_{Ad}} \mathscr{J}(x,c,\pi^1,\pi^2). \tag{5.6.14}$$

using (5.6.13) and (5.6.14), we have (5.6.2) and this gives us $\rho^* = \mathscr{J}(x,c,\pi^{*1},\pi^{*2})$. In particular, this implies that

$$\mathscr{J}(x,c,\pi^1,\pi^{*2}) \leq \mathscr{J}(x,c,\pi^{*1},\pi^{*2}) \leq \mathscr{J}(x,c,\pi^{*1},\pi^2).$$

That is, $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ is a saddle-point equilibrium. □

The next theorem shows that the converse statement of the above theorem is also true.

Theorem 5.6.2. *Suppose Assumption* $(A1')$ *holds. Suppose there exists a saddle-point equilibrium* $(\hat{\pi}^{*1}, \hat{\pi}^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$. *Then* $(\hat{\pi}^{*1}, \hat{\pi}^{*2})$ *is a mini-max selector of (5.6.1).*

Proof. By analogous arguments as in Theorem 5.4.3, in view of Theorem 5.6.1 and definition of saddle-point, we have $\rho^* = \mathscr{J}(x,c,\hat{\pi}^{*1},\hat{\pi}^{*2}) = \sup_{\pi^1 \in \Pi^1_{Ad}} \mathscr{J}(x,c,\pi^1,\hat{\pi}^{*2}) =$

$\inf_{\pi^2 \in \Pi^2_{Ad}} \mathscr{J}(x, c, \hat{\pi}^{*1}, \pi^2)$. Now arguing as in Theorem 5.6.1, it is easy to see that for $\hat{\pi}^{*1} \in \Pi^1_{SM}$ there exists $(\rho^{\hat{\pi}^{*1}}, \psi^*_{\hat{\pi}^{*1}}) \in \mathbb{R}_+ \times \mathscr{C}^+(S)$ with $\psi^*_{\hat{\pi}^{*1}} > 0$ such that

$$e^{\rho^{\hat{\pi}^{*1}}} \psi^*_{\hat{\pi}^{*1}}(x) = \inf_{\nu \in \mathcal{P}(B(x))} \left[e^{c(x, \hat{\pi}^{*1}(x), \nu)} \int_S \psi^*_{\hat{\pi}^{*1}}(y) P(dy|x, \hat{\pi}^{*1}(x), \nu) \right], \tag{5.6.15}$$

and $\rho^{\hat{\pi}^{*1}} = \inf_{\pi^2 \in \Pi^2_{ad}} \mathscr{J}(x, c, \hat{\pi}^{*1}, \pi^2) = \rho^*$. Thus for π^{*2} as in (5.6.11), we have

$$e^{\rho^{\hat{\pi}^{*1}}} \psi^*_{\hat{\pi}^{*1}}(x) \leq \left[e^{c(x, \hat{\pi}^{*1}(x), \pi^{*2}(x))} \int_S \psi^*_{\hat{\pi}^{*1}}(y) P(dy|x, \hat{\pi}^{*1}(x), \pi^{*2}(x)) \right]. \tag{5.6.16}$$

Also, from (5.6.11), we deduce that

$$e^{\rho^*} \psi^*(x) \geq \left[e^{c(x, \hat{\pi}^{*1}(x), \pi^{*2}(x))} \int_S \psi^*(y) P(dy|x, \hat{\pi}^{*1}(x), \pi^{*2}(x)) \right]. \tag{5.6.17}$$

Let $\hat{k}_8 = \min_S \dfrac{\psi^*}{\psi^*_{\hat{\pi}^{*1}}}$, thus we have $(\psi^* - \hat{k}_8 \psi^*_{\hat{\pi}^{*1}}) \geq 0$ in S and for some $\hat{x}_0 \in S$, $\psi^*(\hat{x}_0) - \hat{k}_8 \psi^*_{\hat{\pi}^{*1}}(\hat{x}_0) = 0$. Now, from (5.6.16) and (5.6.17), we deduce that

$$0 = \int_S (\psi^* - \hat{k}_8 \psi^*_{\hat{\pi}^{*1}})(y) P(dy|\hat{x}_0, \hat{\pi}^{*1}(\hat{x}_0), \pi^{*2}(\hat{x}_0)).$$

Then we claim that $\psi^* \equiv \hat{k}_8 \psi^*_{\hat{\pi}^{*1}}$. If not, then since support$(P(dy|x, a, b)) = S \; \forall (x, a, b) \in K$, it follows that

$$\int_S (\psi^* - \hat{k}_8 \psi^*_{\hat{\pi}^{*1}})(y) P(dy|\hat{x}_0, \hat{\pi}^{*1}(\hat{x}_0), \pi^{*2}(\hat{x}_0)) > 0. \tag{5.6.18}$$

So, we arrive at a contradiction and so, $\psi^* \equiv \hat{k}_8 \psi^*_{\hat{\pi}^{*1}}$. Hence from (5.6.1) and (5.6.15), it is easy to see that $\hat{\pi}^{*1}$ is an outer maximizing selector of (5.6.1). Similarly, one can show that $\hat{\pi}^{*2}$ is an outer minimizing selector of (5.6.1). This completes the proof. □

5.7 Conclusions

We have studied a risk-sensitive zero-sum stochastic game with ergodic cost criterion on countable/compact state space where the admissible action spaces ($A(x)$ and $B(x)$) are compact metric spaces. Under certain assumptions, we have established the existence of a saddle-point equilibrium and have completely characterized the same. Instead of employing the traditional vanishing discount asymptotics, we have pursued a direct approach involving the principal eigenpair of the corresponding Shapley equation.

CHAPTER 6

Nonzero-sum risk-sensitive continuous-time stochastic games with ergodic costs

6.1 Introduction

We consider a nonzero-sum stochastic game on the infinite time horizon for continuous-time Markov decision processes (CTMDPs) on a denumerable state space. The performance evaluation criterion is exponential of integral cost which addresses the decision makers (i.e., players') attitude towards risk. In other words, we address the problem of nonzero-sum risk-sensitive stochastic games involving continuous-time Markov decision processes. In discrete-time and discrete state space the risk-sensitive zero-sum stochastic games with bounded cost and transition rates have been studied by Basu and Ghosh [12] and nonzero-sum games in [13]. For CTMDPs, zero-sum stochastic games with risk-sensitive costs for bounded cost and bounded transition rates have been studied in [45]. One can see [48], [110], and the references therein for finite-horizon risk-sensitive stochastic games for CTMDPs, where [48] deals with the zero-sum game and [110] deals with the nonzero-sum game. Recently risk-sensitive CTMDPs for ergodic cost criterion have been studied in [21], [47], [52], [81], [82]. In the above five papers, the authors have studied risk-sensitive stochastic optimal control problems, where the controller is trying to control the state dynamics by choosing appropriate controls. When there is more than one controller the stochastic control problems become stochastic game problems. In this chapter, we have extended the results of the above five papers from one controller case to a multi-controller case where the controllers are non-cooperative. More specifically, we study ergodic nonzero sum risk-sensitive stochastic (non-cooperative) games along the line of the article [21], where the authors studied risk-sensitive discrete/continuous-time ergodic control problems for controlled Markov processes with countable state space. Using the principal eigenvalue approach, under a Lyapunov type stability assumption, we have shown that the corresponding system of coupled HJB equations admits a solution which in turn leads to the existence of Nash equilibrium in stationary strategies. Also, exploiting the stochastic representation of principal eigenfunction we completely characterize all possible Nash equilibria in the space of stationary Markov strategies.

Our main contribution to this chapter is the following. We establish the existence and characterization of Nash equilibria under a blanket Lyapunov type stability assumption. To be more specific, we study ergodic nonzero sum risk-sensitive stochastic games for CTMDPs having the following features: (a) the transition and the cost rates may be unbounded (b) state space is countable (c) at any state of the system the space of admissible actions is compact (d) the strategies are (state) feedback. To our knowledge, these results are new in the literature of ergodic non-zero sum risk-sensitive games for CTMDPs.

The rest of this chapter is organized as follows: Section 6.2 deals with the problem description and preliminaries. The ergodic cost criterion is analyzed in Section 6.3. Under a Lyapunov type stability assumption(s), we first establish the existence of a solution to the corresponding coupled Hamilton-Jacobi-Bellman (HJB) equations. This in turn leads to the existence of a Nash equilibrium in stationary strategies (see Theorem 6.4.1). In Section 6.4, we present an illustrative example. The content of this chapter is based on the published article [40].

6.2 The game model

For the sake of notational simplicity, we treat a two-player game. The N-player game for $N \geq 3$, is analogous. The continuous-time two-person nonzero-sum stochastic game model consists of the following elements

$$\{S, A, B, (A(i) \subset A, B(i) \subset B, i \in S), q(j|i,a,b), c_1(i,a,b), c_2(i,a,b)\}, \tag{6.2.1}$$

where each component is described below:

- S, called the state space, is assumed to be the set of all positive integers endowed with the discrete topology, i.e. $S =: \{1, 2, \cdots\}$.
- A and B are the action sets for players 1 and 2, respectively. The action spaces A and B are assumed to be Borel spaces with the Borel σ-algebras $\mathcal{B}(A)$ and $\mathcal{B}(B)$, respectively.
- For each $i \in S$, $A(i) \in \mathcal{B}(A)$ and $B(i) \in \mathcal{B}(B)$ denote the sets of admissible actions for players 1 and 2 in state i, respectively. Let $K := \{(i,a,b)|i \in S, a \in A(i), b \in B(i)\}$, which is a Borel subset of $S \times A \times B$.
  Throughout this chapter, we assume that

(A0)(a) For each $i \in S$, the admissible action spaces $A(i)$ and $B(i)$, are nonempty and compact subsets of A and B, respectively.

- The transition rates $q(j|i,a,b), (a,b) \in A(i) \times B(i), i,j \in S$, satisfy the condition $q(j|i,a,b) \geq 0$ for all $i \neq j$, $(a,b) \in A(i) \times B(i)$. Also, we assume that:
  (A0)(b) The transition rates $q(j|i,a,b)$ are conservative, i.e.,
  $$\sum_{j \in S} q(j|i,a,b) = 0 \text{ for } i \in S \text{ and } (a,b) \in A(i) \times B(i)$$
  and satisfy the following stability condition
  $$q_i := \sup_{(a,b) \in A(i) \times B(i)} [-q(i|i,a,b)] < \infty\,.$$

- Finally, the measurable function $c_k : K \to \mathbb{R}_+$ denotes the cost rate function for player k, $k = 1, 2$.

The game is played as follows. The players observe continuously the current state of the system. When the system is in state $i \in S$ at time $t \geq 0$, the players independently choose actions $a(t) \in A(i)$ and $b(t) \in B(i)$ according to some strategies, respectively. As a consequence of this, the following happens:

- player 1 (resp. 2) pays an immediate cost at rate $c_1(i, a(t), b(t))$ (resp. $c_2(i, a(t), b(t))$);

- the system stays in state i for a random time, with rate of leaving i given by $-q(i|i, a(t), b(t))$, and then jumps to a new state $j \neq i$ with the probability determined by $\dfrac{q(j|i, a(t), b(t))}{-q(i|i, a(t), b(t))}$ (see Proposition B. 8 in [59, p. 205] for details).

The whole process then repeats from the new state j. Cost accumulates throughout the course of the game. The planning horizon is infinite, and each player wants to minimize his infinite-horizon risk-sensitive cost with respect to some performance criterion $\rho_k^{\pi^1,\pi^2}$, $k = 1, 2$, which in our present case is defined by (6.2.2), below. To formalize what is described above, below we describe the construction of continuous-time Markov decision processes (CTMDPs) under admissible feedback strategies. We consider a continuous-time Markov decision processes (CTMDPs) $\{\xi_t\}_{t \geq 0}$ with state space S and controlled rate matrix $\Pi_{a,b} = (q(j|i,a,b))$. Here we describe the construction of continuous-time Markov decision processes (CTMDPs) under possibly history-dependent controls. The construction of the underlying CTMDPs ξ_t (as in [63], [75], [99]) is same as Chapter 3, p. 42-43.

To complete the specification of a risk-sensitive stochastic game problem, we need, of course, to introduce an optimality criterion. This requires defining the class of strategies as below.

Definition 6.2.1. *An admissible feedback strategy for player 1, denoted by* $\pi^1 = \{\pi^1(t)\}_{t\geq 0}$*, is a transition probability* $\pi^1(da|\omega,t)$ *from* $(\Omega\times[0,\infty),\mathscr{P})$ *onto* $(A_\Delta,\mathcal{B}(A_\Delta))$*, such that* $\pi^1(A(\xi_{t-}(\omega))|\omega,t)=1$*. Using appropriate projections of the transition kernel* π^1*, an admissible feedback strategy for player 1, determines and is, in turn, determined by a sequence* $\{\pi^1_k, k\geq 0\}$ *of the stochastic kernel on A such that*

$$\begin{aligned}\pi^1(t)(\omega) &= \pi^1(da|\omega,t)\\ &= I_{\{t=0\}}(t)\pi^1_0(da|i^{'}_0,0)+\sum_{k\geq 0} I_{\{T_k<t\leq T_{k+1}\}}\pi^1_k(da|i^{'}_0,\theta_1,i^{'}_1,\ldots,\theta_k,i^{'}_k,t-T_k)\\ &\quad + I_{\{t\geq T_\infty\}}\delta_{a_\Delta}(da),\end{aligned}$$

where $\pi^1_0(da|i^{'}_0,0)$ *is a stochastic kernel on A given S such that* $\pi^1_0(A(i^{'}_0)|i^{'}_0,0)=1$*,* $\pi^1_k(k\geq 1)$ *are stochastic kernels on A given* $(S\times(0,\infty))^{k+1}$ *such that* $\pi^1_k(A(i^{'}_k)|i^{'}_0,\theta_1,i^{'}_1,\cdots,\theta_k,i^{'}_k,t-T_k)=1$*, and* $\delta_{a_\Delta}(da)$ *denotes the Dirac measure at the point* a_Δ*.*

For more details see [64, Definition 2.1, Remark 2.2], [110], [119]. The set of all admissible feedback strategies for player 1 is denoted by Π^1_{Ad}. A strategy $\pi^1\in\Pi^1_{Ad}$ for player 1, is called a Markov if $\pi^1(t)(\omega)=\pi^1(\xi_{t-}(\omega),t)$ i.e., $\pi^1(da|\omega,t)=\pi^1(da|\xi_{t-}(\omega),t)$ for every $\omega\in\Omega$ and $t\geq 0$, where $\xi_{t-}(\omega):=\lim_{s\uparrow t}\xi_s(\omega)$. A Markov strategy π^1 is called a stationary Markov strategy if π^1 does not have an explicit dependence on time. For notational simplicity, we would not write ω anywhere throughout the rest of this chapter. We denote by Π^1_M and Π^1_{SM} the family of all Markov strategies and stationary Markov strategies, respectively, for player 1. The sets of admissible feedback strategies Π^2_{Ad}, Markov strategies Π^2_M and stationary strategies Π^2_{SM} for player 2 are defined analogously.

Remark 6.2.2. *As in [Chapter 3, Remark 3.2.2], in the definition of strategies we do not include the entire history of the game, i.e., past and present states, past sojourn times, and past actions taken by the players. In our game model, each player's admissible strategies include only past and present states and past sojourn times. Hence such strategies are called feedback strategies.*

To avoid possible explosion of the state process $\{\xi_t\}_{t\geq 0}$, we make the following Lyapunov stability assumption imposed on the transition rates, which had been widely used in CTMDPs; see, for instance, [53], [62], [63], [64] and the references therein.

(*A*1) There exists a non-constant function $\tilde{V}:S\to[1,\infty)$ such that

(i) $\sum_{j\in S}\tilde{V}(j)q(j|i,a,b)\leq C_1\tilde{V}(i)+C_2$ for all $(a,b)\in A(i)\times B(i)$ and $i\in S$ with some constants $C_1\neq 0$, $C_2\geq 0$;

(ii) $q_i \leq C_3 \tilde{V}(i)$ for all $i \in S$ with some positive constant C_3.

For the rest of this chapter Assumption, $(A1)$ is in force. Note that if $\sup_{i \in S} q_i < \infty$ then Assumption $(A1)$ holds. In this case, we can choose $\tilde{V}$ to be a suitable constant. Also note that under Assumption $(A1)$, for any initial state $i \in S$ and any pair of strategies $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$, Theorem 4.27 in [76] yields the existence of a unique probability measure denoted by $P_i^{\pi^1,\pi^2}$ on $(\Omega, \mathscr{F})$. Let $E_i^{\pi^1,\pi^2}$ be the expectation operator with respect to $P_i^{\pi^1,\pi^2}$. Also, from [59, p.13-15], we know that $\{\xi_t\}_{t \geq 0}$ is a Markov process under any $(\pi^1, \pi^2) \in \Pi^1_M \times \Pi^2_M$ (in fact, strong Markov).

For any compact metric space A, let $\mathcal{P}(A)$ denote the space of probability measures on A with the topology of weak convergence. Since $A(i)$, $B(i)$ are compact sets, we have $\mathcal{P}(A(i))$ and $\mathcal{P}(B(i))$ are compact metric spaces. For each $i, j \in S$, $\mu \in \mathcal{P}(A(i))$ and $\nu \in \mathcal{P}(B(i))$, the associated transition and cost rates are defined, respectively, as follows:

$$q(j|i,\mu,\nu) := \int_{\mathcal{P}(A(i))} \int_{\mathcal{P}(B(i))} q(j|i,a,b)\mu(da)\nu(db),$$

$$c_k(i,\mu,\nu) := \int_{\mathcal{P}(A(i))} \int_{\mathcal{P}(B(i))} c_k(i,a,b)\mu(da)\nu(db).$$

(by abuse of notation we use the same notation q and c_k). Note that for $\pi^1 \in \Pi^1_{SM}$ can be identified with a map $\pi^1 : S \to \mathcal{P}(A(i))$ such that for each $j \in S$, $\pi^1(j) \in \mathcal{P}(A(j))$ for each $j \in S$. Similarly, this conclusion also holds for $\pi^2 \in \Pi^2_{SM}$. Thus, we have $\Pi^1_{SM} = \Pi_{i \in S}\mathcal{P}(A(i))$ and $\Pi^2_{SM} = \Pi_{i \in S}\mathcal{P}(B(i))$ i.e., the sets Π^1_{SM} and Π^2_{SM} are endowed with the product topology. Therefore by Tychonoff theorem, the sets Π^1_{SM} and Π^2_{SM} are compact metric spaces.
Let $\mathcal{M}(A(i)), i = 1, 2, \ldots,$ and $\mathcal{M}(B(i)), i = 1, 2, \ldots,$ be the spaces of finite signed measures on $A(i)$ and $B(i)$, respectively, endowed with the topology of weak convergence. Then $\mathcal{M}(A(i))$ and $\mathcal{M}(B(i))$ are locally convex topological vector spaces that are metrizable [96]. Thus for $\Pi_{i \in S}\mathcal{M}(A(i))$ and $\Pi_{i \in S}\mathcal{M}(B(i))$ are locally convex topological vector spaces which are metrizable as well. Moreover, Π^1_{SM} and Π^2_{SM} are compact, convex subset of $\Pi_{i \in S}\mathcal{M}(A(i))$ and $\Pi_{i \in S}\mathcal{M}(B(i))$, respectively. For more details along these lines, we refer to [34].
We list the commonly used notations below.

- Given any real-valued function $\mathcal{V} \geq 1$ on S, we define a Banach space $(L^\infty_{\mathcal{V}}(S), \|\cdot\|^\infty_{\mathcal{V}})$ of $\mathcal{V}$-weighted functions by

$$L^\infty_{\mathcal{V}}(S) = \left\{ u : S \to \mathbb{R} \mid \|u\|^\infty_{\mathcal{V}} := \sup_{i \in S} \frac{|u(i)|}{\mathcal{V}(i)} < \infty \right\}.$$

- $L^{1,\infty}_{\mathcal{V}}$ denotes the subset of $L^\infty_{\mathcal{V}}$ consists of function u such that $\|u\|^\infty_{\mathcal{V}} \leq 1$.

For a pair of admissible strategies $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$, the risk-sensitive ergodic cost for player k, $k = 1, 2$, is given by

$$\rho_k^{\pi^1,\pi^2}(i) \ := \ \limsup_{T\to\infty} \frac{1}{T} \log E_i^{\pi^1,\pi^2}\left[e^{\int_0^T c_k(\xi_t,\pi^1(t),\pi^2(t))dt}\right], \tag{6.2.2}$$

where ξ_t is the CTMDP corresponding to $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$ and $E_i^{\pi^1,\pi^2}$ denotes the expectation with respect to the law of the process ξ_t with initial condition $\xi_0 = i$.

Definition 6.2.3. *A function $f : S \to \mathbb{R}$ is said to be norm-like if for every $k \in \mathbb{R}$, the set $\{i : f(i) \leq k\}$ is either empty or finite.*

Definition 6.2.4. *A time-homogeneous continuous-time Markov process $\{\xi_t\}$ with rate matrix $Q = [q(j|i)]$ is irreducible if for any $i, j \in S$, $i \neq j$, there exist distinct states $i_1, i_2 \cdots, i_k \in S$ satisfying $q(i_1|i) \cdots q(j|i_k) > 0$ (see, [59], p. 107).*

Since we are allowing our transition and cost rates to be unbounded, to guarantee the finiteness of $\rho_k^{\pi^1,\pi^2}$ for $k = 1, 2$, we make the following Lyapunov stability Assumption.

(*A*2) We assume that the CTMDP $\{\xi_t\}_{t\geq 0}$ is irreducible under every pair of stationary Markov strategies $(\pi^1, \pi^2) \in \Pi^1_{SM} \times \Pi^2_{SM}$. Furthermore, suppose there exists a constant $C_4 > 0$ and a function $V : S \to [1, \infty)$ such that one of the following holds.

(a) **When the running cost is bounded:** For some positive constant $\gamma > \max\{\|c_1\|_\infty, \ \|c_2\|_\infty\}$ and a finite set $\mathscr{K}$ it holds that

$$\sup_{(a,b)\in A(i)\times B(i)} \sum_{j\in S} V(j)q(j|i,a,b) \leq C_4 I_{\mathscr{K}}(i) - \gamma V(i) \ \forall i \in S,$$

where $\|c_k\|_\infty := \sup_{(i,a,b)\in K} c_k(i,a,b)$ for $k = 1, 2$.

(b) **When the running cost is unbounded:** For some norm-like function $\ell : S \to \mathbb{R}_+$ and a finite set $\mathscr{K}$ it holds that

$$\sup_{(a,b)\in A(i)\times B(i)} \sum_{j\in S} V(j)q(j|i,a,b) \leq C_4 I_{\mathscr{K}}(i) - \ell(i)V(i) \ \forall i \in S.$$

Also, the functions $\ell(\cdot) - \max_{(a,b)\in A(\cdot)\times B(\cdot)} c_k(\cdot, a, b), \ k = 1, 2,$ are norm-like.

This type of Foster-Lyapunov condition on the dynamics is quite common in the literature to study the continuous-time risk-sensitive ergodic control problems, for example, see, [6] [7] for controlled diffusion case and [21], [52] for Markov chain case.

Definition 6.2.5. *A pair of strategies* $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$ *is called a Nash equilibrium if*

$$\rho_1^{\pi^{*1},\pi^{*2}}(i) \;\leq\; \rho_1^{\pi^{1},\pi^{*2}}(i) \text{ for all } \pi^1 \in \Pi^1_{Ad} \text{ and } i \in S$$

and

$$\rho_2^{\pi^{*1},\pi^{*2}}(i) \;\leq\; \rho_2^{\pi^{*1},\pi^{2}}(i) \text{ for all } \pi^2 \in \Pi^2_{Ad} \text{ and } i \in S.$$

We wish to establish the existence of a Nash equilibrium in stationary strategies. To ensure the existence of a Nash equilibrium, we assume the following:

(*A*3)

(i) For any fixed $i, j \in S$, k=1,2 , $q(j|i,a,b)$ and $c_k(i,a,b)$ are continuous in $(a,b) \in A(i) \times B(i)$.

(ii) $\sum_{j\in S} V(j)q(j|i,a,b)$ is continuous in $(a,b) \in A(i) \times B(i)$ for any given $i \in S$, where V is as Assumption $(A2)$.

(iii) There exists $i_0 \in S$ such that $q(j|i_0,a,b) > 0$ for all $j \neq i_0$ and for all $(a,b) \in A(j) \times B(j)$.

It is also possible to consider another type of condition instead Assumption $(A3)$(iii). We refer to Remark 6.3.3 for further discussion.

We wish to establish the existence of a Nash equilibrium in stationary strategies.

We now procedure to establish Nash equilibrium in stationary strategies for a nonzero-sum game. To this end, we first outline a standard procedure for establishing the existence of a Nash equilibrium. Suppose player 2 announces that he/she is going to employ a strategy $\pi^2 \in \Pi^2_{SM}$. In such a scenario, player 1 attempts to minimize

$$\rho_1^{\pi^1,\pi^2}(i) \;=\; \limsup_{T\to\infty} \frac{1}{T} \log E_i^{\pi^1,\pi^2}\left[e^{\int_0^T c_1(\xi_t,\pi^1(t),\pi^2(\xi_{t-}))dt}\right],$$

over $\pi^1 \in \Pi^1_{Ad}$. Thus for player 1, it is a continuous-time Markov decision problem (CTMDP) with risk-sensitive ergodic cost. This problem has been studied in [21], [47], [81], [82]. In particular under certain assumptions, it is shown in [21], [81], [82], that the following Hamilton-Jacobi-Bellman (HJB) equation

$$\begin{cases} \rho_1\, \hat{\psi}_1(i) = \inf\limits_{\mu\in\mathcal{P}(A(i))} \Big[\sum\limits_{j\in S} q(j|i,\mu,\pi^2(i))\hat{\psi}_1(j) + c_1(i,\mu,\pi^2(i))\hat{\psi}_1(i)\Big] \\ \hat{\psi}_1(\hat{i}_0) = 1, \end{cases}$$

has a suitable solution $(\rho_1, \hat{\psi}_1)$, where ρ_1 is a scalar and $\hat{\psi}_1 : S \to \mathbb{R}$ has suitable growth rate; and $\hat{i}_0$ is some fixed state in S. Furthermore it is shown in [21], [81], [82], that

$$\rho_1 \;=\; \inf_{\pi^1\in\Pi^1_{Ad}} \limsup_{T\to\infty} \frac{1}{T} \log E_i^{\pi^1,\pi^2}\left[e^{\int_0^T c_1(\xi_t,\pi^1(t),\pi^2(\xi_{t-}))dt}\right],$$

and if $\pi^{*1} \in \Pi^1_{SM}$ is such that for all $i \in S$

$$\begin{aligned}&\inf_{\mu \in \mathcal{P}(A(i))} \Big[\sum_{j\in S} q(j|i,\mu,\pi^2(i))\hat{\psi}_1(j) + c_1(i,\mu,\pi^2(i))\hat{\psi}_1(i)\Big]\\ = &\sum_{j\in S} q(j|i,\pi^{*1}(i),\pi^2(i))\hat{\psi}_1(j) + c_1(i,\pi^{*1}(i),\pi^2(i))\hat{\psi}_1(i),\end{aligned}$$

then $\pi^{*1} \in \Pi^1_{SM}$ is an optimal control for player 1, i.e., for any $i \in S$

$$\rho_1 \;=\; \limsup_{T\to\infty} \frac{1}{T} \log E_i^{\pi^{*1},\pi^2}\Big[e^{\int_0^T c_1(\xi_t,\pi^{*1}(\xi_{t-}),\pi^2(\xi_{t-}))dt}\Big].$$

In [82], the ergodic case is treated via the limit of the corresponding finite-horizon risk-sensitive continuous-time MDP. For the latter, the HJB equation is an infinite system of coupled ODEs. Then as the length of the horizon tends to ∞, the above equation is derived using limiting horizon asymptotics. In [81], the existence of ergodic optimal control is established by using the vanishing discount approach. In [21], the ergodic case is studied directly by an approach involving the principal eigenpair associated with the above equation. In this chapter, we follow the approach of [21].
In view of the foregoing it follows that given that player 2 is using the strategy $\pi^2 \in \Pi^2_{SM}$, $\pi^{*1} \in \Pi^1_{SM}$ is an optimal response for player 1. Clearly π^{*1} depends on π^2 and moreover there may be several optimal responses for player 1 in Π^1_{SM}. Analogous results hold for player 2 if player 1 announces that he is going to use a strategy $\pi^1 \in \Pi^1_{SM}$. Hence given a pair of strategies $(\pi^1,\pi^2) \in \Pi^1_{SM} \times \Pi^2_{SM}$, we can find a set of pairs of optimal responses $\{(\pi^{*1},\pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}\}$ via the appropriate pair of HJB equations described above. This defines a set-valued map. Clearly, any fixed point of this set-valued map is a Nash equilibrium.

The above discussion leads to the following procedure for finding a pair of Nash equilibrium strategies. Suppose that there exists a pair of stationary strategies $(\pi^{*1},\pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$, a pair of scalars (ρ_1^*,ρ_2^*) and a pair of functions $(\hat{\psi}_1^*,\hat{\psi}_2^*)$ with appropriate growth conditions, satisfying the following coupled HJB equations:

$$\begin{cases}\rho_1^*\ \hat{\psi}_1^*(i) = \displaystyle\inf_{\mu\in\mathcal{P}(A(i))}\Big[\sum_{j\in S} q(j|i,\mu,\pi^{*2}(i))\hat{\psi}_1^*(j) + c_1(i,\mu,\pi^{*2}(i))\hat{\psi}_1^*(i)\Big]\\ \qquad\qquad = \displaystyle\sum_{j\in S} q(j|i,\pi^{*1}(i),\pi^{*2}(i))\hat{\psi}_1^*(j) + c_1(i,\pi^{*1}(i),\pi^{*2}(i))\hat{\psi}_1^*(i)\\ \hat{\psi}_1^*(\hat{i}_0) = 1,\\ \rho_2^*\ \hat{\psi}_2^*(i) = \displaystyle\inf_{\nu\in\mathcal{P}(B(i))}\Big[\sum_{j\in S} q(j|i,\pi^{*1}(i),\nu)\hat{\psi}_2^*(j) + c_2(i,\pi^{*1}(i),\nu)\hat{\psi}_2^*(i)\Big]\\ \qquad\qquad = \displaystyle\sum_{j\in S} q(j|i,\pi^{*1}(i),\pi^{*2}(i))\hat{\psi}_2^*(j) + c_2(i,\pi^{*1}(i),\pi^{*2}(i))\hat{\psi}_2^*(i)\\ \hat{\psi}_2^*(\hat{i}_0) = 1,\end{cases}$$

where as before $\hat{i}_0 \in S$ is a fixed state. Then it can be shown that (π^{*1}, π^{*2}) is a pair of Nash equilibrium and (ρ_1^*, ρ_2^*) is the pair of corresponding Nash values. Thus the main result of this chapter is to establish that the above coupled HJB equations have suitable solutions.

Remark 6.2.6. *Note that similar stochastic optimal control problems have been studied in [47], [82] for bounded cost and bounded transition rates. But in our game model transition and cost rates are allowed to be unbounded. Analogous MDP problems are treated in [21].*

6.3 Coupled HJB Equations and Existence of Nash Equilibrium

By the definition of weak convergence of probability measures, one can easily get the following result, which will be crucial for the existence of Nash equilibrium; for details, we refer to [56, Lemma 7.2].

Lemma 6.3.1. *Under Assumptions (A0)-(A3), the functions*

$$c_k(i, \mu, \nu),\ k = 1, 2 \ \text{ and } \sum_{j \in S} q(j|i, \mu, \nu)\phi(j)$$

are continuous on $\mathcal{P}(A(i)) \times \mathcal{P}(B(i))$ for each fixed $\phi \in L_V^\infty(S)$ and $i \in S$.

For any finite set $\mathscr{D} \subset S$, we define

$$\mathcal{B}_{\mathscr{D}} = \{f : S \to \mathbb{R} \mid f \text{ is a Borel measurable function and } f(i) = 0 \ \forall\, i \in \mathscr{D}^c\}.$$

Also, $\mathcal{B}_{\mathscr{D}}^+ \subset \mathcal{B}_{\mathscr{D}}$ denotes the cone of all nonnegative functions vanishing outside $\mathscr{D}$.

Let $\mathscr{D}_n \subset S$ be an increasing sequence of finite sets such that $\cup_n \mathscr{D}_n = S$ such that $i_0 \in \mathscr{D}_n$ for each $n \geq 1$, where $i_0 \in S$ is a fixed state as in Assumption $(A3)$. Also, we denote $\succeq$ as the partial ordering in $\mathcal{B}_{\mathscr{D}_n}$ with respect to the closed cone $\mathcal{B}_{\mathscr{D}_n}^+$, i.e., for $f, g \in \mathcal{B}_{\mathscr{D}_n}$, $f \succeq g$ if and only if $f - g \in \mathcal{B}_{\mathscr{D}_n}^+$. Now using Kreĭn-Rutman theorem we prove the existence of an eigenpair to a Dirichlet problem in $\mathscr{D}_n$ for each $n \in \mathbb{N}$. In the next lemma, we show the existence of eigenpairs to certain equations in $\mathscr{D}_n$ for each $n \in \mathbb{N}$.

Lemma 6.3.2. *Suppose that Assumptions (A0)-(A3) are satisfied. Then for each $n \in \mathbb{N}$, the following hold.*

1. *For* $\hat{\pi}^2 \in \Pi^2_{SM}$, *there exists an eigenpair* $(\rho_{1,n}, \psi_{1,n}) \in \mathbb{R} \times \mathcal{B}^+_{\mathscr{D}_n}$, *satisfying*

$$\begin{cases} \rho_{1,n}\psi_{1,n}(i) = \inf\limits_{\mu \in \mathcal{P}(A(i))} \left[\sum\limits_{j \in S} \psi_{1,n}(j) q(j|i, \mu, \hat{\pi}^2(i)) + c_1(i, \mu, \hat{\pi}^2(i))\psi_{1,n}(i) \right] \text{ for } i \in \mathscr{D}_n, \\ \quad \psi_{1,n}(i_0) = 1. \end{cases} \tag{6.3.1}$$

Moreover, we have

$$0 \le \liminf_{n\to\infty} \rho_{1,n} \le \limsup_{n\to\infty} \rho_{1,n} \le \inf_{\pi^1 \in \Pi^1_{Ad}} \limsup_{T\to\infty} \frac{1}{T} \log E^{\pi^1, \hat{\pi}^2}_{i_0} \left[e^{\int_0^T c_1(\xi_t, \pi^1(t), \hat{\pi}^2(\xi_{t-}))dt} \right], \tag{6.3.2}$$

and $\{\rho_{1,n}\}$ *is a bounded sequence.*

2. *Similarly, for* $\hat{\pi}^1 \in \Pi^1_{SM}$, *there exists an eigenpair* $(\rho_{2,n}, \psi_{2,n}) \in \mathbb{R} \times \mathcal{B}^+_{\mathscr{D}_n}$, *satisfying*

$$\begin{cases} \rho_{2,n}\psi_{2,n}(i) = \inf\limits_{\nu \in \mathcal{P}(B(i))} \left[\sum\limits_{j \in S} \psi_{2,n}(j) q(j|i, \hat{\pi}^1(i), \nu) + c_2(i, \hat{\pi}^1(i), \nu)\psi_{2,n}(i) \right] \text{ for } i \in \mathscr{D}_n, \\ \quad \psi_{2,n}(i_0) = 1. \end{cases} \tag{6.3.3}$$

Moreover, we have

$$0 \le \liminf_{n\to\infty} \rho_{2,n} \le \limsup_{n\to\infty} \rho_{2,n} \le \inf_{\pi^2 \in \Pi^2_{Ad}} \limsup_{T\to\infty} \frac{1}{T} \log E^{\hat{\pi}^1, \pi^2}_{i_0} \left[e^{\int_0^T c_2(\xi_t, \hat{\pi}^1(\xi_{t-}), \pi^2(t))dt} \right], \tag{6.3.4}$$

and $\{\rho_{2,n}\}$ *is a bounded sequence.*

Proof. We prove part (1); part (2) follows by analogous arguments. Fix $\hat{\pi}^2 \in \Pi^2_{SM}$. Let $\delta > 0$. Set $\tilde{c}_1(i, \mu, \nu) = c_1(i, \mu, \nu) - k_n - \delta$, where $k_n = \sup\{c_1(i, \mu, \nu) \mid i \in \mathscr{D}_n, \mu \in \mathcal{P}(A(i)), \nu \in \mathcal{P}(B(i))\}$. From [21, Proposition 3.1], it is easy to see that for each $g \in \mathcal{B}_{\mathscr{D}_n}$ the following equation

$$-g(i) = \inf_{\mu \in \mathcal{P}(A(i))} \left[\sum_{j \in S} \phi_1(j) q(j|i, \mu, \hat{\pi}^2(i)) + \tilde{c}_1(i, \mu, \hat{\pi}^2(i))\phi_1(i) \right] \text{ for } i \in \mathscr{D}_n,$$

admits a unique solution $\phi_1 \in \mathcal{B}_{\mathscr{D}_n}$ and ϕ_1 is given by

$$\phi_1(i) = \inf_{\pi^1 \in \Pi^1_{Ad}} E^{\pi^1, \hat{\pi}^2}_i \left[\int_0^{\tau(\mathscr{D}_n)} e^{\int_0^t \tilde{c}_1(\xi_s, \pi^1(s), \hat{\pi}^2(\xi_{s-}))ds} g(\xi_t) dt \right], \ i \in S,$$

where $\tau(\mathscr{D}_n) := \inf\{t > 0 : \xi_t \notin \mathscr{D}_n\}$. Therefore, the operator $T : \mathcal{B}_{\mathscr{D}_n} \to \mathcal{B}_{\mathscr{D}_n}$ given by

$$T(g)(i) := \phi_1(i) = \inf_{\pi^1 \in \Pi^1_{Ad}} E^{\pi^1, \hat{\pi}^2}_i \left[\int_0^{\tau(\mathscr{D}_n)} e^{\int_0^t \tilde{c}_1(\xi_s, \pi^1(s), \hat{\pi}^2(\xi_{s-}))ds} g(\xi_t) dt \right], \ i \in \mathscr{D}_n, g \in \mathcal{B}_{\mathscr{D}_n}$$

with $T(g)(i) = 0$ for $i \in \mathscr{D}^c_n$ is well defined. Then by similar arguments as in [21, Lemma 3.1], the map T is order-preserving, 1-homogeneous, completely continuous and for some nonzero function $g \in \mathcal{B}^+_{\mathscr{D}_n}$, there exists $M > 0$, such that $MT(g) \succeq g$, i.e., it satisfies all conditions of Kreĭn-Rutman theorem. Hence by a version of nonlinear Kreĭn-Rutman theorem [4, Section 3.1], there exist nontrivial $\psi_{1,n} \in \mathcal{B}^+_{\mathscr{D}_n}$ and $\lambda_{\mathscr{D}_n} > 0$, satisfying $T\psi_{1,n} = \lambda_{\mathscr{D}_n}\psi_{1,n}$. Let $\tilde{\rho}_{1,n} = -[\lambda_{\mathscr{D}_n}]^{-1}$. Then we have that the pair $(\tilde{\rho}_{1,n}, \psi_{1,n})$ satisfies

$$\inf_{\mu\in\mathcal{P}(A(i))}\Big[\sum_{j\in S} q(j|i,\mu,\hat{\pi}^2(i))\psi_{1,n}(j) + \tilde{c}_1(i,\mu,\hat{\pi}^2(i))\psi_{1,n}(i)\Big] = \tilde{\rho}_{1,n}\psi_{1,n}(i), \ \forall i \in \mathscr{D}_n\,.$$

Now, let $\rho_{1,n} = \tilde{\rho}_{1,n} + k_n + \delta$, then it is easy to see that the pair $(\rho_{1,n}, \psi_{1,n})$ satisfies the following equation

$$\rho_{1,n}\psi_{1,n}(i) = \inf_{\mu\in\mathcal{P}(A(i))}\Big[\sum_{j\in S}\psi_{1,n}(j)q(j|i,\mu,\hat{\pi}^2(i)) + c_1(i,\mu,\hat{\pi}^2(i))\psi_{1,n}(i)\Big] \text{ for } i \in \mathscr{D}_n, \psi_{1,n} \gneq 0.$$

From Assumption $(A3)$(iii) and using the above equation we have $\psi_{1,n}(i_0) > 0$. Thus by normalizing $\psi_{1,n}$ we obtain $\psi_{1,n}(i_0) = 1$. Therefore, it follows that the pair $(\rho_{1,n}, \psi_{1,n})$ satisfies the required HJB equation (6.3.1).

Now, following [21, Lemma 3.3] one can show that $\rho_{1,n}$ satisfies (6.3.2) and $\{\rho_{1,n}\}$ is a bounded sequence. □

Next by taking limit $n \to \infty$ we show that the limiting equations admit eigenpairs in appropriate spaces. In particular, we have the following theorem.

Theorem 6.3.1. *Suppose that Assumptions* $(A0)$*-*$(A3)$ *are satisfied. Then the following hold.*

1. *For* $\hat{\pi}^2 \in \Pi^2_{SM}$*, there exists a unique principal eigenpair* $(\rho_1, \psi_1) \in \mathbb{R}_+ \times L^{1,\infty}_{V}$, $\psi_1 > 0$*, satisfying*

$$\begin{cases} \rho_1\psi_1(i) = \inf\limits_{\mu\in\mathcal{P}(A(i))}\Big[\sum\limits_{j\in S}\psi_1(j)q(j|i,\mu,\hat{\pi}^2(i)) + c_1(i,\mu,\hat{\pi}^2(i))\psi_1(i)\Big] \text{ for } i \in S, \\ \psi_1(i_0) = 1. \end{cases} \tag{6.3.5}$$

Moreover, we have

$$\rho_1 = \inf_{\pi^1\in\Pi^1_{Ad}}\limsup_{T\to\infty}\frac{1}{T}\log E_i^{\pi^1,\hat{\pi}^2}\Big[e^{\int_0^T c_1(\xi_t,\pi^1(t),\hat{\pi}^2(\xi_{t-}))dt}\Big] (:= \rho_1^{\hat{\pi}^2} = \inf_{\pi^1\in\Pi^1_{Ad}}\rho_1^{\pi^1,\hat{\pi}^2}), \tag{6.3.6}$$

and there exists a finite set $\mathscr{B}_1 \supset \mathscr{K}$*, such that*

$$\psi_1(i) = \inf_{\pi^1\in\Pi^1_{SM}} E_i^{\pi^1,\hat{\pi}^2}\Big[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}(c_1(\xi_t,\pi^1(\xi_{t-}),\hat{\pi}^2(\xi_{t-}))-\rho_1)dt}\psi_1(\xi_{\hat{\tau}(\mathscr{B}_1)})\Big] (:= \psi_1^{\hat{\pi}^2}(i))\ \forall i \in \mathscr{B}_1^c, \tag{6.3.7}$$

where $\hat{\tau}(\mathscr{B}_1) = \tau(\mathscr{B}_1^c) = \inf\{t : \xi_t \in \mathscr{B}_1\} =: \tilde{\tau}_1$.

2. *Similarly, for* $\hat{\pi}^1 \in \Pi^1_{SM}$*, there exists a unique principal eigenpair* $(\rho_2, \psi_2) \in \mathbb{R}_+ \times L^{1,\infty}_V$*,* $\psi_2 > 0$ *satisfying*

$$\begin{cases} \rho_2\psi_2(i) = \inf_{\nu\in\mathcal{P}(B(i))}\left[\sum_{j\in S}\psi_2(j)q(j|i,\hat{\pi}^1(i),\nu) + c_2(i,\hat{\pi}^1(i),\nu)\psi_2(i)\right] \text{ for } i \in S, \\ \psi_2(i_0) = 1. \end{cases} \tag{6.3.8}$$

Moreover, we have

$$\rho_2 = \inf_{\pi^2\in\Pi^2_{Ad}} \limsup_{T\to\infty} \frac{1}{T}\log E_i^{\hat{\pi}^1,\pi^2}\left[e^{\int_0^T c_2(\xi_t,\hat{\pi}^1(\xi_{t-}),\pi^2(t))dt}\right] (:= \rho_2^{\hat{\pi}^1} = \inf_{\pi^2\in\Pi^2_{Ad}} \rho_2^{\hat{\pi}^1,\pi^2}), \tag{6.3.9}$$

and there exists a finite set $\mathscr{B}_2 \supset \mathscr{K}$*, such that*

$$\psi_2(i) = \inf_{\pi^2\in\Pi^2_{SM}} E_i^{\hat{\pi}^1,\pi^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_2)}(c_2(\xi_t,\hat{\pi}^1(\xi_{t-}),\pi^2(\xi_{t-}))-\rho_2)dt}\psi_2(\xi_{\hat{\tau}(\mathscr{B}_2)})\right] (:= \psi_2^{\hat{\pi}^1}(i))\ \forall i \in \mathscr{B}_2^c, \tag{6.3.10}$$

where $\hat{\tau}(\mathscr{B}_2) = \tau(\mathscr{B}_2^c) = \inf\{t : \xi_t \in \mathscr{B}_2\} =: \tilde{\tau}_2$.

Proof. Since $c_1 \geq 0$, using Assumption $(A2)$, we deduce that there exists a finite set $\mathscr{B}_1$ containing $\mathscr{K}$ such that

- under Assumption $(A2)$(a), since $\gamma > \|c_1\|_\infty$, we have

$$\sup_{(a,b)\in A(i)\times B(i)} c_1(i,a,b) - \rho_{1,n} < \gamma \quad \forall\ i \in \mathscr{B}_1^c \quad \text{and all } n \text{ large enough}.$$

- under Assumption $(A2)$(b), since the function $\ell(\cdot) - \max_{(a,b)\in A(\cdot)\times B(\cdot)} c_1(\cdot,a,b)$ is norm-like, we have

$$\sup_{(a,b)\in A(i)\times B(i)} c_1(i,a,b) - \rho_{1,n} < \ell(i) \quad \forall\ i \in \mathscr{B}_1^c \quad \text{and all } n \text{ large enough}.$$

Let $\pi^1 \in \Pi^1_{SM}$. Then applying Itô-Dynkin formula, from Assumption $(A2)$, we prove the following estimates:

- Under Assumption $(A2)$(a):

$$E_i^{\pi^1,\hat{\pi}^2}\left[e^{\hat{\tau}(\mathscr{B}_1)\gamma}V(\xi_{\hat{\tau}(\mathscr{B}_1)})\right] \leq V(i) \quad \forall i \in \mathscr{B}_1^c. \tag{6.3.11}$$

- Under Assumption $(A2)$(b):

$$E_i^{\pi^1,\hat{\pi}^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}\ell(\xi_t)dt}V(\xi_{\hat{\tau}(\mathscr{B}_1)})\right] \leq V(i) \quad \forall i \in \mathscr{B}_1^c. \tag{6.3.12}$$

It is easy to see that the proof of (6.3.11) is analogous to that the proof of (6.3.12) when we replace ℓ with γ. So, we prove only (6.3.12). Suppose Assumption $(A2)$(b) holds. Let n be large enough so that $\mathscr{B}_1 \subset \mathscr{D}_n$. Applying Dynkin's formula [59, Appendix C.3], for $i \in \mathscr{B}_1^c \cap \mathscr{D}_n$ and $T > 0$, we have

$$\begin{aligned}
&E_i^{\pi^1,\hat{\pi}^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)\wedge T\wedge\tau(\mathscr{D}_n)}\ell(\xi_s)ds}V(\xi_{\hat{\tau}(\mathscr{B}_1)\wedge T\wedge\tau(\mathscr{D}_n)})\right] - V(i)\\
&= E_i^{\pi^1,\hat{\pi}^2}\left[\int_0^{\hat{\tau}(\mathscr{B}_1)\wedge T\wedge\tau(\mathscr{D}_n)} e^{\int_0^t \ell(\xi_s)ds}[\ell(\xi_t)V(\xi_t) + \sum_{j\in S} q(j|\xi_t,\pi^1(\xi_{t-}),\hat{\pi}^2(\xi_{t-}))V(j)]dt\right]\\
&\leq E_i^{\pi^1,\hat{\pi}^2}\left[\int_0^{\hat{\tau}(\mathscr{B}_1)\wedge T\wedge\tau(\mathscr{D}_n)} e^{\int_0^t \ell(\xi_s)ds}C_4 I_{\mathscr{K}}(\xi_t)dt\right] = 0,
\end{aligned}$$

where $\tau(\mathscr{D}_n) = \inf\{t > 0 : \xi_t \notin \mathscr{D}_n\}$ (as defined in Lemma 6.3.2). Now by Fatou's lemma, taking first $n \to \infty$ and then $T \to \infty$, we get (6.3.12). Now we scale $\psi_{1,n}$ in such a way that it touches V from below. Define

$$\hat{\theta}_n = \sup\{k > 0 : (V - k\psi_{1,n}) > 0 \text{ in } S\}.$$

Then we see that $\hat{\theta}_n$ is finite as $\psi_{1,n}$ vanishes in $\mathscr{D}_n^c$ and $\psi_{1,n} \gneq 0$. Also, it is easy to see that $\hat{\theta}_n\psi_{1,n} \leq V$. We claim that if we replace $\psi_{1,n}$ by $\hat{\theta}_n\psi_{1,n}$, then $\psi_{1,n}$ touches V inside $\mathscr{B}_1$. If not, then for some state $\hat{i} \in \mathscr{B}_1^c$, $(V - \psi_{1,n})(\hat{i}) = 0$ and $V - \psi_{1,n} > 0$ in $\mathscr{B}_1 \cup \mathscr{D}_n^c$. Then by Dynkin formula, we get (under Assumption $(A2)$(b))

$$\begin{aligned}
\psi_{1,n}(\hat{i}) &\leq E_{\hat{i}}^{\pi^1,\hat{\pi}^2}\left[e^{\int_0^{T\wedge\hat{\tau}(\mathscr{B}_1)}(c(\xi_s,\pi^1(\xi_{s-}),\hat{\pi}^2(\xi_{s-}))-\rho_{1,n})ds}\psi_{1,n}(\xi_{T\wedge\hat{\tau}(\mathscr{B}_1)})I_{\{T\wedge\hat{\tau}(\mathscr{B}_1)<\tau(\mathscr{D}_n)\}}\right]\\
&\leq E_{\hat{i}}^{\pi^1,\hat{\pi}^2}\left[e^{\int_0^{T\wedge\hat{\tau}(\mathscr{B}_1)}\ell(\xi_s)ds}\psi_{1,n}(\xi_{T\wedge\hat{\tau}(\mathscr{B}_1)})I_{\{T\wedge\hat{\tau}(\mathscr{B}_1)<\tau(\mathscr{D}_n)\}}\right].
\end{aligned}$$

Since $\psi_{1,n} \leq V$, in view of (6.3.12), by the dominated convergence theorem, taking $T \to \infty$, we get

$$\psi_{1,n}(\hat{i}) \leq E_{\hat{i}}^{\pi^1,\hat{\pi}^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}\ell(\xi_s)ds}\psi_{1,n}(\xi_{\hat{\tau}(\mathscr{B}_1)})\right].$$

Using this and (6.3.12), we have

$$0 = (V - \psi_{1,n})(\hat{i}) \geq E_{\hat{i}}^{\pi^1,\hat{\pi}^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}\ell(\xi_s)ds}(V - \psi_{1,n})(\xi_{\hat{\tau}(\mathscr{B}_1)})\right] > 0.$$

Hence we arrive at a contradiction. Thus $\psi_{1,n}$ touches V inside $\mathscr{B}_1$. A similar conclusion holds under Assumption $(A2)$(a). Therefore, $\psi_{1,n} \leq V$ and at some point $\hat{i}^* \in \mathscr{B}_1$, $\psi_{1,n}(\hat{i}^*) = V(\hat{i}^*)$.

Since $\psi_{1,n} \leq V$ for all n large enough, by diagonalization argument, we deduce that along a suitable subsequence $\psi_{1,n}(i) \to \psi_1(i)$ for all $i \in S$, for some $\psi_1 \in L_V^{1,\infty}$. Also, from

Lemma 6.3.2, we have $\{\rho_{1,n}\}$ is a bounded sequence. Thus along a further subsequence we have $\rho_{1,n} \to \rho_1$ as $n \to \infty$. Let $\tilde{\pi}^1_n \in \Pi^1_{SM}$ be a minimizing selector of (6.3.1), i.e., we have

$$\rho_{1,n}\psi_{1,n}(i) = \left[\sum_{j\in S} \psi_{1,n}(j)q(j|i,\tilde{\pi}^1_n(i),\hat{\pi}^2(i)) + c_1(i,\tilde{\pi}^1_n(i),\hat{\pi}^2(i))\psi_{1,n}(i)\right] \text{ for } i \in \mathscr{D}_n\,. \tag{6.3.13}$$

Since Π^1_{SM} is compact along further subsequence $\tilde{\pi}^1_n \to \tilde{\pi}^1$ in Π^1_{SM}. Therefore, by generalized Fatou's lemma [69, Lemma 8.3.7], letting $n \to \infty$, from (6.3.13) it follows that

$$\rho_1\psi_1(i) \geq \inf_{\mu\in\mathcal{P}(A(i))} \left[\sum_{j\in S} \psi_1(j)q(j|i,\mu,\hat{\pi}^2(i)) + c_1(i,\mu,\hat{\pi}^2(i))\psi_1(i)\right] \text{ for } i \in S\,. \tag{6.3.14}$$

Also, from (6.3.1), for any $\mu \in \mathcal{P}(A(i))$, we have

$$\rho_{1,n}\psi_{1,n}(i) \leq \left[\sum_{j\in S} \psi_{1,n}(j)q(j|i,\mu,\hat{\pi}^2(i)) + c_1(i,\mu,\hat{\pi}^2(i))\psi_{1,n}(i)\right] \text{ for } i \in \mathscr{D}_n\,.$$

Since $\psi_{1,n} \leq V$, by the dominated convergence theorem, letting $n \to \infty$ we deduce

$$\rho_1\psi_1(i) \leq \left[\sum_{j\in S} \psi_1(j)q(j|i,\mu,\hat{\pi}^2(i)) + c_1(i,\mu,\hat{\pi}^2(i))\psi_1(i)\right]. \tag{6.3.15}$$

Therefore, combining (6.3.14) and (6.3.15), it follows that the pair $(\rho_1, \psi_1) \in \mathbb{R}_+ \times L^{1,\infty}_V$, $\psi_1 \geq 0$ satisfies

$$\rho_1\psi_1(i) = \inf_{\mu\in\mathcal{P}(A(i))} \left[\sum_{j\in S} \psi_1(j)q(j|i,\mu,\hat{\pi}^2(i)) + c_1(i,\mu,\hat{\pi}^2(i))\psi_1(i)\right] \text{ for } i \in S.$$

Since at some point in $\mathscr{B}_1$ we have $(V - \psi_{1,n}) = 0$, for all large n, we have $(V - \psi_1)(\hat{i}^*) = 0$ for some $\hat{i}^* \in \mathscr{B}_1$. Since $V \geq 1$, it is clear that ψ_1 is nontrivial. Now we claim that $\psi_1 > 0$. If not, we must have $\psi_1(\tilde{i}) = 0$ for some $\tilde{i} \in S$. Then, for any minimizing selector $\tilde{\pi}^{*1} \in \Pi^1_{SM}$ of (6.3.5), it follows that

$$\rho_1\psi_1(\tilde{i}) = \left[\sum_{j\in S} \psi_1(j)q(j|\tilde{i},\tilde{\pi}^{*1}(\tilde{i}),\hat{\pi}^2(\tilde{i})) + c(\tilde{i},\tilde{\pi}^{*1}(\tilde{i}),\hat{\pi}^2(\tilde{i}))\psi_1(\tilde{i})\right].$$

This implies

$$\sum_{j\neq\tilde{i}} \psi_1(j)q(j|\tilde{i},\tilde{\pi}^{*1}(\tilde{i}),\hat{\pi}^2(\tilde{i})) = 0.$$

Since the Markov chain ξ_t is irreducible under $(\tilde{\pi}^{*1}, \hat{\pi}^2) \in \Pi^1_{SM} \times \Pi^2_{SM}$, from the above equation, it follows that $\psi_1 \equiv 0$. So, we arrive at a contradiction. This proves that (ρ_1, ψ_1) is an eigenpair to (6.3.5).

By truncating the running cost c_1, one can show that ρ_1 satisfies (6.3.6) (see, [21, Lemma 3.5]) . Next we prove the stochastic representation (6.3.7).

Applying Itô-Dynkin formula for any minimizing selector π^{*1} of (6.3.5) and any $T > 0$, we have

$$\psi_1(i) = E_i^{\pi^{*1},\hat{\pi}^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)\wedge T}(c_1(\xi_t,\pi^{*1}(\xi_{t-}),\hat{\pi}^2(\xi_{t-}))-\rho_1)dt}\psi_1(\xi_{\hat{\tau}(\mathscr{B}_1)\wedge T})\right] \;\forall i \in \mathscr{B}_1^c\,.$$

Then applying Fatou's lemma, by taking $T \to \infty$, we get

$$\begin{aligned}\psi_1(i) &\geq E_i^{\pi^{*1},\hat{\pi}^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}(c_1(\xi_t,\pi^{*1}(\xi_{t-}),\hat{\pi}^2(\xi_{t-}))-\rho_1)dt}\psi_1(\xi_{\hat{\tau}(\mathscr{B}_1)})\right]\\ &\geq \inf_{\pi^1\in\Pi_{SM}^1} E_i^{\pi^1,\hat{\pi}^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}(c_1(\xi_t,\pi^1(\xi_{t-}),\hat{\pi}^2(\xi_{t-}))-\rho_1)dt}\psi_1(\xi_{\hat{\tau}(\mathscr{B}_1)})\right] \;\forall i \in \mathscr{B}_1^c\,. \qquad (6.3.16)\end{aligned}$$

Again, by applying Itô-Dynkin formula, from (6.3.1) for any $\pi^1 \in \Pi_{SM}^1$, $T > 0$ and $i \in \mathscr{D}_n \cap \mathscr{B}_1^c$ it follows that

$$\begin{aligned}\psi_{1,n}(i) &\leq E_i^{\pi^1,\hat{\pi}^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)\wedge\tau(\mathscr{D}_n)\wedge T}(c_1(\xi_t,\pi^1(\xi_{t-}),\hat{\pi}^2(\xi_{t-}))-\rho_{1,n})dt}\psi_{1,n}(\xi_{\hat{\tau}(\mathscr{B}_1)\wedge\tau(\mathscr{D}_n)\wedge T})\right]\\ &\leq E_i^{\pi^1,\hat{\pi}^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}(c_1(\xi_t,\pi^1(\xi_{t-}),\hat{\pi}^2(\xi_{t-}))-\rho_{1,n})dt}\psi_{1,n}(\xi_{\hat{\tau}(\mathscr{B}_1)})I_{\{\hat{\tau}(\mathscr{B}_1)\leq\tau(\mathscr{D}_n)\wedge T\}}\right]\\ &\quad+ E_i^{\pi^1,\hat{\pi}^2}\left[e^{\int_0^{T}(c_1(\xi_t,\pi^1(\xi_{t-}),\hat{\pi}^2(\xi_{t-}))-\rho_{1,n})dt}\psi_{1,n}(\xi_T)I_{\{T\leq\hat{\tau}(\mathscr{B}_1)\wedge\tau(\mathscr{D}_n)\}}\right]. \qquad (6.3.17)\end{aligned}$$

Under Assumption $(A2)$(a), the estimate (6.3.11) and the fact that $\psi_{1,n} \leq V$ (from the construction of $\hat{\theta}_n$, it is clear that if we replace $\psi_{1,n}$ by $\hat{\theta}_n\psi_{1,n}$, we get this inequality), we have

$$\begin{aligned}&E_i^{\pi^1,\hat{\pi}^2}\left[e^{\int_0^{T}(c_1(\xi_t,\pi^1(\xi_{t-}),\hat{\pi}^2(\xi_{t-}))-\rho_{1,n})dt}\psi_{1,n}(\xi_T)I_{\{T\leq\hat{\tau}(\mathscr{B}_1)\wedge\tau(\mathscr{D}_n)\}}\right]\\ &\leq e^{(\|c_1\|_\infty-\rho_{1,n}-\gamma)T}E_i^{\pi^1,\hat{\pi}^2}\left[e^{T\gamma}V(\xi_T)I_{\{T\leq\hat{\tau}(\mathscr{B}_1)\wedge\tau(\mathscr{D}_n)\}}\right]\\ &\leq e^{(\|c_1\|_\infty-\rho_{1,n}-\gamma)T}V(i)\,.\end{aligned}$$

Thus, letting $T \to \infty$ from (6.3.17) we get

$$\psi_{1,n}(i) \leq E_i^{\pi^1,\hat{\pi}^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}(c_1(\xi_t,\pi^1(\xi_{t-})),\hat{\pi}^2(\xi_{t-}))-\rho_{1,n})dt}\psi_{1,n}(\xi_{\hat{\tau}(\mathscr{B}_1)})I_{\{\hat{\tau}(\mathscr{B}_1)\leq\tau(\mathscr{D}_n)\}}\right].$$

Again, since $\psi_{1,n} \leq V$, using (6.3.11) and applying the dominated convergence theorem it follows that

$$\psi_1(i) \leq E_i^{\pi^1,\hat{\pi}^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}(c_1(\xi_t,\pi^1(\xi_{t-}),\hat{\pi}^2(\xi_{t-}))-\rho_1)dt}\psi_1(\xi_{\hat{\tau}(\mathscr{B}_1)})\right] \;\forall i \in \mathscr{B}_1^c\,. \qquad (6.3.18)$$

Since $\pi^1 \in \Pi^1_{SM}$ is arbitrary, combining (6.3.16) and (6.3.18), we obtain (6.3.7). Also, it is clear from the proof that for any minimizing selector π^{*1} of (6.3.5) we have

$$\psi_1(i) = E_i^{\pi^{*1},\hat{\pi}^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}(c_1(\xi_t,\pi^{*1}(\xi_{t-}),\hat{\pi}^2(\xi_{t-}))-\rho_1)dt}\psi_1(\xi_{\hat{\tau}(\mathscr{B}_1)})\right] \quad \forall i \in \mathscr{B}_1^c\,. \tag{6.3.19}$$

Using (6.3.12) it is easy to check that the same conclusion holds under Assumption $(A2)$(b).

Now exploiting the stochastic representation (6.3.7), we show that $(\rho_1, \psi_1) \in \mathbb{R}_+ \times L_V^{1,\infty}$ is the minimal eigenpair. Suppose $(\hat{\rho}_1, \hat{\psi}_1) \in \mathbb{R}_+ \times L_V^{1,\infty}$, $\hat{\psi}_1 > 0$ is an eigenpair satisfying

$$\begin{cases} \hat{\rho}_1\hat{\psi}_1(i) = \inf\limits_{\mu \in \mathcal{P}(A(i))}\left[\sum\limits_{j\in S}\hat{\psi}_1(j)q(j|i,\mu,\hat{\pi}^2(i)) + c_1(i,\mu,\hat{\pi}^2(i))\hat{\psi}_1(i)\right] \text{ for } i \in S, \\ \hat{\psi}_1(i_0) = 1. \end{cases} \tag{6.3.20}$$

We want to show that $\rho_1 \leq \hat{\rho}_1$. If not suppose that $\rho_1 > \hat{\rho}_1$. Then, for any minimizing selector $\hat{\pi}^{*1}$ of (6.3.20), applying Itô-Dynkin formula and Fatou's lemma, we obtain

$$\hat{\psi}_1(i) \geq E_i^{\hat{\pi}^{*1},\hat{\pi}^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}(c_1(\xi_t,\hat{\pi}^{*1}(\xi_{t-}),\hat{\pi}^2(\xi_{t-}))-\hat{\rho}_1)dt}\hat{\psi}_1(\xi_{\hat{\tau}(\mathscr{B}_1)})\right] \quad \forall i \in \mathscr{B}_1^c\,. \tag{6.3.21}$$

On the other hand from (6.3.7), we have

$$\psi_1(i) \leq E_i^{\hat{\pi}^{*1},\hat{\pi}^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}(c_1(\xi_t,\hat{\pi}^{*1}(\xi_{t-}),\hat{\pi}^2(\xi_{t-}))-\rho_1)dt}\psi_1(\xi_{\hat{\tau}(\mathscr{B}_1)})\right] \quad \forall i \in \mathscr{B}_1^c\,. \tag{6.3.22}$$

Let $\hat{\kappa} := \min_{\mathscr{B}_1}\frac{\hat{\psi}_1}{\psi_1}$. Hence, from (6.3.21) and (6.3.22) it follows that $(\hat{\psi}_1 - \hat{\kappa}\psi_1) \geq 0$ in S and $(\hat{\psi}_1 - \hat{\kappa}\psi_1)(\tilde{i}_0) = 0$ for some $\tilde{i}_0 \in \mathscr{B}_1$. Now, combining (6.3.5) and (6.3.20) we deduce that

$$\left[\sum_{j\neq\tilde{i}_0}(\hat{\psi}_1 - \hat{\kappa}\psi_1)(j)q(j|\tilde{i}_0,\hat{\pi}^{*1}(\tilde{i}_0),\hat{\pi}^2(\tilde{i}_0))\right] \equiv 0\,. \tag{6.3.23}$$

Since ξ_t is irreducible under $(\hat{\pi}^{*1}, \hat{\pi}^2)$, in view of (6.3.23) it is clear that $(\hat{\psi}_1 - \hat{\kappa}\psi_1) \equiv 0$. Again, since $\hat{\psi}_1(i_0) = \psi_1(i_0) = 1$, we get $\hat{\psi}_1 \equiv \psi_1$. But, this is a contradiction to the fact that $\rho_1 > \hat{\rho}_1$. Thus we deduce that $(\rho_1, \psi_1) \in \mathbb{R}_+ \times L_V^{1,\infty}$ is the minimal eigenpair. Following the above argument one can show that any eigenfunction satisfying (6.3.7) is unique up to scalar multiplication. Also, by a similar argument, one can show that there exists a minimal eigenpair $(\rho_2, \psi_2) \in \mathbb{R}_+ \times L_V^{1,\infty}$ satisfying (6.3.8), (6.3.9) and (6.3.10). This completes the proof. □

Remark 6.3.3. *We can replace Assumption $(A3)$(iii) by other similar assumption. For example, if the killed process communicates with every state from i_0 before leaving the domain $\mathscr{D}_n$, for large n, then our method applies. More precisely, for every $\mathscr{D}_n$, $(\pi^1, \pi^2) \in$*

$\Pi^1_{SM}\times\Pi^2_{SM}$ *and for every* $j\in\mathscr{D}_n\backslash\{i_0\}$*, if there exists distinct* $i_1,i_2,\cdots,i_m\in\mathscr{D}_n\backslash\{i_0\}$ *satisfying*

$$q(i_1|i_0,\pi^1(i_0),\pi^2(i_0))q(i_2|i_1,\pi^1(i_1),\pi^2(i_1))\cdots q(j|i_m,\pi^1(i_m),\pi^2(i_m))>0,$$

then we get $\psi_{1,n}(i_0)>0$ *in* $\mathscr{D}_n$ *(see Lemma 6.3.2). Also, the conclusion of Theorem 6.3.1 holds.*

To proceed further we establish some technical results needed later.

Lemma 6.3.4. *Suppose Assumptions (A0)-(A3) hold. Then the maps* $\tilde{\pi}^1\to\psi_2^{\tilde{\pi}^1}$ *from* $\Pi^1_{SM}\to L^\infty_V(S)$, $\tilde{\pi}^1\to\rho_2^{\tilde{\pi}^1}$ *from* $\Pi^1_{SM}\to\mathbb{R}_+$, $\tilde{\pi}^2\to\psi_1^{\tilde{\pi}^2}$ *from* $\Pi^2_{SM}\to L^\infty_V(S)$*, and* $\tilde{\pi}^2\to\rho_1^{\tilde{\pi}^2}$ *from* $\Pi^2_{SM}\to\mathbb{R}_+$ *are continuous.*

Proof. Let $\{\pi^{2,n}\}$ be a sequence in Π^2_{SM} such that $\pi^{2,n}\to\tilde{\pi}^2$ in Π^2_{SM}, i.e., for each $i\in S$, $\pi^{2,n}(i)\to\tilde{\pi}^2(i)$ in $\mathcal{P}(B(i))$. Now by Theorem 6.3.1, there exists $(\rho_1^{\pi^{2,n}},\ \psi_1^{\pi^{2,n}})\in\mathbb{R}_+\times L^{1,\infty}_V$, $\psi_1^{\pi^{2,n}}>0$ satisfying

$$\rho_1^{\pi^{2,n}}\psi_1^{\pi^{2,n}}(i)=\inf_{\mu\in\mathcal{P}(A(i))}\Big[\sum_{j\in S}\psi_1^{\pi^{2,n}}(j)q(j|i,\mu,\pi^{2,n}(i))+c_1(i,\mu,\pi^{2,n}(i))\psi_1^{\pi^{2,n}}(i)\Big],\quad(6.3.24)$$

with $\psi_1^{\pi^{2,n}}(i_0)=1$. Now, since $\psi_1^{\pi^{2,n}}\in L^{1,\infty}_V$, by a standard diagonalization argument, there exists a function $\psi_1^*\in L^{1,\infty}_V$ such that $\psi_1^{\pi^{2,n}}(i)\to\psi_1^*(i)$ as $n\to\infty$ for all $i\in S$. Also, $\{\rho_1^{\pi^{2,n}}\}$ is a bounded sequence. Hence, along a suitable subsequence (without loss of generality denoting by the same notation) $\rho_1^{\pi^{2,n}}\to\rho_1^*$. Now from (6.3.24), for any $\mu\in\mathcal{P}(A(i))$ we deduce that

$$\rho_1^{\pi^{2,n}}\psi_1^{\pi^{2,n}}(i)\leq\Big[\sum_{j\in S}\psi_1^{\pi^{2,n}}q(j|i,\mu,\pi^{2,n}(i))+c_1(i,\mu,\pi^{2,n}(i))\psi_1^{\pi^{2,n}}(i)\Big].$$

This implies that

$$\rho_1^{\pi^{2,n}}\psi_1^{\pi_{2,n}}(i)-\psi_1^{\pi^{2,n}}(i)q(i|i,\mu,\pi^{2,n}(i))\leq\Big[\sum_{j\neq i}\psi_1^{\pi^{2,n}}(j)q(j|i,\mu,\pi^{2,n}(i))+c_1(i,\mu,\pi^{2,n}(i))\psi_1^{\pi^{2,n}}(i)\Big].\quad(6.3.25)$$

Note that

$$\sum_{j\neq i}\psi_1^{\pi^{2,n}}(j)q(j|i,\mu,\pi^{2,n}(i))\leq\sum_{j\neq i}V(j)q(j|i,\mu,\pi^{2,n}(i)).\quad(6.3.26)$$

Thus, using Lemma 6.3.1, generalized Fatou's lemma in [69, Lemma 8.3.7] and taking $n\to\infty$ in (6.3.25), we get

$$\rho_1^*\psi_1^*(i)\leq\Big[\sum_{j\in S}\psi_1^*(j)q(j|i,\mu,\tilde{\pi}^2(i))+c_1(i,\mu,\tilde{\pi}^2(i))\psi_1^*(i)\Big].$$

Hence,

$$\rho_1^*\psi_1^*(i) \le \inf_{\mu\in\mathcal{P}(A(i))}\Big[\sum_{j\in S}\psi_1^*(j)q(j|i,\mu,\tilde{\pi}^2(i)) + c_1(i,\mu,\tilde{\pi}^2(i))\psi_1^*(i)\Big]. \tag{6.3.27}$$

Let $\pi^{*1,n}\in\Pi_{SM}^1$ be a minimizing selector of (6.3.24), i.e.,

$$\rho_1^{\pi^{2,n}}\psi_1^{\pi_{2,n}}(i) = \Big[\sum_{j\in S}\psi_1^{\pi^{2,n}}(j)q(j|i,\pi^{*1,n}(i),\pi^{2,n}(i)) + c_1(i,\pi^{*1,n}(i),\pi^{2,n}(i))\psi_1^{\pi^{2,n}}(i)\Big]. \tag{6.3.28}$$

Since Π_{SM}^1 is compact under the product topology, there exists $\pi^{*1}\in\Pi_{SM}^1$ such that along a subsequence (without loss of generality denoting by the same notation) $\pi^{*1,n}\to\pi^{*1}$.

Now, using Lemma 6.3.1, the dominated convergence theorem and passing $n\to\infty$ in (6.3.28), we obtain

$$\rho_1^*\psi_1^*(i) = \Big[\sum_{j\in S}\psi_1^*(j)q(j|i,\pi^{*1}(i),\tilde{\pi}^2(i)) + c_1(i,\pi^{*1}(i),\tilde{\pi}^2(i))\psi_1^*(i)\Big].$$

Therefore

$$\rho_1^*\psi_1^*(i) \ge \inf_{\mu\in\mathcal{P}(A(i))}\Big[\sum_{j\in S}\psi_1^*(j)q(j|i,\mu,\tilde{\pi}^2(i)) + c_1(i,\mu,\tilde{\pi}^2(i))\psi_1^*(i)\Big]. \tag{6.3.29}$$

Hence, from (6.3.27), and (6.3.29), it follows that

$$\rho_1^*\psi_1^*(i) = \inf_{\mu\in\mathcal{P}(A(i))}\Big[\sum_{j\in S}\psi_1^*(j)q(j|i,\mu,\tilde{\pi}^2(i)) + c_1(i,\mu,\tilde{\pi}^2(i))\psi_1^*(i)\Big]. \tag{6.3.30}$$

Since $\rho_1^{\tilde{\pi}^2}$ is the minimal eigenvalue corresponding to $\tilde{\pi}^2$ of (6.3.30), we have $\rho_1^*\ge\rho_1^{\tilde{\pi}^2}$. Suppose $\rho_1^*>\rho_1^{\tilde{\pi}^2}$. Now, from Theorem 6.3.1, for any minimizing $\hat{\pi}^1\in\Pi_{SM}^1$ of (6.3.5), there exists a finite set $\mathscr{B}_1\supset\mathscr{K}$, such that

$$\psi_1(i) = E_i^{\hat{\pi}^1,\tilde{\pi}^2}\Big[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}(c_1(t,\hat{\pi}^1(\xi_{t-}),\tilde{\pi}^2(\xi_{t-}))-\rho_1^{\tilde{\pi}^2})dt}\psi_1(\xi_{\hat{\tau}(\mathscr{B}_1)})\Big]\ \forall i\in\mathscr{B}_1^c, \tag{6.3.31}$$

where $\hat{\tau}(\mathscr{B}_1)$ is a stopping time define as in Theorem 6.3.1. Since $\rho_1^*>\rho_1^{\tilde{\pi}^2}$, by similar arguments as in [21, Lemma 3.4] we deduce that

$$\psi_1^*(i) \le E_i^{\hat{\pi}^1,\tilde{\pi}^2}\Big[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}(c_1(\xi_t,\hat{\pi}^1(\xi_{t-}),\tilde{\pi}^2(\xi_{t-}))-\rho_1^{\tilde{\pi}^2})dt}\psi_1^*(\xi_{\hat{\tau}(\mathscr{B}_1)})\Big]\ \forall i\in\mathscr{B}_1^c. \tag{6.3.32}$$

From (6.3.31) and (6.3.32), we obtain

$$(\psi_1-\psi_1^*)(i) \ge E_i^{\hat{\pi}^1,\tilde{\pi}^2}\Big[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}(c_1(\xi_t,\hat{\pi}^1(\xi_{t-}),\tilde{\pi}^2(\xi_{t-}))-\rho_1^{\tilde{\pi}^2})dt}(\psi_1-\psi_1^*)(\xi_{(\hat{\tau}(\mathscr{B}_1))})\Big]\ \forall i\in\mathscr{B}_1^c. \tag{6.3.33}$$

Now choosing an appropriate constant θ (e.g., $\theta = \max_{\mathscr{B}_1} \frac{\psi_1}{\psi_1^*}$), we have $(\psi_1 - \theta\psi_1^*) \geq 0$ in $\mathscr{B}_1$ and for some $\hat{i}_0 \in \mathscr{B}_1$, $(\psi_1 - \theta\psi_1^*)(\hat{i}_0) = 0$. Thus, in view of (6.3.33), we get $(\psi_1 - \theta\psi_1^*) \geq 0$ in S. Now combining (6.3.5) and (6.3.30), we get

$$\begin{aligned}&\rho_1^{\tilde{\pi}^2}(\psi_1 - \theta\psi_1^*)(\hat{i}_0)\\ &\qquad \geq \Big[\sum_{j\in S}(\psi_1 - \theta\psi_1^*)(j)q(j|\hat{i}_0, \tilde{\pi}^1(\hat{i}_0), \tilde{\pi}^2(\hat{i}_0)) + c_1(\hat{i}_0, \tilde{\pi}^1(\hat{i}_0), \tilde{\pi}^2(\hat{i}_0))(\psi_1 - \theta\psi_1^*)(\hat{i}_0)\Big].\end{aligned}$$

This implies that

$$\sum_{j\neq \hat{i}_0}(\psi_1 - \theta\psi_1^*)(j)q(j|\hat{i}_0, \tilde{\pi}^1(\hat{i}_0), \tilde{\pi}^2(\hat{i}_0)) = 0\,. \tag{6.3.34}$$

Since, $\{\xi_t\}_{t\geq 0}$ is irreducible under $(\tilde{\pi}^1, \tilde{\pi}^2) \in \Pi^1_{SM} \times \Pi^2_{SM}$, from (6.3.34) it follows that $\psi_1 \equiv \theta\psi_1^*$. But, this is a contradiction to the fact that $\rho_1^* > \rho_1^{\tilde{\pi}^2}$. Hence, we deduce that $\rho_1^* = \rho_1^{\tilde{\pi}^2}$. This proves the continuity of the map $\tilde{\pi}^2 \to \rho_1^{\tilde{\pi}^2}$. Since $\psi_1^{\tilde{\pi}^{2,n}}(i_0) = 1$ for all $n \geq 1$, we have $\psi_1^*(i_0) = 1$. Hence by Theorem 6.3.1, we have ψ_1^* is the unique solution of (6.3.5). Thus $\psi_1^* = \psi_1^{\tilde{\pi}^2}$. This proves the continuity of the map $\tilde{\pi}_2 \to \psi_1^{\tilde{\pi}^2}$. The continuity of other maps follows by a similar argument. □

Fix $\tilde{\pi}^2 \in \Pi^2_{SM}$. For each $i \in S$, $\mu \in \mathcal{P}(A(i))$, set

$$\tilde{F}_1(i, \mu, \tilde{\pi}^2(i)) = \Big[\sum_{j\in S}\psi_1^{\tilde{\pi}^2}(j)q(j|i, \mu, \tilde{\pi}^2(i)) + c_1(i, \mu, \tilde{\pi}^2(i))\psi_1^{\tilde{\pi}^2}(i)\Big],$$

where $\psi_1^{\tilde{\pi}^2}$ is the solution of (6.3.5) corresponding to the strategy $\tilde{\pi}^2 \in \Pi^2_{SM}$. Let

$$\tilde{H}(\tilde{\pi}^2) = \Big\{\hat{\pi}^{*1} \in \Pi^1_{SM} : \tilde{F}_1(i, \hat{\pi}^{*1}(i), \tilde{\pi}^2(i)) = \inf_{\mu\in\mathcal{P}(A(i))}\tilde{F}_1(i, \mu, \tilde{\pi}^2(i)) \ \forall\, i \in S\Big\}.$$

By Lemma 6.3.1, we know that the functions $c_1(i, \mu, \tilde{\pi}^2(i))\psi_1^{\tilde{\pi}^2}(i)$ and $\sum_{j\in S}\psi_1^{\tilde{\pi}^2}(j)q(j|i, \mu, \tilde{\pi}^2(i))$ are continuous on $\mathcal{P}(A(i)) \times \mathcal{P}(B(i))$ for each $i \in S$. Also since $\mathcal{P}(A(i))$ is compact for each $i \in S$, it is easy to see that $\tilde{H}(\tilde{\pi}_2)$ is a non empty subset of Π^1_{SM}. From the definition of $\tilde{H}(\tilde{\pi}^2)$ and the topology of Π^1_{SM}, it is clear that $\tilde{H}(\tilde{\pi}^2)$ is convex and closed. Since Π^1_{SM} is a compact metric space under the product topology, it follows that $\tilde{H}(\tilde{\pi}^2)$ is also compact. Similarly, for $i \in S$, $\tilde{\pi}^1 \in \Pi^1_{SM}$, $\nu \in \mathcal{P}(B(i))$, we set

$$\tilde{F}_2(i, \tilde{\pi}^1(i), \nu) = \Big[\sum_{j\in S}\psi_2^{\tilde{\pi}^1}(j)q(j|i, \tilde{\pi}^1(i), \nu) + c_2(i, \tilde{\pi}^1(i), \nu)\psi_2^{\tilde{\pi}^1}(i)\Big],\ i \in S,$$

where $\psi_2^{\tilde{\pi}^1}$ is the solution of (6.3.8) corresponding to the strategy $\tilde{\pi}^1 \in \Pi^1_{SM}$. Let

$$\tilde{H}(\tilde{\pi}^1) = \Big\{\hat{\pi}^{*2} \in \Pi^2_{SM} : \tilde{F}_2(i, \tilde{\pi}^1(i), \hat{\pi}^{*2}(i)) = \inf_{\nu\in\mathcal{P}(B(i))}\tilde{F}_2(i, \tilde{\pi}^1(i), \nu) \ \forall\, i \in S\Big\}.$$

Then by analogous arguments, $\tilde{H}(\hat{\pi}^1)$ is a nonempty, convex and compact subset of Π^2_{SM}. Next set

$$\tilde{H}(\hat{\pi}^1, \hat{\pi}^2) = \tilde{H}(\hat{\pi}^2) \times \tilde{H}(\hat{\pi}^1).$$

From the above argument it is clear that $\tilde{H}(\hat{\pi}^1, \hat{\pi}^2)$ is a nonempty, convex, and compact subset of $\Pi^1_{SM} \times \Pi^2_{SM}$. Therefore we may define a map from $\Pi^1_{SM} \times \Pi^2_{SM} \to 2^{\Pi^1_{SM} \times \Pi^2_{SM}}$.

6.4 Existence of Nash equilibria

Next lemma proves the upper semicontinuity of a certain set-valued map. This result will be useful in establishing the existence of a Nash equilibrium in the space of stationary Markov strategies.

Lemma 6.4.1. *Suppose Assumptions (A0)-(A3) hold. Then the map* $(\hat{\pi}^1, \hat{\pi}^2) \to \tilde{H}(\hat{\pi}^1, \hat{\pi}^2)$ *from* $\Pi^1_{SM} \times \Pi^2_{SM} \to 2^{\Pi^1_{SM} \times \Pi^2_{SM}}$ *is upper semicontinuous.*

Proof. Let $\{(\pi^1_m, \pi^2_m)\} \in \Pi^1_{SM} \times \Pi^2_{SM}$ and $(\pi^1_m, \pi^2_m) \to (\hat{\pi}^1, \hat{\pi}^2)$ in $\Pi^1_{SM} \times \Pi^2_{SM}$, i.e., for each $i \in S$, $(\pi^1_m(i), \pi^2_m(i)) \to (\hat{\pi}^1(i), \hat{\pi}^2(i))$ in $\mathcal{P}(A(i)) \times \mathcal{P}(B(i))$. Let $\bar{\pi}^1_m \in \tilde{H}(\pi^2_m)$. Then $\{\bar{\pi}^1_m\} \subset \Pi^1_{SM}$. Since Π^1_{SM} is compact, it has a convergent subsequence (denoted by the same sequence by abuse of notation), such that

$$\bar{\pi}^1_m \to \bar{\pi}^1 \text{ in } \Pi^1_{SM}.$$

Then $(\bar{\pi}^1_m, \pi^2_m) \to (\bar{\pi}^1, \hat{\pi}^2)$ in $\Pi^1_{SM} \times \Pi^2_{SM}$. Note that

$$\sum_{j \neq i} q(j|i, \bar{\pi}^1_m(i), \pi^2_m(i)) \psi_1^{\pi^2_m}(j) \leq \sum_{j \neq i} q(j|i, \bar{\pi}^1_m(i), \pi^2_m(i)) V(j).$$

Recall that by Lemma 6.3.4 the maps $\hat{\pi}^1 \to \psi_2^{\hat{\pi}^1}$, $\hat{\pi}^2 \to \psi_1^{\hat{\pi}^2}$, $\hat{\pi}^1 \to \rho_2^{\hat{\pi}^1}$, $\hat{\pi}^2 \to \rho_1^{\hat{\pi}^2}$ are continuous. Thus by generalized Fatou's lemma [69, Lemma 8.3.7], Assumption (A3) and the (product) topology of Π^k_{SM}, $k = 1, 2$, it follows that for each $i \in S$,

$$\sum_{j \in S} q(j|i, \bar{\pi}^1_m(i), \pi^2_m(i)) \psi_1^{\pi^2_m}(j) + c_1(i, \bar{\pi}^1_m(i), \pi^2_m(i)) \psi_1^{\pi^2_m}(i)$$

converges to

$$\sum_{j \in S} q(j|i, \bar{\pi}^1(i), \hat{\pi}^2(i)) \psi_1^{\hat{\pi}^2}(j) + c_1(i, \bar{\pi}^1(i), \hat{\pi}^2(i)) \psi_1^{\hat{\pi}^2}(i).$$

Hence

$$\lim_{m \to \infty} \tilde{F}_1(i, \bar{\pi}^1_m(i), \pi^2_m(i)) = \tilde{F}_1(i, \bar{\pi}^1(i), \hat{\pi}^2(i)). \tag{6.4.1}$$

Now fix $\tilde{\pi}^1 \in \Pi^1_{SM}$ and consider the sequence $\{(\tilde{\pi}^1, \pi^2_m)\}$. Using analogous arguments as above, we conclude that

$$\lim_{m\to\infty} \tilde{F}_1(i, \tilde{\pi}^1(i), \pi^2_m(i)) = \tilde{F}_1(i, \tilde{\pi}^1(i), \hat{\pi}^2(i)). \tag{6.4.2}$$

Since $\bar{\pi}^1_m \in \tilde{H}(\pi^2_m)$, for any m we have

$$\tilde{F}_1(i, \tilde{\pi}^1(i), \pi^2_m(i)) \geq \tilde{F}_1(i, \bar{\pi}^1_m(i), \pi^2_m(i)).$$

Thus, in view of (6.4.1) and (6.4.2), taking $m \to \infty$ in the above equation, for any $\tilde{\pi}^1 \in \Pi^1_{SM}$ we get

$$\tilde{F}_1(i, \tilde{\pi}^1(i), \hat{\pi}^2(i)) \geq \tilde{F}_1(i, \bar{\pi}^1(i), \hat{\pi}^2(i)).$$

Therefore, $\bar{\pi}^1 \in \tilde{H}(\hat{\pi}^2)$. Suppose $\bar{\pi}^2_m \in \tilde{H}(\pi^1_m)$ and along a subsequence $\bar{\pi}^2_m \to \bar{\pi}^2$ in Π^2_{SM}. Then, by similar arguments as above one can show that $\bar{\pi}^2 \in \tilde{H}(\hat{\pi}^1)$. This proves that the map $(\hat{\pi}^1, \hat{\pi}^2) \to \tilde{H}(\hat{\pi}^1, \hat{\pi}^2)$ is upper semicontinuous. □

Theorem 6.4.1. *Suppose that Assumptions* (A0)-(A3) *are satisfied. Then there exists a Nash equilibrium in the space of stationary Markov strategies* $\Pi^1_{SM} \times \Pi^2_{SM}$.

Proof. Since $\tilde{H}(\hat{\pi}^{*1}, \hat{\pi}^{*2})$ is non-empty, compact and convex, using Lemma 6.4.1 and Fan's fixed point theorem [33], it follows that there exists a fixed point $(\hat{\pi}^{*1}, \hat{\pi}^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$, for the map $(\hat{\pi}^1, \hat{\pi}^2) \to \tilde{H}(\hat{\pi}^1, \hat{\pi}^2)$ from $\Pi^1_{SM} \times \Pi^2_{SM} \to 2^{\Pi^1_{SM} \times \Pi^2_{SM}}$, i.e.,

$$(\hat{\pi}^{*1}, \hat{\pi}^{*2}) \in \tilde{H}(\hat{\pi}^{*1}, \hat{\pi}^{*2}).$$

This implies that $(\rho_1^{\hat{\pi}^{*2}}, \psi_1^{\hat{\pi}^{*2}})$, $(\rho_2^{\hat{\pi}^{*1}}, \psi_2^{\hat{\pi}^{*1}})$ satisfy the following coupled HJB equations:

$$\begin{cases} \rho_1^{\hat{\pi}^{*2}} \psi_1^{\hat{\pi}^{*}_2}(i) & = \inf_{\mu \in \mathcal{P}(A(i))} \left[\sum_{j\in S} q(j|i, \mu, \hat{\pi}^{*2}(i)) \psi_1^{\hat{\pi}^{*2}}(j) + c_1(i, \mu, \hat{\pi}^{*2}(i)) \psi_1^{\hat{\pi}^{*2}}(i) \right] \\ & = \left[\sum_{j\in S} q(j|i, \hat{\pi}^{*1}(i), \hat{\pi}^{*2}(i)) \psi_1^{\hat{\pi}^{*2}}(j) + c_1(i, \hat{\pi}^{*1}(i), \hat{\pi}^{*2}(i)) \psi_1^{\hat{\pi}^{*2}}(i) \right], \\ \psi_1^{\hat{\pi}^{*2}}(i_0) = 1 & \end{cases} \tag{6.4.3}$$

and

$$\begin{cases} \rho_2^{\hat{\pi}^{*1}} \psi_2^{\hat{\pi}^{*1}}(i) & = \inf_{\nu \in \mathcal{P}(B(i))} \left[\sum_{j\in S} q(j|i, \hat{\pi}^{*1}(i), \nu) \psi_2^{\hat{\pi}^{*1}}(j) + c_2(i, \hat{\pi}^{*1}(i), \nu) \psi_2^{\hat{\pi}^{*1}}(i) \right] \\ & = \left[\sum_{j\in S} q(j|i, \hat{\pi}^{*1}(i), \hat{\pi}^{*2}(i)) \psi_2^{\hat{\pi}^{*1}}(j) + c_2(i, \hat{\pi}^{*1}(i), \hat{\pi}^{*2}(i)) \psi_2^{\hat{\pi}^{*1}}(i) \right], \\ \psi_2^{\hat{\pi}^{*2}}(i_0) = 1. & \end{cases} \tag{6.4.4}$$

Now by Theorem 6.3.1, from (6.4.3), it follows that

$$\rho_1^{\hat{\pi}^{*2}} = \inf_{\pi^1 \in \Pi^1_{Ad}} \rho_1^{\pi^1, \hat{\pi}^{*2}} = \rho_1^{\hat{\pi}^{*1}, \hat{\pi}^{*2}}. \tag{6.4.5}$$

Similarly, from (6.4.4), we have

$$\rho_2^{\hat{\pi}^{*1}} = \inf_{\pi^2 \in \Pi^2_{Ad}} \rho_2^{\hat{\pi}^{*1}, \pi^2} = \rho_2^{\hat{\pi}^{*1}, \hat{\pi}^{*2}}. \tag{6.4.6}$$

Thus, from equations (6.4.5) and (6.4.6), we get

$$\rho_1^{\pi^1, \hat{\pi}^{*2}} \geq \rho_1^{\hat{\pi}^{*1}, \hat{\pi}^{*2}}, \ \forall\ \pi^1 \in \Pi^1_{Ad},$$

$$\rho_2^{\hat{\pi}^{*1}, \pi^2} \geq \rho_2^{\hat{\pi}^{*1}, \hat{\pi}^{*2}}, \ \forall\ \pi^2 \in \Pi^2_{Ad}.$$

Hence $(\hat{\pi}^{*1}, \hat{\pi}^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ is a Nash equilibrium. This completes the proof. □

The above theorem establishes a Nash equilibrium belonging to the space of stationary Markov strategies. Note that the equilibrium thus obtained is a Nash equilibrium among all admissible strategies. However, the equilibrium need not be unique. In case the set-valued map (of optimal responses) admits a unique fixed point then the Nash-equilibrium will be unique. For uniqueness, stringent conditions may be required on the transition rates and cost functions. Next, we prove a converse of Theorem 6.4.1.

Theorem 6.4.2. *Suppose Assumptions (A0)-(A3) hold. If* $(\underline{\pi}^{*1}, \underline{\pi}^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ *is a Nash equilibrium, i.e.,*

$$\rho_1^{\pi^1, \underline{\pi}^{*2}} \geq \rho_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}, \ \forall\ \pi^1 \in \Pi^1_{Ad},$$

$$\rho_2^{\underline{\pi}^{*1}, \pi^2} \geq \rho_2^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}, \ \forall\ \pi^2 \in \Pi^2_{Ad}.$$

Then $\underline{\pi}^{*1} \in \Pi^1_{SM}$ *is a minimizing selector of (6.3.5) (corresponding to fixed strategy* $\underline{\pi}^{*2} \in \Pi^2_{SM}$ *of player 2) and* $\underline{\pi}^{*2} \in \Pi^2_{SM}$ *is a minimizing selector of (6.3.8) (corresponding to fixed strategy* $\underline{\pi}^{*1} \in \Pi^1_{SM}$ *of player 1).*

Proof. Applying analogous arguments as in [21, Lemma 3.4 and Remark 3.1], one can prove that for the given pair $(\underline{\pi}^{*1}, \underline{\pi}^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$, there exists a eigenpair $(\rho_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}, \psi_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}) \in \mathbb{R} \times L^\infty_V$, $\psi_1^{\underline{\pi}^{*1}} > 0$ and $\rho_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}} \geq 0$ satisfying

$$\begin{cases} \rho_1^{\underline{\pi}^{*1} \underline{\pi}^{*2}} \psi_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}(i) = \sum_{j \in S} q(j|i, \underline{\pi}^{*1}(i), \underline{\pi}^{*2}(i)) \psi_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}(j) + c_1(i, \underline{\pi}^{*1}(i), \underline{\pi}^{*2}(i)) \psi_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}(i), \\ \psi_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}(i_0) = 1. \end{cases} \tag{6.4.7}$$

Also, for given $\underline{\pi}^{*2} \in \Pi^2_{SM}$, there exists a minimal eigenpair $(\rho_1^{\underline{\pi}^{*2}}, \psi_1^{\underline{\pi}^{*2}}) \in \mathbb{R}_+ \times L^\infty_V$, $\psi_1^{\underline{\pi}^{*2}} > 0$, satisfying

$$\begin{cases} \rho_1^{\underline{\pi}^{*2}} \psi_1^{\underline{\pi}^{*2}}(i) = \inf\limits_{\mu \in \mathcal{P}(A(i))} \Big[\sum\limits_{j \in S} q(j|i, \mu, \underline{\pi}^{*2}(i)) \psi_1^{\underline{\pi}^{*2}}(j) + c_1(i, \mu, \underline{\pi}^{*2}(i)) \psi_1^{\underline{\pi}^{*2}}(i) \Big], \\ \psi_1^{\underline{\pi}^{*2}}(i_0) = 1. \end{cases} \tag{6.4.8}$$

Since $\rho_1^{\underline{\pi}^{*2}}$ is a minimal eigenvalue of (6.4.8), corresponding to $\underline{\pi}^{*2}$, we have

$$\rho_1^{\underline{\pi}^{*2}} = \inf_{\pi^1 \in \Pi^1_{Ad}} \rho_1^{\pi^1, \underline{\pi}^{*2}}. \tag{6.4.9}$$

Also, we have

$$\rho_1^{\pi^1, \underline{\pi}^{*2}} \geq \rho_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}, \ \forall \ \pi^1 \in \Pi^1_{Ad}.$$

Hence,

$$\inf_{\pi^1 \in \Pi^1_{Ad}} \rho_1^{\pi^1, \underline{\pi}^{*2}} \geq \rho_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}. \tag{6.4.10}$$

So, by (6.4.9) and (6.4.10), we obtain

$$\rho_1^{\underline{\pi}^{*2}} \geq \rho_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}.$$

Also, from (6.4.9), we have

$$\rho_1^{\underline{\pi}^{*2}} \leq \rho_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}.$$

Hence, we deduce that

$$\rho_1^{\underline{\pi}^{*2}} = \rho_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}. \tag{6.4.11}$$

Now, applying Ito-Dynkin formula, from (6.4.7), it follows that

$$\psi_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}(i) = E_i^{\underline{\pi}^{*1}, \underline{\pi}^{*2}} \Big[e^{\int_0^{T \wedge \hat{\tau}(\mathscr{B}_1)} (c_1(\xi_t, \underline{\pi}^{*1}(\xi_{t-}), \underline{\pi}^{*2}(\xi_{t-})) - \rho_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}) dt} \psi_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}(\xi_{T \wedge \hat{\tau}(\mathscr{B}_1)}) \Big] \ \forall i \in \mathscr{B}_1^c,$$

where $\mathscr{B}_1$ is as in Theorem 6.3.1. Now, by Fatou's Lemma, taking $T \to \infty$ in the above equation, we get

$$\psi_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}(i) \geq E_i^{\underline{\pi}^{*1}, \underline{\pi}^{*2}} \Big[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)} (c_1(\xi_t, \underline{\pi}^{*1}(\xi_{t-}), \underline{\pi}^{*2}(\xi_{t-})) - \rho_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}) dt} \psi_1^{\underline{\pi}^{*1}, \underline{\pi}^{*2}}(\xi_{\hat{\tau}(\mathscr{B}_1)}) \Big] \ \forall i \in \mathscr{B}_1^c. \tag{6.4.12}$$

Again, using (6.4.8), from Theorem 6.3.1, it follows that

$$\psi_1^{\underline{\pi}^{*2}}(i) \leq E_i^{\underline{\pi}^{*1}, \underline{\pi}^{*2}} \Big[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)} (c_1(\xi_t, \underline{\pi}^{*1}(\xi_{t-}), \underline{\pi}^{*2}(\xi_{t-})) - \rho_1^{\underline{\pi}^{*2}}) dt} \psi_1^{\underline{\pi}^{*2}}(\xi_{\hat{\tau}(\mathscr{B}_1)}) \Big] \ \forall i \in \mathscr{B}_1^c. \tag{6.4.13}$$

So, by (6.4.12) and (6.4.13), we obtain

$$\begin{aligned} &\psi_1^{\underline{\pi}^{*1},\underline{\pi}^{*2}}(i) - \psi_1^{\underline{\pi}^{*2}}(i) \\ &\qquad \geq E_i^{\underline{\pi}^{*1},\underline{\pi}^{*2}}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)}(c_1(\xi_t,\underline{\pi}^{*1}(\xi_{t-}),\underline{\pi}^{*2}(\xi_{t-}))-\rho_1^{\underline{\pi}^{*2}})dt}(\psi_1^{\underline{\pi}^{*1},\underline{\pi}^{*2}} - \psi_1^{\underline{\pi}^{*2}})(\xi_{\hat{\tau}(\mathscr{B}_1)})\right] \ \forall i \in \mathscr{B}_1^c. \end{aligned} \tag{6.4.14}$$

Now arguing as in the proof of Lemma 6.3.4, we obtain $\psi_1^{\underline{\pi}^{*1},\underline{\pi}^{*2}}(i) \equiv \psi_1^{\underline{\pi}^{*2}}$. Thus, from (6.4.7) and (6.4.8) it follows that $\underline{\pi}^{*1}$ is a minimizing selector of (6.3.5) (for fixed strategy $\underline{\pi}^{*2} \in \Pi_{SM}^2$ of player 2). Following similar arguments one can show that $\underline{\pi}^{*2}$ is a minimizing selector of (6.3.8) (for fixed strategy $\underline{\pi}^{*1} \in \Pi_{SM}^1$ of player 1). This completes the proof. ☐

6.5 Example

In this section, we present an illustrative where transition rates are unbounded and cost rates are nonnegative and unbounded.

Example 6.5.1. *Consider a shop that deals with only one type of product for buying and selling. Suppose there are two workers, say, player 1 and player 2 for buying and selling the products, respectively. The number of stocks in the shop is a finite subset of the set of natural numbers* $\mathbb{N}$ *at each time* $t \geq 0$. *There are 'natural' buying and selling rates, say* $\tilde{\mu}$ *and* λ, *respectively, and buying parameters* h_1 *controlled by player 1 and selling parameters* h_2 *controlled by player 2. When the state of the system is* $i \in S := \{1, 2, \cdots\}$ *(i.e., number of items in the shop), player 1 takes an action* a *from a given set* $A(i)$, *which may increase* $(h_1(i,a) \geq 0)$ *or decrease* $(h_1(i,a) \leq 0)$ *the buying rate. These actions produce a payoff denoted by* $r_1(i,a)$ *per unit of time. Similarly, if the state is* $i \in S$, *player 2 takes an action* b *from a set* $B(i)$ *to decrease* $(h_2(i,b) \leq 0)$ *or to increase* $(h_2(i,b) \geq 0)$ *the selling rate. These actions result in a payoff denoted by* $r_2(i,b)$ *per unit of time. We assume that when the stock of items in the shop becomes 1, the first player may buy any number of stocks of that item as much as he/she likes depending upon the availability of cash. In addition, we assume that player* $k, (k = 1, 2)$ *'gets' a reward* $r_k(i) := p_k i$ *or incurs a cost* $r_k(i) := p_k i$ *for each unit of time during which the system remains in the state* $i \in S$, *where* $p_k > 0$ *is a fixed reward fee, and* $p_k < 0$, *a fixed cost fee, per stock, from the owner.*
We next formulate this model as a continuous-time Markov game. The corresponding transition rate $q(j|i,a,b)$ *and payoff rate* $c_k(i,a,b)$ *for player* k, $(k = 1, 2)$ *are given as*

follows: for $(1,a,b) \in K$ *(*K *as in the game model section 5.2.*

$$q(j|1,a,b) > 0 \ \forall j \geq 2, \ \textit{such that} \sum_{j\in S} q(j|1,a,b) = 0, \ \textit{and} \ q(j|1,a,b) \leq e^{-2\theta j} \ \forall \ j \geq 2, \tag{6.5.1}$$

where $\theta > 0$ *is a constant.*
Also, for $(i,a,b) \in K$ *with* $i \geq 2$*,*

$$q(j|i,a,b) = \begin{cases} \lambda i + h_2(i,b), & \textit{if } j = i-1 \\ -\tilde{\mu} i - \lambda i - h_1(i,a) - h_2(i,b), & \textit{if } j = i \\ \tilde{\mu} i + h_1(i,a), & \textit{if } j = i+1 \\ 0, & \textit{otherwise.} \end{cases}$$

$$c_1(i,a,b) := ip_1 - r_1(i,a), \ c_2(i,a,b) = ip_2 - r_2(i,b) \ \ \textit{for} \ \ (i,a,b) \in K. \tag{6.5.2}$$

We now investigate conditions under which there exists a Nash equilibrium. To this end we make the following assumptions:

(I) For each $i \in S$*,* $A(i) = B(i) = [-L, L]$*,* $L > 0$ *is a constant.*

(II) Let $\lambda e^{-\theta} > \tilde{\mu} > 0$*,* $\tilde{\mu} i + h_1(i,a) \geq 0$ *and* $\lambda i + h_2(i,b) \geq 0$ *for all* $(i,a,b) \in K$ *with* $i \geq 2$*.*

(III) The functions $h_1(i,a)$*,* $h_2(i,b)$*,* $r_1(i,a)$*,* $r_2(i,b)$*, and* $q(1|1,a,b)$ *are continuous in* (a,b) *for each fixed* $i \in S$*. Suppose there exists a finite set* $\mathscr{K}$ *such that* $h_1(i,a) = \frac{a}{e^{\theta i}} I_{\mathscr{K}}(i)$*,* $h_2(i,b) = \frac{b}{e^{\theta i}} I_{\mathscr{K}}(i)$ *and* $1 \in \mathscr{K}$*. Also assume that* $\inf_{(a,b)\in A(\cdot)\times B(\cdot)} r_1(\cdot,a)$ *and* $\inf_{(a,b)\in A(\cdot)\times B(\cdot)} r_2(\cdot,b)$ *are norm like functions.*

(IV) Suppose $ip_1 - r_1(i,a) \geq 0$*,* $ip_2 - r_2(i,b) \geq 0$ $\forall i \in S, (a,b) \in A(i) \times B(i)$ *and* $(1-e^{-\theta})\lambda + (1-e^{\theta})\tilde{\mu} > p_k$ *for* $k = 1,2$*.*

Proposition 6.5.2. *Under conditions (I)-(IV), the above controlled system satisfies the Assumptions* (A0)*-*(A3)*. Hence by Theorem 6.4.1, there exists a Nash equilibrium.*

Proof. Take a Lyapunov function as $V(i) := e^{\theta i}$ for $i \in S$ for some $\theta > 0$ as described earlier. Then, we have $V(i) \geq 1$ for all $i \in S$. Now for each $i \geq 2$, and $(a,b) \in A(i)\times B(i)$, we have

$$\begin{aligned} \sum_{j\in S} q(j|i,a,b)V(j) &= q(i-1|i,a,b)V(i-1) + V(i)q(i|i,a,b) + V(i+1)q(i+1|i,a,b) \\ &= e^{\theta i}\Big[(\lambda i + h_2(i,b))e^{-\theta} - (i\tilde{\mu} + \lambda i + h_1(i,a) + h_2(i,b)) + (\tilde{\mu} i + h_1(i,a))e^{\theta}\Big] \end{aligned}$$

$$
\begin{aligned}
&= e^{\theta i} i \left[\tilde{\mu}(e^{\theta}-1) + \lambda(e^{-\theta}-1) + \frac{e^{\theta}h_1(i,a) + e^{-\theta}h_2(i,b) - h_1(i,a) - h_2(i,b)}{i} \right] \\
&= iV(i)[\tilde{\mu}(e^{\theta}-1) + \lambda(e^{-\theta}-1)] + \left[a(e^{\theta}-1) + b(e^{-\theta}-1) \right] I_{\mathscr{K}}(i) \\
&\leq iV(i)[\tilde{\mu}(e^{\theta}-1) + \lambda(e^{-\theta}-1)] + L(e^{\theta}-1) I_{\mathscr{K}}(i). \qquad (6.5.3)
\end{aligned}
$$

Now for every $\theta > 0$, we know

$$\lambda(e^{-\theta}-1) + \tilde{\mu}(e^{\theta}-1) < 0 \Leftrightarrow \tilde{\mu} < \lambda e^{-\theta}.$$

Let $[\tilde{\mu}(e^{\theta}-1) + \lambda(e^{-\theta}-1)] = -\alpha$ for some $\alpha > 0$. Also, let $\ell(i) = i\alpha$ and $C_4 = \max\left\{ L(e^{\theta}-1), \frac{e^{-2\theta}}{1-e^{-\theta}} \right\}$ (see (6.5.5)). Then for $i \geq 2$,

$$\sup_{(a,b)\in A(i)\times B(i)} \sum_{j\in S} V(j) q(j|i,a,b) \leq C_4 I_{\mathscr{K}}(i) - \ell(i)V(i) \ \forall i \in S. \qquad (6.5.4)$$

Also, we have

$$\sum_{j\in S} q(j|1,a,b)V(j) < q(1|1,a,b)e^{\theta} + \sum_{j\geq 2} e^{-2\theta j} e^{\theta j} \leq q(1|1,a,b)e^{\theta} + \frac{e^{-2\theta}}{1-e^{-\theta}} < \infty. \qquad (6.5.5)$$

Since $-\ell(i) < 1$ for all $i \in S$. Hence from (6.5.4) and (6.5.5), for $i \geq 1$, we have

$$\sum_{j\in S} q(j|i,a,b)V(j) \leq C_1 \mathcal{V}(i) + C_2, \quad \text{where } C_1 = 1 \ \text{ and } C_2 = C_4. \qquad (6.5.6)$$

For $i \geq 2$,

$$
\begin{aligned}
-q(i|i,a,b) &= \tilde{\mu} i + \lambda i + h_1(i,a) + h_2(i,b) \\
&\leq i(\tilde{\mu} + \lambda) + 2L \\
&\leq \frac{1}{\theta}(\tilde{\mu} + \lambda)V(i) + 2LV(i) \\
&= [2L + (\tilde{\mu} + \lambda)\frac{1}{\theta}]V(i) \\
&= C_3 V(i). \qquad (6.5.7)
\end{aligned}
$$

Take $W = \tilde{W} = \mathcal{V}$. Now

$$
\begin{aligned}
\ell(i) - \sup_{(a,b)\in A(i)\times B(i)} c_1(i,a,b) &= \alpha i - i p_1 + \inf_{a\in A(i)} r_1(i,a) \\
&= i\beta_1 + \inf_{a\in A(i)} r_1(i,a). \qquad (6.5.8)
\end{aligned}
$$

Similarly,

$$\ell(i) - \sup_{(a,b)\in A(i)\times B(i)} c_2(i,a,b) = i\beta_2 + \inf_{b\in B(i)} r_2(i,b). \qquad (6.5.9)$$

We see that from condition (IV), that $\beta_k = \alpha - p_k \geq 0$. So, $\ell(i) - \sup_{(a,b)\in A(i)\times B(i)} c_k(i,a,b)$ is norm-like function for $k = 1, 2$. Now by (6.5.6), we say Assumption $(A1)$(i) holds. Also by (6.5.1) and (6.5.7), Assumption $(A1)$(ii) is verified.
Now we verify Assumption $(A2)$. By (6.5.4), (6.5.5) and (6.5.8), it is easy to see that Assumption $(A2)$ is satisfied.
Now by condition (III) and (6.5.2), we say $c_k(i,a,b)$ and $q(j|i,a,b)$ are continuous in $(a,b) \in A(i) \times B(i)$ for each fixed $i, j \in S$ and for $k = 1, 2$. So, Assumption $(A3)$(i) is verified. By (6.5.3) and (6.5.5) and condition (III), we say that Assumption $(A3)$(ii) is verified. Also, from (6.5.1) it is easy to see that Assumption $(A3)$(iii) is satisfied.
Hence by Theorem 6.4.1 there exists a Nash equilibrium for this controlled process. □

Remark 6.5.3. *It should be noted that here we assume when the number of stock in the shop is one and the players independently choose action according to some strategies* $(\pi^1, \pi^2) \in \Pi^1_{SM} \times \Pi^2_{SM}$*, respectively, then with a positive probability first player may buy any number of stocks of the item, i.e.,* $q(j|1, \pi^1(1), \pi^2(1)) > 0$ *for all* $j \in S$*. In view of Remark 6.3.3, one can weaken this Assumption. For any* $i \in \mathscr{D}_n$ *(for* n *large enough)* $i >> 1$*, the player may increase the number of stock in the shop from* 1 *to* i *in* m *number of steps in* $\mathscr{D}_n$*, i.e., there exists a finite sequence of states* $i_0, i_1, \ldots, i_m$ *connecting* $i_0 = 1$ *to* $i_m = i$*, satisfying*

$$q(i_1|i_0, \pi^1(i_0), \pi^2(i_0))q(i_2|i_1, \pi^1(i_1), \pi^2(i_1)) \cdots q(i|i_{m-1}, \pi^1(i_{m-1}), \pi^2(i_{m-1})) > 0\,,$$

where $i_0, i_1, \ldots, i_m \in \mathscr{D}_n$*.*

CHAPTER 7

Zero-Sum Stochastic Games in Continuous-time with Risk-Sensitive Average Cost Criterion on a Countable State Space

7.1 Introduction

In this chapter, we consider a risk-sensitive ergodic zero-sum game for continuous-time Markov decision processes (CTMDPs). In a zero-sum game, one player is trying to minimize her/his cost and the other player is trying to maximize the same.

The risk-sensitive ergodic cost stochastic optimal control problems for CTMDPs are first considered in [47]. Subsequently, these problems are studied extensively in the literature due to their applications in finance and large deviation theory. Recently, there has been extensive work on risk-sensitive ergodic cost criterion problems for CTMDPs; see, for example, [21], [52], [81], [82], [94] and the references therein. The risk-sensitive stochastic zero-sum games for controlled Markov processes have been studied in [12], [16], [45], [48], [109]. The corresponding nonzero-sum games are studied in [13], [77], [113]. In [12], [16], zero-sum risk-sensitive stochastic games for discrete-time controlled Markov decision processes with bounded cost are studied. Both papers first treat the discounted cost criterion and then study the ergodic cost criterion using vanishing discount asymptotics. The results of [12] are extended from countable state space to the general state space case in [16]. In this respect, we mention that the authors in the paper [22] studied the risk-sensitive zero-sum ergodic game problems for controlled diffusion processes in $\mathbb{R}^d$. Using the eigenvalue approach, they have completely characterized all possible saddle-point equilibrium in the space of stationary Markov strategies. The zero-sum risk-sensitive ergodic games are studied in [45] and discounted risk-sensitive zero-sum games were studied in [95] for CTMDPs with bounded cost and transition rates. But the boundedness requirement restricts its scope of applications, since in many real-life situations the reward/cost and transition rates are unbounded, for example in queueing, telecommunication, and population processes. In [21], the authors studied risk-sensitive stochastic optimal control problems for discrete/continuous-time Markov decision processes with ergodic cost criterion. They considered a cost minimization problem. In this chapter, we have extended their results from one-controller case to two controller cases with strictly opposite interests for CMDP. Also, this chapter is a generalization of the results of [45] to the case with unbounded cost. In particular, we have studied the risk-sensitive zero-sum stochastic game problem for continuous-time Markov decision process. To analyze our game problem we have used the eigenvalue approach (as in [21], [22]), which enables us to obtain a complete characterization of all possible saddle-point equilibrium in the space of stationary Markov strategies. More specifically, in this chapter, we study zero-sum ergodic risk-sensitive stochastic games for CTMDPs with the following features: (a) transition and cost rates may be unbounded (b) the state space is countable (c) the space of admissible actions at any state is compact (d) the strategies may be state feedback, i.e., at any point of time, the strategies are based on the past and present state. To the best of our knowledge, this chapter is the first work that deals with infinite-horizon continuous-time zero-sum risk-sensitive stochastic games for ergodic criteria on a countable state space for unbounded transition and cost rates. Under a Lyapunov stability condition, we prove the existence of a saddle-point equilibrium in the class of stationary strategies. We do not employ the traditional vanishing discount asymptotics. Instead, we treat the associated HJI equation as an eigenvalue problem. Using Kreĭn-Rutman theorem, we first prove that the corresponding HJI equation has a unique solution for any finite subset of the state space. Then using a Lyapunov stability condition, we establish the existence of a unique solution for the corresponding HJI equation on the whole state space. Also we give a complete characterization of saddle-point equilibrium in terms of the corresponding HJI equation.

The rest of this chapter is organized as follows. Section 7.2 contains the problem description and certain assumptions. In this section, we also write down the associated risk-sensitive optimality equation (HJI equation) which is an eigenvalue problem associated with an appropriate operator. In the next section we study the corresponding Dirichlet eigenvalue problem on a finite subset of S. Section 7.4 deals with the limiting analysis of the Dirichlet eigenvalue problems. This enables us to establish an appropriate solution of the HJI equation which in turn establishes the existence of a saddle-point equilibrium. Finally, we completely characterize all possible saddle-point equilibria in the class of stationary Markov strategies. In Section 7.5, we present an illustrative example. Our example is similar to the one in [56]. The content of this chapter is based on the published article [42].

7.2 The game model

In this section, we introduce the continuous-time zero-sum stochastic game model described by the following elements

$$\{S, A, B, (A(i) \subset A, B(i) \subset B, i \in S), q(\cdot|i,a,b), c(i,a,b)\}, \tag{7.2.1}$$

where

- S, called the state space, is the set of all nonnegative integers.
- A and B are action spaces for players 1 and 2, respectively. The action spaces A and B are assumed to be Borel spaces with the Borel σ-algebras $\mathcal{B}(A)$ and $\mathcal{B}(B)$, respectively.
- For each $i \in S$, $A(i) \in \mathcal{B}(A)$ and $B(i) \in \mathcal{B}(B)$ denote the sets of admissible actions for players 1 and 2 at the state i, respectively. Let $K := \{(i,a,b)|i \in S, a \in A(i), b \in B(i)\}$, which is a Borel subset of $S \times A \times B$. Throughout this chapter, we assume that the admissible action spaces $A(i) (\subset A)$ and $B(i) (\subset B)$ are compact for each i.
- Given any $(i,a,b) \in K$, the transition rate $q(j|i,a,b)$ is a signed kernel on S such that $q(j|i,a,b) \geq 0$ for all $j, i \in S$ with $j \neq i$. Moreover, we assume that $q(j|i,a,b)$ satisfies the following conservative and stable conditions: for any $i \in S$,

$$\sum_{j \in S} q(j|i,a,b) = 0 \text{ for all } (a,b) \in A(i) \times B(i) \quad \text{and}$$
$$q^*(i) := \sup_{(a,b) \in A(i) \times B(i)} q(i,a,b) < \infty, \tag{7.2.2}$$

  where $q(i,a,b) := -q(i|i,a,b) \geq 0$.
- Finally, the measurable function $c : K \to \mathbb{R}_+$ denotes the cost rate (representing cost for player 2 and the reward for player 1).

The evoluation of the game and the construction of the underlying CTMDPs ξ_t under admissible feedback strategies (as in [63], [75], [99]) is same as Chapter 3, p. 42-43.

To complete the specification of a risk-sensitive stochastic game problem, we need, of course, to introduce an optimality criterion. This requires to define the class of strategies as below.

Definition 7.2.1. *An admissible feedback strategy for player 1, denoted by* $\pi^1 = \{\pi^1(t)\}_{t \geq 0}$*, is a transition probability* $\pi^1(da|\omega, t)$ *from* $(\Omega \times [0,\infty), \mathscr{P})$ *onto* $(A_\Delta, \mathcal{B}(A_\Delta))$*,* *such that* $\pi^1(A(\xi_{t-}(\omega))|\omega, t) = 1$. *Using appropriate projections of the transition kernel* π^1*, an admissible feedback strategy for player 1, determines and is, in turn, determined by a sequence* $\{\pi^1_k, k \geq 0\}$ *of the stochastic kernel on* A *such that*

$$\begin{aligned}\pi^1(t)(\omega) &= \pi^1(da|\omega, t)\\ &= I_{\{t=0\}}(t)\pi^1_0(da|i'_0, 0) + \sum_{k\geq 0} I_{\{T_k<t\leq T_{k+1}\}}\pi^1_k(da|i'_0, \theta_1, i'_1, \ldots, \theta_k, i'_k, t - T_k)\\ &\quad + I_{\{t\geq T_\infty\}}\delta_{a_\Delta}(da),\end{aligned}$$

where $\pi^1_0(da|i'_0, 0)$ *is a stochastic kernel on* A *given* S *such that* $\pi^1_0(A(i'_0)|i'_0, 0) = 1$, $\pi^1_k(k \geq 1)$ *are stochastic kernels on* A *given* $(S \times (0,\infty))^{k+1}$ *such that* $\pi^1_k(A(i'_k)|i'_0, \theta_1, i'_1, \cdots, \theta_k, i'_k, t - T_k) = 1$*, and* $\delta_{a_\Delta}(da)$ *denotes the Dirac measure at the point* a_Δ.

For more details see [64, Definition 2.1, Remark 2.2], [99], [119]. The set of all admissible feedback strategies for player 1 is denoted by Π^1_{Ad}. A strategy $\pi^1 \in \Pi^1_{Ad}$ for player 1, is called a Markov if $\pi^1(t)(\omega) = \pi^1(\xi_{t-}(\omega), t)$ i.e., $\pi^1(da|\omega, t) = \pi^1(da|\xi_{t-}(\omega), t)$ for every $\omega \in \Omega$ and $t \geq 0$, where $\xi_{t-}(\omega) := \lim_{s\uparrow t} \xi_s(\omega)$. A Markov strategy π^1 is called a stationary Markov strategy if π^1 does not have an explicit dependence on time. For notational simplicity, we would not write ω anywhere throughout the rest of this chapter. We denote by Π^1_M and Π^1_{SM} the family of all Markov strategies and stationary Markov strategies, respectively, for player 1. The sets of admissible feedback strategies Π^2_{Ad}, Markov strategies Π^2_M and stationary strategies Π^2_{SM} for player 2 are defined analogously.

Remark 7.2.2. *As in [Chapter 3, Remark 3.2.2], in the definition of strategies we do not include the entire history of the game, i.e., past and present states, past sojourn times, and past actions taken by the players. In our game model each player's admissible strategies include only past and present states and past sojourn times. Hence such strategies are called feedback strategies.*

For any compact metric space Y, let $\mathcal{P}(Y)$ denote the space of probability measures on $\mathcal{B}(Y)$ with Prohorov topology. Since for each $i \in S$, $A(i)$ and $B(i)$ are compact sets, $\mathcal{P}(A(i))$ and $\mathcal{P}(B(i))$ are compact metric spaces. For each $i, j \in S$, $\mu \in \mathcal{P}(A(i))$ and $\nu \in \mathcal{P}(B(i))$, the associated cost and the transition rates are defined, respectively, as follows:

$$c(i, \mu, \nu) := \int_{B(i)} \int_{A(i)} c(i, a, b)\mu(da)\nu(db),$$

$$q(j|i, \mu, \nu) := \int_{B(i)} \int_{A(i)} q(j|i, a, b)\mu(da)\nu(db).$$

(by abuse of notation we use the same notation c and q). Note that for $\pi^1 \in \Pi^1_{SM}$ can be identified with a map $\pi^1 : S \to \mathcal{P}(A(i))$ such that for each $j \in S$, $\pi^1(j) \in \mathcal{P}(A(j))$

for each $j \in S$. Similarly, this conclusion also holds for $\pi^2 \in \Pi^2_{SM}$. Thus, we have $\Pi^1_{SM} = \Pi_{i\in S}\mathcal{P}(A(i))$ and $\Pi^2_{SM} = \Pi_{i\in S}\mathcal{P}(B(i))$ i.e., the sets Π^1_{SM} and Π^2_{SM} are endowed with the product topology. Therefore by Tychonoff theorem, the sets Π^1_{SM} and Π^2_{SM} are compact metric spaces. Note that under Assumption $(A1)$ (to be given shortly on p. 171) for any initial state $i \in S$ and any pair of strategies $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$, Theorem 4.27 in [76] yields the existence of a unique probability measure denoted by $P_i^{\pi^1,\pi^2}$ on $(\Omega, \mathscr{F})$ and the state process $\{\xi_t\}$ is a right-continuous process with left limits (rcll). Let $E_i^{\pi^1,\pi^2}$ be the expectation operator with respect to $P_i^{\pi^1,\pi^2}$. Also, from [59, p. 13-15], we know that $\{\xi_t\}_{t\geq 0}$ is a Markov process under any $(\pi^1, \pi^2) \in \Pi^1_M \times \Pi^2_M$ (in fact, strong Markov).

Now we give the definition of the risk-sensitive ergodic cost criterion for zero-sum continuous-time games. Since the risk-sensitive parameter remains fixed throughout we assume without any loss of generality that the risk-sensitivity coefficient $\theta = 1$. For each $i \in S$ and any $(\pi^1, \pi^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$, the risk-sensitive ergodic cost criterion is given by

$$J(i, c, \pi^1, \pi^2) := \limsup_{T\to\infty} \frac{1}{T} \log E_i^{\pi^1,\pi^2}\left[e^{\int_0^T c(\xi_t, \pi^1(t), \pi^2(t))dt}\right]. \tag{7.2.3}$$

Player 1 tries to maximize the above over his/her admissible strategies whereas player 2 tries to minimize the same.

Now we define the lower/upper value of the game. The functions on S defined by $L(i) := \sup_{\pi^1\in\Pi^1_{Ad}} \inf_{\pi^2\in\Pi^2_{Ad}} J(i, c, \pi^1, \pi^2)$ and $U(i) := \inf_{\pi^2\in\Pi^2_{Ad}} \sup_{\pi^1\in\Pi^1_{Ad}} J(i, c, \pi^1, \pi^2)$ are called, respectively, the lower value and the upper value of the game. It is easy to see that

$$L(i) \leq U(i) \text{ for all } i \in S.$$

Definition 7.2.3. *If $L(i) = U(i)$ for all $i \in S$, then the common function is called the value of the game and is denoted by $J^*(i)$.*

Definition 7.2.4. *Suppose that the game admits a value J^*. Then a strategy π^{*1} in Π^1_{Ad} is said to be optimal for player 1 if*

$$\inf_{\pi^2\in\Pi^2_{Ad}} J(i, c, \pi^{*1}, \pi^2) = J^*(i) \text{ for all } i \in S.$$

Similarly, $\pi^{*2} \in \Pi^2_{Ad}$ is optimal for player 2 if

$$\sup_{\pi^1\in\Pi^1_{Ad}} J(i, c, \pi^1, \pi^{*2}) = J^*(i) \text{ for all } i \in S.$$

If $\pi^{*k} \in \Pi^k_{Ad}$ is optimal for player k (k=1,2), then (π^{*1}, π^{*2}) is called a pair of optimal strategies and also called a saddle-point equilibrium.

Next, we list the commonly used notations below:

- Given any real-valued function $\mathcal{V} \geq 1$ on S, we define a Banach space $(L^{\infty}_{\mathcal{V}}(S), \|\cdot\|^{\infty}_{\mathcal{V}})$ of $\mathcal{V}$-weighted functions by
$$L^{\infty}_{\mathcal{V}} = \left\{ u : S \to \mathbb{R} \mid \|u\|^{\infty}_{\mathcal{V}}(S) := \sup_{i \in S} \frac{|u(i)|}{\mathcal{V}(i)} < \infty \right\}.$$
- For any finite set $\mathscr{B} \subset S$, $\hat{\tau}(\mathscr{B}) := \inf\{t > 0 : \xi_t \in \mathscr{B}\}$.

Let us now describe the zero-sum stochastic games. Suppose c_1, c_2 are cost functions of player 1 and player 2, respectively as described in Chapter 6. Then, as we have mentioned in Chapter 1 that, in zero-sum games, $c_1(i,a,b) + c_2(i,a,b) = 0$ for all $(i,a,b) \in K$. However, owing to the multiplicative nature of the evaluation criterion, we cannot say that the sum of the risk-sensitive ergodic costs is zero. In this case, if we set $c_1 = -c_2 = c$, then for $\theta > 0$, minimizing player is risk-averse whereas maximizing player is risk-seeking. Hence one must necessarily study the zero-sum case via Nash equilibria. Here we consider a particular type of a non-standard zero-sum risk-sensitive stochastic game that assumes $c_1 = c_2 = c$, for more information see, [90].

Our main goal is to establish the existence of a saddle-point equilibrium among the class of admissible feedback strategies. To this end, following [12] and [45], we investigate the HJI equation given by

$$\begin{aligned} \rho\psi(i) &= \sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \left[\sum_{j \in S} \psi(j) q(j|i,\mu,\nu) + c(i,\mu,\nu)\psi(i) \right] \\ &= \inf_{\nu \in \mathcal{P}(B(i))} \sup_{\mu \in \mathcal{P}(A(i))} \left[\sum_{j \in S} \psi(j) q(j|i,\mu,\nu) + c(i,\mu,\nu)\psi(i) \right]. \end{aligned} \tag{7.2.4}$$

Here ρ is a scalar and ψ is an appropriate function. The above is clearly an eigenvalue problem related to a nonlinear operator on an appropriate space. By a nonlinear version of Kreĭn-Rutman theorem, we first show that the Dirichlet eigenvalue problem associated with the above equation admits a solution in the space of bounded functions. Then by using a suitable limiting argument we show that the above HJI equation admits a principal eigenpair in an appropriate space. Finally exploiting the HJI equation, we completely characterize all possible saddle-point equilibria in the space of stationary Markov strategies. This is a brief outline of our procedure of establishing a saddle-point equilibrium. The details now follow.

Since the transition rates (i.e., $q(j|i,a,b)$) may be unbounded, to avoid the explosion of the state process $\{\xi_t, t \geq 0\}$, the following assumption is imposed on the transition rates, which are widely used in CTMDPs; see, for instance, [62], [63], [64] and references therein.

(*A*1) There exists a non-constant function $\tilde{V} : S \to [1, \infty)$ such that

(i) $\sum_{j\in S}\tilde{V}(j)q(j|i,a,b) \leq C_1\tilde{V}(i)+C_2$ for all $(a,b)\in A(i)\times B(i)$ and $i\in S$ with some constants $C_1\neq 0$, $C_2\geq 0$;

(ii) $q_i \leq C_3\tilde{V}(i)$ for all $i\in S$ with some positive constant C_3.

Throughout the rest of this chapter, we are going to assume that Assumption $(A1)$ holds. Note that if $\sup_{i\in S} q^*(i) < \infty$ then Assumption $(A1)$ holds trivially. In this case, we can choose $\tilde{V}$ to be a suitable constant.

Definition 7.2.5. *A function $f : S \to \mathbb{R}$ is said to be norm-like if for every $k\in\mathbb{R}$, the set $\{i\in S : f(i)\leq k\}$ is either empty or finite.*

Definition 7.2.6. *A time-homogeneous Markov process $\{\xi\}$ with rate matrix $Q=[q(j|i)]$ is irreducible if for any $i,j\in S$, $i\neq j$, there exist distinct states $i_1, i_2\cdots, i_k\in S$ satisfying $q(i_1|i)\cdots q(j|i_k)>0$ (see, [59], p. 107).*

Since we are allowing our transition and cost rates to be unbounded, to guarantee the finiteness of $J(i,c,\pi^1,\pi^2)$, we make the following Assumption.

($A2$) We assume that the CTMDP $\{\xi_t\}_{t\geq 0}$ is irreducible under every pair of stationary Markov strategies $(\pi^1,\pi^2)\in\Pi^1_{SM}\times\Pi^2_{SM}$. Assume that the cost function c is bounded below. Thus without loss of generality, we assume that $c\geq 0$. Furthermore, suppose there exists a constant $C>0$, a finite set $\mathscr{K}$, and a Lyapunov function $V : S\to[1,\infty)$ such that one of the following hold.

(a) **When the running cost is bounded:** For some positive constant $\gamma > \|c\|_\infty$, we have the following stability condition

$$\sup_{(a,b)\in A(i)\times B(i)}\sum_{j\in S}V(j)q(j|i,a,b)\leq CI_{\mathscr{K}}(i)-\gamma V(i)\ \forall i\in S, \tag{7.2.5}$$

where $\|c\|_\infty := \sup_{(i,a,b)\in K} c(i,a,b)$.

(b) **When the running cost is unbounded:** For some norm-like function $\ell : S\to \mathbb{R}_+$, the function $\ell(\cdot)-\max_{(a,b)\in A(\cdot)\times B(\cdot)} c(\cdot,a,b)$ is norm-like and we have the following stability condition

$$\sup_{(a,b)\in A(i)\times B(i)}\sum_{j\in S}V(j)q(j|i,a,b)\leq CI_{\mathscr{K}}(i)-\ell(i)V(i)\ \forall i\in S. \tag{7.2.6}$$

This type of Foster-Lyapunov condition on the dynamics is quite common in the literature to study continuous-time risk-sensitive ergodic control problems. For example, see [6], [7] for the controlled diffusion case and [21], [52] for the Markov decision process case.

We wish to establish the existence of a saddle-point equilibrium in the class of all stationary strategies. In view of this, we also use the following assumptions. Let $i_0 \in S$ be a fixed point (a reference state).

$(A3)$

(i) For any fixed $i, j \in S$ the functions $q(j|i,a,b)$ and $c(i,a,b)$ are continuous in $(a,b) \in A(i) \times B(i)$.

(ii) The sum $\sum_{j\in S} V(j) q(j|i,a,b)$ is continuous in $(a,b) \in A(i) \times B(i)$ for any given $i \in S$, where V is as in Assumption $(A2)$.

(iii) There exists a state $i_0 \in S$ such that any state can be reached from i_0, i.e., $q(j|i_0,a,b) > 0$ for all $j \neq i_0$ and for all $(a,b) \in A(i_0) \times B(i_0)$.

Note that Assumption $(A3)$((i)-(ii)) are quite routine assumptions for controlled Markov decision processes (MDPs). Here Assumption $(A3)$(iii) is very important to show that the limit of the sequence of Dirichlet eigenfunctions is nontrivial (see Lemma 7.3.3 and Lemma 7.4.1 below). It is important to note that we can weaken the Assumption $(A3)$(iii), see Remark 7.4.2 for further discussion.

7.3 Dirichlet Eigenvalue Problems

First, we recall a version of the nonlinear Kreĭn-Rutman theorem from [4, Section 3.1], (cf. [80]) which we use to study Dirichlet eigenvalue problems associated with the HJI equation. Let $\hat{\mathscr{X}}$ be an ordered Banach space. In what follows $\succeq$ denotes a partial ordering in $\hat{\mathscr{X}}$ with respect to a positive cone $\hat{\mathscr{C}}$ $(\subset \hat{\mathscr{X}})$, that is for $x, y \in \hat{\mathscr{X}}$, $x \succeq y \Leftrightarrow x - y \in \hat{\mathscr{C}}$. Also, recall that if a map $\tilde{T} : \hat{\mathscr{X}} \to \hat{\mathscr{X}}$ is continuous and compact, it is called completely continuous.

Theorem 7.3.1. *Let $\hat{\mathscr{X}}$ be an ordered Banach space and $\hat{\mathscr{C}} \subset \hat{\mathscr{X}}$ a positive, nonempty closed cone that satisfies $\hat{\mathscr{C}} - \hat{\mathscr{C}} = \hat{\mathscr{X}}$. Let $\succeq$ be the order with respect to the cone $\hat{\mathscr{C}}$ described above. Let $\tilde{T} : \hat{\mathscr{X}} \to \hat{\mathscr{X}}$ be an order-preserving, completely continuous, 1-homogeneous map with the property that if for some nonzero $\zeta \in \hat{\mathscr{C}}$ and $N > 0$, we have $N\tilde{T}(\zeta) \succeq \zeta$. Then there exist a nontrivial $f \in \hat{\mathscr{C}}$ and $\tilde{\lambda} > 0$ satisfying $\tilde{T} f = \tilde{\lambda} f$.*

For any finite set $\mathscr{D} \subset S$, we define $\mathcal{B}_{\mathscr{D}} = \{f : S \to \mathbb{R} \mid f \text{ is a function and } f(i) = 0 \ \forall\, i \in \mathscr{D}^c\}$. $\mathcal{B}^+_{\mathscr{D}} \subset \mathcal{B}_{\mathscr{D}}$ denotes the cone of all nonnegative functions vanishing outside $\mathscr{D}$. For any function $f \in \mathcal{B}_{\mathscr{D}}$, $\|f\|_{\mathscr{D}} = \max\{|f(i)| : i \in \mathscr{D}\}$.

Let $\mathscr{D}_n \subset S$ be an increasing sequence of finite subsets such that $\cup_{i=0}^{\infty} \mathscr{D}_n = S$ and $i_0 \in \mathscr{D}_n$ for all $n \in \mathbb{N}$, where i_0 is as in Assumption $(A3)$(iii). Define

$$\tau_n := \tau(\mathscr{D}_n) := \inf\{t \geq 0 : \xi_t \notin \mathscr{D}_n\},$$

the first exit time from $\mathscr{D}_n$. Now we prove the existence of a unique solution to the following equation which will pave the way to establish the existence of a Dirichlet eigenpair associated with the HJI equation.

Proposition 7.3.1. *Suppose Assumption $(A3)$ holds. Let $\tilde{c} : K \to \mathbb{R}$ be a function continuous in $(a, b) \in A(i) \times B(i)$ for each fixed $i \in S$. Suppose the function $\tilde{c}$ satisfies the relation $\tilde{c} < -\delta$ in $\mathscr{D}_n$ for some $\delta > 0$ and $n \in \mathbb{N}$. Then for any $g \in \mathcal{B}_{\mathscr{D}_n}$ there exists a unique $\varphi \in \mathcal{B}_{\mathscr{D}_n}$ satisfying the following nonlinear equation*

$$\begin{aligned} -g(i) &= \sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \left[\sum_{j \in S} \varphi(j) q(j|i, \mu, \nu) + \tilde{c}(i, \mu, \nu)\varphi(i) \right] \\ &= \inf_{\nu \in \mathcal{P}(B(i))} \sup_{\mu \in \mathcal{P}(A(i))} \left[\sum_{j \in S} \varphi(j) q(j|i, \mu, \nu) + \tilde{c}(i, \mu, \nu)\varphi(i) \right] \ \forall i \in \mathscr{D}_n, \end{aligned} \tag{7.3.1}$$

with $\varphi(i) = 0$ for all $i \in \mathscr{D}_n^c$. Moreover, the unique solution of the above equation satisfies

$$\begin{aligned} \varphi(i) &= \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} E_i^{\pi^1, \pi^2} \left[\int_0^{\tau_n} e^{\int_0^t \tilde{c}(\xi_s, \pi^1(s), \pi^2(s)) ds} g(\xi_t) dt \right] \\ &= \inf_{\pi^2 \in \Pi^2_{Ad}} \sup_{\pi^1 \in \Pi^1_{Ad}} E_i^{\pi^1, \pi^2} \left[\int_0^{\tau_n} e^{\int_0^t \tilde{c}(\xi_s, \pi^1(s), \pi^2(s)) ds} g(\xi_t) dt \right] \ \forall i \in S, \end{aligned} \tag{7.3.2}$$

where as before $\tau_n = \inf\{t \geq 0 : \xi_t \notin \mathscr{D}_n\}$.

Proof. Let $(y_i)_{i \in \mathscr{D}_n}$ be a sequence in $\mathbb{R}$. Fix $i \in \mathscr{D}_n$. Let $x \in \mathbb{R}$ and let $F : \mathbb{R} \to \mathbb{R}$ be defined by

$$F(x) = \sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \left[\sum_{j \in \mathscr{D}_n \setminus \{i\}} y_j q(j|i, \mu, \nu) + \Big(q(i|i, \mu, \nu) + \tilde{c}(i, \mu, \nu) \Big) x \right], i \in \mathscr{D}_n. \tag{7.3.3}$$

Suppose $x_2 > x_1$. Let $\varepsilon > 0$. Then there exists $\pi^1_\varepsilon \in \Pi^1_{SM}$ for which the following holds

$$\begin{aligned} F(x_1) - F(x_2) = &\sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \left[\sum_{j \in \mathscr{D}_n \setminus \{i\}} y_j q(j|i, \mu, \nu) + \Big(q(i|i, \mu, \nu) + \tilde{c}(i, \mu, \nu) \Big) x_1 \right] \\ &- \sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \left[\sum_{j \in \mathscr{D}_n \setminus \{i\}} y_j q(j|i, \mu, \nu) + \Big(q(i|i, \mu, \nu) + \tilde{c}(i, \mu, \nu) \Big) x_2 \right] \end{aligned}$$

$$
\begin{aligned}
&\geq \inf_{\nu \in \mathcal{P}(B(i))} \Bigg[\sum_{j \in \mathscr{D}_n \setminus \{i\}} y_j q(j|i, \pi^1_\varepsilon(i), \nu) + \Big(q(i|i, \pi^1_\varepsilon(i), \nu) + \tilde{c}(i, \pi^1_\varepsilon(i), \nu) \Big) x_1 \Bigg] \\
&\quad - \inf_{\nu \in \mathcal{P}(B(i))} \Bigg[\sum_{j \in \mathscr{D}_n \setminus \{i\}} y_j q(j|i, \pi^1_\varepsilon(i), \nu) + \Big(q(i|i, \pi^1_\varepsilon(i), \nu) + \tilde{c}(i, \pi^1_\varepsilon(i), \nu) \Big) x_2 + \varepsilon \Bigg] \\
&\geq \inf_{\nu \in \mathcal{P}(B(i))} \Bigg[\Big(q(i|i, \pi^1_\varepsilon(i), \nu) + \tilde{c}(i, \pi^1_\varepsilon(i), \nu) \Big) (x_1 - x_2) \Bigg] - \varepsilon \\
&\geq \inf_{\nu \in \mathcal{P}(B(i))} \Big[-\tilde{c}(i, \pi^1_\varepsilon(i), \nu)(x_2 - x_1) \Big] - \varepsilon \\
&> \delta(x_2 - x_1) - \varepsilon.
\end{aligned}
$$

Since $\varepsilon > 0$ is arbitrary we get $F(x_1) > F(x_2)$. Also, we see that $\lim_{x \to +\infty} F(x) = -\infty$ and $\lim_{x \to -\infty} F(x) = +\infty$. Since F is continuous in x, for every $z \in \mathbb{R}$, there exists a unique x satisfying $F(x) = z$. Now using the definition of F, for fixed $g \in \mathcal{B}_{\mathscr{D}_n}$, we can define a map $\hat{T} : \mathcal{B}_{\mathscr{D}_n} \to \mathcal{B}_{\mathscr{D}_n}$ satisfying

$$
\sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \Bigg[\sum_{j \in \mathscr{D}_n \setminus \{i\}} \tilde{\phi}(j) q(j|i, \mu, \nu) + \Big(q(i|i, \mu, \nu) + \tilde{c}(i, \mu, \nu) \Big) (\hat{T}\tilde{\phi}(i)) \Bigg] = -g(i), \ i \in \mathscr{D}_n. \tag{7.3.4}
$$

Let $\tilde{\phi}_1, \tilde{\phi}_2 \in \mathcal{B}_{\mathscr{D}_n}$. Also, let $\tilde{\pi}^1$ be an outer maximizing selector of

$$
\sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \Bigg[\sum_{j \in \mathscr{D}_n \setminus \{i\}} \tilde{\phi}_2(j) q(j|i, \mu, \nu) + \Big(q(i|i, \mu, \nu) + \tilde{c}(i, \mu, \nu) \Big) \hat{T}\tilde{\phi}_2(i) \Bigg]
$$

i.e., the supremum in the above expression is attained at $\tilde{\pi}^1$. Assumption $(A3)$, ensures the existence of such a selector. It then follows that

$$
\begin{aligned}
0 = &\sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \Bigg[\sum_{j \in \mathscr{D}_n \setminus \{i\}} \tilde{\phi}_1(j) q(j|i, \mu, \nu) + \Big(q(i|i, \mu, \nu) + \tilde{c}(i, \mu, \nu) \Big) \hat{T}\tilde{\phi}_1(i) \Bigg] \\
&- \sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \Bigg[\sum_{j \in \mathscr{D}_n \setminus \{i\}} \tilde{\phi}_2(j) q(j|i, \mu, \nu) + \Big(q(i|i, \mu, \nu) + \tilde{c}(i, \mu, \nu) \Big) \hat{T}\tilde{\phi}_2(i) \Bigg] \\
\geq &\inf_{\nu \in \mathcal{P}(B(i))} \Bigg[\sum_{j \in \mathscr{D}_n \setminus \{i\}} \tilde{\phi}_1(j) q(j|i, \tilde{\pi}^1(i), \nu) + \Big(q(i|i, \tilde{\pi}^1(i), \nu) + \tilde{c}(i, \tilde{\pi}^1(i), \nu) \Big) \hat{T}\tilde{\phi}_1(i) \Bigg] \\
&- \inf_{\nu \in \mathcal{P}(B(i))} \Bigg[\sum_{j \in \mathscr{D}_n \setminus \{i\}} \tilde{\phi}_2(j) q(j|i, \tilde{\pi}^1(i), \nu) + \Big(q(i|i, \tilde{\pi}^1(i), \nu) + \tilde{c}(i, \tilde{\pi}^1(i), \nu) \Big) \hat{T}\tilde{\phi}_2(i) \Bigg] \\
\geq &\inf_{\nu \in \mathcal{P}(B(i))} \Bigg[\sum_{j \in \mathscr{D}_n \setminus \{i\}} (\tilde{\phi}_1(j) - \tilde{\phi}_2(j)) q(j|i, \tilde{\pi}^1(i), \nu) \\
&\qquad + \Big(q(i|i, \tilde{\pi}^1(i), \nu) \tilde{c}(i, \tilde{\pi}^1(i), \nu) \Big) (\hat{T}\tilde{\phi}_1(i) - \hat{T}\tilde{\phi}_2(i)) \Bigg].
\end{aligned}
$$

Now let the infimum of the RHS (of the above) attain at π^{*2}. Then

$$\|\tilde{\phi}_1-\tilde{\phi}_2\|_{\mathscr{D}_n} q(i|i,\tilde{\pi}^1(i),\pi^{*2}(i))+\Big(q(i|i,\tilde{\pi}^1(i),\pi^{*2}(i))+\tilde{c}(i,\tilde{\pi}^1(i),\pi^{*2}(i))\Big)(\hat{T}\tilde{\phi}_1(i)-\hat{T}\tilde{\phi}_2(i)) \leq 0.$$

Hence, we deduce that

$$(\hat{T}\tilde{\phi}_2(i)-\hat{T}\tilde{\phi}_1(i)) \leq \sup_{\mu\in\mathcal{P}(A(i))}\sup_{\nu\in\mathcal{P}(B(i))}\frac{-q(i|i,\mu,\nu)}{-q(i|i,\mu,\nu)-\tilde{c}(i,\mu,\nu)}\|\tilde{\phi}_1-\tilde{\phi}_2\|_{\mathscr{D}_n}.$$

Now in the above calculation, interchanging $\tilde{\phi}_1, \tilde{\phi}_2$, it follows that

$$\|\hat{T}\tilde{\phi}_1-\hat{T}\tilde{\phi}_2\|_{\mathscr{D}_n} \leq \alpha_1\|\tilde{\phi}_1-\tilde{\phi}_2\|_{\mathscr{D}_n},$$

where α_1 is a positive constant less than 1. This implies that $\hat{T}$ is a contraction map. Thus, by Banach fixed point theorem, there exists a unique $\varphi\in\mathcal{B}_{\mathscr{D}_n}$ such that $\hat{T}(\varphi)=\varphi$. Now by Fan's minimax theorem, see [33, Theorem 3], we have

$$\begin{aligned}
&\sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[\sum_{j\in S}\varphi(j)q(j|i,\mu,\nu)+\tilde{c}(i,\mu,\nu)\varphi(i)\right]\\
&=\inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\left[\sum_{j\in S}\varphi(j)q(j|i,\mu,\nu)+\tilde{c}(i,\mu,\nu)\varphi(i)\right].
\end{aligned}$$

This proves that (7.3.1) admits a unique solution. Now by using Dynkin formula as in [59, Appendix C.3], for any $(\pi^1,\pi^2)\in\Pi^1_{Ad}\times\Pi^2_{Ad}$ and $T>0$, we get

$$\begin{aligned}
&E_i^{\pi^1,\pi^2}\left[e^{\int_0^{T\wedge\tau_n}\tilde{c}(\xi_s,\pi^1(s),\pi^2(s))ds}\varphi(\xi_{T\wedge\tau_n})\right]-\varphi(i)\\
&=E_i^{\pi^1,\pi^2}\left[\int_0^{T\wedge\tau_n}e^{\int_0^t\tilde{c}(\xi_s,\pi^1(s),\pi^2(s))ds}\Big(\tilde{c}(\xi_t,\pi^1(t),\pi^2(t))\varphi(\xi_t)+\sum_{j\in S}\varphi(j)q(j|\xi_t,\pi^1(t),\pi^2(t))\Big)dt\right].
\end{aligned} \tag{7.3.5}$$

Using the compactness of $A(i), B(i)$ and the continuity of $\tilde{c}, q$, there exists a pair of selectors $(\pi^{*1}_n,\pi^{*2}_n)\in\Pi^1_{SM}\times\Pi^2_{SM}$ (i.e., a mini-max selector) satisfying

$$\begin{aligned}
-g(i)&=\inf_{\nu\in\mathcal{P}(B(i))}\left[\sum_{j\in S}\varphi(j)q(j|i,\pi^{*1}_n(i),\nu)+\tilde{c}(i,\pi^{*1}_n(i),\nu)\varphi(i)\right]\\
&=\sup_{\mu\in\mathcal{P}(A(i))}\left[\sum_{j\in S}\varphi(j)q(j|i,\mu,\pi^{*2}_n(i))+\tilde{c}(i,\mu,\pi^{*2}_n(i))\varphi(i)\right].
\end{aligned} \tag{7.3.6}$$

Then, using (7.3.5) and (7.3.6), we obtain

$$E_i^{\pi^{*1}_n,\pi^2}\left[\int_0^{T\wedge\tau_n}e^{\int_0^t\tilde{c}(\xi_s,\pi^{*1}_n(\xi_{s-}),\pi^2(s))ds}g(\xi_t)dt\right]\geq -E_i^{\pi^{*1}_n,\pi^2}\left[e^{\int_0^{T\wedge\tau_n}\tilde{c}(\xi_s,\pi^{*1}_n(\xi_{s-}),\pi^2(s))ds}\varphi(\xi_{T\wedge\tau_n})\right]+\varphi(i).$$

Using the dominated convergence theorem, taking $T \to \infty$ in the above equation, we get

$$E_i^{\pi_n^{*1},\pi^2}\left[\int_0^{\tau_n} e^{\int_0^t \tilde{c}(\xi_s,\pi_n^{*1}(\xi_{s-}),\pi^2(s))ds} g(\xi_t)dt\right] \geq -E_i^{\pi_n^{*1},\pi^2}\left[e^{\int_0^{\tau_n} \tilde{c}(\xi_s,\pi_n^{*1}(\xi_{s-}),\pi^2(s))ds}\varphi(\xi_{\tau_n})\right] + \varphi(i).$$

Hence

$$\varphi(i) \leq E_i^{\pi_n^{*1},\pi^2}\left[\int_0^{\tau_n} e^{\int_0^t \tilde{c}(\xi_s,\pi_n^{*1}(\xi_{s-}),\pi^2(s))ds} g(\xi_t)dt\right].$$

Since $\pi^2 \in \Pi_{Ad}^2$ is arbitrary,

$$\varphi(i) \leq \inf_{\pi^2\in\Pi_{Ad}^2} E_i^{\pi_n^{*1},\pi^2}\left[\int_0^{\tau_n} e^{\int_0^t \tilde{c}(\xi_s,\pi_n^{*1}(\xi_{s-}),\pi^2(s))ds} g(\xi_t)dt\right]. \tag{7.3.7}$$

Similarly, using (7.3.5), (7.3.6), and Fatou's Lemma, we get

$$\varphi(i) \geq \sup_{\pi^1\in\Pi_{Ad}^1} E_i^{\pi^1,\pi_n^{*2}}\left[\int_0^{\tau_n} e^{\int_0^t \tilde{c}(\xi_s,\pi^1(s),\pi_n^{*2}(\xi_{s-}))ds} g(\xi_t)dt\right]. \tag{7.3.8}$$

Using (7.3.7) and (7.3.8), we obtain

$$\begin{aligned}\varphi(i) &= \inf_{\pi^2\in\Pi_{Ad}^2}\sup_{\pi^1\in\Pi_{Ad}^1} E_i^{\pi^1,\pi^2}\left[\int_0^{\tau_n} e^{\int_0^t \tilde{c}(\xi_s,\pi^1(s),\pi^2(s))ds} g(\xi_t)dt\right]\\ &= \sup_{\pi^1\in\Pi_{Ad}^1}\inf_{\pi^2\in\Pi_{Ad}^2} E_i^{\pi^1,\pi^2}\left[\int_0^{\tau_n} e^{\int_0^t \tilde{c}(\xi_s,\pi^1(s),\pi^2(s))ds} g(\xi_t)dt\right],\ i \in S.\end{aligned}$$

This completes the proof. □

Lemma 7.3.2. *Suppose Assumption (A2) holds. Consider a finite subset* $\mathscr{B}$ *of* S *such that* $\mathscr{K} \subset \mathscr{B}$*, where* $\mathscr{K}$ *is as in Assumption (A2). Let* $\hat{\tau}(\mathscr{B}) = \inf\{t > 0 : \xi_t \in \mathscr{B}\}$*. Then for any pair of strategies* $(\pi^1,\pi^2) \in \Pi_{Ad}^1 \times \Pi_{Ad}^2$*, the following results hold.*

(i) When Assumption (A2)(a) holds:

$$E_i^{\pi^1,\pi^2}\left[e^{\gamma\hat{\tau}(\mathscr{B})}V(\xi_{\hat{\tau}(\mathscr{B})})\right] \leq V(i)\ \forall\ i \in \mathscr{B}^c. \tag{7.3.9}$$

(ii) When Assumption (A2)(b) holds:

$$E_i^{\pi^1,\pi^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B})} \ell(\xi_s)ds}V(\xi_{\hat{\tau}(\mathscr{B})})\right] \leq V(i)\ \forall\ i \in \mathscr{B}^c. \tag{7.3.10}$$

Proof. It is easy to see that the proof of (i) is analogous to the proof of (ii) when we replace ℓ with γ. So, we prove only part (ii). Suppose Assumption (A2)(b) holds. Let

n be large enough so that $\mathscr{B} \subset \mathscr{D}_n$. Applying Dynkin's formula [59, Appendix C.3], for $i \in \mathscr{B}^c \cap \mathscr{D}_n$ we have

$$
\begin{aligned}
& E_i^{\pi^1,\pi^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B})\wedge T\wedge\tau_n}\ell(\xi_s)ds}V(\hat{\tau}(\mathscr{B})\wedge T\wedge\tau_n)\right] - V(i) \\
&= E_i^{\pi^1,\pi^2}\left[\int_0^{\hat{\tau}(\mathscr{B})\wedge T\wedge\tau_n} e^{\int_0^t \ell(\xi_s)ds}\left[\ell(\xi_t)V(\xi_t) + \sum_{j\in S} q(j|\pi_t,\pi^1(t),\pi^2(t))V(j)\right] dt\right] \\
&\leq E_i^{\pi^1,\pi^2}\left[\int_0^{\hat{\tau}(\mathscr{B})\wedge T\wedge\tau_n} e^{\int_0^t \ell(\xi_s)ds} C I_{\mathscr{K}}(\xi_t)dt\right] = 0,
\end{aligned}
$$

where $\tau_n = \inf\{t \geq 0 : \xi_t \notin \mathscr{D}_n\}$. Now by Fatou's lemma, taking first $n \to \infty$ and then $T \to \infty$, we get the required result.

□

Now applying the nonlinear version of Kreĭn-Rutman theorem, we show that Dirichlet eigenpair exists in the space of bounded functions.

Lemma 7.3.3. *Suppose Assumptions (A1)-(A3) hold. Then for $n \in \mathbb{N}$, there exists a pair $(\rho_n, \psi_n) \in \mathbb{R} \times \mathcal{B}^+_{\mathscr{D}_n}$, $\psi_n \gneq 0$ for the following Dirichlet nonlinear eigenequation*

$$
\begin{aligned}
\rho_n\psi_n(i) &= \sup_{\mu\in\mathcal{P}(A(i))} \inf_{\nu\in\mathcal{P}(B(i))} \left[\sum_{j\in S}\psi_n(j)q(j|i,\mu,\nu) + c(i,\mu,\nu)\psi_n(i)\right] \\
&= \inf_{\nu\in\mathcal{P}(B(i))} \sup_{\mu\in\mathcal{P}(A(i))} \left[\sum_{j\in S}\psi_n(j)q(j|i,\mu,\nu) + c(i,\mu,\nu)\psi_n(i)\right]. \qquad (7.3.11)
\end{aligned}
$$

Also, for each $i \in S$ such that $\psi_n(i) > 0$, we have

$$
\rho_n \leq \sup_{\pi^1\in\Pi^1_{Ad}} \inf_{\pi^2\in\Pi^2_{Ad}} \limsup_{T\to\infty} \frac{1}{T} \log E_i^{\pi^1,\pi^2}\left[e^{\int_0^T c(\xi_t,\pi^1(t),\pi^2(t))dt}\right]. \qquad (7.3.12)
$$

Additionally the sequence $\{\rho_n\}$ is bounded satisfying $\liminf_{n\to\infty}\rho_n \geq 0$.

Proof. Let $\delta > 0$. Set $\tilde{c}(i,a,b) = c(i,a,b) - k_n - \delta$, where $k_n = \sup\{c(i,a,b)\,|\, i \in \mathscr{D}_n, a \in A(i), b \in B(i)\}$.

Let $\tilde{T} : \mathcal{B}_{\mathscr{D}_n} \to \mathcal{B}_{\mathscr{D}_n}$ be an operator defined as

$$
\tilde{T}(g)(i) := \sup_{\pi^1\in\Pi^1_{Ad}} \inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^1,\pi^2}\left[\int_0^{\tau_n} e^{\int_0^t \tilde{c}(\xi_s,\pi^1(s),\pi^2(s))ds} g(\xi_t)dt\right],\ i \in \mathscr{D}_n, \qquad (7.3.13)
$$

with $\tilde{T}(g)(i) = 0$ for $i \in \mathscr{D}_n^c$. Let $g_1, g_2 \in \mathcal{B}_{\mathscr{D}_n}$ such that $g_1 \succeq g_2$, i.e., $g_1(i) \geq g_2(i)$ for each i. Also, let $\tilde{T}(g_1) = \hat{\varphi}_1$ and $\tilde{T}(g_2) = \hat{\varphi}_2$. Then from the Proposition 7.3.1, we obtain

$$
-g_2(i) = \sup_{\mu\in\mathcal{P}(A(i))} \inf_{\nu\in\mathcal{P}(B(i))} \left[\sum_{j\in S}\hat{\varphi}_2(j)q(j|i,\mu,\nu) + \tilde{c}(i,\mu,\nu)\hat{\varphi}_2(i)\right]
$$

$$= \inf_{\nu\in\mathcal{P}(B(i))}\Big[\sum_{j\in S}\hat{\varphi}_2(j)q(j|i,\hat{\pi}^{*1}(i),\nu)+\tilde{c}(i,\hat{\pi}^{*1}(i),\nu)\hat{\varphi}_2(i)\Big]\ \forall i\in\mathscr{D}_n\,,$$

where $\hat{\pi}^{*1}\in\Pi^1_{SM}$. Also, from the Proposition 7.3.1, we have

$$\hat{\varphi}_2(i)=\tilde{T}(g_2)(i)=\inf_{\pi^2\in\Pi^2_{Ad}}E_i^{\hat{\pi}^{*1},\pi^2}\Big[\int_0^{\tau_n}e^{\int_0^t\tilde{c}(\xi_s,\hat{\pi}^{*1}(\xi_{s-}),\pi^2(s))ds}g_2(\xi_t)dt\Big]\,.$$

Thus, we deduce that

$$\begin{aligned}\tilde{T}(g_1)(i)-\tilde{T}(g_2)(i)&=\sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}}E_i^{\pi^1,\pi^2}\Big[\int_0^{\tau_n}e^{\int_0^t\tilde{c}(\xi_s,\pi^1(s),\pi^2(s))ds}g_1(\xi_t)dt\Big]\\&\quad-\sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}}E_i^{\pi^1,\pi^2}\Big[\int_0^{\tau_n}e^{\int_0^t\tilde{c}(\xi_s,\pi^1(s),\pi^2(s))ds}g_2(\xi_t)dt\Big]\\&\geq\inf_{\pi^2\in\Pi^2_{Ad}}E_i^{\hat{\pi}^{*1},\pi^2}\Big[\int_0^{\tau_n}e^{\int_0^t\tilde{c}(\xi_s,\hat{\pi}^{*1}(\xi_{s-}),\pi^2(s))ds}g_1(\xi_t)dt\Big]\\&\quad-\inf_{\pi^2\in\Pi^2_{Ad}}E_i^{\hat{\pi}^{*1},\pi^2}\Big[\int_0^{\tau_n}e^{\int_0^t\tilde{c}(\xi_s,\hat{\pi}^{*1}(\xi_{s-}),\pi^2(s))ds}g_2(\xi_t)dt\Big]\\&\geq\inf_{\pi^2\in\Pi^2_{Ad}}E_i^{\hat{\pi}^{*1},\pi^2}\Big[\int_0^{\tau_n}e^{\int_0^t\tilde{c}(\xi_s,\hat{\pi}^{*1}(\xi_{s-}),\pi^2(s))ds}(g_1(\xi_t)-g_2(\xi_t))dt\Big].\end{aligned}$$

This gives us $\tilde{T}(g_1)\succeq\tilde{T}(g_2)$. Clearly $\tilde{T}(\lambda g)=\lambda\tilde{T}(g)$ for all $\lambda\geq 0$. Since $\tilde{c}<-\delta$, there exists a constant $\alpha_2>0$ such that

$$\|\tilde{T}(\hat{g}_1)-\tilde{T}(\hat{g}_2)\|_{\mathscr{D}_n}\leq\alpha_2\|\hat{g}_1-\hat{g}_2\|_{\mathscr{D}_n},\quad\text{for any }\hat{g}_1,\hat{g}_2\in\mathcal{B}_{\mathscr{D}_n}.$$

Thus $\tilde{T}$ is continuous. Let $\{g_m\}$ be a bounded sequence in $\mathcal{B}_{\mathscr{D}_n}$. Then from (7.3.13), for some constant $\alpha_3>0$ such that $\|\tilde{T}g_m\|_{\mathscr{D}_n}\leq\alpha_3$. Now using a standard diagonalization procedure, there exists a subsequence of $\{\tilde{T}g_m\}$, (denoting by the same sequence without loss of generality) and a function $\phi\in\mathcal{B}_{\mathscr{D}_n}$ such that $\|\tilde{T}g_m-\phi\|_{\mathscr{D}_n}\to 0$ as $m\to\infty$. Hence the map $\tilde{T}:\mathcal{B}_{\mathscr{D}_n}\to\mathcal{B}_{\mathscr{D}_n}$ is compact. Therefore it is completely continuous. Let $g\in\mathcal{B}_{\mathscr{D}_n}$ such that $g(i_0)=1$ and $g(j)=0$ for all $j\neq i_0$. Then by (7.3.13), we have

$$\begin{aligned}\tilde{T}(g)(i_0)&\geq\sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}}E_{i_0}^{\pi^1,\pi^2}\Big[\int_0^{T_1}e^{\int_0^t\tilde{c}(\xi_s,\pi^1(s),\pi^2(s))ds}g(\xi_t)dt\Big]\\&\geq\frac{g(i_0)}{\|\tilde{c}\|_{\mathscr{D}_n}}\sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}}E_{i_0}^{\pi^1,\pi^2}\Big[1-e^{-\|\tilde{c}\|_{\mathscr{D}_n}T_1}\Big]\\&=g(i_0)\frac{1}{\|\tilde{c}\|_{\mathscr{D}_n}+q^*(i_0)},\end{aligned}$$

where T_1 is the first jump time (clearly, $T_1\leq\tau_n$). Thus $N\tilde{T}(g)\succeq g$ where $N=\|\tilde{c}\|_{\mathscr{D}_n}+q^*(i_0)>0$. Therefore by Kreĭn-Rutman Theorem 7.3.1, there exists a nontrivial $\psi_n\in\mathcal{B}^+_{\mathscr{D}_n}$ where $\psi_n\neq 0$ and a constant $\lambda_{\mathscr{D}_n}>0$ such that $\tilde{T}(\psi_n)=\lambda_{\mathscr{D}_n}\psi_n$, i.e.,

$$\tilde{\rho}_n\psi_n(i)=\sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\Big[\psi_n(j)q(j|i,\mu,\nu)+\tilde{c}(i,\mu,\nu)\psi_n(i)\Big]\ \forall i\in\mathscr{D}_n,$$

where $\tilde{\rho}_n = -[\lambda_{\mathscr{D}_n}]^{-1}$. Therefore in terms of c, we have

$$\rho_n \psi_n(i) = \sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \Big[\psi_n(j) q(j|i,\mu,\nu) + c(i,\mu,\nu)\psi_n(i) \Big] \ \forall i \in \mathscr{D}_n,$$

where $\rho_n = \tilde{\rho}_n + k_n + \delta$. Now by Fan's minimax theorem, see [33, Theorem 3], we have

$$\begin{aligned} \rho_n \psi_n(i) &= \sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \Big[\psi_n(j) q(j|i,\mu,\nu) + c(i,\mu,\nu)\psi_n(i) \Big] \\ &= \inf_{\nu \in \mathcal{P}(B(i))} \sup_{\mu \in \mathcal{P}(A(i))} \Big[\psi_n(j) q(j|i,\mu,\nu) + c(i,\mu,\nu)\psi_n(i) \Big] \ \forall i \in \mathscr{D}_n. \end{aligned}$$

This proves that (7.3.11) admits a unique solution. As before by the continuity of c, q and the compactness of $A(i)$, there exists $\pi_n^{*1} \in \Pi_{SM}^1$ such that (7.3.11), can be written as

$$\rho_n \psi_n(i) = \inf_{\nu \in \mathcal{P}(B(i))} \Big[\psi_n(j) q(j|i,\pi_n^{*1}(i),\nu) + c(i,\pi_n^{*1}(i),\nu)\psi_n(i) \Big] \ \forall i \in \mathscr{D}_n. \tag{7.3.14}$$

Now applying Dynkin's formula (see [21, Lemma 3.1]) and using (7.3.14), we deduce that

$$\begin{aligned} \psi_n(i) &\le E_i^{\pi_n^{*1},\pi^2} \Big[e^{\int_0^T (c(\xi_s,\pi_n^{*1}(\xi_{s-}),\pi^2(s)) - \rho_n) ds} \psi_n(\xi_T) I_{\{T<\tau_n\}} \Big] \\ &\le (\sup_{\mathscr{D}_n} \psi_n) E_i^{\pi_n^{*1},\pi^2} \Big[e^{\int_0^T (c(\xi_s,\pi_n^{*1}(\xi_{s-}),\pi^2(s)) - \rho_n) ds} \Big]. \end{aligned} \tag{7.3.15}$$

If $\psi_n(i) > 0$ then by taking logarithm on the both sides in (7.3.15), dividing by T and letting $T \to \infty$, we get

$$\rho_n \le \limsup_{T\to\infty} \frac{1}{T} \log E_i^{\pi_n^{*1},\pi^2} \Big[e^{\int_0^T c(\xi_s,\pi_n^{*1}(\xi_{s-}),\pi^2(s)) ds} \Big].$$

Since $\pi^2 \in \Pi_{Ad}^2$ is arbitrary, we obtain

$$\begin{aligned} \rho_n &\le \inf_{\pi^2 \in \Pi_{Ad}^2} \limsup_{T\to\infty} \frac{1}{T} \log E_i^{\pi_n^{*1},\pi^2} \Big[e^{\int_0^T c(\xi_s,\pi_n^{*1}(\xi_{s-}),\pi^2(s)) ds} \Big] \\ &\le \sup_{\pi^1 \in \Pi_{Ad}^1} \inf_{\pi^2 \in \Pi_{Ad}^2} \limsup_{T\to\infty} \frac{1}{T} \log E_i^{\pi^1,\pi^2} \Big[e^{\int_0^T c(\xi_s,\pi^1(s),\pi^2(s)) ds} \Big]. \end{aligned}$$

We now show that $J(i,c,\pi^1,\pi^2)$ is finite for every $(\pi^1,\pi^2) \in \Pi_{Ad}^1 \times \Pi_{Ad}^2$ and $i \in S$. We only provide proof under Assumption $(A2)$(b) and the proof under Assumption $(A2)$(a) would be analogous. Now from (7.2.6) we have

$$\sup_{(a,b)\in A(i)\times B(i)} \sum_{j\in S} V(j) q(j|i,a,b) \le (C - \ell(i)) V(i) \ \forall i \in S. \tag{7.3.16}$$

Then by the Dynkin formula, we get

$$E_i^{\pi^1,\pi^2}\left[e^{\int_0^{T\wedge\tau_n}(\ell(\xi_t)-C)dt}V(\xi_{T\wedge\tau_n})\right] \leq V(i)\ \forall i\in S. \tag{7.3.17}$$

By Fatou's lemma, taking $n\to\infty$ in (7.3.17), we obtain

$$E_i^{\pi^1,\pi^2}\left[e^{\int_0^{T}(\ell(\xi_t)-C)dt}V(\xi_{T})\right] \leq V(i)\ \forall i\in S.$$

Now, since $V\geq 1$, taking logarithm on both sides in the above equation, dividing both sides by T, we have

$$\frac{1}{T}\log E_i^{\pi^1,\pi^2}\left[e^{\int_0^{T}\ell(\xi_t)dt}V(\xi_{T})\right] - C \leq \frac{1}{T}\log V(i)\ \forall i\in S.$$

Hence, letting $T\to\infty$, it follows that

$$J(i,\ell,\pi^1,\pi^2)\leq C \quad \text{for all } i\in S.$$

Since, $\ell - \sup_{(a,b)\in A(i)\times B(i)} c(\cdot,a,b)$ is norm-like, we have $\sup_{(a,b)\in A(i)\times B(i)} c(i,a,b)\leq \ell(i)+k_1\ \ \forall$ $i\in S$ for some constant k_1. This implies that

$$J(i,c,\pi^1,\pi^2)\leq C+k_1\ \ \forall(\pi^1,\pi^2)\in\Pi^1_{Ad}\times\Pi^2_{Ad}, \forall i\in S. \tag{7.3.18}$$

It is clear from (7.3.12) and (7.3.18) that ρ_n has an upper bound. Next, we prove that ρ_n is bounded below. By using Assumption $(A3)$(iii) and (7.3.11), we have $\psi_n(i_0)>0$. Thus normalizing ψ_n, we have $\psi_n(i_0)=1$. Also, since $c\geq 0$, by (7.3.11) we get

$$\rho_n \geq \sup_{\mu\in\mathcal{P}(A(i_0))}\inf_{\nu\in\mathcal{P}(B(i_0))}\left[\sum_{j\in S}\psi_n(j)q(j|i_0,\mu,\nu)\right] \geq \sup_{\mu\in\mathcal{P}((i_0))}\inf_{\nu\in\mathcal{P}(B(i_0))} q(i_0|i_0,\mu,\nu).$$

So, $\{\rho_n\}$ is bounded below. Now we claim that $\hat{\rho}:=\liminf_{n\to\infty}\rho_n\geq 0$. If not, then on the contrary, $\hat{\rho}<0$. So, along some subsequence, we have (with an abuse of notation, we use the same sequence) $\rho_n\to\hat{\rho}$, as $n\to\infty$ and for large n, $\rho_n<0$. Let π_n^{*2} be outer minimizing selector of (7.3.11). Thus, using (7.3.11), for large enough n, we have

$$\begin{aligned}
0>\rho_n\psi_n(i_0) &= \sup_{\mu\in\mathcal{P}(A(i_0))}\left[\sum_{j\in S}\psi_n(j)q(j|i_0,\mu,\pi_n^{*2}(i_0)) + c(i_0,\mu,\pi_n^{*2}(i_0))\psi_n(i_0)\right]\\
&\geq \sup_{\mu\in\mathcal{P}(A(i_0))}\left[\sum_{j\in S}\psi_n(j)q(j|i_0,\mu,\pi_n^{*2}(i_0))\right]\\
&\geq \sum_{j\in S}\psi_n(j)q(j|i_0,\mu,\pi_n^{*2}(i_0)).
\end{aligned}$$

Now by Assumption $(A3)$(iii), from the above equation, we get

$$\psi_n(j)\leq \frac{-q(i_0|i_0,\mu,\pi_n^{*2}(i_0))}{q(j|i_0,\mu,\pi_n^{*2}(i_0))} \leq \sup_{\mu\in\mathcal{P}(A(i_0))}\sup_{\nu\in\mathcal{P}(B(i_0))}\frac{-q(i_0|i_0,\mu,\nu)}{q(j|i_0,\mu,\nu)} \quad \text{for}\ \ j\neq i_0.$$

As before by a suitable diagonalization procedure, there exists a subsequence (denoting by the same sequence by abuse of notation) and a function ψ with $\psi(i_0) = 1$ such that $\psi_n(i) \to \psi(i)$, as $n \to \infty$ for all $i \in S$. By our assumption $B(i)$ is compact for each $i \in S$ and π_n^{*2} is outer minimizing selector of (7.3.11). Therefore there exists a strategy $\pi^{*2} \in \Pi_{SM}^2$, such that $\pi_n^{*2}(i) \to \pi^{*2}(i)$, for all $i \in S$, as $n \to \infty$. Hence we have

$$\rho_n \psi_n(i) \geq \sum_{j \in S} \psi_n(j) q(j|i, \mu, \pi_n^{*2}(i)) + c(i, \mu, \pi_n^{*2}(i)) \psi_n(i). \tag{7.3.19}$$

So, taking $n \to \infty$ in the above equation, we get

$$0 > \hat{\rho}\psi(i) \geq \sum_{j \in S} \psi(j) q(j|i, \mu, \pi^{*2}(i)) + c(i, \mu, \pi^{*2}(i))\psi(i) \geq \sum_{j \in S} \psi(j) q(j|i, \mu, \pi^{*2}(i)). \tag{7.3.20}$$

Let $\pi^1 \in \Pi_{SM}^1$. Applying Dynkin formula and using (7.3.20), we obtain

$$E_i^{\pi^1, \pi^{*2}}[\psi(\xi_{t \wedge \tau_n})] - \psi(i) = E_i^{\pi^1, \pi^{*2}}\left[\int_0^{t \wedge \tau_n} \sum_{j \in S} \psi(j) q(j|\xi_s, \pi^1(\xi_{s-}), \pi^{*2}(\xi_{s-})) ds\right] \leq 0.$$

Now, using the dominated convergence theorem, taking $n \to \infty$, it follows that

$$E_i^{\pi^1, \pi^{*2}}[\psi(\xi_t)] \leq \psi(i).$$

So, with respect to the canonical filtration of ξ, $\{\psi(\xi_t)\}$ is supermartingale. So, by Doob's martingale convergence theorem as $t \to \infty$, $\psi(\xi_t)$ converges. Now by Assumption $(A2)$, ξ is recurrent. Thus the skeleton process $\{\xi_n : n \in \mathbb{N}\}$ is also recurrent (see for details [3, Proposition 5.1.1]). This implies, that the process $\{\xi_n : n \in \mathbb{N}\}$ visits every state of S infinitely often. Since $\psi(i_0) = 1$, this is possible only if $\psi \equiv 1$. Since $c \geq 0$, this contradicts (7.3.20). Thus, $\liminf_{n \to \infty} \rho_n \geq 0$. □

7.4 Existence of risk-sensitive average optimal strategies

We begin this section by taking limit $n \to \infty$ from (7.3.11) to show that the limiting equation, which is the HJI equation, admits a positive eigenpair. Subsequently, we prove that any mini-max selector of the associated HJI equation is a saddle-point equilibrium. Also, exploiting the stochastic representation of the eigenfunction, we completely characterize all possible saddle-point equilibrium in the space of stationary Markov strategies.

Lemma 7.4.1. *Suppose Assumptions (A1)-(A3) hold. Then there exists* $(\rho, \psi^*) \in \mathbb{R}_+ \times L_V^\infty(S)$ *with* $\psi^* > 0$, *such that*

$$\rho \psi^*(i) = \sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \left[\sum_{j \in S} \psi^*(j) q(j|i, \mu, \nu) + c(i, \mu, \nu) \psi^*(i)\right]$$

$$= \inf_{\nu\in\mathcal{P}(B(i))}\ \sup_{\mu\in\mathcal{P}(A(i))}\left[\sum_{j\in S}\psi^*(j)q(j|i,\mu,\nu)+c(i,\mu,\nu)\psi^*(i)\right],\ i\in S. \tag{7.4.1}$$

Also, the solution (ρ,ψ^*) has the following characteristic.

(i) $\rho \le \inf_{i\in S}\ \sup_{\pi^1\in\Pi^1_{Ad}}\ \inf_{\pi^2\in\Pi^2_{Ad}}\ \limsup_{T\to\infty}\frac{1}{T}\ln E_i^{\pi^1,\pi^2}\left[e^{\int_0^T c(\xi_t,\pi^1(t),\pi^2(t))dt}\right].$

(ii) For any mini-max selector $(\pi^{*1},\pi^{*2})\in\Pi^1_{SM}\times\Pi^2_{SM}$ of (7.4.1), we have

$$\begin{aligned}\psi^*(i) &= \sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^{*2}}\left[e^{\int_0^{\hat\tau(\mathscr{B})}(c(\xi_t,\pi^1(t),\pi^{*2}(\xi_{t-}))-\rho)dt}\psi^*(\xi_{\hat\tau(\mathscr{B})})\right]\\ &= \inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^{*1},\pi^2}\left[e^{\int_0^{\hat\tau(\mathscr{B})}(c(\xi_t,\pi^{*1}(\xi_{t-}),\pi^2(t))-\rho)dt}\psi^*(\xi_{\hat\tau(\mathscr{B})})\right]\ \forall i\in\mathscr{B}^c\,,\end{aligned} \tag{7.4.2}$$

for some finite set $\mathscr{B}\supset\mathscr{K}$.

Proof. Using Assumption $(A2)$ and the fact that $c\ge 0$, there exists a finite set $\mathscr{B}$ containing $\mathscr{K}$ such that the following hold.

- When Assumption $(A2)$(a) holds: since $\gamma>\|c\|_\infty$,

$$\sup_{(a,b)\in A(i)\times B(i)} c(i,a,b)-\rho_n<\gamma\ \forall i\in\mathscr{B}^c,\ \text{for all } n \text{ large}. \tag{7.4.3}$$

- When Assumption $(A2)$(b) holds: since the function $\ell(\cdot)-\max_{(a,b)\in A(\cdot)\times B(\cdot)} c(\cdot,a,b)$ is norm-like,

$$\sup_{(a,b)\in A(i)\times B(i)} c(i,a,b)-\rho_n<\ell(i)\ \forall i\in\mathscr{B}^c,\ \text{for all } n \text{ large}. \tag{7.4.4}$$

Now we scale ψ_n in such a way that it touches V from below. Define

$$\hat\theta_n=\sup\{k>0:(V-k\psi_n)>0 \text{ in } S\}.$$

Then we see that $\hat\theta_n$ is finite as ψ_n vanishes in $\mathscr{D}_n^c$ and $\psi_n\gneq 0$. We claim that if we replace ψ_n by $\hat\theta_n\psi_n$, then ψ_n touches V inside $\mathscr{B}$. If not, then for some state $\hat i\in\mathscr{B}^c$, $(V-\psi_n)(\hat i)=0$ and $V-\psi_n>0$ in $\mathscr{B}\cup\mathscr{D}_n^c$. Let $\pi_n^{*1}\in\Pi^1_{SM}$ be an outer maximizing selector of (7.3.11). Then by Dynkin formula, we get (under Assumption $(A2)$(b))

$$\begin{aligned}\psi_n(\hat i) &\le E_{\hat i}^{\pi_n^{*1},\pi^2}\left[e^{\int_0^{T\wedge\hat\tau(\mathscr{B})}(c(\xi_s,\pi_n^{*1}(\xi_{s-}),\pi^2(s))-\rho_n)ds}\psi_n(\xi_{T\wedge\hat\tau(\mathscr{B})})I_{\{T\wedge\hat\tau(\mathscr{B})<\tau_n\}}\right]\\ &\le E_{\hat i}^{\pi_n^{*1},\pi^2}\left[e^{\int_0^{T\wedge\hat\tau(\mathscr{B})}\ell(\xi_s)ds}\psi_n\left(\xi_{T\wedge\hat\tau(\mathscr{B})}\right)I_{\{T\wedge\hat\tau(\mathscr{B})<\tau_n\}}\right].\end{aligned}$$

Since $\psi_n \leq V$, in view of Lemma 7.3.2, by the dominated convergence theorem, taking $T \to \infty$, we get

$$\psi_n(\hat{i}) \leq E_{\hat{i}}^{\pi_n^{*1},\pi^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B})}\ell(\xi_s)ds}\psi_n\left(\xi_{\hat{\tau}(\mathscr{B})}\right)\right].$$

Using this and (7.3.10), we have

$$0 = (V - \psi_n)(\hat{i}) \geq E_{\hat{i}}^{\pi_n^{*1},\pi^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B})}\ell(\xi_s)ds}(V - \psi_n)(\xi_{\hat{\tau}(\mathscr{B})})\right] > 0.$$

Hence we arrive at a contradiction. Thus ψ_n touches V inside $\mathscr{B}$. A similar conclusion holds under Assumption $(A2)$(a). Now, since $\psi_n \leq V$ for all large n, by a suitable diagonalization procedure, there exists a subsequence (by abuse of notation, we use the same sequence) such that, $\psi_n \to \psi^*$ for all $i \in S$, as $n \to \infty$, and $\psi^* \leq V$. Also, since by Lemma 7.3.3, the sequence $\{\rho_n\}$ is bounded and $\liminf_{n\to\infty} \rho_n \geq 0$, we can find a subsequence (by abuse of notation we use the same sequence) and some $\rho \geq 0$ such that $\rho_n \to \rho$ as $n \to \infty$. Thus as before, there exists a mini-max selector $(\pi_n^{*1}, \pi_n^{*2}) \in \Pi_{SM}^1 \times \Pi_{SM}^2$ of (7.3.11), i.e.,

$$\begin{aligned}
&\inf_{\nu\in\mathcal{P}(B(i))}\left[\sum_{j\in S}\psi_n(j)q(j|i,\pi_n^{*1}(i),\nu) + c(i,\pi_n^{*1}(i),\nu)\psi_n(i)\right]\\
&= \sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[\sum_{j\in S}\psi_n(j)q(j|i,\mu,\nu) + c(i,\mu,\nu)\psi_n(i)\right]\\
&= \inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in\mathcal{P}(A(i))}\left[\sum_{j\in S}\psi_n(j)q(j|i,\mu,\nu) + c(i,\mu,\nu)\psi_n(i)\right]\\
&= \sup_{\mu\in\mathcal{P}(A(i))}\left[\sum_{j\in S}\psi_n(j)q(j|i,\mu,\pi_n^{*2}(i)) + c(i,\mu,\pi_n^{*2}(i))\psi_n(i)\right].
\end{aligned} \tag{7.4.5}$$

Hence,

$$\rho_n\psi_n(i) \leq \left[\sum_{j\in S}\psi_n(j)q(j|i,\pi_n^{*1}(i),\nu) + c(i,\pi_n^{*1}(i),\nu)\psi_n(i)\right].$$

The above implies

$$\rho_n\psi_n(i) - \psi_n(i)q(i|i,\pi_n^{*1}(i),\nu) \leq \left[\sum_{j\neq i}\psi_n(j)q(j|i,\pi_n^{*1}(i),\nu) + c(i,\pi_n^{*1}(i),\nu)\psi_n(i)\right]. \tag{7.4.6}$$

Now, since $\psi_n(i) \leq V(i)$ for all $i \in S$, we have

$$\sum_{j\neq i}\psi(j)q(j|i,\pi_n^{*1}(i),\nu) \leq \sum_{j\neq i}V(j)q(j|i,\pi_n^{*1}(i),\nu). \tag{7.4.7}$$

Also, since Π^1_{SM} and Π^2_{SM} are compact there exist $\pi^{*1} \in \Pi^1_{SM}$ and $\pi^{*2} \in \Pi^2_{SM}$ such that $\pi^{*1}_n \to \pi^{*1}$ and $\pi^{*2}_n \to \pi^{*2}$ as $n \to \infty$. Under given assumptions, from [56, Lemma 7.2] it is clear that the functions $c(i, \mu, \nu)$, and $\sum_{j \in S} q(j|i, \mu, \nu)u(j)$ are continuous at (μ, ν) on $\mathcal{P}(A(i)) \times \mathcal{P}(B(i))$ for each fixed $u \in L^\infty_V(S)$, $i \in S$. Therefore by the dominated convergence theorem, letting $n \to \infty$ in (7.4.6), we obtain

$$\rho\psi^*(i) \le \sum_{j \in S} \psi^*(j)q(j|i, \pi^{*1}(i), \nu) + c(i, \pi^{*1}(i), \nu)\psi^*(i).$$

Hence we have

$$\begin{aligned}\rho\psi^*(i) &\le \inf_{\nu \in \mathcal{P}(B(i))} \Big[\sum_{j \in S} \psi^*(j)q(j|i, \pi^{*1}(i), \nu) + c(i, \pi^{*1}(i), \nu)\psi^*(i)\Big]. \\ &\le \sup_{\mu \in \mathcal{P}(A(i))} \inf_{\nu \in \mathcal{P}(B(i))} \Big[\sum_{j \in S} \psi^*(j)q(j|i, \mu, \nu) + c(i, \mu, \nu)\psi^*(i)\Big]. \end{aligned} \tag{7.4.8}$$

By similar arguments using (7.4.5) and extended Fatou's lemma [69, Lemma 8.3.7], we get

$$\begin{aligned}\rho\psi^*(i) &\ge \sup_{\mu \in \mathcal{P}(A(i))} \Big[\sum_{j \in S} \psi^*(j)q(j|i, \mu, \pi^{*2}(i)) + c(i, \mu, \pi^{*2}(i))\psi^*(i)\Big] \\ &\ge \inf_{\nu \in \mathcal{P}(B(i))} \sup_{\mu \in \mathcal{P}(A(i))} \Big[\sum_{j \in S} \psi^*(j)q(j|i, \mu, \nu) + c(i, \mu, \nu)\psi^*(i)\Big]. \end{aligned} \tag{7.4.9}$$

Hence by (7.4.8) and (7.4.9), we get (7.4.1). Since at some point in $\mathscr{B}$ we have $(V - \psi_n) = 0$, for all large n. It follows that $(V - \psi^*)(i^*) = 0$ for some $i^* \in \mathscr{B}$. Since $V \ge 1$, it is clear that ψ^* is nontrivial. Now we claim that $\psi^* > 0$. If not, then we must have $\psi^*(\tilde{i}) = 0$ for some $\tilde{i} \in S$. Again as before, there exists a pair of a mini-max selector $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ such that form (7.4.1), we have

$$\rho\psi^*(\tilde{i}) = \Big[\sum_{j \in S} \psi^*(j)q(j|\tilde{i}, \pi^{*1}(\tilde{i}), \pi^{*2}(\tilde{i})) + c(\tilde{i}, \pi^{*1}(\tilde{i}), \pi^{*2}(\tilde{i}))\psi^*(\tilde{i})\Big]. \tag{7.4.10}$$

This implies

$$\sum_{j \neq \tilde{i}} \psi^*(j)q(j|\tilde{i}, \pi^{*1}(\tilde{i}), \pi^{*2}(\tilde{i})) = 0.$$

Since the Markov chain ξ is irreducible under $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$, from the above equation, it follows that $\psi^* \equiv 0$. So, we arrive at a contradiction. This proves the claim. Now we prove (i) and (ii).

(i) Since $\psi^* > 0$ and $\psi_n(i) \to \psi^*(i)$ as $n \to \infty$, we have $\psi_n > 0$ for all large enough n. So, using (7.3.12), we have $\lim_{n \to \infty} \rho_n \le \sup_{\pi^1 \in \Pi^1_{Ad}} \inf_{\pi^2 \in \Pi^2_{Ad}} J(i, c, \pi^1, \pi^2)$ for all $i \in S$.

(ii) Using compactness-continuity assumptions, there exists a pair of strategies (a mini-max selector) $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ (as in (7.4.5)) satisfying

$$\begin{aligned}\rho\psi^*(i) &= \sup_{\mu\in\mathcal{P}(A(i))}\left[\sum_{j\in S}\psi^*(j)q(j|i,\mu,\pi^{*2}(i)) + c(i,\mu,\pi^{*2}(i))\psi^*(i)\right]\\ &= \inf_{\nu\in\mathcal{P}(B(i))}\left[\sum_{j\in S}\psi^*(j)q(j|i,\pi^{*1}(i),\nu) + c(i,\pi^{*1}(i),\nu)\psi^*(i)\right]. \qquad (7.4.11)\end{aligned}$$

Using (7.4.11), Lemma 7.3.2, and Dynkin's formula, we have

$$\psi^*(i) \geq E_i^{\pi^1,\pi^{*2}}\left[e^{\int_0^{\hat{\tau}(\mathscr{B})\wedge T}(c(\xi_t,\pi^1(t),\pi^{*2}(\xi_{t-}))-\rho)dt}\psi^*(\xi_{\hat{\tau}(\mathscr{B})\wedge T})\right] \ \forall i \in \mathscr{B}^c.$$

By Fatou's lemma taking $T \to \infty$, we get

$$\psi^*(i) \geq E_i^{\pi^1,\pi^{*2}}\left[e^{\int_0^{\hat{\tau}(\mathscr{B})}(c(\xi_t,\pi^1(t),\pi^{*2}(\xi_{t-}))-\rho)dt}\psi^*(\xi_{\hat{\tau}(\mathscr{B})})\right], \ \forall i \in \mathscr{B}^c. \qquad (7.4.12)$$

Hence,

$$\begin{aligned}\psi^*(i) &\geq \sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^{*2}}\left[e^{\int_0^{\hat{\tau}(\mathscr{B})}(c(\xi_t,\pi^1(t),\pi^{*2}(\xi_{t-}))-\rho)dt}\psi^*(\xi_{\hat{\tau}(\mathscr{B})})\right]\\ &\geq \inf_{\pi^2\in\Pi^2_{Ad}}\sup_{\pi^1\in\Pi^1_{Ad}} E_i^{\pi^1,\pi^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B})}(c(\xi_t,\pi^1(t),\pi^2(t))-\rho)dt}\psi^*(\xi_{\hat{\tau}(\mathscr{B})})\right], \ \forall i \in \mathscr{B}^c. \qquad (7.4.13)\end{aligned}$$

Also, using (7.4.11), Lemma 7.3.2, and Dynkin's formula, we obtain

$$\psi^*(i) \leq E_i^{\pi^{*1},\pi^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B})\wedge T}(c(\xi_t,\pi^{*1}(\xi_{t-}),\pi^2(t))-\rho)dt}\psi^*(\xi_{\hat{\tau}(\mathscr{B})\wedge T})\right] \ \forall i \in \mathscr{B}^c.$$

Since $\psi^* \leq V$, using the estimates as in Lemma 7.3.2, taking $T \to \infty$, by the dominated convergence theorem it follows that

$$\psi^*(i) \leq E_i^{\pi^{*1},\pi^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B})}(c(\xi_t,\pi^{*1}(\xi_{t-}),\pi^2(t))-\rho)dt}\psi^*(\xi_{\hat{\tau}(\mathscr{B})})\right] \ \forall i \in \mathscr{B}^c. \qquad (7.4.14)$$

Hence

$$\begin{aligned}\psi^*(i) &\leq \inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^{*1},\pi^2}\left[e^{\int_0^{\tilde{\tau}(\mathscr{B})}(c(\xi_t,\pi^{*1}(\xi_{t-}),\pi^2(t))-\rho)dt}\psi^*(\xi_{\hat{\tau}(\mathscr{B})})\right]\\ &\leq \sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}} E_i^{\pi^1,\pi^2}\left[e^{\int_0^{\hat{\tau}(\mathscr{B})}(c(\xi_t,\pi^1(t),\pi^2(t))-\rho)dt}\psi^*(\xi_{\hat{\tau}(\mathscr{B})})\right], \ \forall i \in \mathscr{B}^c. \qquad (7.4.15)\end{aligned}$$

From (7.4.13) and (7.4.15), we obtain (7.4.2). □

We now establish the following result which brings us just a step away from proving the existence of a pair of saddle-point strategies.

Theorem 7.4.1. *Suppose Assumptions (A1)-(A3) hold. Then for any mini-max selector* $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ *of (7.4.1), i.e., for any pair* $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ *satisfying*

$$\begin{aligned}
&\inf_{\nu\in\mathcal{P}(B(i))}\left[\sum_{j\in S}\psi^*(j)q(j|i,\pi^{*1}(i),\nu)+c(i,\pi^{*1}(i),\nu)\psi^*(i)\right]\\
&= \sup_{\mu\in\mathcal{P}(A(i))}\inf_{\nu\in\mathcal{P}(B(i))}\left[\sum_{j\in S}\psi^*(j)q(j|i,\mu,\nu)+c(i,\mu,\nu)\psi^*(i)\right]\\
&= \inf_{\nu\in\mathcal{P}(B(i))}\sup_{\mu\in P(A(i))}\left[\sum_{j\in S}\psi^*(j)q(j|i,\mu,\nu)+c(i,\mu,\nu)\psi^*(i)\right]\\
&= \sup_{\mu\in\mathcal{P}(A(i))}\left[\sum_{j\in S}\psi^*(j)q(j|i,\mu,\pi^{*2}(i))+c(i,\mu,\pi^{*2}(i))\psi^*(i)\right],\ i\in S, \qquad (7.4.16)
\end{aligned}$$

we have

$$\begin{aligned}
\rho &= \inf_{i\in S}\sup_{\pi^1\in\Pi^1_{Ad}}\limsup_{T\to\infty}\frac{1}{T}\log E_i^{\pi^1,\pi^{*2}}\left[e^{\int_0^T c(\xi_t,\pi^1(t),\pi^{*2}(\xi_{t-}))dt}\right]\\
&= \inf_{i\in S}\sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}}\limsup_{T\to\infty}\frac{1}{T}\log E_i^{\pi^1,\pi^2}\left[e^{\int_0^T c(\xi_t,\pi^1(t),\pi^2(t))dt}\right]\\
&= \inf_{i\in S}\inf_{\pi^2\in\Pi^2_{Ad}}\sup_{\pi^1\in\Pi^1_{Ad}}\limsup_{T\to\infty}\frac{1}{T}\log E_i^{\pi^1,\pi^2}\left[e^{\int_0^T c(\xi_t,\pi^1(t),\pi^2(t))dt}\right]. \qquad (7.4.17)
\end{aligned}$$

Proof. We perturb the cost function as follows.

- If Assumption $(A2)$(a) holds: We define for $(a,b)\in A(i)\times B(i)$, $i\in S$, $\hat{c}_m(i,a,b) = c(i,a,b)I_{\mathscr{D}_m}(i)+(\|c\|_\infty+\alpha_3)I_{\mathscr{D}^c_m}$. Here $\alpha_3>0$, is a small number satisfying $\|c\|_\infty+\alpha_3<\gamma$. Note that $\|\hat{c}_m\|_\infty<\gamma$.
- If Assumption $(A2)$(b) holds: We define for $(a,b)\in A(i)\times B(i)$, $i\in S$, $\hat{c}_m(i,a,b) = c(i,a,b)+\frac{1}{m}[\ell(i)-\sup_{(a,b)\in A(i)\times B(i)}c(i,a,b)]_+$. Note that the function
$$\left[\ell(\cdot)-\sup_{(a,b)\in A(\cdot)\times B(\cdot)}c(\cdot,a,b)\right]_+$$
is a norm-like function. Also, it is easy to see that for large enough m, $\ell(\cdot)-\sup_{(a,b)\in A(\cdot)\times B(\cdot)}\hat{c}_m(\cdot,a,b)$ is norm-like.

In view of Lemma 7.4.1, it is clear that for $\pi^{*2}\in\Pi^2_{SM}$, there exists $(\tilde{\psi}_m,\tilde{\rho}_m)\in L^\infty_V(S)\times\mathbb{R}_+$, $\tilde{\psi}_m>0$ satisfying

$$\tilde{\rho}_m\tilde{\psi}_m(i) = \sup_{\mu\in\mathcal{P}(A(i))}\left[\sum_{j\in S}\tilde{\psi}_m(j)q(j|i,\mu,\pi^{*2}(i))+\hat{c}_m(i,\mu,\pi^{*2}(i))\tilde{\psi}_m(i)\right] \qquad (7.4.18)$$

such that

$$0 \leq \tilde{\rho}_m \leq \sup_{\pi^1 \in \Pi^1_{Ad}} \limsup_{T \to \infty} \frac{1}{T} \log E_i^{\pi^1, \pi^{*2}} \left[e^{\int_0^T \hat{c}_m(\xi_t, \pi^1(t), \pi^{*2}(\xi_{t-}))dt} \right]. \tag{7.4.19}$$

Also, for some finite set $\mathscr{B}_1 \supset \mathscr{B} \supset \mathscr{K}$, we have

$$\tilde{\psi}_m(i) = \sup_{\pi^1 \in \Pi^1_{Ad}} E_i^{\pi^1, \pi^{*2}} \left[e^{\int_0^{\hat{\tau}(\mathscr{B}_1)} (\hat{c}_m(\xi_t, \pi^1(t), \pi^{*2}(\xi_{t-})) - \tilde{\rho}_m)dt} \tilde{\psi}_m(\xi_{\hat{\tau}(\mathscr{B}_1)}) \right], \ i \in \mathscr{B}_1^c. \tag{7.4.20}$$

Now from the proof of Lemma 7.4.1, we have a finite set $\tilde{\mathscr{B}}$, depending on m, containing $\mathscr{K}$ such that the following cases happen:

- Under Assumption $(A2)$(a): From (7.4.19), we have $\tilde{\rho}_m \leq \|\hat{c}_m\|_\infty$. Thus, for $i \in \mathscr{D}_m^c$, it follows that $\hat{c}_m(i,a,b) - \tilde{\rho}_m \geq 0$ for all $(a,b) \in A(i) \times B(i)$. Consequently, we may take $\tilde{\mathscr{B}} = \mathscr{D}_m$ such that $\hat{c}_m(i,a,b) - \tilde{\rho}_m \geq 0$ in $\tilde{\mathscr{B}}^c$ for all $(a,b) \in A(i) \times B(i)$.
- Under Assumption $(A2)$(b): since $\hat{c}_m$ is norm-like function, we can choose suitable finite set $\tilde{\mathscr{B}}$ such that $(\hat{c}_m(i,a,b) - \tilde{\rho}_m) \geq 0$ in $\tilde{\mathscr{B}}^c$ for all $(a,b) \in A(i) \times B(i)$.

For any $\pi^1 \in \Pi^1_{Ad}$, applying Dynkin formula and using (7.4.18) and Lemma 7.3.2, we get

$$\tilde{\psi}_m(i) \geq E_i^{\pi^1, \pi^{*2}} \left[e^{\int_0^{\hat{\tau}(\tilde{\mathscr{B}}) \wedge T} (\hat{c}_m(\xi_t, \pi^1(t), \pi^{*2}(\xi_{t-})) - \tilde{\rho}_m)dt} \tilde{\psi}_m(\xi_{\hat{\tau}(\tilde{\mathscr{B}}) \wedge T}) \right].$$

Since for $i \in \tilde{\mathscr{B}}^c$, $\hat{c}_m(i,a,b) - \tilde{\rho}_m \geq 0$, by Fatou's lemma taking $T \to \infty$, we get

$$\tilde{\psi}_m(i) \geq E_i^{\pi^1, \pi^{*2}} \left[e^{\int_0^{\hat{\tau}(\tilde{\mathscr{B}})} (\hat{c}_m(\xi_t, \pi^1(t), \pi^{*2}(\xi_{t-})) - \tilde{\rho}_m)dt} \tilde{\psi}_m(\xi_{\hat{\tau}(\tilde{\mathscr{B}})}) \right] \geq (\min_{\tilde{\mathscr{B}}} \tilde{\psi}_m) \ \forall \ i \in \tilde{\mathscr{B}}^c.$$

This implies that $\tilde{\psi}_m$ has a lower bound. Now, applying Dynkin formula, and using (7.4.18) and Lemma 7.3.2, we deduce that

$$\tilde{\psi}_m(i) \geq E_i^{\pi^1, \pi^{*2}} \left[e^{\int_0^{T \wedge \tau_N} (\hat{c}_m(\xi_t, \pi^1(t), \pi^{*2}(\xi_{t-})) - \tilde{\rho}_m)dt} \tilde{\psi}_m(\xi_{T \wedge \tau_N}) \right],$$

for any $i \in S$, where $\tau_N := \inf\{t \geq 0 : \xi_t \notin \{1, 2, \cdots, N\}\}$, $N \in \mathbb{N}$. By Fatou's lemma taking $N \to \infty$, we get

$$\begin{aligned} \tilde{\psi}_m(i) &\geq E_i^{\pi^1, \pi^{*2}} \left[e^{\int_0^T (\hat{c}_m(\xi_t, \pi^1(t), \pi^{*2}(\xi_{t-})) - \tilde{\rho}_m)dt} \tilde{\psi}_m(\xi_T) \right] \\ &\geq \min_{\tilde{\mathscr{B}}} \tilde{\psi}_m E_i^{\pi^1, \pi^{*2}} \left[e^{\int_0^T (\hat{c}_m(\xi_t, \pi^1(t), \pi^{*2}(\xi_{t-})) - \tilde{\rho}_m)dt} \right]. \end{aligned}$$

Now since $\tilde{\mathscr{B}}$ is a finite set and $\tilde{\psi}_m > 0$, $\min_{\tilde{\mathscr{B}}} \tilde{\psi}_m > 0$ is also finite. Also, the logarithm is an increasing function. Thus, taking logarithm on both sides, we have

$$\log \tilde{\psi}_m(i) \geq \log \min_{\tilde{\mathscr{B}}} \tilde{\psi}_m + \log E_i^{\pi^1, \pi^{*2}} \left[e^{\int_0^T \hat{c}_m(\xi_t, \pi^1(t), \pi^{*2}(\xi_{t-}))dt} \right] - T\tilde{\rho}_m.$$

Thus dividing by T both sides and letting $T\to\infty$, we obtain

$$\tilde{\rho}_m \geq \limsup_{T\to\infty}\frac{1}{T}\log E_i^{\pi^1,\pi^{*2}}\left[e^{\int_0^T \hat{c}_m(\xi_t,\pi^1(t),\pi^{*2}(\xi_{t-}))dt}\right].$$

Since $\pi^1\in\Pi^1_{Ad}$ arbitrary, it follows that

$$\begin{aligned}\tilde{\rho}_m &\geq \sup_{\pi^1\in\Pi^1_{Ad}}\limsup_{T\to\infty}\frac{1}{T}\log E_i^{\pi^1,\pi^{*2}}\left[e^{\int_0^T \hat{c}_m(\xi_t,\pi^1(t),\pi^{*2}(\xi_{t-}))dt}\right]\\ &\geq \sup_{\pi^1\in\Pi^1_{Ad}}\limsup_{T\to\infty}\frac{1}{T}\log E_i^{\pi^1,\pi^{*2}}\left[e^{\int_0^T c(\xi_t,\pi^1(t),\pi^{*2}(\xi_{t-}))dt}\right].\end{aligned}$$

Using this and (7.4.19), we get $\sup_{\pi^1\in\Pi^1_{Ad}} J(i,c,\pi^1,\pi^{*2}) \leq \sup_{\pi^1\in\Pi^1_{Ad}} J(i,\hat{c}_m,\pi^1,\pi^{*2}) = \tilde{\rho}_m$ for all m. From the definition of $\hat{c}_m$, it is easy to see that $\tilde{\rho}_m$ is a decreasing sequence that has a lower bound. Now by similar arguments as in Lemma 7.4.1, it follows that there exists a pair $(\tilde{\rho},\tilde{\psi})$ such that $\tilde{\rho}_m\to\tilde{\rho}$ and $\tilde{\psi}_m\to\tilde{\psi}$ as $m\to\infty$. As in Lemma 7.4.1, by taking $m\to\infty$ in (7.4.18), we get

$$\tilde{\rho}\tilde{\psi}(i) = \sup_{\mu\in\mathcal{P}(A(i))}\left[\sum_{j\in S}\tilde{\psi}(j)q(j|i,\mu,\pi^{*2}(i)) + c(i,\mu,\pi^{*2}(i))\tilde{\psi}(i)\right]. \tag{7.4.21}$$

Also, we have $\tilde{\rho}\geq \sup_{\pi^1\in\Pi^1_{Ad}} J(i,c,\pi^1,\pi^{*2})\geq\rho$. Now, we want to show that $\tilde{\rho}=\rho$. Let $\tilde{\pi}^{*1}$ be a maximizing selector in (7.4.21). Thus

$$\tilde{\rho}\tilde{\psi}(i) = \left[\sum_{j\in S}\tilde{\psi}(j)q(j|i,\tilde{\pi}^{*1}(i),\pi^{*2}(i)) + c(i,\tilde{\pi}^{*1}(i),\pi^{*2}(i))\tilde{\psi}(i)\right]. \tag{7.4.22}$$

In view of estimates in Lemma 7.3.2, applying Dynkin's formula and the dominated convergence theorem, from (7.4.22) we deduce that there exists a finite set $\mathscr{B}_2\supset\mathscr{B}_1$ such that

$$\tilde{\psi}(i)\leq E_i^{\tilde{\pi}^{*1},\pi^{*2}}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_2)}(c(\xi_t,\tilde{\pi}^{*1}(\xi_{t-}),\pi^{*2}(\xi_{t-}))-\tilde{\rho})dt}\tilde{\psi}(\xi_{\hat{\tau}(\mathscr{B}_2)})\right],\ \forall\ i\in\mathscr{B}_2^c. \tag{7.4.23}$$

Since $\tilde{\rho}\geq\rho$, arguing as in Lemma 7.4.1 (see, (7.4.2)) for $\mathscr{B}_2\supset\mathscr{B}$ we have

$$\psi^*(i)\geq E_i^{\tilde{\pi}^{*1},\pi^{*2}}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_2)}(c(\xi_t,\tilde{\pi}^{*1}(\xi_{t-}),\pi^{*2}(\xi_{t-}))-\tilde{\rho})dt}\psi^*(\xi_{\hat{\tau}(\mathscr{B}_2)})\right]\ \forall\ i\in\mathscr{B}_2^c. \tag{7.4.24}$$

Now we choose an appropriate constant κ (e.g., $\kappa=\min_{\mathscr{B}_2}\frac{\psi^*}{\tilde{\psi}}$), so that $(\psi^*-\kappa\tilde{\psi})\geq 0$ in $\mathscr{B}_2$ and for some $\hat{i}_0\in\mathscr{B}_2$, $(\psi^*-\kappa\tilde{\psi})(\hat{i}_0)=0$. From (7.4.23) and (7.4.24), we get

$$\psi^*(i)-\kappa\tilde{\psi}(i)\geq E_i^{\tilde{\pi}^{*1},\pi^{*2}}\left[e^{\int_0^{\hat{\tau}(\mathscr{B}_2)}(c(\xi_t,\tilde{\pi}^{*1}(\xi_{t-}),\pi^{*2}(\xi_{t-}))-\tilde{\rho})dt}(\psi^*-\kappa\tilde{\psi})(\xi_{\hat{\tau}(\mathscr{B}_2)})\right]\ \forall\ i\in\mathscr{B}_2^c. \tag{7.4.25}$$

From the above expression it is easy to see that $(\psi^* - \kappa\tilde{\psi}) \geq 0$ in S. Now using (7.4.16), (7.4.22) and the fact that $\tilde{\rho} \geq \rho$, we get

$$\tilde{\rho}(\psi^* - \kappa\tilde{\psi})(\hat{i}_0) \geq \Big[\sum_{j\in S}(\psi^* - \kappa\tilde{\psi})(j)q(j|\hat{i}_0, \tilde{\pi}^{*1}(\hat{i}_0), \pi^{*2}(\hat{i}_0)) + c(\hat{i}_0, \tilde{\pi}^{*1}(\hat{i}_0), \pi^{*2}(\hat{i}_0))(\psi^* - \kappa\tilde{\psi})(\hat{i}_0)\Big].$$

This implies that

$$\sum_{j\neq\hat{i}_0}(\psi^* - \kappa\tilde{\psi})(j)q(j|\hat{i}_0, \tilde{\pi}^{*1}(\hat{i}_0), \pi^{*2}(\hat{i}_0)) = 0\,. \tag{7.4.26}$$

Since the Markov chain ξ is irreducible under $(\tilde{\pi}^{*1}, \pi^{*2})$, by (7.4.26), we have $\psi^* = \kappa\tilde{\psi}$ in S. From (7.4.16) and (7.4.21) it follows that $\tilde{\rho} = \rho$. This proves (7.4.17). □

In the next theorem, we show that any mini-max selector of (7.4.1) is a saddle-point equilibrium.

Theorem 7.4.2. *Suppose Assumptions (A1)-(A3) hold. Then any mini-max selector* $(\pi^{*1}, \pi^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ *of (7.4.1) is a saddle-point equilibrium.*

Proof. Arguing as in Lemma 7.4.1 and Theorem 7.4.1, there exists $(\rho^{\pi^{*1},\pi^{*2}}, \psi^{\pi^{*1},\pi^{*2}}) \in \mathbb{R}_+ \times L^\infty_V(S)$ with $\psi^{\pi^{*1},\pi^{*2}} > 0$ satisfying

$$\rho^{\pi^{*1},\pi^{*2}}\psi^{\pi^{*1},\pi^{*2}}(i) = \Big[\sum_{j\in S}\psi^{\pi^{*1},\pi^{*2}}(j)q(j|i, \pi^{*1}(i), \pi^{*2}(i)) + c(i, \pi^{*1}(i), \pi^{*2}(i))\psi^{\pi^{*1},\pi^{*2}}(i)\Big]. \tag{7.4.27}$$

Furthermore $\rho^{\pi^{*1},\pi^{*2}} = J(i, c, \pi^{*1}, \pi^{*2})$ and for some finite set $\mathscr{B} \supset \mathscr{K}$ (without loss of generality denoting by the same notation)

$$\psi^{\pi^{*1},\pi^{*2}}(i) = E_i^{\pi^{*1},\pi^{*2}}\Big[e^{\int_0^{\hat{\tau}(\mathscr{B})}(c(\xi_t,\pi^{*1}(\xi_{t-}),\pi^{*2}(\xi_{t-}))-\rho^{\pi^{*1},\pi^{*2}})dt}\psi^{\pi^{*1},\pi^{*2}}(\xi_{\hat{\tau}(\mathscr{B})})\Big] \quad \forall i \in \mathscr{B}^c. \tag{7.4.28}$$

Thus, from (7.4.17) it is clear that $\rho^{\pi^{*1},\pi^{*2}} \leq \rho$. Now, following similar arguments as in Theorem 7.4.1 it is easy to see that $\rho^{\pi^{*1},\pi^{*2}} = \rho$. This implies that $J(i, c, \pi^1, \pi^{*2}) \leq \rho^{\pi^{*1},\pi^{*2}}$ for all $\pi^1 \in \Pi^1_{Ad}$. Next, from [21] it is clear that if we consider the minimization problem

$$\min_{\pi^2\in\Pi^2_{Ad}} J(i, c, \pi^{*1}, \pi^2),$$

then the optimal control exists in the space of stationary Markov strategies. Thus to complete the proof, it is enough to show that $J(i, c, \pi^{*1}, \pi^{*2}) \leq J(i, c, \pi^{*1}, \pi^2)$ for any $\pi^2 \in \Pi^2_{SM}$. If not suppose that $J(i, c, \pi^{*1}, \pi^{*2}) > J(i, c, \pi^{*1}, \pi^2)$ for some $\pi^2 \in \Pi^2_{SM}$. We know that for $\pi^2 \in \Pi^2_{SM}$, there exists $(\rho^{\pi^{*1},\pi^2}, \psi^{\pi^{*1},\pi^2}) \in \mathbb{R}_+ \times L^\infty_V(S)$ with $\psi^{\pi^{*1},\pi^2} > 0$ satisfying

$$\rho^{\pi^{*1},\pi^2}\psi^{\pi^{*1},\pi^2}(i) = \Big[\sum_{j\in S}\psi^{\pi^{*1},\pi^2}(j)q(j|i,\pi^{*1}(i),\pi^2(i)) + c(i,\pi^{*1}(i),\pi^2(i))\psi^{\pi^{*1},\pi^2}(i)\Big], \tag{7.4.29}$$

also we have $\rho^{\pi^{*1},\pi^2} = J(i,c,\pi^{*1},\pi^2)$ and for some finite set $\mathscr{B}$ $(\supset \mathscr{K})$

$$\psi^{\pi^{*1},\pi^2}(i) = E_i^{\pi^{*1},\pi^2}\Big[e^{\int_0^{\hat\tau(\mathscr{B})}(c(\xi_t,\pi^{*1}(\xi_{t-}),\pi^2(\xi_{t-}))-\rho^{\pi^{*1},\pi^2})dt}\psi^{\pi^{*1},\pi^2}(\xi_{\hat\tau(\mathscr{B})})\Big] \ \forall i \in \mathscr{B}^c. \tag{7.4.30}$$

From (7.4.2), we deduce that

$$\psi^*(i) \le E_i^{\pi^{*1},\pi^2}\Big[e^{\int_0^{\hat\tau(\mathscr{B})}(c(\xi_t,\pi^{*1}(\xi_{t-}),\pi^2(\xi_{t-}))-\rho)dt}\psi^*(\xi_{\hat\tau(\mathscr{B})})\Big], \ \forall i \in \mathscr{B}^c. \tag{7.4.31}$$

Now, as in Theorem 7.4.1, using (7.4.30) and (7.4.31) one can deduce that $\psi^{\pi^{*1},\pi^2} = \eta\psi^*$ for some positive constant η. Thus, in view of (7.4.1) and (7.4.29), it follows that $\rho \le \rho^{\pi^{*1},\pi^2}$, i.e., $J(i,c,\pi^{*1},\pi^{*2}) \le J(i,c,\pi^{*1},\pi^2)$, which is a contradiction. This completes the proof. □

Next, we prove the converse of the above theorem.

Theorem 7.4.3. *Suppose Assumptions (A1)-(A3) hold. If there exists a saddle-point equilibrium* $(\hat\pi^{*1},\hat\pi^{*2}) \in \Pi^1_{SM}\times\Pi^2_{SM}$ *, i.e.,*

$$J(i,c,\pi^1,\hat\pi^{*2}) \le J(i,c,\hat\pi^{*1},\hat\pi^{*2}) \le J(i,c,\hat\pi^{*1},\pi^2),$$

for all $i \in S$ *,* $\pi^1 \in \Pi^1_{Ad}$ *and* $\pi^2 \in \Pi^2_{Ad}$ *. Then* $(\hat\pi^{*1},\hat\pi^{*2})$ *is a mini-max selector of (7.4.1).*

Proof. From Theorem 7.4.2, we deduce that

$$\begin{aligned}\rho = \inf_{\pi^2\in\Pi^2_{Ad}}\sup_{\pi^1\in\Pi^1_{Ad}} J(i,c,\pi^1,\pi^2) &\le \sup_{\pi^1\in\Pi^1_{Ad}} J(i,c,\pi^1,\hat\pi^{*2}) \le J(i,c,\hat\pi^{*1},\hat\pi^{*2})\\ &\le \inf_{\pi^2\in\Pi^2_{Ad}} J(i,c,\hat\pi^{*1},\pi^2) \le \sup_{\pi^1\in\Pi^1_{Ad}}\inf_{\pi^2\in\Pi^2_{Ad}} J(i,c,\pi^1,\pi^2) = \rho\,.\end{aligned}$$

This implies that $\rho = J(i,c,\hat\pi^{*1},\hat\pi^{*2})$ and $\rho = \inf_{\pi^2\in\Pi^2_{Ad}} J(i,c,\hat\pi^{*1},\pi^2)$. Arguing as in Lemma 7.4.1 and using Theorem 7.4.1, it follows that for $\hat\pi^{*1} \in \Pi^1_{SM}$ there exists $(\rho_{\hat\pi^{*1}},\psi^*_{\hat\pi^{*1}}) \in \mathbb{R}_+\times L^\infty_V(S)$ with $\psi^*_{\hat\pi^{*1}} > 0$ such that

$$\rho_{\hat\pi^{*1}}\psi^*_{\hat\pi^{*1}}(i) = \inf_{\nu\in\mathcal{P}(B(i))}\Big[\sum_{j\in S}\psi^*_{\hat\pi^{*1}}(j)q(j|i,\hat\pi^{*1}(i),\nu) + c(i,\hat\pi^{*1}(i),\nu)\psi^*_{\hat\pi^{*1}}(i)\Big], \tag{7.4.32}$$

and $\rho_{\hat{\pi}^{*1}} = \rho$ (since $\rho = \inf_{\pi^2 \in \Pi^2_{Ad}} J(i, c, \hat{\pi}^{*1}, \pi^2)$). Let (π^{*1}, π^{*2}) be any mini-max selector of (7.4.1). Then from the above, we get

$$\rho_{\hat{\pi}^{*1}} \psi^*_{\hat{\pi}^{*1}}(i) \leq \left[\sum_{j \in S} \psi^*_{\hat{\pi}^{*1}}(j) q(j|i, \hat{\pi}^{*1}(i), \pi^{*2}(i)) + c(i, \hat{\pi}^{*1}(i), \pi^{*2}(i)) \psi^*_{\hat{\pi}^{*1}}(i) \right]. \tag{7.4.33}$$

Again arguing as in Lemma 7.4.1, for some $\mathscr{B} \supset \mathscr{K}$ we have

$$\psi^*_{\hat{\pi}^{*1}}(i) \leq E_i^{\hat{\pi}^{*1}, \pi^{*2}} \left[e^{\int_0^{\hat{\tau}(\mathscr{B})} (c(\xi_t, \hat{\pi}^{*1}(\xi_{t-}), \pi^{*2}(\xi_{t-})) - \rho) dt} \psi^*_{\hat{\pi}^{*1}}(\xi_{\hat{\tau}(\mathscr{B})}) \right] \ \forall i \in \mathscr{B}^c. \tag{7.4.34}$$

Since, (π^{*1}, π^{*2}) is a mini-max selector of (7.4.1), we have

$$\rho \psi^*(i) \geq \left[\sum_{j \in S} \psi^*(j) q(j|i, \hat{\pi}^{*1}(i), \pi^{*2}(i)) + c(i, \hat{\pi}^{*1}(i), \pi^{*2}(i)) \psi^*(i) \right].$$

Thus, by applying Dynkin's formula and Fatou's lemma, we obtain

$$\psi^*(i) \geq E_i^{\hat{\pi}^{*1}, \pi^{*2}} \left[e^{\int_0^{\hat{\tau}(\mathscr{B})} (c(\xi_t, \hat{\pi}^{*1}(\xi_{t-}), \pi^{*2}(\xi_{t-})) - \rho) dt} \psi^*(\xi_{\hat{\tau}(\mathscr{B})}) \right] \ \forall i \in \mathscr{B}^c. \tag{7.4.35}$$

Using (7.4.34) and (7.4.35), and following the arguments as in Theorem 7.4.1 one can show that $\psi^* = \hat{\eta} \psi^*_{\hat{\pi}^{*1}}$ for some positive constant $\hat{\eta}$. Therefore, combining (7.4.1) and (7.4.32) it is easy to see that $\hat{\pi}^{*1}$ is an outer maximizing selector of (7.4.1). By similar arguments we can show that $\hat{\pi}^{*2}$ is an outer minimizing selector of (7.4.1). This completes the proof. □

Remark 7.4.2. *We can replace Assumption* (A3)*(iii) by other similar assumption. For example, if the killed process communicates with every state in* $\mathscr{D}_n$ *from* i_0 *before leaving the domain* $\mathscr{D}_n$*, for large* n*, then our method applies. More precisely for every* $\mathscr{D}_n$*,* $(\pi^1, \pi^2) \in \Pi^1_{SM} \times \Pi^2_{SM}$ *and for every* $j \in \mathscr{D}_n \backslash \{i_0\}$*, if there exists distinct* $i_1, i_2, \cdots, i_m \in \mathscr{D}_n \backslash \{i_0\}$ *satisfying*

$$q(i_1|i_0, \pi^1(i_0), \pi^2(i_0)) q(i_2|i_1, \pi^1(i_1), \pi^2(i_1)) \cdots q(j|i_m, \pi^1(i_m), \pi^2(i_m)) > 0,$$

then we get $\psi_n(i_0) > 0$ *in* $\mathscr{D}_n$ *(see Lemma 7.3.3). Also, the conclusion of Theorem 7.4.1 and Theorem 7.4.2 hold.*

7.5 Example

In this section, an illustrative example is presented. In our model the transition rate is unbounded, and the cost rate is nonnegative and unbounded.

Example 7.5.1. *Consider a controlled birth-death system arising in a single server queue wherein the state variable denotes the total population (i.e. the number of customers in the system) at each time* $t \geq 0$. *In this system there are 'natural' arrival and departure rates, say* $\hat{\lambda}$ *and* $\hat{\mu}$, *respectively. Here player 1 controls departure parameters* $\hat{h}_1$ *and player 2 controls arrival parameters* $\hat{h}_2$. *At any time* t, *when the state of the system is* $i \in S := \{0, 1, \cdots\}$, *player 1 takes an action* a *from a given set* $A(i)$ *(which is a closed, bounded interval of* $\mathbb{R}$*). This action may increase* $(\hat{h}_1(i,a) \geq 0)$ *or decrease* $(\hat{h}_1(i,a) \leq 0)$, *the departure rate and these actions result in a cost for this player denoted by* $\hat{c}_1(i,a)$ *per unit time. Similarly, if the state is* $i \in \{0, 1, \cdots\}$, *player 2 takes an action* b *from a set* $B(i)$ *(which is a closed and bounded interval of* $\mathbb{R}$*) to increase* $(\hat{h}_2(i,b) \geq 0)$ *or to decrease* $(\hat{h}_2(i,b) \leq 0)$, *the arrival rate and these actions produce a cost for this player denoted by* $\hat{c}_2(i,b)$ *per unit time. Also, in addition, assume that player 1 'owns' the system and he/she gets a reward* $\hat{p} \cdot i$ *for each unit of time during which the system remains in the state* $i \in S$, *where* $\hat{p} > 0$ *is a fixed reward fee per customer. We also assume that when the state of the system reaches state* $i = 0$, *any number of arrivals may occur. When there is no customer in the system, (i.e.,* $i = 0$*), control of departure is unnecessary.*

We next formulate this model as a continuous-time Markov game. The corresponding transition rate $q(j|i,a,b)$ *and the reward rate* $c(i,a,b)$ *for player 1(cost for player 2) are given as follows: for* $(0,a,b) \in K$ *(*K *as in the game model (7.2.1)). We take*

$$q(j|0,a,b) = \frac{\alpha}{(j+3)^4} \quad \text{for all } j \geq 1, \text{ such that } \sum_{j \in S} q(j|0,a,b) = 0, \tag{7.5.1}$$

where $\alpha > 0$ *is some constant so that* $q(0|0,a,b) \leq -3$. *Also for* $(i,a,b) \in K$ *with* $i \geq 1$,

$$q(j|i,a,b) = \begin{cases} \hat{\mu}(i+3)^2 + \hat{h}_1(i,a), \ \text{if } j = i-1 \\ -\hat{\lambda}i - \hat{\mu}(i+3)^2 - \hat{h}_2(i,b) - \hat{h}_1(i,a), \ \text{if } j = i \\ \hat{\lambda}i + \hat{h}_2(i,b), \ \text{if } j = i+1 \\ 0, \ \text{otherwise.} \end{cases}$$

$$c(i,a,b) := \hat{p} \cdot i - \hat{c}_1(i,a) + \hat{c}_2(i,b) \quad \text{for} \quad (i,a,b) \in K. \tag{7.5.2}$$

We assume that player 1 is maximizing player and player 2 is minimizing player. Player 1 gets an immediately reward rate $c(i,a,b)$ *and he/she tries to maximize his/her total payoff through the infinite horizon continuous-time ergodic cost criterion and at the same time, player 2 incurs an immediately cost rate* $c(i,a,b)$ *and he/she tries to minimize his/her total payoff through the same cost criterion.*

We now explore conditions under which there exists a pair of optimal strategies. To do so, we make the following assumptions.

(I) Let $\hat{\mu} \geq \max\{\hat{\lambda}, 2\}$, $\hat{\lambda} i + \hat{h}_2(i,b) > 0$, and $\hat{\mu}(i+3)^2 + \hat{h}_1(i,a) > 0$ for all $(i,a,b) \in K$ with $i \geq 1$; and assume that $\hat{h}_2(0,b) > 0$ and $\hat{h}_1(0,a) = 0$ for all $(a,b) \in A(i) \times B(i)$.

(II) The functions $\hat{h}_1(\cdot,\cdot) : S \times A \to [-\hat{\mu}, \hat{\mu}]$, $\hat{h}_2(\cdot,\cdot) : S \times B \to [-\hat{\lambda}, \hat{\lambda}]$, $\hat{c}_1(i,a)$, and $\hat{c}_2(i,b)$ are continuous with their respective variables for each fixed $i \in S$. Also, assume that $\min\limits_{(a,b)\in A(i)\times B(i)} [\hat{c}_1(i,a) - \hat{c}_2(i,b)]$ is norm-like function and $\hat{p} \cdot i - \hat{c}_1(i,a) + \hat{c}_2(i,b) \geq 0$ for $(i,a,b) \in K$. Here we take $\hat{p} < 1$.

Proposition 7.5.2. *Under conditions (I)-(II), the above controlled birth-death system satisfies the Assumptions (A1)-(A3). Hence by Theorem 7.4.2, there exists a pair of optimal strategies.*

Proof. Consider the Lyapunov function given by

$$V(i) := (i+3)^2 \quad \text{for} \quad i \in S.$$

We have $V(i) \geq 1$ for all $i \in S$. Now for each $i \geq 1$, and $(a,b) \in A(i) \times B(i)$, we have

$$\begin{aligned}
\sum_{j\in S} q(j|i,a,b)V(j) &= q(i-1|i,a,b)V(i-1) + V(i)q(i|i,a,b) + V(i+1)q(i+1|i,a,b) \\
&= (i+2)^2[\hat{\mu}(i+3)^2 + \hat{h}_1(i,a)] - (i+3)^2[\hat{\lambda} i + \hat{\mu}(i+3)^2 + \hat{h}_2(i,b) + \hat{h}_1(i,a)] \\
&\qquad + (i+4)^2[\hat{\lambda} i + \hat{h}_2(i,b)] \\
&= -[\hat{\mu}(i+3)^2 + \hat{h}_1(i,a)](2i+5) + (\hat{\lambda} i + \hat{h}_2(i,b)(2i+7) \\
&= -\hat{\mu}(i+3)^2(i+3+i+2) - \hat{h}_1(i,a)(2i+5) + (\hat{\lambda} i + \hat{h}_2(i,b)(2i+7) \\
&= -\hat{\mu}(i+3)^2(i+3) - \hat{\mu}(i+3)^2(i+2) - \hat{h}_1(i,a)(2i+5) \\
&\qquad + (\hat{\lambda} i + \hat{h}_2(i,b)(2i+7) \\
&\leq \hat{\mu}(i+3)(i+3)^2 - \hat{\mu}(i+3)^2(i+2) + \hat{\mu}(2i+5) + \hat{\mu} i(2i+7) + \hat{\mu}(2i+7) \\
&\quad (\text{ since } -\hat{h}_1(i,a) \leq \hat{\mu},\ \hat{\lambda} \leq \hat{\mu},\ \hat{h}_2(i,b) \leq \hat{\lambda} \leq \hat{\mu}, \text{ by conditions (I) and (II)}) \\
&= -\frac{\hat{\mu}}{2}(i+3)V(i) + \Big\{ -\frac{\hat{\mu}}{2}(i+3)(i+3)^2 - \hat{\mu}(i+3)^2(i+2) + \hat{\mu}(2i+5) \\
&\qquad + \hat{\mu} i(2i+7) + \hat{\mu}(2i+7) \Big\} \\
&\leq -\frac{\hat{\mu}}{2}(i+3)V(i) \ (\text{ since the term within the second bracket is negative}) \\
&\leq -(i+3)V(i) \ (\text{ by condition (I), since } \hat{\mu} \geq 2) \\
&= -\ell(i)V(i) \qquad (7.5.3)
\end{aligned}$$

where $\ell(i) = i + 3$. For $i = 0$,

$$\sum_{j\in S} q(j|i,a,b)V(j) = 9q(0|0,a,b) + \sum_{j\geq 1} q(j|0,a,b)(j+3)^2$$

$$\leq CI_{\mathscr{K}}(0) - \ell(0)V(0), \tag{7.5.4}$$

where $\mathscr{K} = \{0\}$ and $C = \frac{\alpha\pi^2}{6}$. Now

$$\begin{aligned} \sup_{(a,b)\in A(i)\times B(i)} q(i,a,b) &\leq \sup_{(a,b)\in A(i)\times B(i)} \Big\{ \hat{\lambda} i + \hat{\mu}(i+3)^2 + |\hat{h}_1(i,a)| + |\hat{h}_2(i,b)| \Big\} \\ &\leq \hat{\mu} i + 2\hat{\mu} + \hat{\mu}(i+3)^2 \\ &\leq 2(i+3)^2\hat{\mu} \leq b_2 V(i) \ \forall\ i \geq 1, \end{aligned} \tag{7.5.5}$$

where $b_2 = \max\{2\hat{\mu}, \sum_{j\geq 1} \frac{\alpha}{(j+3)^4}\}$. From (7.5.3) and (7.5.4), for all $i \in S$, we get

$$\sum_{j\in S} q(j|i,a,b)V(j) \leq b_0 V(i) + b_1, \tag{7.5.6}$$

where $b_1 = C$ and $b_0 = 1$. Now

$$\ell(i) - \max_{(a,b)\in A(i)\times B(i)} c(i,a,b) = 3 + (1-\hat{p})i + \min_{(a,b)\in A(i)\times B(i)} [\hat{c}_1(i,a) - \hat{c}_2(i,b)]. \tag{7.5.7}$$

From (7.5.5) and (7.5.6), Assumption ($A1$) is verified. By the condition (II), equations (7.5.3), (7.5.4), (7.5.7), it is easy to see that Assumption ($A2$) is verified. By (7.5.1), (7.5.2), the condition (II), and the definition of q as defined above, Assumption ($A3$)(i) is verified. By (7.5.3) and (7.5.4), Assumption ($A3$)(ii) is verified. Hence by Theorem 7.4.2, it follows that there exists an optimal pair of stationary strategies for this controlled Birth-Death process. □

Remark 7.5.3. *Note that, in view of Remark 7.4.2, one can relax the condition (7.5.1).*

CHAPTER 8

Conclusions

The main intention of this thesis is to study risk-sensitive stochastic control and stochastic games with various cost criteria. The observations, the contributions, and the scope of the future work related to this thesis in each chapter are highlighted below:

- ☞ First, we have investigated risk-sensitive discounted control problems for continuous-time jump Markov decision processes taking values in general state space. The transition rates of underlying continuous-time jump Markov processes and the cost rates are allowed to be unbounded. Under certain Lyapunov conditions, the existence and uniqueness of the solution to the Hamilton-Jacobi-Bellman (HJB) equation have been established. Also, we have proved an optimal risk-sensitive control in the class of Markov controls and have completely characterized the optimal control.

- ☞ After then, a two-person zero-sum stochastic game for controlled continuous-time Markov decision processes with risk-sensitive discounted cost criterion on countable state space has been studied. The transition and cost rates are possibly unbounded. We have established the existence of the value of the game and saddle-point equilibrium in the class of admissible feedback strategies under a Foster-Lyapunov condition by studying the corresponding Hamilton-Jacobi-Isaacs equation. The study of the corresponding nonzero-sum game problems on countable state space is still an open problem.

- ☞ Next, we have considered zero-sum stochastic games for controlled continuous time Markov processes on a general state space with risk-sensitive discounted cost criterion. The transition and cost rates are possibly unbounded. Under a stability assumption, we have proven the existence of the value of the game and a saddle-point equilibrium in the class of Markov strategies and have given a characterization of this saddle-point equilibrium in terms of the corresponding Hamilton-Jacobi-Isaacs (HJI) equation. Also, to illustrate the results and assumptions an example has been provided. The study of the corresponding nonzero-sum game problems on Borel state space is still an open problem.

☞ After that, zero-sum stochastic games for controlled discrete-time Markov decision processes with risk-sensitive average cost criterion with countable/compact state space and Borel action spaces have been studied. The payoff function is nonnegative and possibly unbounded for countable state space case and for compact state space case it is a real-valued and bounded function. For countable state space case, under a certain Lyapunov type stability assumption on the dynamics, we have established the existence of the value and a saddle-point equilibrium. For compact state space case, these results have been established without any Lyapunov type stability assumptions. Using the stochastic representation of the principal eigenfunction of the associated optimality equation, all possible saddle-point strategies in the class of stationary Markov strategies have been completely characterized. Also, an illustrative example has been presented.

It will be interesting to study the problem for locally compact state space. The major issue in extending our results from compact state space to locally compact state space is to prove the existence of a principal eigenpair to the associated Shapley equation. Following the arguments of Chapter 5 (also from [2]), it is easy to see that the principal eigenpair exists for Dirichlet eigenvalue problems (in bounded domains). But the main difficulty here is to take the limit from the bounded domain to the unbounded domain . In countable state space, we use a diagonalization argument to get the limit. This argument does not work in the general state space case. As it is pointed out in [2], that it will be a very challenging work to extend our results from compact state space to a locally compact state space. One possible way to overcome these difficulties is by imposing stronger equicontinuity conditions on the transition kernels (as in Assumption $(A1^{'})$(iii) of Chapter 5), which will enable us to use Arzela-Ascoli theorem to pass to the limit . Also, it would be an interesting works to study the same problem on countable/locally compact state space under near monotone running cost condition.

☞ Subsequently, we have analyzed nonzero-sum stochastic games for continuous-time Markov decision processes on a denumerable state space with risk-sensitive ergodic cost criterion. Transition rates and cost rates are allowed to be unbounded. Under a Lyapunov type stability assumption, we have shown that the corresponding system of coupled HJB equations admits a solution that leads to the existence of a Nash equilibrium in stationary strategies. This has been established using an approach involving principal eigenvalues associated with the HJB equations. Furthermore, exploiting appropriate stochastic representation of principal eigenfunctions, Nash equilibria in the space of stationary Markov strategies have been completely characterized. The corresponding problems on locally compact state space state space are still open. The major issue in extending our results from countable state space to locally compact state space is to prove the existence of a principal eigenpair to the associated Shapley equation. Following the arguments of Chapter 5 (also from [21]), it is easy to see that the principal eigenpair exists for Dirichlet eigenvalue problems (in bounded domains). But the main difficulty here is to take the limit from the bounded domain to the unbounded domain. In countable state space, we use a diagonalization argument to get the limit. This argument does not work in the general state space case. One possible way to overcome these difficulties is by imposing stronger equicontinuity conditions on the transition kernels (as in Assumption $(A1')$(iii) of Chapter 5), which will enable us to use Arzela-Ascoli theorem to pass to the limit. Also, it would be an interesting works to study the same problem on countable/locally compact state space under near monotone running cost condition.

- ☞ Finally, a zero-sum stochastic game in continuous-time with controlled Markov decision processes and with risk-sensitive average cost criterion has been considered. Here the transition and the cost rates may be unbounded. We have proven the existence of the value of the game and a saddle-point equilibrium in the class of all stationary strategies under a Lyapunov stability condition. This is accomplished by establishing the existence of a principal eigenpair for the corresponding Hamilton-Jacobi-Isaacs (HJI) equation. This, in turn, has been established by using a nonlinear version of Kreĭn-Rutman theorem. Then a characterization of the saddle-point equilibrium in terms of the corresponding HJI equation has been obtained. Finally, a controlled population system has been used to illustrate our results. The corresponding problems on locally compact state space are still open. The major issue in extending our results from countable state space to locally compact state space is to prove the existence of a principal eigenpair to the associated Shapley equation. Following the arguments of Chapter 5 (also from [21]), it is easy to see that the principal eigenpair exists for Dirichlet eigenvalue problems (in bounded domains). But the main difficulty here is to take the limit from the bounded domain to the unbounded domain. In countable state space, we use a diagonalization argument to get the limit. This argument does not work in the general state space case. One possible way to overcome these difficulties is by imposing stronger equicontinuity conditions on the transition kernels (as in Assumption $(A1^{'})$(iii) of Chapter 5), which will enable us to use Arzela-Ascoli theorem to pass to the limit . Also, it would be an interesting works to study the same problem on countable/locally compact state space under near monotone running cost condition.

Bibliography

[1] Altman, E. Flow control using the theory of zero sum Markov games. *IEEE Trans. Automat. Control 39*, 4 (1994), 814–818.

[2] Anantharam, V., and Borkar, V. S. A variational formula for risk-sensitive reward. *SIAM J. Control Optim. 55*, 2 (2017), 961–988.

[3] Anderson, W. J. *Continuous-time Markov chains.* Springer Series in Statistics: Probability and its Applications. Springer-Verlag, New York, 1991. An applications-oriented approach.

[4] Arapostathis, A. A counterexample to a nonlinear version of the Kreĭn-Rutman theorem by R. Mahadevan. *Nonlinear Anal. 171* (2018), 170–176.

[5] Arapostathis, A., and Biswas, A. Infinite horizon risk-sensitive control of diffusions without any blanket stability assumptions. *Stochastic Process. Appl. 128*, 5 (2018), 1485–1524.

[6] Arapostathis, A., Biswas, A., and Pradhan, S. On the policy improvement algorithm for ergodic risk-sensitive control. *Proc. Roy. Soc. Edinburgh Sect. A 151*, 4 (2021), 1305–1330.

[7] Arapostathis, A., Biswas, A., and Saha, S. Strict monotonicity of principal eigenvalues of elliptic operators in $\mathbb{R}^d$ and risk-sensitive control. *J. Math. Pures Appl. (9) 124* (2019), 169–219.

[8] Arapostathis, A., Borkar, V. S., Fernández-Gaucherand, E., Ghosh, M. K., and Marcus, S. I. Discrete-time controlled Markov processes with average cost criterion: a survey. *SIAM J. Control Optim. 31*, 2 (1993), 282–344.

[9] Athreya, K. B., and Lahiri, S. N. *Measure theory and probability theory.* Springer Texts in Statistics. Springer, New York, 2006.

[10] Balaji, S., and Meyn, S. P. Multiplicative ergodicity and large deviations for an irreducible Markov chain. *Stochastic Process. Appl. 90*, 1 (2000), 123–144.

[11] Basu, A., and Ghosh, M. K. Stochastic differential games with multiple modes and applications to portfolio optimization. *Stoch. Anal. Appl. 25*, 4 (2007), 845–867.

[12] Basu, A., and Ghosh, M. K. Zero-sum risk-sensitive stochastic games on a countable state space. *Stochastic Process. Appl. 124*, 1 (2014), 961–983.

[13] Basu, A., and Ghosh, M. K. Nonzero-sum risk-sensitive stochastic games on a countable state space. *Math. Oper. Res. 43*, 2 (2018), 516–532.

[14] Bäuerle, N., and Rieder, U. *Markov decision processes with applications to finance.* Universitext. Springer, Heidelberg, 2011.

[15] Bäuerle, N., and Rieder, U. More risk-sensitive Markov decision processes. *Math. Oper. Res. 39*, 1 (2014), 105–120.

[16] Bäuerle, N., and Rieder, U. Zero-sum risk-sensitive stochastic games. *Stochastic Process. Appl. 127*, 2 (2017), 622–642.

[17] Bertsekas, D. P., and Shreve, S. E. *Stochastic optimal control*, vol. 139 of *Mathematics in Science and Engineering.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1978. The discrete time case.

[18] Bielecki, T. R., and Pliska, S. R. Risk-sensitive dynamic asset management. *Appl. Math. Optim. 39*, 3 (1999), 337–360.

[19] Biswas, A. An eigenvalue approach to the risk sensitive control problem in near monotone case. *Systems Control Lett. 60*, 3 (2011), 181–184.

[20] Biswas, A., Borkar, V. S., and Kumar, K. S. Erratum to: Risk-sensitive control with near monotone cost [mr2679473]. *Appl. Math. Optim. 62*, 2 (2010), 165–167.

[21] Biswas, A., and Pradhan, S. Ergodic risk-sensitive control of Markov processes on countable state space revisited. *ESAIM Control Optim. Calc. Var. 28* (2022), Paper No. 26, 50.

[22] Biswas, A., and Saha, S. Zero-sum stochastic differential games with risk-sensitive cost. *Appl. Math. Optim. 81*, 1 (2020), 113–140.

[23] Borkar, V. S., and Ghosh, M. K. Denumerable state stochastic games with limiting average payoff. *J. Optim. Theory Appl. 76*, 3 (1993), 539–560.

[24] Borkar, V. S., and Meyn, S. P. Risk-sensitive optimal control for Markov decision processes with monotone cost. *Math. Oper. Res. 27*, 1 (2002), 192–209.

[25] Cavazos-Cadena, R., and Fernández-Gaucherand, E. Controlled Markov chains with risk-sensitive criteria: average cost, optimality equations, and optimal solutions. *Math. Methods Oper. Res. 49*, 2 (1999), 299–324.

[26] Cavazos-Cadena, R., and Hernández-Hernández, D. A characterization of the optimal risk-sensitive average cost in finite controlled Markov chains. *Ann. Appl. Probab. 15*, 1A (2005), 175–212.

[27] Cavazos-Cadena, R., and Hernández-Lerma, O. Equivalence of Lyapunov stability criteria in a class of Markov decision processes. *Appl. Math. Optim. 26*, 2 (1992), 113–137.

[28] Cavazos-Cadena, R., Montes-de Oca, R., and Sladký, K. A counterexample on sample-path optimality in stable Markov decision chains with the average reward criterion. *J. Optim. Theory Appl. 163*, 2 (2014), 674–684.

[29] Cavazos-Cadena, R., Montes-de Oca, R., and Sladký, K. Sample-path optimal stationary policies in stable Markov decision chains with the average reward criterion. *J. Appl. Probab. 52*, 2 (2015), 419–440.

[30] Di Masi, G. B., and Stettner, L. Risk-sensitive control of discrete-time Markov processes with infinite horizon. *SIAM J. Control Optim. 38*, 1 (1999), 61–78.

[31] Di Masi, G. B., and Stettner, L. Infinite horizon risk sensitive control of discrete time Markov processes with small risk. *Systems Control Lett. 40*, 1 (2000), 15–20.

[32] Di Masi, G. B., and Stettner, L. Infinite horizon risk sensitive control of discrete time Markov processes under minorization property. *SIAM J. Control Optim. 46*, 1 (2007), 231–252.

[33] Fan, K. Minimax theorems. *Proc. Nat. Acad. Sci. U.S.A. 39* (1953), 42–47.

[34] Federgruen, A. On N-person stochastic games with denumerable state space. *Adv. in Appl. Probab. 10*, 2 (1978), 452–471.

[35] Feinberg, E. A., Kasyanov, P. O., and Zadoianchuk, N. V. Fatou's lemma for weakly converging probabilities. *Theory Probab. Appl. 58*, 4 (2014), 683–689.

[36] Fleming, W. H., and Hernández-Hernández, D. Risk-sensitive control of finite state machines on an infinite horizon. I. *SIAM J. Control Optim. 35*, 5 (1997), 1790–1810.

[37] Fleming, W. H., and McEneaney, W. M. Risk-sensitive control on an infinite time horizon. *SIAM J. Control Optim. 33*, 6 (1995), 1881–1915.

[38] Frobenius, G. Über matrizen aus nicht negativen elementen. *Königliche Akademie der Wissenschaften Sitzungsber, Kön* (1912).

[39] Ghosh, M. K., and Bagchi, A. Stochastic games with average payoff criterion. *Appl. Math. Optim. 38*, 3 (1998), 283–301.

[40] Ghosh, M. K., Golui, S., Pal, C., and Pradhan, S. Nonzero-sum risk-sensitive continuous-time stochastic games with ergodic costs. *Appl. Math. Optim. 86*, 1 (2022), Paper No. 6, 31.

[41] Ghosh, M. K., Golui, S., Pal, C., and Pradhan, S. Discrete-time zero-sum games for Markov chains with risk-sensitive average cost criterion. *Stochastic Process. Appl. 158* (2023), 40–74.

[42] Ghosh, M. K., Golui, S., Pal, C., and Pradhan, S. Zero-sum stochastic games in continuous-time with risk-sensitive average cost criterion on a countable state space. *Math. Control Rel. Fields* (2023).

[43] Ghosh, M. K., Goswami, A., and Kumar, K. S. Portfolio optimization in a semi-Markov modulated market. *Appl. Math. Optim. 60*, 2 (2009), 275–296.

[44] Ghosh, M. K., and Kumar, K. S. A stochastic differential game in the orthrant. *J. Math. Anal. Appl. 265*, 1 (2002), 12–37.

[45] Ghosh, M. K., Kumar, K. S., and Pal, C. Zero-sum risk-sensitive stochastic games for continuous time Markov chains. *Stoch. Anal. Appl. 34*, 5 (2016), 835–851.

[46] Ghosh, M. K., and Pradhan, S. Zero-sum risk-sensitive stochastic differential games with reflecting diffusions in the orthant. *ESAIM Control Optim. Calc. Var. 26* (2020), Paper No. 114, 33.

[47] Ghosh, M. K., and Saha, S. Risk-sensitive control of continuous time Markov chains. *Stochastics 86*, 4 (2014), 655–675.

[48] Golui, S., and Pal, C. Continuous-time zero-sum games for Markov chains with risk-sensitive finite-horizon cost criterion. *Stoch. Anal. Appl. 40*, 1 (2022), 78–95.

[49] Golui, S., and Pal, C. Risk-sensitive discounted cost criterion for continuous-time Markov decision processes on a general state space. *Math. Meth. Oper. Res. 95*, 2 (2022), 219–247.

[50] Golui, S., and Pal, C. Continuous-time zero-sum games for Markov decision processes with discounted risk-sensitive cost criterion on a general state space. *Stoch. Anal. Appl. 41*, 2 (2023), 327–357.

[51] Golui, S., Pal, C., and Saha, S. Continuous-time zero-sum games for Markov decision processes with discounted risk-sensitive cost criterion. *Dyn. Games Appl. 12*, 2 (2022), 485–512.

[52] Guo, X., and Huang, Y. Risk-sensitive average continuous-time Markov decision processes with unbounded transition and cost rates. *J. Appl. Probab. 58*, 2 (2021), 523–550.

[53] Guo, X., Liu, Q., and Zhang, Y. Finite horizon risk-sensitive continuous-time Markov decision processes with unbounded transition and cost rates. *4OR 17*, 4 (2019), 427–442.

[54] Guo, X., and Zhang, Y. On risk-sensitive piecewise deterministic Markov decision processes. *Appl. Math. Optim. 81*, 3 (2020), 685–710.

[55] Guo, X. P., and Hernández-Lerma, O. Continuous-time controlled Markov chains. *Ann. Appl. Probab. 13*, 1 (2003), 363–388.

[56] Guo, X. P., and Hernández-Lerma, O. Zero-sum games for continuous-time Markov chains with unbounded transition and average payoff rates. *J. Appl. Probab. 40*, 2 (2003), 327–345.

[57] Guo, X. P., and Hernández-Lerma, O. Nonzero-sum games for continuous-time Markov chains with unbounded discounted payoffs. *J. Appl. Probab. 42*, 2 (2005), 303–320.

[58] Guo, X. P., and Hernández-Lerma, O. Zero-sum games for continuous-time jump Markov processes in Polish spaces: discounted payoffs. *Adv. in Appl. Probab. 39*, 3 (2007), 645–668.

[59] Guo, X. P., and Hernández-Lerma, O. *Continuous-time Markov decision processes*, vol. 62 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2009. Theory and applications.

[60] Guo, X. P., Hernández-Lerma, O., and Prieto-Rumeau, T. A survey of recent results on continuous-time Markov decision processes. *Top 14*, 2 (2006), 177–261. With comments and a rejoinder by the authors.

[61] Guo, X. P., Huang, Y., and Song, X. Linear programming and constrained average optimality for general continuous-time Markov decision processes in history-dependent policies. *SIAM J. Control Optim. 50*, 1 (2012), 23–47.

[62] Guo, X. P., and Liao, Z.-W. Risk-sensitive discounted continuous-time Markov decision processes with unbounded rates. *SIAM J. Control Optim. 57*, 6 (2019), 3857–3883.

[63] Guo, X. P., and Piunovskiy, A. Discounted continuous-time Markov decision processes with constraints: unbounded transition and loss rates. *Math. Oper. Res. 36*, 1 (2011), 105–132.

[64] Guo, X. P., and Song, X. Discounted continuous-time constrained Markov decision processes in Polish spaces. *Ann. Appl. Probab. 21*, 5 (2011), 2016–2049.

[65] Guo, X. P., and Zhang, J. Risk-sensitive continuous-time Markov decision processes with unbounded rates and Borel spaces. *Discrete Event Dyn. Syst. 29*, 4 (2019), 445–471.

[66] Hernández-Hernández, D., and Marcus, S. I. Risk sensitive control of Markov processes in countable state space. *Systems Control Lett. 29*, 3 (1996), 147–155.

[67] Hernández-Hernández, D., and Marcus, S. I. Existence of risk-sensitive optimal stationary policies for controlled Markov processes. *Appl. Math. Optim. 40*, 3 (1999), 273–285.

[68] Hernández-Lerma, O. *Adaptive Markov control processes*, vol. 79 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1989.

[69] Hernández-Lerma, O., and Lasserre, J. B. *Further topics on discrete-time Markov control processes*, vol. 42 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1999.

[70] Howard, R. A. *Dynamic programming and Markov processes*. Technology Press of M.I.T., Cambridge, Mass.; John Wiley & Sons, Inc., New York-London, 1960.

[71] Howard, R. A., and Matheson, J. E. Risk-sensitive Markov decision processes. *Management Sci. 18* (1971/72), 356–369.

[72] Huang, Y. Finite horizon continuous-time Markov decision processes with mean and variance criteria. *Discrete Event Dyn. Syst. 28*, 4 (2018), 539–564.

[73] Jacobson, D. H. Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games. *IEEE Trans. Automatic Control AC-18*, 2 (1973), 124–131.

[74] Jaśkiewicz, A. Average optimality for risk-sensitive control with general state space. *Ann. Appl. Probab. 17*, 2 (2007), 654–675.

[75] Kitaev, M. Y. Semi-Markov and jump Markov controllable models. Average cost criterion. *Teor. Veroyatnost. i Primenen. 30*, 2 (1985), 252–268.

[76] Kitaev, M. Y., and Rykov, V. V. *Controlled queueing systems.* CRC Press, Boca Raton, FL, 1995.

[77] Klompstra, M. B. Nash equilibria in risk-sensitive dynamic games. *IEEE Trans. Automat. Control 45*, 7 (2000), 1397–1401.

[78] Kontoyiannis, I., and Meyn, S. P. Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab. 13*, 1 (2003), 304–362.

[79] Kontoyiannis, I., and Meyn, S. P. Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electron. J. Probab. 10* (2005), no. 3, 61–123.

[80] Kreĭn, M. G., and Rutman, M. A. Linear operators leaving invariant a cone in a Banach space. *Amer. Math. Soc. Translation 1950*, 26 (1950), 128.

[81] Kumar, K. S., and Pal, C. Risk-sensitive control of pure jump process on countable space with near monotone cost. *Appl. Math. Optim. 68*, 3 (2013), 311–331.

[82] Kumar, K. S., and Pal, C. Risk-sensitive ergodic control of continuous time Markov processes with denumerable state space. *Stoch. Anal. Appl. 33*, 5 (2015), 863–881.

[83] Lal, A. K., and Sinha, S. Zero-sum two-person semi-Markov games. *J. Appl. Probab. 29*, 1 (1992), 56–72.

[84] Lemmens, B., and Nussbaum, R. *Nonlinear Perron-Frobenius theory*, vol. 189 of *Cambridge Tracts in Mathematics.* Cambridge University Press, Cambridge, 2012.

[85] Levy, Y. J., and Solan, E. Stochastic games. In *Complex social and behavioral systems—game theory and agent-based models*, Encycl. Complex. Syst. Sci. Springer, New York, [2020] ©2020, pp. 229–250.

[86] Mahadevan, R. A note on a non-linear Krein-Rutman theorem. *Nonlinear Anal. 67*, 11 (2007), 3084–3090.

[87] Meyn, S. P., and Tweedie, R. L. *Markov chains and stochastic stability.* Communications and Control Engineering Series. Springer-Verlag London, Ltd., London, 1993.

[88] Neyman, A. Continuous-time stochastic games. *Games Econom. Behav. 104* (2017), 92–130.

[89] Nowak, A. S. Measurable selection theorems for minimax stochastic optimization problems. *SIAM J. Control Optim. 23*, 3 (1985), 466–476.

[90] Nowak, A. S. Notes on risk-sensitive Nash equilibria. In *Advances in dynamic games*, vol. 7 of *Ann. Internat. Soc. Dynam. Games.* Birkhäuser Boston, Boston, MA, 2005, pp. 95–109.

[91] Nussbaum, R. D. Eigenvectors of nonlinear positive operators and the linear Kreĭn-Rutman theorem. In *Fixed point theory (Sherbrooke, Que., 1980)*, vol. 886 of *Lecture Notes in Math.* Springer, Berlin-New York, 1981, pp. 309–330.

[92] Ogiwara, T. Nonlinear Perron-Frobenius problem for order-preserving mappings. I. *Proc. Japan Acad. Ser. A Math. Sci. 69*, 8 (1993), 312–316.

[93] Ogiwara, T. Nonlinear Perron-Frobenius problem on an ordered Banach space. *Japan. J. Math. (N.S.) 21*, 1 (1995), 43–103.

[94] Pal, C., and Pradhan, S. Risk sensitive control of pure jump processes on a general state space. *Stochastics 91*, 2 (2019), 155–174.

[95] Pal, C., and Pradhan, S. Zero-sum games for pure jump processes with risk-sensitive discounted cost criteria. *J. Dyn. Games 9*, 1 (2022), 13–25.

[96] Parthasarathy, K. R. *Probability measures on metric spaces.* AMS Chelsea Publishing, Providence, RI, 2005. Reprint of the 1967 original.

[97] Parthasarathy, T., and Sinha, S. Existence of stationary equilibrium strategies in non-zero-sum discounted stochastic games with uncountable state space and state-independent transitions. *Internat. J. Game Theory 18*, 2 (1989), 189–194.

[98] Perron, O. Zur Theorie der Matrices. *Math. Ann. 64*, 2 (1907), 248–263.

[99] Piunovskiy, A., and Zhang, Y. Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach. *SIAM J. Control Optim. 49*, 5 (2011), 2032–2061.

[100] Piunovskiy, A., and Zhang, Y. *Continuous-time Markov decision processes—Borel space models and general control strategies*, vol. 97 of *Probability Theory and Stochastic Modelling.* Springer, Cham, [2020] ©2020. With a foreword by Albert Nikolaevich Shiryaev.

[101] Prieto-Rumeau, T., and Hernández-Lerma, O. *Selected topics on continuous-time controlled Markov chains and Markov games*, vol. 5 of *ICP Advanced Texts in Mathematics.* Imperial College Press, London, 2012.

[102] Puterman, M. L. *Markov decision processes: discrete stochastic dynamic programming.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, 1994. A Wiley-Interscience Publication.

[103] Ross, S. M. *Applied probability models with optimization applications.* Dover Publications, Inc., New York, 1992. Reprint of the 1970 original.

[104] Shapley, L. S. Stochastic games. In *Stochastic games and applications (Stony Brook, NY, 1999)*, vol. 570 of *NATO Sci. Ser. C Math. Phys. Sci.* Kluwer Acad. Publ., Dordrecht, 2003, pp. 1–7. Reprint of Proc. Nat. Acad. Sci. U.S.A. **39** (1953), 1095–1100 [0061807].

[105] Sussmann, H. J., and Willems, J. C. 300 years of optimal control: from the brachystochrone to the maximum principle. *IEEE Control Systems Magazine 17*, 3 (1997), 32–44.

[106] von Neumann, J., and Morgenstern, O. *Theory of Games and Economic Behavior.* Princeton University Press, Princeton, N.J., 1944.

[107] Vrieze, K. Zero-sum stochastic games. A survey. *CWI Quarterly 2*, 2 (1989), 147–170.

[108] Wei, Q. Continuous-time Markov decision processes with risk-sensitive finite-horizon cost criterion. *Math. Methods Oper. Res. 84*, 3 (2016), 461–487.

[109] Wei, Q. Zero-sum games for continuous-time Markov jump processes with risk-sensitive finite-horizon cost criterion. *Oper. Res. Lett. 46*, 1 (2018), 69–75.

[110] Wei, Q. Nonzero-sum risk-sensitive finite-horizon continuous-time stochastic games. *Statist. Probab. Lett. 147* (2019), 96–104.

[111] Wei, Q., and Chen, X. Risk-sensitive average continuous-time Markov decision processes with unbounded rates. *Optimization 68*, 4 (2019), 773–800.

[112] Wei, Q., and Chen, X. Nonzero-sum games for continuous-time jump processes under the expected average payoff criterion. *Appl. Math. Optim. 83*, 2 (2021), 915–938.

[113] Wei, Q., and Chen, X. Nonzero-sum risk-sensitive average stochastic games: the case of unbounded costs. *Dyn. Games Appl. 11*, 4 (2021), 835–862.

[114] Whittle, P. *Risk-sensitive optimal control.* Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons, Ltd., Chichester, 1990.

[115] Yushkevich, A. A. On reducing a jump controllable markov model to a model with discrete time. *Theory of Probab. Appl. 25*, 1 (1980), 58–69.

[116] Zaanen, A. C. *Introduction to operator theory in Riesz spaces.* Springer-Verlag, Berlin, 1997.

[117] Zhang, J., Guo, X. P., and Xia, L. Zero-sum risk-sensitive continuous-time stochastic games with unbounded payoff and transition rates and borel spaces. *arXiv preprint arXiv:2103.04057* (2021).

[118] Zhang, W., and Guo, X. P. Nonzero-sum games for continuous-time Markov chains with unbounded transition and average payoff rates. *Sci. China Math. 55*, 11 (2012), 2405–2416.

[119] Zhang, Y. Average optimality for continuous-time Markov decision processes under weak continuity conditions. *J. Appl. Probab. 51*, 4 (2014), 954–970.

[120] Zhang, Y. Continuous-time Markov decision processes with exponential utility. *SIAM J. Control Optim. 55*, 4 (2017), 2636–2660.

List of Published Papers that are included in the Thesis

The content of this thesis is based on the following published articles.

1. Golui, S., and Pal, C. Risk-sensitive discounted cost criterion for continuous-time Markov decision processes on a general state space, *Mathematical Methods of Operations Research*, 95, 2 (2022), 219-247, https://doi.org/10.1007/s00186-022-00779-9.

2. Golui, S., Pal, C., and Saha, S. Continuous-time zero-sum games for Markov decision processes with discounted risk-sensitive cost criterion, *Dynamic Games And Applications*, 12, 2 (2022), 485-512, https://doi.org/10.1007/s13235-021-00391-2.

3. Golui, S., and Pal, C. Continuous-time zero-sum games for Markov decision processes with discounted risk-sensitive cost criterion on a general state space, *Stochastic Analysis and Appliciations*, 41 (2021), 327-357, https://doi.org/10.1080/07362994.2021.2013889.

4. Ghosh, M. K., Golui, S., Pal, C., and Pradhan, S. Discrete-time zero-sum games for Markov chains with risk-sensitive average cost criterion, *Stochastic Processes and their Applications*, 158 (2023), 40–74, https://doi.org/10.1016/j.spa.2022.12.009.

5. Ghosh, M. K., Golui, S., Pal, C., and Pradhan, S. Nonzero-sum risk-sensitive continuous-time stochastic games with ergodic costs, *Applied Mathematics and Optimization*, 86, 6 (2022), https://doi.org/10.1007/s00245-022-09878-9.
6. Ghosh, M. K., Golui, S., Pal, C., and Pradhan, S. Zero-sum games for continuous-time Markov decision processes with risk-sensitive average cost criterion, *Mathematical Control Related Fields*, (2023), doi: 10.3934/mcrf.2023003.

List of other Papers not included in this thesis

1. Golui, S., and Pal, C. Continuous-time zero-sum games for Markov chains with risk-sensitive finite-horizon cost criterion, *Stochastic Analysis and Appliciations*, 40, 1 (2022), 78-95.

2. Golui, S., Pal, C., Rangaswamy, M., and, Sobhanan, A. Optimal Control for Production Inventory System with Various Cost Criteria, (2022), https://doi.org/10.48550/arXiv.2210.15251.